11-1-2005

Vol. 5, No. 2 (Full Issue)

JMASM Editors

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

Recommended Citation

Editors, JMASM (2005) "Vol. 5, No. 2 (Full Issue)," *Journal of Modern Applied Statistical Methods*: Vol. 5 : Iss. 2 , Article 35.

DOI: 10.22237/jmasm/1162355640

Available at: <http://digitalcommons.wayne.edu/jmasm/vol5/iss2/35>

This Full Issue is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

```

do i1=1,4
  j(1)=i1
  do i2=1,4
    j(2)=i2
    do i3=1,4
      j(3)=i3
      do i4=1,4
        j(4)=i4
        if (j(1) .eq. j(2) .or. j(1) .eq. j(3) .or. j(1) .eq. j(4)) cycle
        if (j(2) .eq. j(3) .or. j(2) .eq. j(4)) cycle
        if (j(3) .eq. j(4)) cycle
        print*,j(1),j(2),j(3),j(4)
      end do
    end do
  end do
end do

```

Journal of Modern Applied Statistical Methods

Invited Article

283 – 308

**Robert J. Boik,
Ben Haaland**

Second-Order Accurate Inference on Simple,
Partial, and Multiple Correlations

Editor

Shlomo S. Sawilowsky
Evaluation & Research
Wayne State University

Associate Editors

Harvey Keselman
Department of Psychology
University of Manitoba

Bruno D. Zumbo
Measurement, Evaluation, & Research
University of British Columbia

Assistant Editors

Vance W. Berger
Biometry Research Group
National Cancer Institute

Todd C. Headrick
Educational Psychology
Southern Illinois University
Carbondale

Alan Klockars
Educational Psychology
University of Washington

Buy Direct and Save!

Quantitative Methods in Education and the Behavioral Sciences: Issues, Research, and Teaching

(sponsored by the American Educational Research Association's Special Interest Group:
Educational Statisticians)

Series Editor

Ronald C. Serlin, *University of Wisconsin-Madison*

Real Data Analysis

Edited by **Shlomo S. Sawilowsky**, *Wayne State University*

The invited authors of this edited volume have been prolific in the arena of Real Data Analysis (RDA) as it applies to the social and behavioral sciences, especially in the disciplines of education and psychology. Combined, this brain trust represents 3,247 articles in refereed journals, 127 books published, US \$45.3 Million in extramural research funding, 34 teaching and 92 research awards, serve(d) as Editor/Assistant Editor/Editorial Board Member for 95 peer reviewed journals, and provide(d) ad hoc reviews for 362 journals. Their enormous footprint on real data analysis is showcased for professors, researchers, educators, administrators, and graduate students in the second text in the AERA/SIG ES Quantitative Methods series.

CONTENTS: Preface, *Shlomo S. Sawilowsky*. **PART I: FOUNDATIONS.** The Co-Evolution of Statistics and Hz, *Joseph M. Hilbe*. Effective Sample Size: A Crucial Concept, *Thomas R. Knapp*. Advances in Missing Data Methods and Implications for Educational Research, *Chao-Ying Joanne Peng, Michael Harwell, Show-Mann Liou, Lee H. Ehman*. Methods for Simulating Real World Data for the Psycho-Educational Sciences, *Todd Christopher Headrick*. How and Why I Use Real, Messy Data to Investigate Theory and Inform Decision Making, *Ted Micceri*. **PART II: STATISTICAL METHODS.** Using E-Mail Messages to Help Students Prepare for a Statistics Exam, *Schuyler Huck*. Randomization Tests: Statistical Tools for Assessing the Effects of Educational Interventions When Resources are Scarce, *Joel R. Levin*. A Skipped Multivariate Measure of Location: One- And Two-Sample Hypothesis Testing, *Rand R. Wilcox, H. J. Keselman*. Robust Step-Down Tests for Multivariate Group Differences, *Lisa M. Lix, Ian Clara, Aynsle Hinds, Charles Bernstein*. Dunn-Sidak Critical Values and *p* Values, *Roger E. Kirk, Joel Hetzer*. Controlling Experiment-wise Type I Errors: Good Advice for Simultaneous and Sequential Hypothesis Testing, *Shlomo S. Sawilowsky, Patric R. Spence*. Robustness and Power of Ordinal d for Paired Data, *Du Feng*. Factorial ANOVA in SPSS: Fixed-, Random-, and Mixed-Effects Models, *Richard G. Lomax, Stacy Hughey Surman*. ANOVA: Effect Sizes, Simulating Interaction vs. Main Effects, and a Modified ANOVA Table, *Shlomo S. Sawilowsky*. ANCOVA and Quasi-Experimental Design: The Legacy of Campbell and Stanley, *Shlomo S. Sawilowsky*. **PART III: MEASUREMENT:** Thinking About Item Response Theory from a Logistic Regression Perspective: A Focus on Polytomous Models, *Amery D. Wu, Bruno D. Zumbo*. Some Practical Uses of Item Response Time to Improve the Quality of Low-Stakes Achievement Test Data, *Steven L. Wise, Xiaojing Kong*. Using Moving Averages to Detect Exposed Test Items in Computer-Based Testing, *Ning Han, Ronald K. Hambleton*. An Empirical Calibration of the Effects of Multiple Sources of Measurement Error on Reliability Estimates for Individual Differences Measures, *Frank L. Schmidt, Huy Ahn Le*. Latent Structure of Attitudes toward Abortion, *C. Mitchell Dayton*. **PART IV: DATA ANALYSIS.** Hierarchical Linear Models and the Estimation of Students' Mathematics Achievement, *Kathrin A. Parks, Dudley L. Poston, Jr.* Grade Inflation: An Examination at the Institutional Level, *Sharon L. Weinberg*. Using Discrete-Time Survival Analysis to Study Gender Differences in Leaving Mathematics, *Suzanne E. Graham, Judith D. Singer*. Nonparametric procedures for testing for dropout rates on University courses with application to an Italian case study, *Rosa Arboretti Giancristofaro, Fortunato Pesarin, Luigi Salmaso, Aldo Solari*. Nonparametric Approaches for Multivariate Testing with Mixed Variables and for Ranking on Ordered Categorical Variables with an Application to the Evaluation of Ph. D. Programs, *Rosa Arboretti Giancristofaro, Fortunato Pesarin, Luigi Salmaso*. Randomized Replicated Single-case Experiments: Treatment of Pain-related Fear by Graded Exposure *In Vivo*, *Patrick Onghena, Johan W. S. Vlaeyen, Jeroen de Jong*. Whole Brain Correlations: Examining Similarity Across Conditions of Overall Patterns of Neural Activation in fMRI, *Arthur Aron, Susan Whitfield, Wemara Lichty*. Principal Component Analysis of Senate Voting Patterns. *Jan de Leeuw*

In press 2006

Paperback ISBN: 978-1-59311-564-7 \$39.95

Hardcover ISBN: 978-1-59311-565-4 \$73.95

Also Available in the AERA SIG/Educational Statistician Series:

Structural Equation Modeling: A Second Course

2005

Paperback ISBN: 1-59311-014-6 \$39.95

Hardcover ISBN: 1-59311-015-4 \$73.95

Published by:

Information Age Publishing Inc., PO Box 4967 Greenwich, CT 06831

Tel: 203-661-7602 Fax: 203-661-7952 URL: www.infoagepub.com

Journal Of Modern Applied Statistical Methods

Shlomo S. Sawilowsky

Editor

College of Education
Wayne State University

Harvey Keselman

Associate Editor

Department of Psychology
University of Manitoba

Bruno D. Zumbo

Associate Editor

Measurement, Evaluation, & Research Methodology
University of British Columbia

Vance W. Berger

Assistant Editor

Biometry Research Group
National Cancer Institute

Todd C. Headrick

Assistant Editor

Educational Psychology and Special Education
Southern Illinois University-Carbondale

Alan Klockars

Assistant Editor

Educational Psychology
University of Washington

John L. Cuzzocrea

Editorial Assistant

Educational Research
University of Akron

Editorial Board

Subhash Chandra Bagui
Department of Mathematics & Statistics
University of West Florida

J. Jackson Barnette
School of Public Health
University of Alabama at Birmingham

Vincent A. R. Camara
Department of Mathematics
University of South Florida

Ling Chen
Department of Statistics
Florida International University

Christopher W. Chiu
Test Development & Psychometric Rsch
Law School Admission Council, PA

Jai Won Choi
National Center for Health Statistics
Hyattsville, MD

Rahul Dhanda
Forest Pharmaceuticals
New York, NY

John N. Dyer
Dept. of Information System & Logistics
Georgia Southern University

Matthew E. Elam
Dept. of Industrial Engineering
University of Alabama

Mohammed A. El-Saidi
Accounting, Finance, Economics &
Statistics, Ferris State University

Felix Famoye
Department of Mathematics
Central Michigan University

Barbara Foster
Academic Computing Services, UT
Southwestern Medical Center, Dallas

Shiva Gautam
Department of Preventive Medicine
Vanderbilt University

Dominique Haughton
Mathematical Sciences Department
Bentley College

Scott L. Hersherberger
Department of Psychology
California State University, Long Beach

Joseph Hilbe
Departments of Statistics/ Sociology
Arizona State University

Sin-Ho Jung
Dept. of Biostatistics & Bioinformatics
Duke University

Jong-Min Kim
Statistics, Division of Science & Math
University of Minnesota

Harry Khamis
Statistical Consulting Center
Wright State University

Kallappa M. Koti
Food and Drug Administration
Rockville, MD

Tomasz J. Kozubowski
Department of Mathematics
University of Nevada

Kwan R. Lee
GlaxoSmithKline Pharmaceuticals
Collegeville, PA

Hee-Jeong Lim
Dept. of Math & Computer Science
Northern Kentucky University

Balgobin Nandram
Department of Mathematical Sciences
Worcester Polytechnic Institute

J. Sunil Rao
Dept. of Epidemiology & Biostatistics
Case Western Reserve University

Karan P. Singh
University of North Texas Health
Science Center, Fort Worth

Jianguo (Tony) Sun
Department of Statistics
University of Missouri, Columbia

Joshua M. Tebbs
Department of Statistics
Kansas State University

Dimitrios D. Thomakos
Department of Economics
Florida International University

Justin Tobias
Department of Economics
University of California-Irvine

Dawn M. VanLeeuwen
Agricultural & Extension Education
New Mexico State University

David Walker
Educational Tech, Rsrch, & Assessment
Northern Illinois University

J. J. Wang
Dept. of Advanced Educational Studies
California State University, Bakersfield

Dongfeng Wu
Dept. of Mathematics & Statistics
Mississippi State University

Chengjie Xiong
Division of Biostatistics
Washington University in St. Louis

Andrei Yakovlev
Biostatistics and Computational Biology
University of Rochester

Heping Zhang
Dept. of Epidemiology & Public Health
Yale University

INTERNATIONAL

Mohammed Ageel
Dept. of Mathematics, & Graduate School
King Khalid University, Saudi Arabia

Mohammad Fraiwan Al-Saleh
Department of Statistics
Yarmouk University, Irbid-Jordan

Keumhee Chough (K.C.) Carriere
Mathematical & Statistical Sciences
University of Alberta, Canada

Michael B. C. Khoo
Mathematical Sciences
Universiti Sains, Malaysia

Debasis Kundu
Department of Mathematics
Indian Institute of Technology, India

Christos Koukouvinos
Department of Mathematics
National Technical University, Greece

Lisa M. Lix
Dept. of Community Health Sciences
University of Manitoba, Canada

Takis Papaioannou
Statistics and Insurance Science
University of Piraeus, Greece

Nasrollah Saebi
Computing, Information Systems & Math
Kingston University, UK

Keming Yu
Department of Statistics
University of Plymouth, UK

Journal Of Modern Applied Statistical Methods

Invited Articles

283 – 308	Robert J. Boik,
Ben Haaland | Second-Order Accurate Inference on Simple,
Partial, And Multiple Correlations |
|-----------|--|--|

Regular Articles

309 – 316	Rand R. Wilcox	Inferences About The Components Of A
Generalized Additive Model
317 – 322	Markus Neuhäuser,
Ludwig A. Hothorn	Maximum Tests Are Adaptive Tests
323 – 327	Vee Ming Ng	Inference for $P(Y < X)$ For Exponential And
Related Distributions
328 – 331	Sat Gupta,
Javid Shabbir	An Alternative to Warner's Randomized Response
Model
332 – 343	Gibbs Y. Kanyongo	The Influence of Reliability on Four Rules for
Determining the Number of Components to Retain
344 – 346	Marley W. Watkins	Determining Parallel Analysis Criteria
347 – 355	Leming Qu,
Yi-Cheng Tu	Change Point Estimation of Bilevel Functions
356 – 366	Ayman Baklizi,
Amjad AL-Nasser	Large Sample and Bootstrap Intervals for the
Gamma Scale Parameter Based On Grouped Data
367 – 380	B. M. Golam Kibria	Applications of Some Improved Estimators in
Linear Regression
381 – 389	Mohammed Ebrahim	Bootstrap Intervals of the Parameters of
Lognormal Distribution Using Power Rule Model
and Accelerated Life Tests
390 – 394	Seemon Thomas,
S. C. Bagui, D. K. Ghosh,
Alex Thannippara	Determination of Optimal Block Designs
with Pre-assigned Variance for
Elementary Contrasts
395 – 407	Kung-Jong Lui	Interval Estimation of Risk Difference in Simple
Compliance Randomized Trials
408 – 416	H. E. T. Holgersson	Simulation of Non-normal Auto Correlated
Variables

417 – 431	Vicki S. Hertzberg, Fan Xu, Michael Haber	Restricted Quasi-independent Model Resolves Paradoxical Behaviors of Cohen's Kappa
432 – 442	Jamis J. Perrett	A Method For Analyzing Unreplicated Experiments Using Information on the Intraclass Correlation Coefficient
443 – 451	David A. Walker	A Comparison of the Spearman-Brown and Flanagan-Rulon Formulas for Split Half Reliability under Variance Parameter Conditions
452 – 457	James F. Reed, III	AB/BA Crossover Trials - Binary Outcome
458 – 463	Ramalingam Shanmugam, Anwar Hassan, Peer B. Ahmad	Correlation Between the Number of Epileptic and Healthy Children in Family Size that Follows a Size-Based Modified Power Series Distribution
464 – 474	Michael B. C. Khoo, S. Y. Sim	A Robust Exponentially Weighted Moving Average Control Chart for the Process Mean
475 – 489	M.L.Aggarwal, S.Roy Chowdhury, Anita Bansal, Neena Mital	Interaction Graphs for $4r2n-p$ Fractional Factorial Designs
490 – 494	Kenneth Lachlan, Patric R. Spence	Corrections for Type I Error in Social Science Research: A Disconnect Between Theory and Practice
495 – 512	Mammadagha Mammadov, Berna Yazıcı, Şenay Yolacan, Atilla Aslanargun, Ali Fuat Yüzer, Embiya Ağaoğlu	Statistical Methods and Artificial Neural Networks
513 – 528	N. Ismail Abdul Aziz Jemain	A Comparison of Risk Classification Methods for Claim Severity Data
529 – 537	Maria D. S. Paraguas, Anton A. Kamil	Estimation of Meat Demand System in Malaysia: Model Selection Between the Rotterdam model and the Almost Ideal Demand System (AIDS)

Brief Reports

538 – 541	Michael Wolf-Branigin	Supporting and Preparing Future Decision-makers with the Needed Tools
-----------	------------------------------	--

Early Scholars

- 542 – 550 **Betül Kan,** The Individuals Control Chart in Case of Non-
 Berna Yazici Normality

Statistical Software Applications and Review

- 551 – 566 **David A. Heiser** Statistical Tests, Tests of Significance, and Tests of
 a Hypothesis (Excel)

Algorithms and Code

- 567 – 574 **Todd C. Headrick** JMASM24: Numerical Computing for Third-
 Order Power Method Polynomials (Excel)

- 575 – 588 **Sikha Bagui,** JMASM25: Computing Percentiles of Skew-
 Subhash Bagui Normal Distributions (Visual Basic)

Translations, Ephemerals, & Biographies

- 589 – 592 **Shlomo S. Sawilowsky,** Joseph Liouville's 'Mathematical Works Of
 John L. Cuzzocrea Évariste Galois'

- 593 – 595 **John L. Cuzzocrea,** Pietro Paoli, Italian Algebraist
 Shlomo S. Sawilowsky

JMASM is an independent print and electronic journal (<http://tbf.coe.wayne.edu/jmasm>), publishing (1) new statistical tests or procedures, or the comparison of existing statistical tests or procedures, using computer-intensive Monte Carlo, bootstrap, jackknife, or resampling methods, (2) the study of nonparametric, robust, permutation, exact, and approximate randomization methods, and (3) applications of computer programming, preferably in Fortran (all other programming environments are welcome), related to statistical algorithms, pseudo-random number generators, simulation techniques, and self-contained executable code to carry out new or interesting statistical methods.

Editorial Assistant: **John Cuzzocrea** Production Staff: **Christina Gase**
Internet Sponsor: **Paula C. Wood**, Dean, College of Education, Wayne State University

Cushing-Malloy, Inc.	(888) 295-7244 toll-free (Phone)	Sales & Information:
Internet: www.cushing-malloy.com	(734) 663-5731 (Fax)	skehoe@cushing-malloy.com

STATISTICIANS

HAVE YOU VISITED THE

Mathematics Genealogy Project?

The Mathematics Genealogy Project is an ongoing research project tracing the intellectual history of all the mathematical arts and sciences through an individual's Ph.D. advisor and Ph.D. students. Currently we have over 80,000 records in our database. We welcome and encourage all statisticians to join us in this endeavor.

Please visit our web site

<http://genealogy.math.ndsu.nodak.edu>

The information which we collect is the following:

The full name of the individual, the school where he/she earned a Ph.D., the year of the degree, the title of the dissertation, and, MOST IMPORTANTLY, the full name of the advisor(s). E.g., Fuller, Wayne Arthur; Iowa State University; 1959; *A Non-Static Model of the Beef and Pork Economy*; Shepherd, Geoffrey Seddon

For additions or corrections for one or two people a link is available on the site. For contributions of large sets of names, e.g., all graduates of a given university, it is better to send the data in a text file or an MS Word file or an MS Excel file, etc. Send such information to:

harry.coonce@ndsu.nodak.edu

The genealogy project is a not-for-profit endeavor supported by donations from individuals and sales of posters and t-shirts. If you would like to help this cause please send your tax-deductible contribution to: Mathematics Genealogy Project, 300 Minard Hall, P. O. Box 5075, Fargo, North Dakota 58105-5075E

Invited Articles
**Second-Order Accurate Inference on Simple, Partial, and
Multiple Correlations**

Robert J. Boik
Mathematical Sciences

Montana State University–Bozeman

Ben Haaland
Statistics

University of Wisconsin–Madison

This article develops confidence interval procedures for functions of simple, partial, and squared multiple correlation coefficients. It is assumed that the observed multivariate data represent a random sample from a distribution that possesses finite moments, but there is no requirement that the distribution be normal. The coverage error of conventional one-sided large sample intervals decreases at rate $1/\sqrt{n}$ as n increases, where n is an index of sample size. The coverage error of the proposed intervals decreases at rate $1/n$ as n increases. The results of a simulation study that evaluates the performance of the proposed intervals is reported and the intervals are illustrated on a real data set.

Key words: bootstrap, confidence intervals, Cornish-Fisher expansion, Edgeworth expansion, second-order accuracy

Introduction

Accurate inference procedures for simple, partial, and multiple correlation coefficients depend on accurate approximations to the distributions of estimators of these coefficients. In this article, Edgeworth ex-

pansions for these distributions are derived. The Edgeworth expansions, in turn, are used to construct confidence intervals that are more accurate than conventional large sample intervals.

Denote the $p \times p$ sample correlation matrix based on a random sample of size N from a p -variate distribution by $\mathbf{R}$ and denote the corresponding population correlation matrix by $\mathbf{\Delta}$. The $p^2 \times 1$ vectors obtained by stacking the columns of $\mathbf{R}$ and $\mathbf{\Delta}$ are denoted by $\mathbf{r}$ and $\boldsymbol{\rho}$, and are obtained by applying the vec operator. That is, $\mathbf{r} \stackrel{\text{def}}{=} \text{vec } \mathbf{R}$ and $\boldsymbol{\rho} \stackrel{\text{def}}{=} \text{vec } \mathbf{\Delta}$. The exact joint distribution of the components of $\mathbf{r}$,

Robert Boik (rjboik@math.montana.edu) is Professor of Statistics. His research interests include linear models, multivariate analysis, and large sample methods. Ben Haaland (haaland@stat.wisc.edu) is a Ph.D. student. His research interests include multivariate analysis, rates of convergence, and nonparametric techniques.

when sampling from a multivariate normal distribution, was derived by Fisher (1962), but the distribution is very difficult to use in practice because it is expressed in integral form unless $p = 2$. If sample size is sufficiently large, however, then one may substitute the asymptotic distribution of $\mathbf{r}$, but with some loss of accuracy.

In this article, the big O (pronounced oh) notation (Bishop, Fienberg, & Holland, 1975, §14.2) is used to index the magnitude of a quantity. Let u_n be a quantity that depends on $n = N - r_x$, where r_x is a fixed constant and N is sample size. Then, $u_n = O(n^{-k})$ if $n^k|u_n|$ is bounded as $n \rightarrow \infty$. Note that if $u_n = O(n^{-k})$ and $k > 0$, then u_n converges to zero as $n \rightarrow \infty$. Also, the rate of convergence to zero is faster for large k than for small k . An approximation to the distribution of a random variable is said to be j^{th} -order accurate if the difference between the approximating cumulative distribution function (cdf) and the exact cdf has magnitude $O(n^{-j/2})$. For example, the cdf of a sample correlation coefficient, r_{ij} , is $F_{r_{ij}}(t) \stackrel{\text{def}}{=} P(r_{ij} \leq t)$. Suppose that $\hat{F}_{r_{ij}}$ is an estimator of $F_{r_{ij}}$. Then, $\hat{F}_{r_{ij}}$ is first-order accurate if $|F_{r_{ij}}(t) - \hat{F}_{r_{ij}}(t)| = O(n^{-1/2})$, for all t .

Pearson and Filon (1898) showed that the first-order accurate asymptotic joint distribution of the components of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$, when sampling from a multivariate normal distribution, is itself multivariate normal with mean zero and finite covariance matrix. Pearson and Filon also derived expressions for the components of the asymptotic covariance matrix. A matrix expression for the asymptotic covariance matrix was derived by Nell (1985). Niki and Konishi (1984) derived the Edgeworth and Cornish-Fisher expansions for $\sqrt{n}[Z(r) - Z(\rho)]$ with error only $O(n^{-9/2})$ when sampling from a bivariate normal

distribution, where Z is Fisher's (1921) Z transformation. If the requirement of multivariate normality is relaxed, then the distributions of r and $\mathbf{r}$ are substantially more complicated. Fortunately, the asymptotic distribution of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ still is multivariate normal with mean zero and finite covariance matrix whenever the parent distribution has finite fourth-order moments. Expressions for the scalar components of the asymptotic covariance matrix of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ when sampling from non-normal distributions were derived by Hsu (1949) and Steiger and Hakstian (1982, 1983). Corresponding matrix expressions were derived by Browne and Shapiro (1986) and Neudecker and Wesselman (1990). The asymptotic bias of $\mathbf{r}$ when sampling from non-normal distributions was obtained by Boik (1998).

In the bivariate case, Cook (1951) obtained scalar expressions for the moments of r with error $O(n^{-5/2})$. These moments could be used to compute the first three cumulants of $\sqrt{n}(r - \rho)$. The first four cumulants and the corresponding Edgeworth expansion for the distribution of $\sqrt{n}(r - \rho)$ were obtained by Nakagawa and Niki (1992).

Some distributional results for $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ have been obtained under special non-normal conditions. Neudecker (1996), for example, obtained a matrix expression for the asymptotic covariance matrix in the special case of elliptical parent distributions. Also, Yuan and Bentler (1999) gave the asymptotic distribution of correlation and multiple correlation coefficients when the p -vector of random variables can be written as a linear function of $\mathbf{z}v$, where $\mathbf{z}$ is a p -vector of independent components and v is a scalar that is independent of $\mathbf{z}$.

This article focuses on confidence intervals for functions of $\boldsymbol{\Delta}$. Specifically, second-order accurate interval estimators

for simple, partial, and squared multiple correlation coefficients as well as for differences among simple, partial, and squared multiple correlation coefficients are constructed. A confidence interval is said to be j^{th} -order accurate if the difference between the actual coverage and the stated nominal coverage has magnitude $O(n^{-j/2})$. In general, large sample confidence intervals are first-order accurate under mild validity conditions.

First-order accurate confidence intervals for correlation functions are not new. Olkin and Siotani (1976) and Hedges and Olkin (1983) used the delta method (Rao, 1973, §6a.2) together with the asymptotic distribution of $\sqrt{n}(\mathbf{r} - \boldsymbol{\rho})$ when sampling from a multivariate normal distribution to derive the asymptotic distribution of partial and multiple correlation coefficient estimators. Olkin and Finn (1995) used these results to obtain explicit expressions for asymptotic standard errors of estimators of simple, partial, and squared multiple correlation coefficients as well as expressions for standard errors of differences among these coefficients. The major contribution of Olkin and Finn (1995) was that they demonstrated how to use existing moment expressions to compute first-order accurate confidence intervals for correlation functions when sampling from multivariate normal distributions.

To avoid complicated expressions for derivatives, Olkin and Finn (1995) gave confidence intervals for partial correlations and their differences only in the special case when the effects of a single variable are partialled out. Similarly, confidence intervals for differences among squared multiple correlation coefficients (estimated using a single sample) are given only for the special cases when either one or two explanatory variables are employed. Graf and Alf (1999) used numerical derivatives to extend the re-

sults of Olkin and Finn to allow conditioning on any set of variables rather than a single variable and to allow squared multiple correlations to be based on an arbitrary number of explanatory variables. Alf and Graf (1999) gave scalar equations for the standard error of the difference between two squared sample multiple correlations. This article also extends the results of Olkin and Finn (1995), but does so using relatively simple and easily computed (with a computer) matrix expressions for the required derivatives. As in the previous articles, this article relies on derivatives of correlation functions. Unlike Olkin and Finn, however, simple, partial, and multiple correlations are treated as functions of the covariance matrix, $\boldsymbol{\Sigma}$, instead of the correlation matrix, $\boldsymbol{\Delta}$. The advantage of the current approach is simplicity of the resulting expansions.

This article also extends the results of Olkin and Finn (1995), Graf and Alf (1999), and Alf and Graf (1999) to be applicable to non-normal as well as normal distributions. This extension is straightforward. Denote the sample covariance based on a random sample of size N by $\mathbf{S}$. Then, one needs only to replace the asymptotic covariance of $\text{vec } \mathbf{S}$ derived under normality with the asymptotic covariance matrix derived under general conditions.

In summary, the contributions of this article are as follows: (a) easily computed matrix expressions for the required derivatives are given (see Theorems 2 and 3), (b) the proposed intervals are asymptotically distribution free (ADF); and (c) the accuracy of the confidence intervals is extended from first-order to second-order. In addition, the proposed interval estimators are illustrated on a real data set and a small simulation study that evaluates the performance of the proposed intervals is conducted.

Before describing the new results, the Olkin and Finn (1995) article will be briefly revisited. It appears that the methods described in Olkin and Finn have begun to be incorporated into statistical practice (34 citations from 1997 to June 2005, most of which are reports of empirical studies). Unfortunately, the derivatives reported in the article contain several errors. The effects of these errors on the examples given by Olkin and Finn are minimal, but their effects in practice are unknown. A list of corrections to these derivatives is given in the next section.

Corrections to Olkin and Finn

Below is a list of corrections to the derivatives reported in Olkin and Finn (1995).

Model B. Components a_1 and a_3 in model B on page 159 should be the following:

$$a_1 = 2 \left[\rho_{01} (\rho_{12}^2 - \rho_{13}^2) + \rho_{03} \rho_{13} - \rho_{02} \rho_{12} \right. \\ \left. + \rho_{12} \rho_{13} (\rho_{02} \rho_{13} - \rho_{03} \rho_{12}) \right] \\ \div [(1 - \rho_{12}^2)(1 - \rho_{13}^2)] \\ \text{and } a_3 = -2 \left(\frac{\rho_{03} - \rho_{01} \rho_{13}}{1 - \rho_{13}^2} \right).$$

The nominal 95% confidence interval for $\rho_{0(12)}^2 - \rho_{0(13)}^2$ should be $[-0.19, 0.015]$ rather than $[-0.22, 0.018]$.

Model C. Components a_1 and a_3 in model C on page 160 should be the following:

$$a_1 = \sqrt{(1 - \rho_{02}^2)(1 - \rho_{12}^2)} - 1 \text{ and} \\ a_3 = \frac{\rho_{02} - \rho_{01} \rho_{12}}{1 - \rho_{12}^2}.$$

The nominal 95% confidence interval for $\rho_{01} - \rho_{01.2}$ should be $[-0.0018, 0.0073]$ rather than $[-0.0017, 0.0080]$.

Model D. Components a_1 , a_3 , and a_5 in model D on page 161 should be the following:

$$a_1 = \frac{1}{\sqrt{(1 - \rho_{02}^2)(1 - \rho_{12}^2)}} \\ - \frac{1}{\sqrt{(1 - \rho_{03}^2)(1 - \rho_{13}^2)}}, \\ a_3 = \frac{\rho_{13} - \rho_{01} \rho_{03}}{(1 - \rho_{03}^2)^{3/2}(1 - \rho_{13}^2)^{1/2}}, \text{ and} \\ a_5 = \frac{\rho_{03} - \rho_{01} \rho_{13}}{(1 - \rho_{03}^2)^{1/2}(1 - \rho_{13}^2)^{3/2}}.$$

The nominal 95% confidence interval for $\rho_{01.2} - \rho_{01.3}$ should be $[0.020, 0.054]$ rather than $[0.017, 0.057]$.

Notation and Preliminary Remarks

In some applications, an investigator may have one or more fixed explanatory variables (e.g., group or treatment variables) whose effects must be removed before estimating the correlation matrix. A linear model that is suitable for this purpose is described in this section. The sample covariance matrix, $\mathbf{S}$, is based on the residuals from the fitted model. The confidence interval procedures that are constructed later in this article require estimates of second and third-order moments of $\mathbf{S}$. Expressions for the population moments as well as consistent estimators of these moments are described in this article. Sufficient validity conditions were described in Boik (2005, §5.1).

Linear Model

Denote the response vector for the i^{th} subject by $\mathbf{y}_i$ and let $\mathbf{Y}$ be the $N \times p$ response matrix whose i^{th} row is $\mathbf{y}_i'$. That is,

$\mathbf{Y} = (\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_N)'$. The joint model for the N subjects is

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \quad (1)$$

where $\mathbf{X}$ is an $N \times q$ matrix of fixed explanatory variables, and $\mathbf{E}$ is an $N \times p$ random matrix. Denote the i^{th} row of $\mathbf{E}$ by $\boldsymbol{\varepsilon}'_i$. It is assumed that $\boldsymbol{\varepsilon}_i$ for $i = 1, \dots, N$ are independently and identically distributed with mean $E(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$.

The components of $\boldsymbol{\Sigma}$ are denoted by σ_{ij} for $i = 1, \dots, p$ and $j = 1, \dots, p$. Note that the variance of the i^{th} variable is σ_{ii} , not σ_{ii}^2 , and the correlation between the i^{th} and j^{th} variables is $\rho_{ij} = \sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}}$. To represent the matrix of correlations as a function of $\boldsymbol{\Sigma}$, a notation for diagonal matrices is needed. Suppose that $\mathbf{M}$ is a $q \times q$ matrix with positive diagonal components $m_{11}, m_{22}, \dots, m_{qq}$ and that b is a real number. Then the diagonal matrix with diagonal components $m_{11}^b, m_{22}^b, \dots, m_{qq}^b$ is denoted by $(\mathbf{M})_D^b$. That is,

$$(\mathbf{M})_D^b = \sum_{i=1}^q \mathbf{e}_i^q m_{ii}^b \mathbf{e}_i^{q'}, \quad (2)$$

where $\mathbf{e}_i^q$ is the i^{th} column of $\mathbf{I}_q$. Using this notation, the $p \times p$ correlation matrix is

$$\boldsymbol{\Delta} = (\boldsymbol{\Sigma})_D^{-\frac{1}{2}} \boldsymbol{\Sigma} (\boldsymbol{\Sigma})_D^{-\frac{1}{2}}.$$

Estimators of simple, partial, and multiple correlation coefficients will be based on the sample covariance matrix, namely

$$\mathbf{S} = n^{-1} \mathbf{Y}' \mathbf{A} \mathbf{Y}, \quad (3)$$

where $\mathbf{A} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$, $(\)^{-}$ denotes an arbitrary generalized inverse, $n = N - r_x$, and $r_x = \text{rank}(\mathbf{X})$. It is readily shown that if the rows of $\mathbf{Y}$ have finite variance, then $\mathbf{S}$ is unbiased for $\boldsymbol{\Sigma}$.

Moments of the Sample Covariance Matrix

If fourth-order moments of $\mathbf{Y}$ are finite, then the central limit theorem ensures that

$$\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \xrightarrow{\text{dist}} N(\mathbf{0}, \boldsymbol{\Omega}_{22,\infty}), \quad \text{where} \quad (4)$$

$$\boldsymbol{\Omega}_{22,\infty} = \boldsymbol{\Upsilon}_{22} - \boldsymbol{\sigma}\boldsymbol{\sigma}', \quad \boldsymbol{\Upsilon}_{22} = E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \otimes \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i'),$$

$\mathbf{s} \stackrel{\text{def}}{=} \text{vec } \mathbf{S}$, $\boldsymbol{\sigma} \stackrel{\text{def}}{=} \text{vec } \boldsymbol{\Sigma}$, and $\boldsymbol{\varepsilon}_i$ is the i^{th} row of $\mathbf{E}$ in (1). Boik (1998, 2005, eq. 12) showed that if fourth-order moments of $\mathbf{Y}$ are finite, then the finite-sample variance of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ is

$$\boldsymbol{\Omega}_{22,n} \stackrel{\text{def}}{=} \text{Var} [\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})] \quad (5)$$

$$= \frac{c_1}{n} (\boldsymbol{\Upsilon}_{22} - \boldsymbol{\sigma}\boldsymbol{\sigma}') + \left(1 - \frac{c_1}{n}\right) 2\mathbf{N}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}),$$

where $c_1 = \sum_{i=1}^N a_{ii}^2$, a_{ij} is the ij^{th} component of $\mathbf{A}$ in (3), $\boldsymbol{\Upsilon}_{22}$ is defined in (4), $\mathbf{N}_p = (\mathbf{I}_{p^2} + \mathbf{I}_{(p,p)})/2$, and $\mathbf{I}_{(a,b)}$ is the commutation matrix (MacRae, 1974). Magnus and Neudecker (1979, 1999 §3.7) denoted $\mathbf{I}_{(a,b)}$ by $\mathbf{K}_{b,a}$. If data are sampled from a multivariate normal distribution, then the finite-sample variance and an unbiased estimator of the finite-sample variance of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ are

$$\boldsymbol{\Omega}_{22,n} = 2\mathbf{N}_p(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \quad \text{and}$$

$$\begin{aligned} \hat{\boldsymbol{\Omega}}_{22,n} &= \frac{n^2}{(n-1)(n+2)} 2\mathbf{N}_p(\mathbf{S} \otimes \mathbf{S}) \\ &\quad - \frac{2n}{(n-1)(n+2)} \mathbf{ss}', \end{aligned}$$

respectively. From Boik (1998, 2005, eq. 13), an unbiased estimator of $\boldsymbol{\Omega}_{22,n}$ under general conditions is

$$\hat{\boldsymbol{\Omega}}_{22,n} = a_1 \tilde{\boldsymbol{\Upsilon}}_{22} + a_2 2\mathbf{N}_p(\mathbf{S} \otimes \mathbf{S}) \quad (6)$$

$$+ a_3 \mathbf{ss}', \quad \text{where}$$

$$\begin{aligned}
a_1 &= \frac{n^2 c_1}{d}, \quad a_2 = -\frac{n^2(c_1^2 - nc_2)}{(n-1)d}, \\
a_3 &= -\frac{n[2nc_2 + (n-3)c_1^2]}{(n-1)d}, \\
c_2 &= \sum_{i=1}^N \sum_{j=1}^N a_{ij}^4, \quad d = n(n+2)c_2 - 3c_1^2, \\
\tilde{\Upsilon}_{22} &= n^{-1} \sum_{i=1}^N (\tilde{\epsilon}_i \tilde{\epsilon}_i' \otimes \tilde{\epsilon}_i \tilde{\epsilon}_i'),
\end{aligned}$$

c_1 is defined in (5), and $\tilde{\epsilon}_i = \mathbf{Y}' \mathbf{A} \mathbf{e}_i^N$ is the p -vector of residuals from the fitted model for subject i .

Two additional covariance-related quantities are needed in this article, namely

$$\Omega_{42,n} \quad (7)$$

$$\begin{aligned}
&\stackrel{\text{def}}{=} \text{Cov} [\sqrt{n}(\hat{\omega} - \omega), \sqrt{n}(\mathbf{s} - \sigma)] \\
&= n\mathbf{E}[(\hat{\omega} - \omega)(\mathbf{s} - \sigma)'] \text{ and}
\end{aligned}$$

$$\Omega_{222,n} \stackrel{\text{def}}{=} n^{\frac{3}{2}} \mathbf{E}[(\mathbf{s} - \sigma)(\mathbf{s} - \sigma)' \otimes (\mathbf{s} - \sigma)],$$

where $\hat{\omega} = \text{vec } \hat{\Omega}_{22,n}$ and $\omega = \text{vec } \Omega_{22,n}$. Using the results of Boik (2005), it can be shown that if sixth order moments are finite and the column space generated by $\mathbf{X}$ in (1) contains the n -vector of ones, then $\Omega_{42,n}$ and $\sqrt{n}\Omega_{222,n}$ can be written as follows:

$$\begin{aligned}
&\Omega_{42,n} \quad (8) \\
&= \Upsilon_{42} - 2\mathbf{N}_{p^2} (\Omega_{22,n} \otimes \sigma) - (\Upsilon_{22} \otimes \sigma) \\
&\quad - 4\mathbf{N}_{p^2} (\mathbf{N}_p \otimes \mathbf{I}_{p^2}) (\Upsilon_{21}' \otimes \text{vec } \Upsilon_{21}) \\
&\quad + O(n^{-1}) \text{ and } \sqrt{n}\Omega_{222,n} \\
&= \Omega_{42,n} - (\Upsilon_{21} \otimes \Upsilon_{21}) 2\mathbf{N}_p + O(n^{-1}),
\end{aligned}$$

where $\Upsilon_{42} = \mathbf{E}(\epsilon_i \otimes \epsilon_i \otimes \epsilon_i \epsilon_i' \otimes \epsilon_i \epsilon_i')$,

and $\Upsilon_{21} = \mathbf{E}(\epsilon_i \otimes \epsilon_i \epsilon_i')$.

If the data are a random sample from a multivariate normal distribution, then to order $O(n^{-1})$, $\Omega_{42,n}$ and $\sqrt{n}\Omega_{222,n}$ simplify to

$$\Omega_{42,n} \quad (9)$$

$$= (2\mathbf{N}_p \otimes 2\mathbf{N}_p) (\Sigma \otimes \sigma \otimes \Sigma) 2\mathbf{N}_p$$

$$\text{and } \sqrt{n}\Omega_{222,n} = \Omega_{42,n}.$$

It will be seen later in this article that the proposed inference procedures depend on $\Omega_{222,n}$ and $\Omega_{42,n}$ only through functions that can be written as $(\dot{\psi}_\sigma \otimes \dot{\psi}_\sigma \otimes \dot{\psi}_\sigma)' \text{vec } \Omega_{222,n}$ and $(\dot{\psi}_\sigma \otimes \dot{\psi}_\sigma \otimes \dot{\psi}_\sigma)' \text{vec } \Omega_{42,n}$, where $\dot{\psi}_\sigma$ is a $p^2 \times 1$ vector that satisfies $\dot{\psi}_\sigma' \sigma = 0$. Theorem 1 gives consistent estimators of the required quantities. The results in Theorem 1 follow from the consistency of $\hat{\Omega}_{42,n}$ and $\hat{\Omega}_{222,n}$ described by Boik (2005).

Theorem 1. Suppose that $\dot{\psi}_\sigma$ is a $p^2 \times 1$ vector function of Σ that satisfies $\dot{\psi}_\sigma' \sigma = 0$. Then, the required sixth-order quantities can be consistently estimated as

$$(\dot{\psi}_s \otimes \dot{\psi}_s \otimes \dot{\psi}_s)' \text{vec } \hat{\Omega}_{42,n} =$$

$$n^{-1} \sum_{i=1}^N \left[\dot{\psi}_s' (\tilde{\epsilon}_i \otimes \tilde{\epsilon}_i) \right]^3 - 4\dot{\psi}_s' \hat{\Upsilon}_{21} \dot{\Psi}_s \hat{\Upsilon}_{21}' \dot{\psi}_s$$

$$\text{and } \sqrt{n} (\dot{\psi}_s \otimes \dot{\psi}_s \otimes \dot{\psi}_s)' \text{vec } \hat{\Omega}_{222,n}$$

$$= (\dot{\psi}_s \otimes \dot{\psi}_s \otimes \dot{\psi}_s)' \text{vec } \hat{\Omega}_{42,n}$$

$$- 2\dot{\psi}_s' \hat{\Upsilon}_{21} \dot{\Psi}_s \hat{\Upsilon}_{21}' \dot{\psi}_s,$$

$$\text{where } \hat{\mathbf{\Upsilon}}_{21} = n^{-1} \sum_{i=1}^N (\tilde{\boldsymbol{\varepsilon}}_i \otimes \tilde{\boldsymbol{\varepsilon}}_i \tilde{\boldsymbol{\varepsilon}}_i'),$$

$\dot{\boldsymbol{\psi}}_s$ is a consistent estimator of $\dot{\boldsymbol{\psi}}_\sigma$, $\dot{\boldsymbol{\Psi}}_s$ is a $p \times p$ matrix that satisfies $\text{vec } \dot{\boldsymbol{\Psi}}_s = \dot{\boldsymbol{\psi}}_s$, and $\tilde{\boldsymbol{\varepsilon}}_i$ is defined in (6). If the rows of $\mathbf{Y}$ have a multivariate normal distribution, then the required sixth-order quantities can be consistently estimated as

$$\begin{aligned} & (\dot{\boldsymbol{\psi}}_s \otimes \dot{\boldsymbol{\psi}}_s \otimes \dot{\boldsymbol{\psi}}_s)' \text{vec } \hat{\boldsymbol{\Omega}}_{42,n} \quad (10) \\ &= \sqrt{n} (\dot{\boldsymbol{\psi}}_s \otimes \dot{\boldsymbol{\psi}}_s \otimes \dot{\boldsymbol{\psi}}_s)' \text{vec } \hat{\boldsymbol{\Omega}}_{222,n} \\ &= 8 \text{trace} \left[(\mathbf{S} \dot{\boldsymbol{\Psi}}_s)^3 \right]. \quad \square \end{aligned}$$

Specific structures for $\dot{\boldsymbol{\psi}}_\sigma$ and $\dot{\boldsymbol{\psi}}_s$ are described below.

Derivatives

The main theoretical tool used in this article is the Taylor series expansion of random quantities. The first-order linear expansion is the basis of the delta method. To construct second-order accurate interval estimators, it is necessary to include the quadratic term in the expansion as well. Accordingly, both first and second derivatives of correlation functions are required. Derivatives of partial correlations as well as squared multiple correlations are given in this article.

Partial Correlations

Suppose that an investigator is interested in the correlation between variables i and j , controlling for a subset of variables indexed by $\mathbf{k} = (k_1 \ k_2 \ \cdots \ k_q)'$. This partial correlation, denoted as $\rho_{ij \cdot \mathbf{k}}$, is the off diagonal component in the correlation matrix $\boldsymbol{\Delta}_{ij \cdot \mathbf{k}}$, where

$$\boldsymbol{\Delta}_{ij \cdot \mathbf{k}} = \begin{pmatrix} 1 & \rho_{ij \cdot \mathbf{k}} \\ \rho_{ij \cdot \mathbf{k}} & 1 \end{pmatrix} \quad (11)$$

$$= (\boldsymbol{\Sigma}_{ij \cdot \mathbf{k}})_D^{-\frac{1}{2}} \boldsymbol{\Sigma}_{ij \cdot \mathbf{k}} (\boldsymbol{\Sigma}_{ij \cdot \mathbf{k}})_D^{-\frac{1}{2}}, \text{ where}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{ij \cdot \mathbf{k}} &= \begin{pmatrix} \sigma_{ii \cdot \mathbf{k}} & \sigma_{ij \cdot \mathbf{k}} \\ \sigma_{ij \cdot \mathbf{k}} & \sigma_{jj \cdot \mathbf{k}} \end{pmatrix} \\ &= \mathbf{E}_{ij}' [\boldsymbol{\Sigma} - \boldsymbol{\Sigma} \mathbf{H}_k \boldsymbol{\Sigma}] \mathbf{E}_{ij}, \quad \mathbf{E}_{ij} = (\mathbf{e}_i^p \ \mathbf{e}_j^p), \\ \mathbf{E}_k &= (\mathbf{e}_{k_1}^p \ \mathbf{e}_{k_2}^p \ \cdots \ \mathbf{e}_{k_q}^p), \\ \mathbf{H}_k &= \mathbf{E}_k (\mathbf{E}_k' \boldsymbol{\Sigma} \mathbf{E}_k)^{-1} \mathbf{E}_k', \end{aligned}$$

and the remaining terms are defined in (2). It is required that neither i nor j be a component of $\mathbf{k}$; otherwise $\rho_{ij \cdot \mathbf{k}} = 0$ and no inference is necessary. The parameter $\rho_{ij \cdot \mathbf{k}}$ can be estimated by $r_{ij \cdot \mathbf{k}}$, where $r_{ij \cdot \mathbf{k}}$ is obtained by substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$ in (11).

Theorem 2 gives first and second derivatives of $\rho_{ij \cdot \mathbf{k}}$ with respect to $\boldsymbol{\sigma}$. The derivatives are given without proof because they follow from established matrix calculus results (Graham, 1981).

Theorem 2. *The first two derivatives of $\rho_{ij \cdot \mathbf{k}}$ with respect to $\boldsymbol{\sigma}$ are*

$$\begin{aligned} \dot{\rho}_{ij \cdot \mathbf{k}} &\stackrel{\text{def}}{=} \frac{\partial \rho_{ij \cdot \mathbf{k}}}{\partial \boldsymbol{\sigma}} = \mathbf{N}_p(\boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}}) \\ &- [(\boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{i \cdot \mathbf{k}}) + (\boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}})] \rho_{ij \cdot \mathbf{k}}/2 \text{ and} \\ \ddot{\rho}_{ij \cdot \mathbf{k}} &\stackrel{\text{def}}{=} \frac{\partial^2 \rho_{ij \cdot \mathbf{k}}}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}} \\ &= \mathbf{N}_{p^2}(\boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}}) \rho_{ij \cdot \mathbf{k}}/2 \\ &+ [(\boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{i \cdot \mathbf{k}}) \\ &+ (\boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}})] 3\rho_{ij \cdot \mathbf{k}}/4 \\ &- (\mathbf{N}_p \otimes \mathbf{N}_p) \mathbf{N}_{p^2} [(\boldsymbol{\gamma}_{i \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}} \otimes \boldsymbol{\gamma}_{j \cdot \mathbf{k}}) \end{aligned}$$

$$+ (\gamma_{j \cdot \mathbf{k}} \otimes \gamma_{i \cdot \mathbf{k}} \otimes \gamma_{i \cdot \mathbf{k}} \otimes \gamma_{i \cdot \mathbf{k}}) \Big] \\ -2 (\mathbf{N}_p \otimes \mathbf{N}_p) (\mathbf{I}_p \otimes \mathbf{h}_k \otimes \mathbf{I}_p) \dot{\rho}_{ij \cdot \mathbf{k}}, \text{ where}$$

$$\gamma_{i \cdot \mathbf{k}} = [\mathbf{I}_p - \mathbf{H}_k \boldsymbol{\Sigma}] \mathbf{e}_i^p / \sqrt{\sigma_{ii \cdot \mathbf{k}}},$$

$$\mathbf{h}_k = \text{vec}(\mathbf{H}_k),$$

$\mathbf{H}_k$ is defined in (11), $\mathbf{N}_p$ is defined in (5), and $\sigma_{ii \cdot \mathbf{k}}$ is defined in (11). If $\mathbf{k}$ is empty, then $\rho_{ij \cdot \mathbf{k}}$ becomes the simple correlation ρ_{ij} and $\gamma_{i \cdot \mathbf{k}}$ becomes $\gamma_i = \mathbf{e}_i^p / \sqrt{\sigma_{ii}}$. $\square$

Multiple Correlations

Suppose that an investigator has interest in the multiple correlation between variable i and variables indexed by $\mathbf{k} = (k_1 \ k_2 \ \dots \ k_q)'$. The squared multiple correlation, denoted as $\rho_{i(\mathbf{k})}^2$, can be written as

$$\rho_{i(\mathbf{k})}^2 = \mathbf{e}_i^{p'} \boldsymbol{\Sigma} \mathbf{H}_k \boldsymbol{\Sigma} \mathbf{e}_i^p / \sigma_{ii}, \quad (12)$$

where $\mathbf{e}_i^p$ is defined in (2), and $\mathbf{H}_k$ is defined in (11). It is required that i not be a component of $\mathbf{k}$; otherwise $\rho_{i(\mathbf{k})}^2 = 1$ and no inference is necessary. The parameter $\rho_{i(\mathbf{k})}^2$ can be estimated by $r_{i(\mathbf{k})}^2$, where $r_{i(\mathbf{k})}^2$ is obtained by substituting $\mathbf{S}$ for $\boldsymbol{\Sigma}$ in (12).

Theorem 3 gives first and second derivatives of $\rho_{i(\mathbf{k})}^2$ with respect to $\boldsymbol{\sigma}$. The derivatives are given without proof because they follow from established matrix calculus results (Graham, 1981).

Theorem 3. *The first two derivatives of $\rho_{i(\mathbf{k})}^2$ with respect to $\boldsymbol{\sigma}$ are*

$$\dot{\rho}_{i(\mathbf{k})}^2 \stackrel{\text{def}}{=} \frac{\partial \rho_{i(\mathbf{k})}^2}{\partial \boldsymbol{\sigma}} = (\gamma_i \otimes \gamma_i) (1 - \rho_{i(\mathbf{k})}^2) \\ - (\gamma_{i(\mathbf{k})} \otimes \gamma_{i(\mathbf{k})}) \text{ and} \\ \ddot{\rho}_{i(\mathbf{k})}^2 \stackrel{\text{def}}{=} \frac{\partial^2 \rho_{i(\mathbf{k})}^2}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}}$$

$$= 2\mathbf{N}_{p^2} (\gamma_{i(\mathbf{k})} \otimes \gamma_{i(\mathbf{k})} \otimes \gamma_i \otimes \gamma_i)$$

$$+ 2 (\mathbf{N}_p \otimes \mathbf{N}_p) (\gamma_{i(\mathbf{k})} \otimes \mathbf{h}_k \otimes \gamma_{i(\mathbf{k})})$$

$$- 2 (\gamma_i \otimes \gamma_i \otimes \gamma_i \otimes \gamma_i) (1 - \rho_{i(\mathbf{k})}^2),$$

$$\text{where } \gamma_i = \mathbf{e}_i^p / \sqrt{\sigma_{ii}},$$

$$\gamma_{i(\mathbf{k})} = (\mathbf{I}_p - \mathbf{H}_k \boldsymbol{\Sigma}) \mathbf{e}_i^p / \sqrt{\sigma_{ii}},$$

$\mathbf{H}_k$ is defined in (11), and $\mathbf{h}_k$ is defined in Theorem 2. $\square$

Derivatives of Differences

In practice, a linear function of correlation coefficients rather than a correlation coefficient itself could be of interest. Let ψ be a linear function of simple, partial, and/or squared multiple correlation coefficients. For example, if the difference between a simple and a partial correlation is of interest, then $\psi = \rho_{ij} - \rho_{ij \cdot \mathbf{k}}$, where $\mathbf{k}$ is a non-empty vector of index values. By properties of derivatives, the derivative of a linear function is the linear function of the derivatives. Accordingly, Theorems 2 and 3 can be used to obtain the first two derivatives of any linear function, ψ , with respect to $\boldsymbol{\sigma}$.

Denote the first two derivatives of ψ with respect to $\boldsymbol{\sigma}$ by $\dot{\psi}_{\boldsymbol{\sigma}}$ and $\ddot{\psi}_{\boldsymbol{\sigma}}$. That is,

$$\dot{\psi}_{\boldsymbol{\sigma}} \stackrel{\text{def}}{=} \frac{\partial \psi}{\partial \boldsymbol{\sigma}} \text{ and } \ddot{\psi}_{\boldsymbol{\sigma}} \stackrel{\text{def}}{=} \frac{\partial^2 \psi}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}}, \quad (13)$$

where ψ is a linear function of simple, partial, and/or squared multiple correlation coefficients. The specific structure of these derivatives for individual correlation coefficients as well as for linear functions that correspond to extensions of the first four models in Olkin and Finn (1995) are listed below. The fifth model in Olkin and Finn is discussed separately later in this article.

1. Single Partial Correlation Coefficient. If $\psi = \rho_{ij \cdot \mathbf{k}}$, then $\dot{\psi}_\sigma = \dot{\rho}_{ij \cdot \mathbf{k}}$, and $\ddot{\psi}_\sigma = \ddot{\rho}_{ij \cdot \mathbf{k}}$, where the derivatives are given in Theorem 2. If $\mathbf{k}$ is empty, then ψ becomes $\psi = \rho_{ij}$.
2. Difference Between Partial Correlation Coefficients. Denote the set that consists of the components of $\mathbf{k}$ by $\{\mathbf{k}\}$. If $\psi = \rho_{ij \cdot \mathbf{k}_1} - \rho_{tu \cdot \mathbf{k}_0}$, then $\dot{\psi}_\sigma = \dot{\rho}_{ij \cdot \mathbf{k}_1} - \dot{\rho}_{tu \cdot \mathbf{k}_0}$, and $\ddot{\psi}_\sigma = \ddot{\rho}_{ij \cdot \mathbf{k}_1} - \ddot{\rho}_{tu \cdot \mathbf{k}_0}$, where $\mathbf{k}_1$ and/or $\mathbf{k}_0$ could be empty, $\{i, j\} = \{t, u\} \implies \{\mathbf{k}_0\} \neq \{\mathbf{k}_1\}$, and the derivatives are given in Theorem 2. This difference is an extension of Olkin and Finn's Models C and D. Examples include $\psi = \rho_{ij \cdot \mathbf{k}_1} - \rho_{ij \cdot \mathbf{k}_0}$, $\psi = \rho_{ij} - \rho_{ij \cdot \mathbf{k}}$, $\psi = \rho_{ij \cdot \mathbf{k}} - \rho_{tu \cdot \mathbf{k}}$, and $\psi = \rho_{ij} - \rho_{tu}$.
3. Single Squared Multiple Correlation Coefficient. If $\psi = \rho_{i(\mathbf{k})}^2$, then $\dot{\psi}_\sigma = \dot{\rho}_{i(\mathbf{k})}^2$, and $\ddot{\psi}_\sigma = \ddot{\rho}_{i(\mathbf{k})}^2$, where the derivatives are given in Theorem 3.
4. Difference Between Squared Multiple Correlation Coefficients. If $\psi = \rho_{i(\mathbf{k}_1)}^2 - \rho_{j(\mathbf{k}_0)}^2$, then $\dot{\psi}_\sigma = \dot{\rho}_{i(\mathbf{k}_1)}^2 - \dot{\rho}_{j(\mathbf{k}_0)}^2$ and $\ddot{\psi}_\sigma = \ddot{\rho}_{i(\mathbf{k}_1)}^2 - \ddot{\rho}_{j(\mathbf{k}_0)}^2$, where $i = j \implies \{\mathbf{k}_0\} \neq \{\mathbf{k}_1\}$, and the derivatives are given in Theorem 3. This difference is an extension Olkin and Finn's Models A and B.

Second-Order Accurate Confidence Intervals

Let ψ be a linear function of simple, partial, or squared multiple correlation coefficients and denote the estimator obtained by substituting $\mathbf{S}$ for Σ by $\hat{\psi}$. Define σ_ψ^2 as $\sigma_\psi^2 \stackrel{\text{def}}{=} n \text{Var}(\hat{\psi})$. An application of the delta method reveals that

$$\sigma_\psi^2 = \dot{\psi}'_\sigma \Omega_{22,n} \dot{\psi}_\sigma + O(n^{-1}), \quad (14)$$

where $\dot{\psi}_\sigma$ is defined in (13) and $\Omega_{22,n}$ is defined in (5). The central limit theorem ensures that

$$P(\hat{\psi} \leq u) = \Phi\left(\frac{\sqrt{n}(u - \psi)}{\sigma_\psi}\right) + O(n^{-\frac{1}{2}}),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

The proposed confidence intervals are based on the asymptotically pivotal quantity

$$T = \frac{\sqrt{n}(\hat{\psi} - \psi)}{\hat{\sigma}_\psi}, \quad (15)$$

where $\hat{\sigma}_\psi^2$ is a consistent estimator of σ_ψ^2 . The quantity, $\dot{\psi}'_s \hat{\Omega}_{22,n} \dot{\psi}_s$, for example, is consistent for σ_ψ^2 , where $\hat{\Omega}_{22,n}$ is defined in (6) and $\dot{\psi}_s$ is $\dot{\psi}_\sigma$ after substituting $\mathbf{S}$ for Σ . It follows from the central limit Theorem and Slutsky's Theorem (Casella & Berger, 2002, Theorem 5.5.17) that $T \sim N(0, 1)$ to first-order accuracy. Inverting the cdf of T yields first-order accurate one-sided lower and upper intervals:

$$(\hat{\psi} - \hat{\sigma}_\psi z_{1-\alpha}/\sqrt{n}, \infty) \text{ and } (-\infty, \hat{\psi} - \hat{\sigma}_\psi z_\alpha/\sqrt{n})$$

respectively, where z_α is the 100 α percentile of the standard normal distribution. The standard normal critical value z_α can be replaced by the t critical value, $t_{\alpha,n}$, without affecting the asymptotic properties of the intervals.

Edgeworth Expansion of the Distribution of T

To construct a second-order accurate approximation to the distribution of T , it is necessary to obtain the mean (bias), variance, and skewness of T . These quantities and estimators of these quantities are summarized in Theorem 4. A proof is sketched in the Appendix.

Theorem 4. *If ψ is a function of a single covariance matrix, Σ , and $\hat{\psi}$ is the same function of a single sample covariance matrix, $\mathbf{S}$, then the first three cumulants of T in (15) can be written as follows:*

$$E(T) = \frac{\kappa_1}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right),$$

$$\text{Var}(T) = 1 + O\left(n^{-1}\right),$$

$$\text{and } E\left[\left(T - \frac{\kappa_1}{\sqrt{n}}\right)^3\right] = \frac{\kappa_3}{\sqrt{n}} + O\left(n^{-\frac{3}{2}}\right),$$

$$\text{where } \kappa_1 = \frac{m_1}{\sigma_\psi} - \frac{m_{11}}{2\sigma_\psi^3},$$

$$\kappa_3 = \frac{m_3}{\sigma_\psi^3} - 3\frac{m_{11}}{\sigma_\psi^3},$$

$$\frac{m_1}{\sqrt{n}} = E\left[\sqrt{n}\left(\hat{\psi} - \psi\right)\right] + O\left(n^{-\frac{3}{2}}\right),$$

$$\frac{m_3}{\sqrt{n}}$$

$$= E\left\{\left[\sqrt{n}\left(\hat{\psi} - \psi\right) - \frac{m_1}{\sqrt{n}}\right]^3\right\} + O\left(n^{-\frac{3}{2}}\right),$$

$$\text{and } m_{11}$$

$$= n \text{Cov}\left(\hat{\psi} - \psi, \hat{\sigma}_\psi^2 - \sigma_\psi^2\right) + O\left(n^{-1}\right).$$

Specific solutions for m_1 , m_3 , and m_{11} are

$$m_1 = \frac{1}{2}\ddot{\psi}'_\sigma \text{vec } \Omega_{22,n},$$

$$m_3 = \sqrt{n}\left(\dot{\psi}_\sigma \otimes \dot{\psi}_\sigma \otimes \dot{\psi}_\sigma\right)' \text{vec } \Omega_{222,n}$$

$$+ 3\left(\dot{\psi}'_\sigma \Omega_{22,n} \otimes \dot{\psi}'_\sigma \Omega_{22,n}\right) \ddot{\psi}_\sigma, \text{ and}$$

$$m_{11} = \left(\dot{\psi}_\sigma \otimes \dot{\psi}_\sigma \otimes \dot{\psi}_\sigma\right)' \text{vec } \Omega_{42,n}$$

$$+ 2\left(\dot{\psi}'_\sigma \Omega_{22,n} \otimes \dot{\psi}'_\sigma \Omega_{22,n}\right) \ddot{\psi}_\sigma,$$

where $\Omega_{42,n}$ and $\Omega_{222,n}$ are described in (7) and (8), and $\dot{\psi}_\sigma$ and $\ddot{\psi}_\sigma$ are defined in (13). Furthermore, $\dot{\psi}'_\sigma \sigma = \mathbf{0}$ as required in Theorem 1 and κ_1 and κ_2 can be estimated consistently by

$$\hat{\kappa}_1 = \frac{\hat{m}_1}{\hat{\sigma}_\psi} - \frac{\hat{m}_{11}}{2\hat{\sigma}_\psi^3}, \text{ and } \hat{\kappa}_3 = \frac{\hat{m}_3}{\hat{\sigma}_\psi^3} - 3\frac{\hat{m}_{11}}{\hat{\sigma}_\psi^3},$$

$$\text{where } \hat{\sigma}_\psi^2 = \dot{\psi}'_s \hat{\Omega}_{22,n} \dot{\psi}_s + \frac{U}{n},$$

$$\hat{m}_1 = \frac{1}{2}\ddot{\psi}'_s \text{vec } \hat{\Omega}_{22,n},$$

$$\hat{m}_3 = n^{-1} \sum_{i=1}^N \left[\dot{\psi}'_s (\tilde{\varepsilon}_i \otimes \tilde{\varepsilon}_i) \right]^3$$

$$- 6\dot{\psi}'_s \hat{\Upsilon}_{21} \ddot{\psi}_s \hat{\Upsilon}'_{21} \dot{\psi}_s$$

$$+ 3\left(\dot{\psi}'_s \hat{\Omega}_{22,n} \otimes \dot{\psi}'_s \hat{\Omega}_{22,n}\right) \ddot{\psi}_s,$$

$$\hat{m}_{11} = n^{-1} \sum_{i=1}^N \left[\dot{\psi}'_s (\tilde{\varepsilon}_i \otimes \tilde{\varepsilon}_i) \right]^3$$

$$- 4\dot{\psi}'_s \hat{\Upsilon}_{21} \ddot{\psi}_s \hat{\Upsilon}'_{21} \dot{\psi}_s$$

$$+ 2\left(\dot{\psi}'_s \hat{\Omega}_{22,n} \otimes \dot{\psi}'_s \hat{\Omega}_{22,n}\right) \ddot{\psi}_s,$$

$\dot{\psi}_s$ and $\ddot{\psi}_s$ are $\dot{\psi}_\sigma$ and $\ddot{\psi}_\sigma$ after substituting $\mathbf{S}$ for Σ , U is any $O_p(1)$ random variable, and the remaining terms are defined in Theorem 1. $\square$

Asymptotically, the choice of U in Theorem 4 is arbitrary, because $\lim_{n \rightarrow 0} U/n = 0$ and the asymptotic properties of the intervals are not affected by U/n . Nonetheless, the finite-sample properties of the intervals are affected by the choice of U . From experience, it appears that $\dot{\psi}'_s \hat{\Omega}_{22,n} \dot{\psi}_s$ often

underestimates the variance of $\sqrt{n}(\hat{\psi} - \psi)$. Accordingly, choosing U to be a positive quantity might be beneficial. In this article, the quantity U will be chosen by using the second order Taylor series expansion of $\hat{\psi}$ under the condition that $\dot{\psi}_{\sigma} = \mathbf{0}$. Specifically, the second-order Taylor series expansion of $\hat{\psi}$ is

$$\begin{aligned} \sqrt{n}(\hat{\psi} - \psi) &= \dot{\psi}'_{\sigma} \sqrt{n}(\mathbf{s} - \sigma) \\ &+ \frac{1}{2\sqrt{n}} \ddot{\psi}'_{\sigma} [\sqrt{n}(\mathbf{s} - \sigma) \otimes \sqrt{n}(\mathbf{s} - \sigma)] \\ &+ O_p(n^{-1}). \end{aligned}$$

If $\dot{\psi}_{\sigma} = \mathbf{0}$, then the variance of $\sqrt{n}(\hat{\psi} - \psi)$ is

$$\begin{aligned} \text{Var} [\sqrt{n}(\hat{\psi} - \psi)] \\ = \frac{1}{2n} \ddot{\psi}'_{\sigma} (\Omega_{22,n} \otimes \Omega_{22,n}) \ddot{\psi}_{\sigma} + O(n^{-2}). \end{aligned}$$

Accordingly, U will be chosen as $U = \ddot{\psi}'_{\sigma} (\hat{\Omega}_{22,n} \otimes \hat{\Omega}_{22,n}) \ddot{\psi}_{\sigma} / 2$. For this choice, the quantity U/n represents an estimate of one of the two leading terms in the $O(n^{-1})$ component of (14).

The primary device for constructing the interval estimator is a second-order Edgeworth expansion of the distribution of T . This expansion is summarized in Theorem 5. A proof is sketched in the Appendix.

Theorem 5. *The probability density function (pdf) and the cdf of T in (15) can be written as*

$$\begin{aligned} f_T(t) &= \\ \varphi(t) \left[1 + \frac{\kappa_1 t}{\sqrt{n}} + \frac{\kappa_3 (t^3 - 3t)}{6\sqrt{n}} \right] &+ O(n^{-1}) \end{aligned}$$

$$\text{and } F_T(t) \stackrel{\text{def}}{=} P(T \leq t) = \Phi(t)$$

$$- \varphi(t) \left[\frac{\kappa_1}{\sqrt{n}} + \frac{\kappa_3 (t^2 - 1)}{6\sqrt{n}} \right] + O(n^{-1})$$

respectively, where $\kappa_1/\sqrt{n}$ and $\kappa_3/\sqrt{n}$ are the mean and skewness of T (see Theorem 4), and $\varphi(t)$ and $\Phi(t)$ are the standard normal pdf and cdf. $\square$

Proposed Interval Estimators

The expansion of F_T in Theorem 5 can be used to express the percentiles of T in terms of those of a standard normal random variable and to obtain a polynomial transformation of T that has distribution $N(0, 1)$ up to $O(n^{-1})$. These results, known as Cornish-Fisher expansions (Pace and Salvan, 1997, §10.6), are summarized in Corollary 5.1.

Corollary 5.1. *Denote the 100α percentile of T by t_{α} . Then,*

$$t_{\alpha} = z_{\alpha} + \frac{\kappa_1}{\sqrt{n}} + \frac{\kappa_3(z_{\alpha}^2 - 1)}{6\sqrt{n}} + O(n^{-1}) \text{ and}$$

$$t_{\alpha} = \hat{t}_{\alpha} + O_p(n^{-1}),$$

where z_{α} is the 100α percentile of the $N(0, 1)$ distribution,

$$\hat{t}_{\alpha} = z_{\alpha} + \hat{\kappa}_1/\sqrt{n} + \hat{\kappa}_3(z_{\alpha}^2 - 1)/(6\sqrt{n}),$$

and $\hat{\kappa}_1$ and $\hat{\kappa}_3$ are consistent estimators of κ_1 and κ_3 in Theorem 4. The normal critical value, z_{α} , can be replaced by the t critical value, $t_{\alpha,n}$, without affecting the second-order property of $\hat{t}_{\alpha}$. Also, define T_1 and $\hat{T}_1$ as

$$T_1 \stackrel{\text{def}}{=} T - \frac{\kappa_1}{\sqrt{n}} - \frac{\kappa_3(T^2 - 1)}{6\sqrt{n}} \text{ and}$$

$$\hat{T}_1 \stackrel{\text{def}}{=} T - \frac{\hat{\kappa}_1}{\sqrt{n}} - \frac{\hat{\kappa}_3(T^2 - 1)}{6\sqrt{n}}.$$

Then, $P(T_1 \leq t) = \Phi(t) + O(n^{-1})$ and $P(\hat{T}_1 \leq t) = \Phi(t) + O(n^{-1})$. $\square$

A straightforward application of Corollary 5.1 yields second-order accurate intervals for ψ . Note that $\hat{t}_\alpha$ in Corollary 5.1 is a second-order accurate percentile of the distribution of T . Accordingly, $(\hat{\psi} - \hat{t}_{1-\alpha}\hat{\sigma}_\psi/\sqrt{n}, \infty)$ is a second-order accurate $100(1-\alpha)\%$ lower confidence interval for ψ and $(-\infty, \hat{\psi} - \hat{t}_\alpha\hat{\sigma}_\psi/\sqrt{n})$ is a second-order accurate $100(1-\alpha)\%$ upper confidence interval for ψ . One drawback of these intervals, however, is that $\hat{t}_\alpha$ is a quadratic function of z_α and, therefore, is not monotonic in α . The same issue exists if $\hat{T}_1$ is used as an asymptotic pivotal quantity. It is not monotone in T and, therefore, its cdf cannot be inverted for arbitrary α . Methods for correcting this non-monotone deficiency are summarized in Corollaries 5.2 and 5.3. The method in Corollary 5.2 is based on Hall's (1992) cubic transformation, whereas the method in Corollary 5.3 is based on the exponential transformation described by Boik (2006).

Corollary 5.2. Define $\hat{T}_2$ as

$$\hat{T}_2 \stackrel{\text{def}}{=} T - \frac{\hat{\kappa}_1}{\sqrt{n}} - \frac{\hat{\kappa}_3(T^2 - 1)}{6\sqrt{n}} + \frac{\hat{\kappa}_3^2 T^3}{108n}.$$

Then, $\hat{T}_2 = \hat{T}_1 + O_p(n^{-1})$ and $\hat{T}_2$ is monotone in T . Inverting the cdf of $\hat{T}_2$ yields the following lower and upper $100(1-\alpha)\%$ second-order accurate confidence intervals:

$$\left(\hat{\psi} - \frac{\hat{\sigma}_\psi \hat{t}_{2,1-\alpha}}{\sqrt{n}}, \infty \right) \text{ and}$$

$$\left(-\infty, \hat{\psi} - \frac{\hat{\sigma}_\psi \hat{t}_{2,\alpha}}{\sqrt{n}} \right),$$

$$\text{where } \hat{t}_{2,\alpha} = \frac{\hat{\kappa}_1}{\sqrt{n}} + \frac{6\sqrt{n}}{\hat{\kappa}_3} \times$$

$$\left\{ 1 - \left[1 + \frac{\hat{\kappa}_3}{2\sqrt{n}} \left(\frac{\hat{\kappa}_3}{6\sqrt{n}} - z_\alpha \right) \right]^{\frac{1}{3}} \right\}. \quad \square$$

Corollary 5.3. Define $\hat{T}_3$ as

$$\hat{T}_3 \stackrel{\text{def}}{=} T - \frac{\hat{\kappa}_1}{\sqrt{n}} - \frac{\hat{\kappa}_3 \left(T^2 e^{-\frac{\hat{\kappa}_3 T^2}{2}} - 1 \right)}{6\sqrt{n}}, \text{ where}$$

$$\hat{d} = \frac{\hat{\kappa}_3^2 (31 - 7\sqrt{17})}{72n} e^{-\frac{(5-\sqrt{17})}{2}}.$$

Then, $\hat{T}_3 = \hat{T}_1 + O_p(n^{-3/2})$, and is monotone in T . Inverting the cdf of $\hat{T}_3$ yields the following lower and upper $100(1-\alpha)\%$ second-order accurate confidence intervals:

$$\left(\hat{\psi} - \frac{\hat{\sigma}_\psi \hat{t}_{3,1-\alpha}}{\sqrt{n}}, \infty \right) \text{ and}$$

$$\left(-\infty, \hat{\psi} - \frac{\hat{\sigma}_\psi \hat{t}_{3,\alpha}}{\sqrt{n}} \right),$$

where $\hat{t}_{3,\alpha}$ is the solution to $\hat{T}_3 = z_\alpha$ or $\hat{T}_3 = t_{\alpha,n}$ for T . The solution can be computed using the modified Newton method described in the Appendix. $\square$

The intervals described in Corollaries 5.2 and 5.3 require consistent estimators of κ_1 and κ_3 , the mean and skewness of T . If ψ is a function of a single covariance matrix, then the estimators described in Theorem 4 can be used. In some situations, investigators are interested in comparing correlation functions based on two covariance matrices. For example, if a comparison between squared multiple correlation coefficients from two populations is of interest, then one could define ψ as $\psi = \rho_{i(\mathbf{k})}^2(\Sigma_1) - \rho_{i(\mathbf{k})}^2(\Sigma_2)$, where $\rho_{i(\mathbf{k})}^2(\Sigma_u)$ is $\rho_{i(\mathbf{k})}^2$ computed on Σ_u , for $u = 1, 2$. Of course, the comparison need not be restricted to squared multiple correlations. One could, for example, define ψ as $\psi = \rho_{ij \cdot \mathbf{k}}(\Sigma_1) - \rho_{ij \cdot \mathbf{k}}(\Sigma_2)$ or $\psi = \rho_{ij \cdot \mathbf{k}}(\Sigma_1) - \rho_{tu \cdot \mathbf{k}_0}(\Sigma_2)$. These comparisons are extensions of Olkin and Finn's (1995) Model E. The mean and skewness

as well as estimators of the mean and skewness of T for comparisons such as these are given in Theorem 6. A proof is sketched in the Appendix.

Theorem 6. *Let ψ_i be a correlation function computed on $\mathbf{\Sigma}_i$ for $i = 1, 2$ and let $\hat{\psi}_i$ be ψ_i computed on $\mathbf{S}_i$, where $\mathbf{S}_i$ is a sample covariance matrix based on n_i degrees of freedom from sample i , and samples 1 and 2 are independent. Define T as*

$$T = \frac{\sqrt{n_1 n_2} (\hat{\psi} - \psi)}{\sqrt{\hat{\delta}}}, \text{ where}$$

$$\hat{\psi} = \hat{\psi}_1 - \hat{\psi}_2, \quad \psi = \psi_1 - \psi_2,$$

$\hat{\delta} = n_2 \hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2$, and $\hat{\sigma}_i^2$, $\hat{\psi}_{s_i}$, and $\hat{\psi}_{s_i}$ are $\hat{\sigma}_\psi^2$, $\hat{\psi}_s$, and $\hat{\psi}_s$ based on sample i . Let $m_{1;i}$, $m_{3;i}$, and $m_{11;i}$ be m_1 , m_3 , and m_{11} of Theorem 4 for population i . Then, the first three cumulants of T are

$$E(T) = \frac{\kappa_1^*}{\sqrt{\delta}} + O\left(n^{-\frac{3}{2}}\right),$$

$$\text{Var}(T) = 1 + O\left(n^{-1}\right), \text{ and}$$

$$E[T - E(T)]^3 = \frac{\kappa_3^*}{\sqrt{\delta}} + O\left(n^{-\frac{3}{2}}\right),$$

$$\text{where } \delta = n_2 \sigma_1^2 + n_1 \sigma_2^2,$$

$$\kappa_1^* = \frac{n_2 m_{1;1} - n_1 m_{1;2}}{\sqrt{n_1 n_2}} - \frac{n_2^{3/2} m_{11;1}}{2\delta \sqrt{n_1}}$$

$$+ \frac{n_1^{3/2} m_{11;2}}{2\delta \sqrt{n_2}},$$

$$\kappa_3^* = \frac{n_2^{3/2} m_{3;1}}{\delta \sqrt{n_1}} - \frac{n_1^{3/2} m_{3;2}}{\delta \sqrt{n_2}} - 3 \frac{n_2^{5/2} \sigma_1^2 m_{11;1}}{\delta^2 \sqrt{n_1}}$$

$$+ 3 \frac{n_1^{5/2} \sigma_2^2 m_{11;2}}{\delta^2 \sqrt{n_2}},$$

and $n = \min(n_1, n_2)$. Furthermore, consistent estimators of κ_1^* and κ_3^* are obtained by substituting $\hat{\sigma}_i^2$, $\hat{m}_{1;i}$, $\hat{m}_{3;i}$, and $\hat{m}_{11;i}$ for σ_i^2 , $m_{1;i}$, $m_{3;i}$, and $m_{11;i}$, respectively. Estimators of $m_{1;i}$, $m_{3;i}$, and $m_{11;i}$ are given by $\hat{m}_1$, $\hat{m}_3$, and $\hat{m}_{11}$ in Theorem 4 computed on sample i . $\square$

Second-order accurate confidence intervals for ψ in Theorem 6 can be computed by substituting $\hat{\kappa}_1^*/\sqrt{\hat{\delta}}$ and $\hat{\kappa}_3^*/\sqrt{\hat{\delta}}$ for $\kappa_1/\sqrt{n}$ and $\kappa_3/\sqrt{n}$ in Corollaries 5.2 and 5.3.

Simulation Studies

The finite sample properties of the proposed intervals were evaluated by computing the intervals on samples from normal, normal mixture, and affine lognormal distributions. Let $\mathbf{C}$ be a $p \times p$ nonsingular matrix and denote a randomly selected p -vector from the $N(\mathbf{0}, \mathbf{I}_p)$ distribution as $\mathbf{z}$. Denote a randomly selected p -vector from a normal, normal mixture, or affine lognormal distribution with mean $\mathbf{0}$ and covariance $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}'$ as $\mathbf{y}$. Then $\mathbf{z}$ can be transformed to $\mathbf{y}$ by

$$\mathbf{y} = \mathbf{C}\mathbf{z}, \quad \mathbf{y} = \mathbf{C}\mathbf{z} \frac{[U + k(1 - U)]}{\sqrt{\theta + k^2(1 - \theta)}}, \text{ and}$$

$$\mathbf{y} = \mathbf{C} \left(e^{\odot \mathbf{z}m} - \mathbf{1}_p e^{m^2/2} \right) \frac{1}{\sqrt{e^{m^2} (e^{m^2} - 1)}}$$

for the normal, normal mixture, and affine lognormal distributions, respectively; where U is a Bernoulli(θ) random variable distributed independently of $\mathbf{z}$, k and m are scalar constants, and $e^{\odot \mathbf{z}m}$ is a p -vector whose i^{th} component is $e^{z_i m}$.

The parameter values for the normal mixture were $\theta = 0.7$ and $k = 3$. For this choice, marginal kurtosis is 3.488 and multivariate kurtosis (Bilodeau and Brenner, 1999, §13.2) is 1.1626. The normal

mixture is a member of the class of elliptically contoured distributions. Accordingly, asymptotic standard errors based on normal theory are smaller than the correct standard errors by a factor of $1/\sqrt{2.1626}$ and coverage of one-sided nominal $100(1 - \alpha)\%$ normal theory intervals converges to $\Phi(z_{1-\alpha}/\sqrt{2.1626})$, where $z_{1-\alpha}$ is the $100(1 - \alpha)$ percentile of the standard normal distribution. For $\alpha = 0.05$, coverage of normal theory intervals converges to 0.868 rather than 0.95.

The parameter value for the affine lognormal distribution was $m = 0.8326$. For this choice, the marginal skewness and kurtosis for each component of the lognormal random vector $e^{\odot \mathbf{z}m}$ is 4.00 and 38. This distribution is not a member of the class of elliptically contoured distributions.

Generation of Covariance Matrices

To examine the performance of intervals for $\psi = \rho_{ij} - \rho_{ij\cdot\mathbf{k}}$, ten 4-dimensional covariance matrices were constructed such that $\rho_{ij} \in \mathcal{L}_1$, $\rho_{ij\cdot\mathbf{k}} \in \mathcal{L}_1$, and $\rho_{ij} \leq \rho_{ij\cdot\mathbf{k}}$, where $\mathcal{L}_1 = \{\sqrt{1/6}, \sqrt{1/3}, \sqrt{2/3}, \sqrt{5/6}\}$. The covariance matrices were constructed as follows:

$$\Sigma = \mathbf{C}\mathbf{C}', \text{ where } \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & v & w & 1 \end{pmatrix},$$

$$w = \frac{-1 \pm \sqrt{\rho_{ij\cdot\mathbf{k}}^2 (1 - \rho_{ij\cdot\mathbf{k}}^2)}}{1 - 2\rho_{ij\cdot\mathbf{k}}^2},$$

$$v = \frac{-(2 + w) \pm 2\sqrt{h}}{1 - 4\rho_{ij}^2}$$

$$h = 2\rho_{ij}^2(3 + 2w + w^2 - 4\rho_{ij}^2 - 2\rho_{ij}^2 w^2),$$

$$i = 3, j = 4, \text{ and } \mathbf{k} = (1 \ 2)'$$

To examine the performance of intervals for $\psi = \rho_{i(\mathbf{k})}^2 - \rho_{j(\mathbf{k})}^2$, ten 4-dimensional

covariance matrices were constructed such that $\rho_{i(\mathbf{k})}^2 \in \mathcal{L}_2$, $\rho_{j(\mathbf{k})}^2 \in \mathcal{L}_2$, and $\rho_{j(\mathbf{k})}^2 \leq \rho_{i(\mathbf{k})}^2$, where $\mathcal{L}_2 = \{0.2, 0.4, 0.6, 0.8\}$. The covariance matrices were constructed as follows:

$$\Sigma = \mathbf{C}\mathbf{C}', \text{ where } \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ v & v & 1 & 0 \\ w & w & 1 & 1 \end{pmatrix},$$

$$v = \sqrt{\frac{\rho_{i(\mathbf{k})}^2}{2(1 - \rho_{i(\mathbf{k})}^2)}}, \quad w = \sqrt{\frac{\rho_{j(\mathbf{k})}^2}{1 - \rho_{j(\mathbf{k})}^2}},$$

$$i = 3, j = 4, \text{ and } \mathbf{k} = (1 \ 2)'$$

For each covariance matrix, 5,000 realizations of the $N \times 4$ matrix $\mathbf{Y}$ were sampled from each of the three distributions (normal, normal mixture, affine lognormal) and one-sided nominal 95% confidence intervals for ψ were computed. The simulation study was repeated for sample sizes $N \in \{25, 50, 100, 200\}$.

Results

Each of Figures 1–3 displays four panels of 16 plots each. The upper two panels display the coverage of normal theory and ADF confidence intervals for $\psi = \rho_{ij} - \rho_{ij\cdot\mathbf{k}}$. Coverage of one-sided intervals for $\psi = \rho_{i(\mathbf{k})}^2 - \rho_{j(\mathbf{k})}^2$ is displayed in the lower two panels. The 10 plots in the upper triangle of each panel display the coverage of upper one-sided intervals and the 10 plots in the lower triangle display the coverage of lower one-sided intervals. The four plots on the diagonal in each panel display coverage for both lower and upper intervals.

In the upper two panels, row and column labels correspond to ρ_{ij}^2 and $\rho_{ij\cdot\mathbf{k}}^2$, respectively, for upper intervals and to $\rho_{ij\cdot\mathbf{k}}^2$ and ρ_{ij}^2 , respectively, for lower intervals. In the lower two panels, row and column labels correspond to $\rho_{j(\mathbf{k})}^2$ and $\rho_{i(\mathbf{k})}^2$, respectively, for upper intervals and to $\rho_{i(\mathbf{k})}^2$ and

$\rho_{j(\mathbf{k})}^2$, respectively, for lower intervals. In all panels, plot symbols are assigned as follows:

Interval	Symbol
First-Order Lower	$\times$
First-Order Upper	$*$
Second-Order Lower	$\square$
Second-Order Upper	$\circ$

Second-order intervals were based on the exponential method described in Corollary 5.3. Within each plot, the coverage of intervals based on sample sizes of 25, 50, 100, and 200 is plotted from left to right. Figures 2 and 3 display coverage of intervals when sampling from normal mixtures and from affine lognormal distributions.

Figure 1: Coverage when Sampling from Normal Distributions

Figure 2: Coverage when Sampling from Normal Mixture Distributions

It is apparent in Figure 1 that the coverage of second-order intervals is superior to that of first-order intervals when sampling from multivariate normal distributions. Furthermore, the second-order ADF intervals perform nearly as well as do the second-order intervals that are explicitly based on normal theory.

Figure 2 confirms that normal theory based intervals perform poorly when sampling from normal mixtures. As expected from theory, coverage approaches 0.86 as N increases in each plot. The ADF intervals fare much better. Also, the coverage of second-order accurate ADF intervals is equal to or superior to that of first-order

Figure 3: Coverage when Sampling from Affine Lognormal Distributions

accurate ADF intervals whenever the coverage of first-order accurate intervals departs substantially from 0.95. See, for example, plot (4, 1) in the upper right-hand panel of Figure 2. First-order accurate ADF intervals perform well if the bias of $\hat{\psi}$ is small and the distribution of $\hat{\psi}$ is nearly symmetric. In these cases, second-order accu-

rate intervals cannot improve on first-order accurate intervals and can even perform slightly worse if sample size is small. The performance degradation is due to the difficulty of estimating higher-order moments from small samples. Variability in these estimates induces variability in the interval endpoints. Nonetheless, Figure 2 reveals

that the second-order ADF intervals never performed much worse and sometimes performed much better than first-order accurate intervals.

Figure 3 displays coverage of one-sided intervals when sampling from affine lognormal distributions. The performance of normal theory intervals is quite poor. The performance of first-order ADF intervals varies depending on the covariance matrix and coverage is as low as 0.78; see plot (3,4) in the lower right-hand panel. Figure 3 also reveals that second-order ADF intervals never performed much worse and some-

times performed much better than first-order accurate intervals.

Illustration

Heller, Judge, and Watson (2002) conducted a survey of university employees to examine relationships among personality traits, job satisfaction, and life satisfaction. They obtained six broad personality trait measures, three job satisfaction measures, and three life satisfaction measures on each of $N = 134$ subjects. A subset of their data from a random sample of 15 subjects is listed in Table 1.

Table 1: Sample of Size $N = 15$ from Heller

Case	Variable							
	1 N	2 E	3 C	4 CSE	5 PA	6 NA	7 LSO	8 JSO
1	2.75	2.50	3.92	2.44	3.80	2.30	3.00	3.25
1	1.83	4.42	4.25	2.55	4.10	1.30	4.20	4.63
3	2.00	3.25	4.83	2.92	4.00	1.40	2.60	3.88
4	2.17	2.50	4.17	2.53	3.80	1.50	3.60	3.88
5	2.92	3.50	4.08	1.96	3.10	2.20	4.00	3.25
6	1.58	3.67	3.67	2.89	4.00	1.30	4.00	4.00
7	2.00	3.67	4.00	2.68	4.20	1.50	4.00	4.38
8	1.58	4.17	4.75	2.90	4.00	1.30	2.20	2.38
9	2.58	3.17	4.08	2.34	2.80	1.30	3.40	4.38
10	2.92	3.00	3.92	2.21	3.40	1.50	3.60	3.13
11	2.33	2.08	4.67	2.17	2.40	1.70	3.00	3.88
12	2.25	3.67	4.92	2.74	3.90	1.60	3.80	3.00
13	2.08	3.33	3.50	2.39	3.30	1.60	4.00	3.88
14	3.92	2.42	3.00	1.60	2.50	3.10	3.60	4.13
15	1.67	4.17	4.58	2.93	4.30	1.60	4.00	3.63

Notes: N = Neuroticism, E = Extroversion, C = Conscientiousness, CSE = Core Self-Evaluations, PA = Positive Affectivity, NA = Negative Affectivity, LSO = Life Satisfaction—Significant Other Report, JSO = Job Satisfaction—Significant Other Report.

Appreciation is expressed to Danny Heller who provided the raw data for this example. They expected that job and life satisfaction would be positively related, and

that this relationship, in part, would be due to the influence of stable personality traits. Specifically, they hypothesized that (a) zero-order correlation coefficients be-

tween job and life satisfaction would be positive, and that (b) partial correlation coefficients between job and life satisfaction,

controlling for personality traits, would be smaller than the zero-order coefficients.

Table 2: Confidence Intervals Computed on Heller's Data

ψ	Quantity	Normal Theory		ADF	
		$N = 15$	$N = 134$	$N = 15$	$N = 134$
$\rho_{7,8} - \rho_{7,8,\mathbf{k}}$	$r_{7,8}$	0.476	0.477	0.476	0.477
	$r_{7,8,\mathbf{k}}$	0.410	0.398	0.410	0.398
	$\hat{\sigma}_\psi$	0.889	0.473	0.966	0.562
	$\hat{\omega}_1$	-0.086	-0.821	-0.281	-1.597
	$\hat{\omega}_3$	-0.183	-2.171	-0.082	-3.597
	$\hat{t}_{3,0.05}$	-1.803	-1.797	-1.845	-1.937
	$\hat{t}_{3,0.95}$	1.722	1.543	1.679	1.459
	$L(1)$	-0.353	0.110	-0.389	-0.001
	$U(1)$	0.484	0.147	0.520	0.160
	$L(2)$	-0.344	0.016	-0.368	0.008
	$U(2)$	0.494	0.153	0.542	0.174
	$t_{0.05,n}$	-1.761	-1.656	-1.761	-1.656
	$t_{0.95,n}$	1.761	1.656	1.761	1.656
	$r_{7(\mathbf{k})}^2$	0.468	0.249	0.468	0.249
$\rho_{7(\mathbf{k})}^2 - \rho_{8(\mathbf{k})}^2$	$r_{8(\mathbf{k})}^2$	0.263	0.109	0.263	0.109
	$\hat{\sigma}_\psi$	1.191	0.818	1.343	0.822
	$\hat{\omega}_1$	-0.269	-0.110	-0.347	-0.212
	$\hat{\omega}_3$	0.095	-0.010	-0.147	-0.124
	$\hat{t}_{3,0.05}$	-1.984	-1.771	-2.203	-1.924
	$\hat{t}_{3,0.95}$	1.512	1.543	1.392	1.423
	$L(1)$	-0.355	0.023	-0.426	0.023
	$U(1)$	0.766	0.258	0.838	0.259
	$L(2)$	-0.276	0.031	-0.294	0.039
	$U(2)$	0.837	0.266	0.996	0.278

$L(1)$ and $U(1)$ are lower and upper limits of first-order accurate intervals.

$L(2)$ and $U(2)$ are lower and upper limits of second-order accurate intervals.

One way to test hypothesis (b) is to compute a confidence interval for the difference between the simple and partial correlation coefficients. The upper portion of Table 2 displays first- and second-order accurate intervals for $\rho_{7,8} - \rho_{7,8,\mathbf{k}}$, where $\mathbf{k}$ consists of indices $1, \dots, 6$. This interval was not examined by Heller, Judge, and

Watson, but is comparable to intervals that were examined. Intervals based on the subsample of $N = 15$ cases and intervals based on the full sample of $N = 134$ cases are reported in Table 2. Results of intermediate calculations also are given to aid investigators who wish to verify accuracy of computer routines. Endpoints of second-

order accurate ADF intervals are somewhat shifted to the right compared to first-order accurate ADF intervals both for $N = 15$ and for $N = 134$. Notice that the first-order accurate interval based on $N=134$ contains zero whereas the second-order accurate interval does not contain zero.

The lower portion of Table 2 displays intermediate computations and intervals for $\rho_{7(\mathbf{k})}^2 - \rho_{8(\mathbf{k})}^2$ both for the subsample and for the full sample. Again, the second-order intervals are shifted to the right compared to the first-order accurate intervals.

Concluding Remarks

Some caution must be exercised when interpreting the proposed intervals computed on nonnormal data. If the population is multivariate normal, then all relationships among variables are linear and the traditional correlation coefficients adequately summarize the existing relationships. If the population is nonnormal, however, then relationships among variables need not be linear. Traditional correlation coefficients still summarize the linear relationships, but they do not reflect nonlinear components of the existing relationships.

In some cases, confidence interval procedures can be improved by applying a normalizing transformation of the correlation estimator. Let W_n be a statistic computed on a random sample of size N , where $n = N - r$ and r is a fixed positive integer. It is assumed that the moments of W_n are finite at least to order three. Denote the mean and variance of W_n by μ_W and σ_W^2 , respectively. A normalizing transformation of W_n , if it exists, is a function of W_n , say $g(W_n)$, chosen such that the skewness of $\sqrt{n}[g(W_n) - g(\mu_W)]$ has magnitude $O(n^{-3/2})$. Konishi (1981) showed that if data are sampled from a multivariate nor-

mal distribution, then

$$Z_{\rho_{ij}}(r_{ij}) = \frac{1}{2} \ln \left(\frac{1 + r_{ij}}{1 - r_{ij}} \right) \text{ and}$$

$$Z_{\rho_{i(\mathbf{k})}^2}(r_{i(\mathbf{k})}^2) = \frac{1}{2} \ln \left(\frac{1 + \sqrt{r_{i(\mathbf{k})}^2}}{1 - \sqrt{r_{i(\mathbf{k})}^2}} \right)$$

are normalizing transformations for r_{ij} and $r_{i(\mathbf{k})}^2$, respectively. Also,

$$Z_{\rho_{ij \cdot \mathbf{k}}}(r_{ij \cdot \mathbf{k}}) = \frac{1}{2} \ln \left(\frac{1 + r_{ij \cdot \mathbf{k}}}{1 - r_{ij \cdot \mathbf{k}}} \right)$$

is a normalizing transformation for $r_{ij \cdot \mathbf{k}}$. If data are sampled from multivariate normal distributions, then conventional confidence intervals for $\rho_{ij \cdot \mathbf{k}}$ or $\rho_{i(\mathbf{k})}^2$ that are based on the above normalizing transformations are only first-order accurate, however, because the normalizing transformation corrects for skewness but not for bias.

Derivatives of functions of $\rho_{ij \cdot \mathbf{k}}$ or $\rho_{i(\mathbf{k})}^2$ are readily obtained by one or more applications of the chain rule. For simple and partial correlations,

$$\frac{\partial Z_{\rho_{ij \cdot \mathbf{k}}}(\rho_{ij \cdot \mathbf{k}})}{\partial \boldsymbol{\sigma}} = \frac{\dot{\rho}_{ij \cdot \mathbf{k}}}{1 - \rho_{ij \cdot \mathbf{k}}^2} \text{ and}$$

$$\frac{\partial^2 Z_{\rho_{ij \cdot \mathbf{k}}}(\rho_{ij \cdot \mathbf{k}})}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}} = \frac{\ddot{\rho}_{ij \cdot \mathbf{k}}}{1 - \rho_{ij \cdot \mathbf{k}}^2}$$

$$+ \frac{(\dot{\rho}_{ij \cdot \mathbf{k}} \otimes \dot{\rho}_{ij \cdot \mathbf{k}}) 2\rho_{ij \cdot \mathbf{k}}}{(1 - \rho_{ij \cdot \mathbf{k}}^2)^2}.$$

Also, for squared multiple correlation coefficients,

$$\frac{\partial Z_{\rho_{i(\mathbf{k})}^2}(\rho_{i(\mathbf{k})}^2)}{\partial \boldsymbol{\sigma}} = \frac{\dot{\rho}_{i(\mathbf{k})}^2}{2\sqrt{\rho_{i(\mathbf{k})}^2} (1 - \rho_{i(\mathbf{k})}^2)} \text{ and}$$

$$\frac{\partial^2 Z_{\rho_{i(\mathbf{k})}^2}(\rho_{i(\mathbf{k})}^2)}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}} = \frac{\ddot{\rho}_{i(\mathbf{k})}^2}{2\sqrt{\rho_{i(\mathbf{k})}^2} (1 - \rho_{i(\mathbf{k})}^2)}$$

$$+ \frac{(\dot{\rho}_{i(\mathbf{k})}^2 \otimes \dot{\rho}_{i(\mathbf{k})}^2) (3\rho_{i(\mathbf{k})}^2 - 1)}{4 \left(\rho_{i(\mathbf{k})}^2\right)^{\frac{3}{2}} (1 - \rho_{i(\mathbf{k})}^2)^2}.$$

Employing the above transformations when data have not been sampled from multivariate normal distributions does not necessarily reduce skewness. Nonetheless, their use can be advantageous because the endpoints of the confidence intervals obtained by back-transforming the intervals for $Z_{\rho_{ij \cdot \mathbf{k}}}$ are guaranteed to be in $(-1, 1)$ and endpoints obtained by back-transforming the intervals for $Z_{\rho_{i(\mathbf{k})}^2}$ are guaranteed to be in $(0, 1)$. Intervals computed directly on the correlation coefficients need not satisfy this property. In small samples, it might be better to transform $r_{i(\mathbf{k})}^2$ as

$$Z^*(r_{i(\mathbf{k})}^2) = \frac{1}{2} \ln \left(\frac{\sqrt{r_{i(\mathbf{k})}^2}}{1 - \sqrt{r_{i(\mathbf{k})}^2}} \right)$$

because this transformation maps $(0, 1)$ to $(-\infty, \infty)$, whereas the normalizing transformation for $r_{i(\mathbf{k})}^2$ maps $(0, 1)$ to $(0, \infty)$.

Fisher's Z transformation also can be employed when ψ is a difference between correlation coefficients. For example, suppose that $\psi = (\rho_{ij} - \rho_{ij \cdot \mathbf{k}})/2$. The divisor 2 is used so that $\psi \in (-1, 1)$. Then, $\hat{\psi}$ can be transformed by

$$Z_{\psi}(\hat{\psi}) = \frac{1}{2} \ln \left(\frac{1 + \hat{\psi}}{1 - \hat{\psi}} \right).$$

Even if skewness is not reduced by this transformation, the endpoints of the back-transformed interval will be in $(-1, 1)$. The derivatives of this function are

$$\frac{\partial Z_{\psi}(\psi)}{\partial \boldsymbol{\sigma}} = \frac{\dot{\psi}_{\boldsymbol{\sigma}}}{1 - \psi^2} \text{ and}$$

$$\frac{\partial^2 Z_{\psi}(\psi)}{\partial \boldsymbol{\sigma} \otimes \partial \boldsymbol{\sigma}} = \frac{\ddot{\psi}_{\boldsymbol{\sigma}}}{1 - \psi^2} + \frac{(\dot{\psi}_{\boldsymbol{\sigma}} \otimes \dot{\psi}_{\boldsymbol{\sigma}}) 2\psi}{(1 - \psi^2)^2}.$$

The simulation study also examined properties of first- and second-order intervals that are based on Fisher's Z transformation. The results are very similar to those displayed in Figures 1–3. The performance of the second-order intervals was comparable to that of first-order intervals whenever the first-order intervals did not perform too badly. The second-order intervals improved on the first-order intervals in cases where coverage of the first-order intervals deviated substantially from $1 - \alpha$.

An investigator might prefer to bootstrap correlation functions and thereby dispense with the requirement of explicitly estimating the bias, variance, and skewness of the relevant sampling distributions. One-sided percentile bootstrap intervals for functions of simple, partial, and multiple correlations, however, are only first-order accurate. Often these intervals have poor coverage and can be inferior to normal theory intervals even when multivariate normality is violated (Rasmussen, 1987, 1988; Strube, 1988). Also see Efron (1988) for a discussion of this issue. Second-order accurate confidence intervals can be obtained by bootstrapping the asymptotic pivotal quantity T in (15); i.e., percentile- t intervals. Hall, Martin, and Schucany (1989) examined percentile, percentile- t , transformed percentile- t , and coverage-corrected iterated bootstrap intervals for zero-order correlation coefficients. Fisher's transformation was employed for the transformed percentile- t intervals. They found that in very small samples ($N \leq 20$), the iterated intervals as well as the percentile- t intervals worked well. Percentile- t intervals that employed the jackknife to estimate the standard error of the (transformed) sample correlation coefficient were superior to those that employed the delta method. One could

employ jackknife estimates of bias, variance, and skewness in the proposed second-order accurate intervals. This issue remains to be explored.

Abramovitch and Singh (1985) and Zhou and Gao (2000) showed that bootstrapping corrected asymptotic pivotal quantities such as $\hat{T}_2$ in Corollary 5.2 or $\hat{T}_3$ in Corollary 5.3 to obtain quantiles of their respective distributions yields third-order accurate confidence intervals. This bootstrap adjustment procedure was examined for the conditions corresponding to plot (3,4) in the lower right-hand panel of Figure 3. A sample size of $N = 25$ was employed. The results are summarized in

Table 3. The coverage values in the first two lines of Table 3 are those plotted in Figure 3. The remaining lines are based on an additional 2,000 samples. For each of these samples, 1,000 bootstrap samples were drawn and the bootstrap sampling distributions of T and $\hat{T}_3$ were approximated. Quantiles of these distributions were substituted for $t_{\alpha,n}$ in the computation of confidence intervals. Table 3 demonstrates that bootstrapping T yields second-order accurate confidence intervals with coverage similar to those based on $\hat{T}_3$. Bootstrapping $\hat{T}_3$ slightly improved the coverage of the upper interval, but over adjusted the endpoint of the lower interval.

Table 3: Coverage of Bootstrap-Adjusted Confidence Intervals for $\psi = \rho^2_{i(k)} - \rho^2_{i(k)}$ when Sampling from a Lognormal Distribution

Pivotal Quantity	Bootstrap Adjusted	Coverage		Number of Samples
		Lower Intervals	Upper Intervals	
T	No	0.972	0.781	5000
$\hat{T}_3$	No	0.966	0.869	5000
T	No	0.968	0.786	2000
$\hat{T}_3$	No	0.959	0.877	2000
T	Yes	0.930	0.869	2000
$\hat{T}_3$	Yes	0.934	0.886	2000

The intervals constructed in this article have reasonable coverage properties under most of the conditions that were examined. In some conditions, however, the coverage of the second-order accurate intervals is less accurate than desired, even though it represents a substantial improvement over coverage of first-order accurate intervals. A bootstrap adjustment to the second-order accurate intervals might improve coverage slightly, but this issue needs to be examined more thoroughly.

1 **References**

Alf, E.F., & Graf, R.G. (1999). Asymptotic confidence limits for the difference between two squared multiple correlations: A simplified approach. *Psychological Methods*, 4, 70–75.

Abramovitch, L., & Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *The Annals of Statistics*, 13, 116–132.

Bilodeau, M., & Brenner, D. (1999). *Theory of multivariate statistics*. New York: Springer.

- Bishop, Y.M.M., Fienberg, S.F., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press.
- Boik, R.J. (1998). A local parameterization of orthogonal and semi-orthogonal matrices with applications. *Journal of Multivariate Analysis*, 67, 244–276.
- Boik, R.J. (2005). Second-order accurate inference on eigenvalues of covariance and correlation matrices. *Journal of Multivariate Analysis*, 96, 136–171.
- Boik, R.J. (2006). Accurate confidence intervals in regression analyses of non-normal data. *Annals of the Institute of Statistical Mathematics*, accepted.
- Browne, M.W. & Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra & its Applications*, 82, 169–176.
- Casella, G. & Berger, R.L. (2002). *Statistical inference* (Second Ed.). Pacific Grove CA: Duxbury.
- Cook, M.B. (1951). Two applications of bivariate k -statistics. *Biometrika*, 38, 368–376.
- Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, 104, 293–296.
- Fisher, R.A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32.
- Fisher, R.A. (1962). The simultaneous distribution of correlation coefficients. *Sankhyā, A*, 24, 1–8.
- Graf, R.G., & Alf, E.A., Jr. (1999). Correlations redux: Asymptotic confidence limits for partial and squared multiple correlations. *Applied Psychological Measurement*, 23, 116–119.
- Graham, A. (1981). *Kronecker products and matrix differentiation with applications*. Chichester: Ellis Horwood Limited.
- Hall, P. (1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B*, 54, 221–228.
- Hall, P., Martin, M.A., & Schucany, W.R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *Journal of Statistical Computation and Simulation*, 33, 161–172.
- Hedges, L.V., & Olkin, I. (1983). Joint distribution of some indices based on correlation coefficients. In S. Karlin, T. Amemiya, & L. A. Goodman (Eds.), *Studies in econometrics, time series, and multivariate analysis* (pp. 437–454). New York: Academic Press.
- Heller, D., Judge, T.A., & Watson, D. (2002). The confounding role of personality and trait affectivity in the relationship between job and life satisfaction. *Journal of Organizational Behavior*, 23, 815–835.
- Hsu, P.L. (1949). The limiting distribution of functions of sample means and applications to testing hypotheses. In J. Neyman (Ed.), *Proceedings of the Berkeley Symposium on Mathematical Statistics & Probability* (pp. 359–402). Berkeley: University of California Press.

- Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis. *Biometrika*, 68, 647–651.
- MacRae, E.C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *Annals of Statistics*, 2, 337–346.
- Magnus, J.R., & Neudecker, H. (1979). The commutation matrix: some properties and applications. *Annals of Statistics*, 7, 381–394.
- Magnus, J.R., & Neudecker, H. (1999). Matrix differential calculus with applications in statistics and econometrics (Rev. ed.). Chichester: John Wiley & Sons.
- Nakagawa, S., & Niki, N. (1992). Distribution of the sample correlation coefficient for nonnormal populations. *Journal of the Japanese Society of Computational Statistics*, 5, 1–19.
- Neudecker, H. (1996). The asymptotic variance matrices of the sample correlation matrix in elliptical and normal situations and their proportionality. *Linear Algebra & its Applications*, 237/238, 127–132.
- Neudecker, H. & Wesselman, A.M. (1990). The asymptotic variance matrix of the sample correlation matrix. *Linear Algebra & its Applications* 127, 589–599.
- Niki, N., & Konishi, S. (1986). Higher-order asymptotic expansions for the distribution of the sample correlation coefficient. *Communications in Statistics—Simulation and Computation*, 13, 169–182.
- Olkin, I., & Finn, J.D. (1995). Correlations redux. *Psychological Bulletin*, 118, 155–164.
- Olkin, I., & Siotani, M. (1976). Asymptotic distributions of functions of a correlation matrix. In S. Ikeda, T. Hayakawa, H. Hudimoto, M. Okamoto, S. Siotani, & S. Yamamoto (Eds.), *Essays in probability and Statistics* (pp. 235–251). Tokyo: Shinko Tsusho.
- Pace, L., & Salvan, A. (1997). *Principles of statistical inference*. Singapore: World Scientific.
- Pearson, K. & Filon, L.N.G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London, A* 191, 229–311.
- Rao, C.R. (1973). *Linear statistical inference and its applications* (Second ed.). New York: John Wiley & Sons.
- Rasmussen, J.L. (1987). Estimating correlation coefficients: Bootstrap and parametric approaches, *Psychological Bulletin*, 101, 136–139.
- Rasmussen, J.L. (1988). “Bootstrap confidence intervals: Good or Bad”: Comments on Efron (1988) and Strube (1988) and further evaluation. *Psychological Bulletin*, 104, 297–299.
- Steiger, J.H. & Hakstian, A.R. (1982). The asymptotic distribution of the elements of a correlation matrix: Theory and application. *British Journal of Mathematical & Statistical Psychology*, 35, 208–215.

Steiger, J.H. & Hakstian, A.R. (1983). A historical note on the asymptotic distribution of correlations. *British Journal of Mathematical & Statistical Psychology*, 36, 157.

Strube, M.J. (1988). Bootstrap type I error rates for the correlation coefficient: An examination of alternate procedures. *Psychological Bulletin*, 104, 290–292.

Zhou, X.-H., & Gao, S. (2000). One-sided confidence intervals for means of positively skewed distributions. *The American Statistician*, 54, 100–104.

Appendix

Proof of Theorem 4

Taylor series expansions for $\sqrt{n}(\hat{\psi} - \psi)$, $\left[\sqrt{n}(\hat{\psi} - \psi)\right]^2$, and $\left[\sqrt{n}(\hat{\psi} - \psi)\right]^3$ around $\mathbf{s} = \boldsymbol{\sigma}$ can be written as follows.

$$\begin{aligned} \sqrt{n}(\hat{\psi} - \psi) &= \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \\ &+ \frac{1}{2\sqrt{n}} \ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' [\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \otimes \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})] \\ &+ O_p(n^{-1}), \\ \left[\sqrt{n}(\hat{\psi} - \psi)\right]^2 &= \left[\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]^2 \\ &+ O_p(n^{-\frac{1}{2}}), \text{ and} \\ \left[\sqrt{n}(\hat{\psi} - \psi)\right]^3 &= \left[\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]^3 \\ &+ \frac{3}{2\sqrt{n}} \left[\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})\right]^2 \\ &\times \ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' [\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \otimes \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})] + O(n^{-1}). \end{aligned}$$

Furthermore, expanding $\sqrt{n}(\hat{\sigma}_{\psi}^2 - \sigma_{\psi}^2)$ around $\mathbf{s} = \boldsymbol{\sigma}$ reveals that

$$\begin{aligned} \sqrt{n}(\hat{\sigma}_{\psi}^2 - \sigma_{\psi}^2) &= \left(\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}\right)' \sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) + \\ &2 \left[\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})' \otimes \dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \boldsymbol{\Omega}_{22,n}\right] \ddot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}} + O_p(n^{-\frac{1}{2}}), \end{aligned}$$

where $\boldsymbol{\omega} = \text{vec } \boldsymbol{\Omega}_{22,n}$. The claimed result is obtained by taking expectations, using $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \boldsymbol{\sigma} = 0$, and collecting terms of like order.

To verify that $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}' \boldsymbol{\sigma} = 0$, note that $\rho_{ij \cdot \mathbf{k}}$ and $\rho_{i(\mathbf{k})}^2$ are scale invariant functions of $\boldsymbol{\Sigma}$. That is, $\rho_{ij \cdot \mathbf{k}}(\boldsymbol{\Sigma}) = \rho_{ij \cdot \mathbf{k}}(\alpha \boldsymbol{\Sigma})$ and $\rho_{i(\mathbf{k})}^2(\boldsymbol{\Sigma}) = \rho_{i(\mathbf{k})}^2(\alpha \boldsymbol{\Sigma})$, where α is any positive scalar constant. It follows that the derivatives of $\rho_{ij \cdot \mathbf{k}}(\alpha \boldsymbol{\Sigma})$ and $\rho_{i(\mathbf{k})}^2(\alpha \boldsymbol{\Sigma})$ with respect to α each are zero. Using the chain rule,

$$\begin{aligned} 0 &= \frac{\partial \rho_{ij \cdot \mathbf{k}}(\alpha \boldsymbol{\Sigma})}{\partial \alpha} = \left(\frac{\partial \alpha \boldsymbol{\sigma}'}{\partial \alpha}\right) \left(\frac{\partial \rho_{ij \cdot \mathbf{k}}}{\partial \boldsymbol{\sigma} \alpha}\right) \\ &= \boldsymbol{\sigma}' \dot{\boldsymbol{\rho}}_{ij \cdot \mathbf{k}} \left(\frac{1}{\alpha}\right) \text{ and} \\ 0 &= \frac{\partial \rho_{i(\mathbf{k})}^2(\alpha \boldsymbol{\Sigma})}{\partial \alpha} = \left(\frac{\partial \alpha \boldsymbol{\sigma}'}{\partial \alpha}\right) \left(\frac{\partial \rho_{i(\mathbf{k})}^2}{\partial \boldsymbol{\sigma} \alpha}\right) \\ &= \boldsymbol{\sigma}' \dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2 \left(\frac{1}{\alpha}\right). \end{aligned}$$

The result follows because $\dot{\boldsymbol{\psi}}_{\boldsymbol{\sigma}}$ is a linear function of $\dot{\boldsymbol{\rho}}_{ij \cdot \mathbf{k}}$ and/or $\dot{\boldsymbol{\rho}}_{i(\mathbf{k})}^2$ terms.

Proof of Theorem 5

Expand T as

$$\begin{aligned} T &= \frac{\sqrt{n}(\hat{\psi} - \psi)}{\sigma_{\psi}} \times \\ &\left[1 - \frac{1}{2\sqrt{n}\sigma_{\psi}^2} \sqrt{n}(\hat{\sigma}_{\psi}^2 - \sigma_{\psi}^2) + O_p(n^{-1})\right], \end{aligned}$$

and then take expectations of T , T^2 , and T^3 to verify the claimed expressions for the mean, variance, and skewness of T . The sizes of the various remainder terms differ because the expectation of $O_p(1)$ terms that are even-order in $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ or $\sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ have magnitude $O(1)$ whereas the expectation of $O_p(1)$ terms that are odd-order in $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ or $\sqrt{n}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ have magnitude $O(n^{-1/2})$. It follows that the cumulant generating function of T can be expanded as follows:

$$\begin{aligned} C_T(u) &= \ln \{E [e^{iuT}]\} \\ &= \frac{\kappa_1 iu}{\sqrt{n}} - \frac{u^2}{2} + \frac{\kappa_3 (iu)^3}{6\sqrt{n}} + O(n^{-1}). \end{aligned}$$

Inverting the characteristic function, $\exp \{C_T(u)\}$, yields the density function, f_T , and integrating f_T yields the distribution function, F_T .

Inversion of Exponential Functions

This section describes a Modified Newton algorithm for finding the value of $\hat{t}_{3,\alpha}$ that satisfies $h(\hat{t}_{3,\alpha}) = 0$, where

$$\begin{aligned} &h(\hat{t}_{3,\alpha}) \\ &= \hat{t}_{3,\alpha} - \frac{\hat{\kappa}_1}{\sqrt{n}} - \frac{\hat{\kappa}_3 \left(\hat{t}_{3,\alpha}^2 e^{-\hat{d}\hat{t}_{3,\alpha}^2/2} - 1 \right)}{6\sqrt{n}} - z_\alpha, \end{aligned}$$

where $\hat{d}$ is defined in Corollary 5.3. The solution is unique because $h(\hat{t}_{3,\alpha})$ is a monotonic function of $\hat{t}_{3,\alpha}$ by construction. An initial guess for $\hat{t}_{3,\alpha}$ is $\hat{t}_{3,\alpha,0} = \hat{t}_\alpha$, where $\hat{t}_\alpha$ is defined in Corollary 5.2. At iteration $i + 1$ of the algorithm, the value of $\hat{t}_{3,\alpha}$ is

$$\hat{t}_{3,\alpha,i+1} = \hat{t}_{3,\alpha,i} - u_i, \text{ where}$$

$$u_i = \begin{cases} -\hat{t}_{3,\alpha,i} & \text{if } \frac{h(\hat{t}_{3,\alpha,i})}{\hat{t}_{3,\alpha,i} h^{(1)}(\hat{t}_{3,\alpha,i})} < -1, \\ \frac{\hat{t}_{3,\alpha,i}}{2} & \text{if } \frac{h(\hat{t}_{3,\alpha,i})}{\hat{t}_{3,\alpha,i} h^{(1)}(\hat{t}_{3,\alpha,i})} > \frac{1}{2}, \\ \frac{h(\hat{t}_{3,\alpha,i})}{h^{(1)}(\hat{t}_{3,\alpha,i})} & \text{otherwise,} \end{cases}$$

$$\text{and } h^{(1)}(\hat{t}) = 1 - \frac{\hat{\kappa}_3 \hat{t} (2 - \hat{d}\hat{t}^2) e^{-\hat{d}\hat{t}^2/2}}{6\sqrt{n}}.$$

Proof of Theorem 6

Define Z_i as

$$Z_i \stackrel{\text{def}}{=} \frac{\sqrt{n_i} (\hat{\psi}_i - \psi_i)}{\sigma_i},$$

where all terms are defined in Theorem 6. Then, the first-order Taylor series expansion of T in Theorem 6 around $\hat{\sigma}_i^2 = \sigma_i^2$ for $i = 1, 2$ is

$$T = \frac{g^{-\frac{1}{2}} \sigma_1 Z_1 - \sigma_2 Z_2}{(g^{-1} \sigma_1^2 + \sigma_2^2)^{\frac{1}{2}}} \times$$

$$\left[1 - \frac{V_1}{2\sqrt{n_1}} + \frac{V_2}{2\sqrt{n_2}} + O_p(n^{-1}) \right], \text{ where}$$

$$g = \frac{n_1}{n_2}, \quad n = \min(n_1, n_2),$$

$$V_1 = \frac{\sqrt{n_1}(\hat{\sigma}_1^2 - \sigma_1^2)}{\sigma_1^2 + g\sigma_2^2}, \text{ and}$$

$$V_2 = \frac{\sqrt{n_2}(\hat{\sigma}_2^2 - \sigma_2^2)}{g^{-1}\sigma_1^2 + \sigma_2^2}.$$

The claimed results are obtained by using Theorem 4 to obtain expectations of T , $[T - E(T)]^2$, and $[T - E(T)]^3$.

Inferences About the Components of a Generalized Additive Model

Rand R. Wilcox
Department of Psychology
University of Southern California

A method for making inferences about the components of a generalized additive model is described. It is found that a variation of the method, based on means, performs well in simulations. Unlike many other inferential methods, switching from a mean to a 20% trimmed mean was found to offer little or no advantage in terms of both power and controlling the probability of a Type I error.

Key words: Nonparametric regression, smoothers, backfitting algorithm, wild bootstrap

Introduction

When dealing with a regression problem, a standard approach is to assume

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \varepsilon_i, \quad (1)$$

$i=1, \dots, n$, where ε_i is independent of $X_{i1}, \dots, X_{ip}$, $E(\varepsilon)=0$, and then test hypotheses about the unknown parameters $\beta_0, \dots, \beta_p$. This approach seems appropriate when the assumed model, given by (1), is a reasonable approximation of the true regression surface. But experience with smoothers suggests that, at least in some situations, the assumption that Y is linearly related to the p regressors is unsatisfactory, and often it is unclear how to correct this problem when using a parametric approach to modeling the data, particularly when $p > 2$. That is, simple transformations of the regressors might be used, such as taking logarithms, but situations arise where effective transformations are not evident and difficult to discern.

One problem is that smoothers often suggest that over some region of the predictor space, the regression surface is, approximately, a horizontal plane, meaning there is virtually no association at all, but for other regions a curvilinear association appears that can be difficult to model.

Figure 1 provides an example where the goal is to predict reading ability (measured by a word identification score) based on two measures of phonological awareness. Shown is a smooth using the loess method derived by Cleveland and Devlin (1988). Note that for low measures associated with both predictors, the regression surface is nearly flat, but in other regions there appears to be a nonlinear association. (Switching to a robust smooth, namely the running interval smoother in Wilcox, 2003, with the span set to 1.2, results in a plot nearly identical to Figure 1.)

A more flexible approach, when modeling the data, is to use a generalized additive model (Hastie & Tibshirani, 1993). That is, assume that there exists functions $f_1, \dots, f_p$ such that

$$Y = \beta_0 + f_1(X_1) + \cdots + f_p(X_p) + \varepsilon. \quad (2)$$

Rand R. Wilcox (rwilcox@usc.edu) is Professor of Psychology at the University of Southern California.

Figure 1

Of course, equation (2) contains the usual model, given by (1), as a special case. For any fixed j , the goal in this paper is to consider the problem of testing

$$H_0: f_j(X_j) = 0. \quad (3)$$

The general strategy used here is to fit a generalized additive model omitting the j th variable and then check what is essentially a simple extension of the partial residual plot (e.g. Berk & Booth, 1995) for an association. (The approach used here for detecting an association is closely connected to what Berk and Booth call the AMALL method for detecting curvature, which stems from Breiman and Friedman, 1985,

p. 618). That is, test the hypothesis that the regression line between the resulting residuals and X_j is straight and horizontal; this is done with the wild bootstrap method derived by Stute, González-Manteiga and Presedo-Quindimil (1998). Details of the proposed method are given in the next two sections.

A Generalized Additive Fit

There are many ways of fitting the model given by (2) with most methods assuming that the goal is to estimate the mean of Y given $(X_{i1}, \dots, X_{ip})$. The method used here was chosen because it represents a particularly simple way of including virtually any robust measure of location. Robust measures of location are known to have many advantages, versus the mean, for a wide range of situations

(e.g., Hampel, Ronchetti, Rousseeuw and Stahel, 1986; Huber, 1981; Staudte & Sheather, 1990; Wilcox, 2005).

Two fundamental advantages are improved control over the probability of a Type I error in situations where methods based on means perform poorly, and substantial gains in power, even under small departures from normality. Here, however, when using means, good control over the probability of a Type I error is obtained in simulations and using means actually offers higher power. So the method used here provides an interesting example of a situation where a non-robust estimator performs better than a robust estimator in terms of power, even when sampling from a heavy-tailed distribution.

The robust location estimator used here is the 20% trimmed mean which is computed as follows. Let $X_1, \dots, X_m$ be any m values and let $X_{(1)} \leq \dots \leq X_{(m)}$ be the values written in ascending order. Let $g = [2m]$, where $[x]$ is the greatest integer less than or equal to x . Then the 20% trimmed mean is

$$\frac{1}{m-2g} \sum_{i=g+1}^{m-g} X_{(i)}.$$

The reason for choosing 20% trimming, over alternative amounts of trimming, stems from results on efficiency reported by Rosenberger and Gasko (1983).

As explained in Hampel, Ronchetti, Rousseeuw and Stahel (1986), Huber (1981), and Staudte and Sheather (1990), a reasonable alternative to the 20% trimmed mean is some robust M-estimator. The only reason for choosing a 20% trimmed over the better-known robust M-estimators is to avoid division by zero in certain situations to be described.

First consider the one-predictor case ($p=1$). There are many ways of estimating f_1 using so-called smoothers (e.g., Hastie & Tibshirani, 1990; Härdle, 1990). Here, a running interval smoother is used mainly because it is readily extended to robust measures of location such as the 20% trimmed mean. This is not to suggest that other smoothers have no value for

the problem at hand. Rather, the goal is find at least one method that performs well in simulations, and the running interval smoother is relatively easy to implement.

A fairly well-known alternative is the smoother derived by Cleveland (1979) which includes a method of down weighting extreme Y values. One reason for choosing a running interval smoother is that when used with a 20% trimmed mean, it seems to be a bit better at handling moderately large or small outliers, versus Cleveland's method, and it seems to perform reasonably well compared to a variety of other smoothers that might be used (Wilcox, 2005). Again, this is not to suggest that all other smoothers be eliminated from consideration for the problem at hand, but the relative merits of using other smoothers is left for future investigations.

The running interval smoother is applied as follows. Let M be the median of the values $X_1, \dots, X_n$. The median absolute deviation (MAD), based on $X_1, \dots, X_n$, is the median of the n values $|X_1 - M|, \dots, |X_n - M|$. Let $MADN = MAD/.6745$; under normality, $MADN$ estimates σ , the standard deviation. Let κ be some constant that is chosen in a manner to be described. Then the point X is said to be close to X_i if

$$|X_i - X| \leq \kappa \times MADN.$$

The constant κ is called the span. Thus, for normal distributions, X is close to X_i if X is within κ standard deviations of X_i . Let

$$N(X_i) = \{j: |X_j - X_i| \leq \kappa \times MADN\}.$$

That is, $N(X_i)$ indexes all X_j values that are close to X_i . Now consider the random sample $(X_1, Y_1), \dots, (X_n, Y_n)$ and let $\hat{\theta}_i$ be an estimate of some parameter of interest, based on the Y_j values such that $j \in N(X_i)$. That is, use all of the Y_j values for which X_j is close to X_i . Here, as

previously indicated, a mean or 20% trimmed mean is used. So, for example, $\hat{\theta}_i$ might be estimated with the 20% trimmed mean of the Y_j values such that $j \in N(X_i)$. In exploratory work, a good choice for the span is often $\kappa=.8$ or 1, but for the situation at hand an alternative choice is needed.

Virtually any smoother, including the one used here, can be extended to the generalized additive model given by (2) using the backfitting algorithm in Hastie and Tibshirani (1990). Set $k=0$ and let f_j^0 be some initial estimate of f_j ($j=1, \dots, p$). Here, $f_j^0 = S_j(X | Y_j)$, where $S_j(Y | X_j)$ is the running interval smooth based on the j th predictor, ignoring the other predictors under investigation. Next, iterate as follows:

1. Increment k .
2. For each $j, j=1, \dots, p$, let

$$f_j^k = S_j(Y - \sum_{l \neq j} f_l^k | X_j).$$
3. Repeat steps 1 and 2 until convergence.

Finally, estimate β_0 with

$$b_0 = m(Y - \sum f_j^k),$$

where m indicates the measure of location used when computing the smooth, which here is taken to be a 20% trimmed mean or the usual mean.

Testing H_0

For convenience, momentarily assume the goal is to test

$$H_0 : f_1(X_1) = 0.$$

The proposed method begins by fitting the generalized additive model as described in the

previous section using the $X_2, \dots, X_p$ values, ignoring X_1 , yielding

$$\hat{Y}_i = b_0 + \hat{f}_2(X_{i2}) + \dots + \hat{f}_p(X_{ip}),$$

where $\hat{f}_j(X_{ij})$ is the estimate of $f_j(X_{ij})$ based on the backfitting algorithm. Let $r_i = Y_i - \hat{Y}_i$, $i=1, \dots, n$. Then the strategy is to test the hypothesis that when predicting the residuals, given X_1 , the regression is a straight horizontal line. This is done using the wild bootstrap method derived by Stute, González-Manteiga and Presedo-Quindimil (1998). As is evident, the method is readily modified to test (3) for any j .

To elaborate, let $\bar{r}_i$ be the mean or 20% trimmed mean based on the residuals $r_1, \dots, r_n$. Fix j and set $I_i = 1$ if $X_i \leq X_j$, otherwise $I_i = 0$. The notation $X_i \leq X_j$ means that for every k , $k=1, \dots, p$, $X_{ik} \leq X_{jk}$. Let

$$\begin{aligned} R_j &= \frac{1}{\sqrt{n}} \sum I_i (r_i - \bar{r}_i) \\ &= \frac{1}{\sqrt{n}} \sum I_i v_i \end{aligned} \quad (4)$$

where

$$v_i = r_i - \bar{r}_i.$$

The test statistic is the maximum absolute value of all the R_j values. That is, the test statistic is

$$D = \max |R_j|. \quad (5)$$

An appropriate critical value is estimated with the *wild bootstrap method* as follows. Generate $U_1, \dots, U_n$ from a uniform distribution and set

$$V_i = \sqrt{12}(U_i - .5),$$

$$v_i^* = v_i V_i,$$

and

$$r_i^* = \bar{r}_i + v_i^*.$$

Then based on the n pairs of points $(\mathbf{X}_1, r_1^*), \dots, (\mathbf{X}_n, r_n^*)$, compute the test statistic as described in the previous paragraph and label it D^* . Repeat this process B times and label the resulting (bootstrap) test statistics $D_1^*, \dots, D_B^*$. Finally, put these B values in ascending order yielding $D_{(1)}^* \leq \dots \leq D_{(B)}^*$. Then the critical value is $D_{(u)}^*$, where $u=(1-\alpha)B$ rounded to the nearest integer. That is, reject if

$$D \geq D_{(u)}^*.$$

Based on Theorem 1 in Stute et al. (1998), this method is valid under weak assumptions placed on X .

For convenience, when using means, the technique just described will be called method V1. When using a 20% trimmed mean, it will be called method V2. Note that a smooth can be fit using a 20% trimmed mean, but when using the wild bootstrap in conjunction with the resulting residuals, one could use the mean of the residuals, rather than a 20% trimmed mean, when testing H_0 . This will be called method V3.

Choosing the Span

There remains the issue of choosing the span, κ , when fitting the generalized additive model. Preliminary simulations indicated that if the span is too large, the actual probability of a Type I error can exceed the nominal level (cf. Härdle & Mammen, 1993). A proper choice for the span, given n and the amount of trimming, was found to correct this problem. That is, the choice of the span when using means differs from the choice when using a 20% trimmed mean instead. Here the span was determined by assuming that $X_1, \dots, X_p$ and $\mathcal{E}$ have independent standard normal distributions, and

then for a given sample size and depending on whether means or 20% trimmed means were used, κ was determined via simulations so that the actual probability of a Type I error is approximately equal to the nominal level when testing at the .05 level. Then given n , and depending on whether a trimmed mean was to be used, this value for κ was used in the simulations described in the next section. All indications are that the choice for the span does not depend on p for $p=2, 3, 4$ and 5. (Whether this remains true for $p>5$ has not been investigated.) The results are summarized in Table 1.

Table 1: Choices for the span, κ

n	20% trimming	mean
20	1.20	.80
40	1.0	.70
60	.85	.55
80	.75	.50
120	.65	.50
160	.65	.50

Simulation Results

Simulations were used to check the small-sample properties of the proposed method. Observations were generated where the marginal distributions have a g-and-h distribution (Hoaglin, 1985) which includes the normal distribution as a special case. More precisely, observations Z_{ij} ($i=1, \dots, n; j=1, 2$) were initially generated from a multivariate normal distribution having correlation ρ , then the marginal distributions were transformed to

$$X_{ij} = \begin{cases} \frac{\exp(gZ_{ij}) - 1}{g} \exp(hZ_{ij}^2/2), & \text{if } g > 0 \\ Z \exp(hZ_{ij}^2/2), & \text{if } g = 0 \end{cases}$$

where g and h are parameters that determine the third and fourth moments. The four (marginal) g-and-h distributions examined were the standard normal ($g=h=0$), a symmetric heavy-tailed distribution ($g=0, h=.5$), an asymmetric

distribution with relatively light tails ($g=.5$, $h=0$), and an asymmetric distribution with heavy tails ($g=h=.5$). Here, two choices for ρ were considered: 0 and .5. Table 1 shows the theoretical skewness (κ_1) and kurtosis (κ_2) for each distribution considered. When $g>0$ and $h>1/k$, $E(X^k)$ is not defined and the corresponding entry in Table 1 is left blank. Additional properties of the g-and-h distribution are summarized by Hoaglin (1985). Some of these distributions might appear to represent extreme departures from normality, but the idea is that if a method performs reasonably well in these cases, this helps support the notion that they will perform well under conditions found in practice.

Table 2: Some properties of the g-and-h distribution.

g	h	κ_1	κ_2	$\hat{\kappa}_1$	$\hat{\kappa}_2$
0.0	0.0	0.00	3.0	0.00	3.0
0.0	0.5	0.00	—	0.00	11,896.2
0.5	0.0	1.75	8.9	1.81	9.7
0.5	0.5	—	—	120.10	18,393.6

A possible objection to Table 2 when performing simulations is that the distribution of observations generated on a computer does not always have the theoretical skewness and kurtosis values shown. The reason is that observations generated on a computer come from a bounded interval, so the skewness and kurtosis of the distribution will be finite, even when in theory it should be infinite. Accordingly, Table 2 also reports the estimated skewness ($\hat{\kappa}_1$) and kurtosis ($\hat{\kappa}_2$) values based on simulations with 10,000 replications.

Two sets of simulations were run. The first was for $p=3$ with the goal of testing $H_0: f_3(X_3)=0$. The correlation between X_1 and X_2 was taken to be either 0 or .5, and observations were generated according to one of three models: $Y=\varepsilon$, $Y=X_1+X_2+\varepsilon$ and $Y=X_1+2X_2^2+\varepsilon$. Table 3 contains $\hat{\alpha}$, the estimated probability of making a Type I error

when testing at the .05 level with $n=20$, when $\rho=0$. Increasing ρ to .5 had a negligible effect, so for brevity the results are not reported.

It is noted, however, that if $n=20$ and $\rho=.7$, then some effect on the probability of a Type I error results: it tends to decrease somewhat versus situations where $\rho=.5$ or 0. But for $n=40$, this was no longer the case. Introducing curvature had more of an effect, and so results for this case are reported. No situation was found where the estimated probability of a Type I error exceeded .065 when testing at the .05 level, and the lowest estimate was .030 except when $\rho=.7$, in which case, with $n=20$, $\hat{\alpha}$ goes as low as .017.

The second set of simulations was for $p=5$. Again observations were generated according to the models $Y=\varepsilon$, $Y=X_1+X_2+\varepsilon$ and $Y=X_1+2X_2^2+\varepsilon$, only now the goal was to test $H_0: f_5(X_5)=0$. Obviously, with $p=5$, it is difficult to consider the many variations that might arise when the null hypothesis is true. Here, as a partial check on the method, some additional simulations were run assuming normality. If, for example, the models

$$Y = X_1 + X_2 + X_3 + X_4 + \varepsilon$$

and

$$Y = X_1 + X_2 + X_3^2 + X_4^3 + \varepsilon$$

are used to generate the data, the estimated probability of a Type I error when testing at the .05 level was .049 and .045, respectively.

Power

Now consider power. Various situations were considered and it was found that regardless of the distributions used, or the model used to generate the data, method V1 always had higher power than V2, and often the gain in power was substantial. For example, with $n=40$ and $Y=X_1+.5X_3+\varepsilon$, if X_1 , X_2 and ε have independent standard normal random variables, power is .63 for method V1 and .40 for method V2. If instead ε has a symmetric, heavy-tailed distribution ($g=0$ and $h=.5$), now power is .19 and .06 for methods V1 and V2, respectively.

So, based purely on Type I error and power, all indications are that the approach based on means performs well, and there is no known reason for preferring the method based on a trimmed mean. This is not to suggest completely ruling out a trimmed mean, because using a trimmed mean can result in a better fit to the data, but in terms of detecting situations where a component of a generalized additive model differs from zero, using the mean appears to be preferable.

The reason method V2 has relatively low power is evidently related to using a trimmed mean applied to the residuals when using the wild bootstrap method. If method V3 is used instead, the problem of relatively low power, when using a general additive model based on a trimmed mean, is reduced substantially. Table 4 shows power estimates when using V1 versus V3. Often there is little separating the two methods, but even now, V1 has uniformly higher power. This continued to be the case when data were generated from non-linear models.

Conclusion

Of course, simulations cannot prove that a particular method always controls the probability of a Type I error, or that one particular method always has higher power than another. Nevertheless, all indications are that method V1, based on means, dominates in terms of power, and it performs well in terms of controlling the probability of a Type I error under what would seem like fairly extreme departures from normality.

References

- Breiman, L., & Friedman, J. H. (1985). Rejoinder to comments on estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 614–619.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829–836.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally-weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83, 596–610.
- Friedman, J. H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 76, 817–823.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics*. NY: Wiley.
- Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.
- Härdle, W., & Mammen, E. (1993). Comparing non-parametric versus parametric regression fits. *Annals of Statistics*, 21, 1926–1947.
- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. New York: Chapman and Hall.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. Hoaglin, F. Mosteller, & J. Tukey (Eds.) *Exploring data tables, trends, and shapes*. NY: Wiley, p. 461–515.
- Huber, P. J. (1981). *Robust statistics*. New York: Wiley.
- Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. Hoaglin, F. Mosteller, & J. Tukey (Eds.) *Understanding robust and exploratory data analysis*. NY: Wiley, p. 297–336.
- Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. NY: Wiley.
- Stute, W., Gonzalez-Manteiga, W. G., & Presedo-Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93, 141–149.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. (2nd ed.). San Diego, CA: Academic Press.

Table 3: Estimated probability of a Type I error, $n=20$

g	h	$p=3$		$p=5$	
		g	h	$Y = X_1 + X_2 + \varepsilon$	$Y = X_1 + X_2^2 + \varepsilon$
0.0	0.0	0.0	0.0	.051	.065
		0.0	0.5	.041	.048
		0.5	0.0	.044	.062
		0.5	0.5	.033	.037
0.0	0.5	0.0	0.0	.047	.045
		0.0	0.5	.043	.046
		0.5	0.0	.040	.043
		0.5	0.5	.038	.039
0.5	0.0	0.0	0.0	.053	.055
		0.0	0.5	.038	.053
		0.5	0.0	.050	.056
		0.5	0.5	.036	.043
0.5	0.5	0.0	0.0	.043	.040
		0.0	0.5	.039	.043
		0.5	0.0	.045	.043
		0.5	0.5	.036	.043

Table 4: Estimated power, $n=40, p=3, Y = X_1 + .5X_3 + \varepsilon$

g	h	g	h	Method V1	Method V3
0.0	0.0	0.0	0.0	.63	.63
		0.0	0.5	.19	.18
		0.5	0.0	.53	.52
		0.5	0.5	.15	.15
0.0	0.5	0.0	0.0	.64	.61
		0.0	0.5	.29	.27
		0.5	0.0	.59	.37
		0.5	0.5	.26	.24
0.5	0.0	0.0	0.0	.68	.45
		0.0	0.5	.21	.20
		0.5	0.0	.57	.42
		0.5	0.5	.19	.17
0.5	0.5	0.0	0.0	.56	.34
		0.0	0.5	.28	.24
		0.5	0.0	.52	.34
		0.5	0.5	.26	.24

Maximum Tests are Adaptive Permutation Tests

Markus Neuhäuser
Koblenz University of Applied Sciences

Ludwig A. Hothorn
Leibniz University of Hannover

In some areas, e.g., statistical genetics, it is common to apply a maximum test, where the maximum of several competing test statistics is used as a new statistic, and the permutation distribution of the maximum is used for inference. Here, it is shown that maximum tests are special cases of adaptive permutation tests. The 30-year old idea of adaptive statistical tests is more flexible than previously thought when permutation tests are used, and the selector statistic is calculated for every permutation. Because the independence between the selector and the test statistics is no longer needed, the test statistics themselves can be used as selectors. Then, the maximum tests fit into the concept of adaptive tests. In addition to the gained flexibility, maximum tests can be more powerful than classical adaptive tests.

Key words: Adaptive test, maximum test, maximum efficiency robust test, nonparametric statistics

Introduction

In this article the two-sample location problem is considered. Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ denote two random samples. The observations within each sample are independent and identically distributed, and independence between the two samples is assumed. Let F_1 and F_2 be the distribution functions corresponding to populations 1 and 2, respectively. In the location-shift model the distribution functions are the same except perhaps for a change in their locations; that is, $F_1(t) = F_2(t - \theta)$ for every t . The null hypothesis is $H_0: \theta = 0$, whereas the alternative states $\theta \neq 0$.

Markus Neuhäuser is professor of statistics at the Koblenz University of Applied Sciences, RheinAhrCampus Remagen. His main research interests are nonparametric methods and application in life sciences. Ludwig Hothorn is professor for biostatistics at the natural science faculty of Leibniz University of Hannover in Germany. His research interests are multiple comparison procedure and order restricted inference with bio-medical applications. He is council member of the International Biometric Society.

Often, a normal assumption for F_1 and F_2 is not tenable. In this case, a nonparametric test can be performed using a linear rank statistic

$$T = \sum_{i=1}^N g(i) V_i, \text{ where } g(i) \text{ are real valued}$$

scores, and $V_i = 1$ when the i -th smallest of the $N = n + m$ observations is from the first sample and $V_i = 0$ otherwise. There is a variety of different scores and, consequently, it is difficult for the practicing statistician to select a test statistic. A powerful test exists for every distribution, but the real distribution is usually *a priori* unknown and, consequently, one needs a test that has high relative power across the different possible distributions which may be difficult for small sample sizes.

In order to solve this dilemma Randles and Hogg (1973) and Hogg (1974) introduced adaptive statistical tests as a new dimension in distribution-free inference. The basic idea is that the value of a selector statistic decides which of some possible test statistics is applied. To be precise, the concept is based on the following lemma:

- (i) Let $\mathcal{F}$ denote the class of distributions under consideration. Suppose that each of k tests $T_1, \dots, T_k$ is distribution-free over $\mathcal{F}$, that is, $\Pr_{H_0}(T_i \in C_i) \leq \alpha$ for each $F \in \mathcal{F}, i = 1, \dots, k$.

(ii) Let S be some statistic (called a selector statistic) that is, under H_0 , independent of $T_1, \dots, T_k$ for each $F \in \mathcal{F}$. Suppose we use S to decide which test T_i to conduct. Specifically, let M_S denote the set of all values of S with the following decomposition:
 $M_S = D_1 \cup D_2 \cup \dots \cup D_k$, $D_h \cap D_j = \emptyset$ for $h \neq j$, so that $S \in D_i$ corresponds to the decision to use test T_i .

The overall testing procedure is then defined by: If $S \in D_i$ then reject H_0 if $T_i \in C_i$. This two-staged adaptive test is distribution-free under H_0 over the class $\mathcal{F}$, i.e., it maintains the level α for each $F \in \mathcal{F}$.

The proof of this lemma was given e.g. by Randles and Wolfe (1979, p. 388). Usually, tests based on ranks were used together with a selector statistic that depends on the combined ordered sample (Bünig, 1991). The reason is that under the null hypothesis and in case of a continuous distribution, the rank vector is independent of the order statistics (Randles & Wolfe, 1979).

During the last 30 years several adaptive tests were introduced, not only for the two-sample location problem, but also for multi-sample problems and scale tests (Beier & Bünig, 1997, Bünig, 1991, 2000, 2002). Freidlin et al. (2003a) proposed a test where the selector and the test statistics are asymptotically uncorrelated only. Furthermore, the concept of adaptive tests was applied to parametric tests (Neuhäuser & Hothorn, 1997). However, this study focused on nonparametric two-sample location tests.

In 1995, Weerahandi wrote that, “until recently, most of the applications involving nonparametric tests were performed using asymptotic approximations” (p. 78). Therefore, most adaptive tests are constructed of asymptotic tests. Obviously, permutation tests (see e.g. Good, 2000) can be combined to an adaptive test, too, an example is the test introduced by O’Gorman (2001). The aim of this article is to show that permutation tests can offer a large flexibility to the concept of adaptive tests and that a maximum test is an adaptive permutation test.

The Combination of Permutation Tests

On the one hand, one can use the concept of adaptive tests in the classical way. That is, the selector is computed once and the chosen test is performed, now based on the permutation distribution. On the other hand, there is an alternative: the selector may be calculated for each permutation. In this case, a permutation test is carried out using the statistic

$$T_{P1} = \sum_{i=1}^k I(S \in D_i) T_i, \text{ where } I(\cdot) \text{ denotes the}$$

indicator function. With this statistic a permutation test can be performed, and neither the independence between S and the T_i nor the continuousness of the underlying distribution is necessary, in contrast to tests based on the lemma given in the introduction. Note that for a classical adaptive test the distributions have to be continuous for the independence between rank vector and order statistics. In practice, however, ties frequently occur in a variety of settings (see e.g. Coakley & Heise, 1996). Even when the underlying distribution is continuous rounding leads to ties. For example, reaction times may be measured with a time clock graduated in tenths or hundredths of a second. Moreover, it is an advantage of nonparametric rank tests that they can also be applied to ordered categorical data, but when continuousness has to be assumed, this advantage is lost.

Because the independence to the selector is no longer necessary one can use the (standardized) test statistics themselves as selectors. To be precise, one can perform a permutation test based on the statistic,

$$T_{P2} = \sum_{i=1}^k I(T_i = \max(T_1, \dots, T_k)) I(T_i > T_j \ \forall j > i) T_i.$$

The second indicator function is needed because two statistics T_i and T_j (with $i \neq j$) could have an equal value for a given data set. Now, it is easy to see that $T_{P2} = \max(T_1, \dots, T_k)$. Thus, a maximum test may be regarded as an adaptive permutation test.

The use of the maximum of several (standardized) statistics as a new test statistic is

common in rather different testing problems (e.g. Bretz & Hothorn, 2001, Chung & Fraser, 1958, Freidlin & Korn, 2002, Freidlin et al., 1999, 2002, 2003b; Gastwirth & Freidlin, 2000; Hirotsu, 1986; Marozzi, 2004a, 2004b, Neuhäuser & Hothorn, 1999, Neuhäuser et al., 2000, 2004, Zheng et al., 2002). The approach has the advantage that neither a selector statistic nor the specification of which test should be performed for which values of the selector is needed. Furthermore, a maximum test is possible for relatively small sample sizes. In contrast, a classical adaptive test needs a sample size of at least 20 per group to avoid too many misclassifications (Hill et al., 1988, Büning, 1991, p. 238).

Example

As an example, the class of all continuous and symmetric distributions is considered. In this case the following scores $g(i)$ may be useful:

Gastwirth test (short tails):

$$g(i) = \begin{cases} i - \frac{N+1}{4} & \text{for } i \leq \frac{N+1}{4} \\ 0 & \text{for } \frac{N+1}{4} < i < \frac{3(N+1)}{4} \\ i - \frac{3(N+1)}{4} & \text{for } i \geq \frac{3(N+1)}{4} \end{cases}$$

Wilcoxon test (medium to long tails): $g(i) = i$

Median test (very long tails):

$$g(i) = \begin{cases} 1 & \text{for } i > \frac{N+1}{2} \\ 0 & \text{for } i \leq \frac{N+1}{2} \end{cases}.$$

Above, in the parenthesis that type of distribution is indicated for which the test has high power (Büning, 1994). As a selector

$\hat{Q} = \frac{\hat{U}_{0.05} - \hat{L}_{0.05}}{\hat{U}_{0.5} - \hat{L}_{0.5}}$ is chosen as a measure for

tailweight (Hogg, 1974); $\hat{L}_\gamma$ and $\hat{U}_\gamma$ denote the average of the smallest and largest γN order statistics, respectively, in the combined sample. Fractional items are used when γN is not an integer. The longer the tails the greater is $\hat{Q}$. The adaptive test can be defined as follows:

If $\hat{Q} \leq 2$, apply the Gastwirth test,

if $2 < \hat{Q} \leq 7$, apply the Wilcoxon test,

if $\hat{Q} > 7$, apply the Median test.

The maximum test is constructed of the same three statistics. However, because the two-sided alternative $\theta \neq 0$ is considered, the maximum of the absolute values of the standardized statistics is used. Under H_0 , expectation and variance of a linear rank statistic T are

$$E(T) = \frac{m}{N} \sum_{i=1}^N g(i)$$

and

$$Var(T) = \frac{mn}{N^2(N-1)} \left[N \sum_{i=1}^N g^2(i) - \left(\sum_{i=1}^N g(i) \right)^2 \right]$$

(Büning & Trenkler, 1994, pp. 127-130). Let ST_G , ST_W , and ST_M denote the standardized statistic $\frac{T - E(T)}{\sqrt{Var(T)}}$ using Gastwirth, Wilcoxon,

and Median scores, respectively. Then, the test statistic of the maximum test considered here is $T_{\max} = \max(|ST_G|, |ST_W|, |ST_M|)$. Inference is based on the permutation distribution of this maximum.

Table 1 shows type I error rates and powers of the univariate tests, the adaptive test and the maximum test. According to these results, the maximum test is less conservative than the other tests (for $\alpha = 0.05$). This finding also holds for other maximum tests (see e.g.

Table 1. Type I error rates (simulated for the adaptive test) and simulated powers of different permutation tests, the adaptive test and the corresponding maximum test ($n = m = 10$, $\alpha = 0.05$, 10,000 simulation runs for each configuration)

Distribution	θ	Gastwirth test	Wilcoxon test	Median test	Adaptive test	Maximum test
Uniform on (0, 1)	0	0.042	0.043	0.023	0.042	0.049
	0.4	0.880	0.751	0.365	0.758	0.854
Standard normal	0	0.042	0.043	0.023	0.044	0.049
	1.5	0.755	0.854	0.625	0.834	0.835
Cauchy	0	0.042	0.043	0.023	0.039	0.049
	3	0.264	0.683	0.711	0.684	0.742

Neuhäuser et al., 2004). According to Table 1, the maximum test is more powerful than the adaptive test. Moreover, an important point is that the maximum test can be more powerful than the best univariate test, as it is here in the case of the Cauchy distribution. In contrast, the power of an adaptive test is always a weighted average of the powers of the univariate tests, i.e. the power of the adaptive test is always between the best and worst of the powers of the univariate tests.

Conclusion

The use of the maximum of several competing univariate statistics is quite common nowadays, especially in statistical genetics (see the references given above). Here, it is demonstrated that such maximum tests can be integrated within the 30-year old theory of adaptive tests. The distribution of the maximum can be determined by generating all possible permutations. The p-value of the resultant exact permutation test is the proportion of permutations yielding a statistic as supportive or more supportive of the alternative than the originally observed test statistic. When sample sizes are large and/or a multi-sample problem is considered, permutation tests can be performed using a simple random sample from all the possible permutations (see e.g. Good, 2000).

In some applications the correlation between the different statistics is known and the (asymptotic) distribution of the maximum is available, an example is a multiple contrast test where the maximum is multivariate t -distributed (see e.g. Hothorn et al., 1997, Genz & Bretz, 2002). However, to use a standard distribution is not generally a better way than using the permutation distribution. Instead, permutation tests may be preferable for several applications (Ludbrook & Dudley, 1998). Note that an approximation using the asymptotic distribution of a maximum statistic can be poor even when all univariate statistics are asymptotically normal (Freidlin & Korn, 2002).

Some decades ago, permutation tests were “almost never quick ... seldom practical, and often ... not even feasible” (Bradley, 1968, p. 84). Thus, maximum tests based on the permutation distribution could not be carried out. As an alternative method to univariate tests the concept of maximin efficiency robust tests (MERT) was introduced in order to obtain a single robust test statistic from a set of possible statistics (Gastwirth, 1966, 1970). The MERT idea is to maximize the minimum asymptotic efficiency over the possible tests.

Recently, MERTs were compared with the corresponding maximum tests (Freidlin et al., 1999, 2002, 2003b; Freidlin & Korn, 2002, Gastwirth & Freidlin, 2000, Neuhäuser &

Hothorn, 1999, Neuhäuser et al., 2004, Zheng et al., 2002). Such a comparison depends on the minimum correlation ρ^* between two of the univariate tests. When this correlation is small the maximum test is often preferable to the MERT, in particular in case of $\rho^* \leq 0.5$. For $\rho^* \geq 0.7$ there was, however, virtually no difference in their powers (Freidlin et al., 1999, 2002; Freidlin & Korn, 2002, Gastwirth & Freidlin, 2000, Neuhäuser et al., 2004, Zheng et al., 2002). Other linear combinations than the MERT are further alternatives to the maximum test, see e.g. Chi and Tsai (2001).

Instead to use the maximum test statistic one may use the minimum p-value (see e.g. Weichert & Hothorn, 2002). Such a procedure is essentially Tippett's combination, although the latter was introduced for independent tests. However, other combination functions could be used as well, see Pesarin (2001) for an overview of nonparametric combination methodology which is outside the scope of this article. However, irrespective of the method used to combine the different tests, it is often difficult to select them. This is, of course, also the case for the classical adaptive test. On the one hand, statistics with low correlation may be suitable because they focus on different areas of the alternative hypothesis. On the other hand, the penalty for using more than one statistic may also depend on the correlation as the comparison maximum test versus MERT does. Hence, there seems to be no general principle to select the test statistics, but, in contrast to adaptive tests, a maximum test neither needs a selector statistic nor the specification of which test should be performed for which values of the selector.

References

- Beier, F. & Büning, H. (1997). An adaptive test against ordered alternatives. *Computational Statistics and Data Analysis* 25, 441-452.
- Bradley, J. (1968). *Distribution-free statistical tests*. Englewood Cliffs: Prentice-Hall.
- Bretz, F. & Hothorn, L. A. (2001). Testing dose-response relationships with a priori unknown, possibly non-monotonic shapes. *Journal of Biopharmaceutical Statistics* 11, 193-207.
- Büning, H. (1991). *Robuste und adaptive Tests*. Berlin: De Gruyter.
- Büning, H. (1994). Robust and adaptive tests for the two-sample location problem. *OR Spektrum* 16, 33-39.
- Büning, H. (2000). Robustness and power of parametric, nonparametric, robustified and adaptive tests – the multi-sample location problem. *Statistical Papers* 41, 381-407.
- Büning, H. (2002). An adaptive distribution-free test for the general two-sample problem. *Computational Statistics* 17, 297-313.
- Büning, H. & Trenkler, G. (1994). *Nichtparametrische statistische Methoden* (2nd Ed.). Berlin: De Gruyter.
- Chi, Y. & Tsai, M.-H. (2001). Some versatile tests based on the simultaneous use of weighted logrank and weighted Kaplan-Meier statistics. *Communications in Statistics – Theory and Methods* 30, 743-759.
- Chung, J. H. & Fraser, D. A. S. (1958). Randomization tests for a multivariate two-sample problem. *Journal of the American Statistical Association* 53, 729-735.
- Coakley, C. W. & Heise, M. A. (1996). Versions of the sign test in the presence of ties. *Biometrics* 52, 1242-1251.
- Freidlin, B. & Korn, E. L. (2002). A testing procedure for survival data with few responders. *Statistics in Medicine* 21, 65-78.
- Freidlin, B., Miao, W. & Gastwirth J.L. (2003a): On the use of the Shapiro-Wilk test in two-stage adaptive inference for paired data from moderate to very heavy tailed distributions. *Biometrical Journal* 45, 887-900.
- Freidlin, B., Podgor, M. V. & Gastwirth, J. L. (1999). Efficiency robust tests for survival or ordered categorical data. *Biometrics* 55, 883-886.
- Freidlin, B., Zheng, G., Li, Z. & Gastwirth, J.L. (2002): Trend tests for case-control studies of genetic markers: Power, sample size and robustness. *Human Heredity* 53, 146-152.
- Freidlin, B., Zheng, G., Li, Z. & Gastwirth, J. L. (2003b). Efficiency robust tests for mapping quantitative trait loci using extremely discordant sib pairs. *Human Heredity* 55, 117-124.

Gastwirth, J. L. (1966). On robust procedures. *Journal of the American Statistical Association* 61, 929-948.

Gastwirth, J. L. (1970). On robust rank tests. In: Puri, M. L. (ed.) *Nonparametric techniques in statistical inference*. Cambridge University Press, Cambridge, pp. 89-109.

Gastwirth, J. L. & Freidlin, B. (2000). On power and efficiency robust linkage tests for affected sibs. *Annals of Human Genetics* 64, 443-453.

Genz, A. & Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* 11, 950-971.

Good, P. I. (2000): *Permutation tests*. Springer, New York (2nd edition).

Hill, N. J., Padmanabhan, A. R. & Puri, M. L. (1988). Adaptive nonparametric procedures and applications. *Applied Statistics* 37, 205-218.

Hirotsu, C. (1986). Cumulative chi-squared statistic or a toll for testing goodness of fit. *Biometrika* 73, 165-173.

Hogg, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association* 69, 909-927.

Hothorn, L. A., Neuhauser, M. & Koch, H.-F. (1997). Analysis of randomized dose finding studies: Closure test modifications based on multiple contrast tests. *Biometrical Journal* 39, 467-479.

Ludbrook, J. & Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *American Statistician* 52, 127-132.

Marozzi, M. (2004a). A bi-aspect nonparametric test for the two-sample location problem. *Computational Statistics and Data Analysis* 44, 639-648.

Marozzi, M. (2004b) A bi-aspect nonparametric test for the multi-sample location problem. *Computational Statistics and Data Analysis* 46, 81-92.

Neuhauser, M., Büning, H. & Hothorn, L. A. (2004). Maximum test versus adaptive tests for the two-sample location problem. *Journal of Applied Statistics* 31, 215-227.

Neuhauser, M. & Hothorn, L. A. (1997). Adaptive tests for trend. In: Kitsos, C. P., Edler, L. (eds.): *Industrial Statistics: Aims and Computational Aspects*, Physica-Verlag, Heidelberg, pp. 269-273.

Neuhauser, M. & Hothorn, L. A. (1999). An exact Cochran-Armitage test for trend when dose-response shapes are a priori unknown. *Computational Statistics and Data Analysis* 30, 403-412.

Neuhauser, M., Seidel, D., Hothorn, L. A. & Urfer, W. (2000). Robust trend tests with application to toxicology. *Environmental and Ecological Statistics* 7, 43-56.

O'Gorman, T.W. (2001). An adaptive permutation test procedure for several common tests of significance. *Computational Statistics and Data Analysis* 35, 335-350.

Pesarin, F. (2001): *Multivariate permutation tests*. Chichester: Wiley,

Randles, R. H. & Hogg, R. V. (1973). Adaptive distribution-free tests. *Communications in Statistics* 2, 337-356.

Randles, R. H. & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*. New York, N.Y.: Wiley.

Weerahandi, S. (1995). *Exact statistical methods for data analysis*. New York, N.Y.: Springer.

Weichert, M. & Hothorn, L. A. (2002). Robust hybrid tests for the two-sample location problem. *Communications in Statistics – Simulation and Computation* 31, 175-187.

Zheng, G., Freidlin, B. & Gastwirth, J. L. (2002). Robust TDT-type candidate-gene association tests. *Annals of Human Genetics* 66, 145-155.

Inference for $P(Y < X)$ for Exponential and Related Distributions

Vee Ming Ng
School of Engineering Sciences,
Murdoch University, Australia

Some tests and confidence bounds for the reliability parameter $R = P(Y < X)$ are proposed, where X and Y are independent random variables from a two-parameter exponential distribution. The results are based on missing or incomplete data and are applicable to some related distributions.

Key words - Confidence bounds, exponential distribution, missing data, P-value, reliability parameter

Introduction

The problem of estimating and testing the reliability parameter $R = P(Y < X)$ has been widely researched in the literature. The problem originated in the context of reliability of a component of strength X subjected to a stress Y , the component failing if and only if at any time the applied stress is greater than its strength. Other applications for the reliability parameter exists when X and Y have different interpretation, such as when Y is the response for a control group and X is the response for the treatment group. Inference on R shall be considered when X and Y are random variables from a two-parameter exponential distribution. Inference on R for the one-parameter exponential distribution can be found in Enis and Geisser (1971), Tong (1977), and Chao (1982) among others.

Gupta and Gupta (1988) derived and compared some point estimators for R in the case of two independent exponential variables having a common scale parameter. For the case in which the location parameter is common, Bai

and Hong (1992) discussed point and interval estimation of R and Baklizi (2003) compared the performance of several types of asymptotic, approximate, and bootstraps confidence intervals. Ali, Woo, and Pal (2004) considered test and estimation of R when the scale parameters are equal and known and also inference procedures for R which are based on likelihood ratio tests for equality of scale and equality of location parameters.

This article considers some tests and confidence bounds for $P\{Y < X\}$ for the two-parameter exponential distribution with a common but unknown scale parameter and also with a common but unknown location parameter. Exact tests and confidence bounds are derived in situations where data may be missing or incomplete, the situation with complete data being a special case. These results are extended to some related distributions.

Methodology and Results

A two-parameter exponential distribution with parameters (μ, σ) is defined by the probability density function:

$$f(x; \mu, \sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}, \quad x > \mu, \quad \sigma > 0$$

Suppose X and Y are independent exponential random variables with parameters (μ_x, σ_x) and (μ_y, σ_y) and probability density

Vee Ming Ng is lecturer at Murdoch University. His research interests include statistical inference and prediction and multivariate analysis. Email: V.Ng@murdoch.edu.au

functions $f(x; \mu_x, \sigma_x)$ and $f(y; \mu_y, \sigma_y)$ respectively. Then

$$R \equiv P(Y < X) \\ = \int \int_{\mu_y < y < x} f(x; \mu_x, \sigma_x) f(y; \mu_y, \sigma_y) dy dx \\ = \begin{cases} \frac{\sigma_x}{\sigma_x + \sigma_y} e^{-\delta/\sigma_x}, \delta \geq 0 \\ 1 - \frac{\sigma_y}{\sigma_x + \sigma_y} e^{\delta/\sigma_y}, \delta < 0 \end{cases}$$

where $\delta = \mu_y - \mu_x$. Inference on R is considered for two cases: (a) scale parameters are equal and unknown and (b) location parameters are equal and unknown.

Assuming two independent samples of size n and m from the exponential distributions with parameters (μ_x, σ_x) and (μ_y, σ_y) respectively, let $X_q < X_{q+1} < \dots < X_r$ and $Y_l < Y_{l+1} < \dots < Y_p$ denote the ordered observations; some of these could be missing where $q=1, r=n$, and $l=1, p=m$ would correspond to all observations being available.

Let $S_x = \sum_{i=q+1}^r c_i (n-i+1)(X_i - X_{i-1})$ $c_i = 1$

or 0; $S_y = \sum_{j=l+1}^p d_j (m-j+1)(Y_j - Y_{j-1})$

$d_j = 1$ or 0, and $S_p = S_x + S_y$, $v_x = \sum_{i=q+1}^r c_i$,

$v_y = \sum_{j=l+1}^p d_j$, $v = v_x + v_y$. It is well known

that:

- X_q, Y_l, S_x, S_y, S_p are statistically independent (see Tanis (1964), Likes (1974)).
- $2S_x / \sigma_x, 2S_y / \sigma_y, 2S_p / \sigma$ when $\sigma_x = \sigma_y = \sigma$, have chi-square distributions with $2v_x, 2v_y, 2v$ degrees of freedom respectively.

- The probability density functions of the ordered statistics X_q and Y_l can be written, respectively, as

$$f(x; \mu_x, \sigma_x, n, q) \\ = \sum_{j=0}^{q-1} a(n, q, j) \frac{1}{\sigma_x} e^{-n(q,j)(x-\mu_x)/\sigma_x}$$

$$f(y; \mu_y, \sigma_y, m, l) \\ = \sum_{j=0}^{l-1} b(m, l, j) \frac{1}{\sigma_y} e^{-m(l,j)(y-\mu_y)/\sigma_y}$$

where

$$a(n, q, i) = q \binom{n}{q} \binom{q-1}{i} (-1)^i,$$

$$b(m, l, j) = l \binom{m}{l} \binom{l-1}{j} (-1)^j,$$

$$n(q, i) = n - q + i + 1,$$

$$m(l, j) = m - l + j + 1.$$

Test of hypothesis when $\sigma_x = \sigma_y = \sigma$

Suppose that $\sigma_x = \sigma_y = \sigma$ but σ is unknown, then

$$R = \begin{cases} \frac{1}{2} e^{-\lambda}, \lambda \geq 0 \\ 1 - \frac{1}{2} e^{-\lambda}, \lambda < 0 \end{cases}$$

where

$$\lambda = (\mu_y - \mu_x) / \sigma.$$

A test procedure is now derived for testing hypotheses about the reliability parameter R; a similar procedure is considered in Ranganathan and Kale (1979) for a 1-sample reliability problem. Because $P(X < Y) = 1 - R$, it suffices to consider the problem of testing the null hypothesis $H_0: \frac{1}{2} e^{-\lambda} \geq p_0$, against the

alternative $H_1: \frac{1}{2}e^{-\lambda} < p_0$, p_0 being a specified value less than 0.5. As these hypotheses are equivalent to $H_0: \lambda \geq -\ln(2p_0)$ against $H_1: \lambda < -\ln(2p_0)$, consider the test statistic $T = (Y_l - X_q) / S_p$. T is a maximal invariant and its distribution depends only on λ . A large value of T would be evidence against H_0 . Hence, for an observed value t of T , $P(T > t)$ for $t \geq 0$, $\lambda \geq 0$ is the P-value of the test, a small value of which would indicate sufficient evidence against H_0 . In order to get an expression for the P-value, one must first obtain, from the joint of density function of X_q, Y_l, S_p , the joint probability density function of $D = Y_l - X_q$ and S_p , $f(d, s; \delta, \sigma)$, which then yields the joint of density function $f(t, w; \lambda)$ of $T = D / S_p$ and $W = S_p / \sigma$:

$$f(d, s; \delta, \sigma)$$

$$= \left\{ \begin{array}{l} \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i) b(m, l, j) e^{m(l, j) \delta / \sigma} s^{v-1} e^{-[m(l, j) d + s] / \sigma}}{\Gamma(v) [n(q, i) + m(l, j)] \sigma^{v+1}} \\ , d \geq \delta, s \geq 0 \\ \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i) b(m, l, j) e^{-n(q, i) \delta / \sigma} s^{v-1} e^{[n(q, i) d - s] / \sigma}}{\Gamma(v) [n(q, i) + m(l, j)] \sigma^{v+1}} \\ , d < \delta, s \geq 0 \end{array} \right\}$$

$$f(t, w; \lambda)$$

$$= \left\{ \begin{array}{l} \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i) b(m, l, j) e^{m(l, j) \lambda} w^v e^{-w[m(l, j) t + 1]}}{\Gamma(v) [n(q, i) + m(l, j)]} \\ , tw \geq \lambda, w \geq 0 \\ \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i) b(m, l, j) e^{-n(q, i) \lambda} w^v e^{w[n(q, i) t - 1]}}{\Gamma(v) [n(q, i) + m(l, j)]} \\ , tw < \lambda, w \geq 0 \end{array} \right\}$$

The P-value, $P(T > t)$ is obtained from $f(t, w; \lambda)$, $tw \geq \lambda, w \geq 0$ as

$$P(T > t)$$

$$= \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \int_t^\infty \int_{\lambda[m(l, j) + 1/u]}^\infty \frac{a(n, q, i) b(m, l, j) e^{m(l, j) \lambda} z^v e^{-z}}{\Gamma(v) [n(q, i) + m(l, j)] [1 + m(l, j) u]^{v+1}} dz du$$

Integration by parts yields

$$P(T > t) = \sum_{i=0}^{q-1} \sum_{j=0}^{l-1} \frac{a(n, q, i) b(m, l, j)}{m(l, j) [n(q, i) + m(l, j)]} \left\{ \frac{e^{m(l, j) \lambda}}{(1 + m(l, j) t)^v} P(G > \lambda[m(l, j) + 1/t]) \right. \\ \left. + P(G < \lambda/t) \right\}$$

where G the Gamma random variable with shape parameter $v + 1$.

In many situations the first ordered statistics are available i.e. $q = 1, l = 1$ and the above simplifies to

$$P(T > t) = \frac{n}{n + m} \left\{ \frac{e^{m \lambda}}{(1 + mt)^v} P(G > \lambda[m + 1/t]) \right. \\ \left. + P(G < \lambda/t) \right\}$$

Point estimators of R for the case $q = 1, l = 1$ are considered in Gupta and Gupta (1988) where the maximum likelihood estimator of R is obtained with $T/(m + n)$ as an estimator of λ in the equation for R .

Inference when $\mu_x = \mu_y = \mu$

When $\mu_x = \mu_y = \mu$ but μ is unknown then R reduces to

$$\theta \equiv \frac{\sigma_x}{\sigma_x + \sigma_y}$$

Consider the null hypothesis
 $H_0 : \theta \geq q_0$ or equivalently $H_0 : \frac{\sigma_x}{\sigma_y} \geq \frac{q_0}{1-q_0}$

where q_0 is a specified probability. $2S_x / \sigma_x$ and $2S_y / \sigma_y$ are independently distributed as chi-square with $2v_x$ and $2v_y$ degrees of

freedom and $\frac{S_x / (v_x \sigma_x)}{S_y / (v_y \sigma_y)}$ has a F distribution

with v_x and v_y degrees of freedom. Hence, one

can use $F = \frac{S_x / v_x (1-q_0)}{S_y / v_y q_0}$ as the test statistic.

An estimate of θ is
 $\hat{\theta} = \frac{S_x / v_x}{S_x / v_x + S_y / v_y}$. A $(1-\alpha)$ confidence

interval for θ is obtainable from the F distribution with v_x and v_y degrees of freedom

via $P\{F_l < \frac{S_x \sigma_y v_y}{S_y \sigma_x v_x} < F_u\}$ where F_l and F_u

satisfies $1-\alpha = P\{F_l < F < F_u\}$. The confidence interval can be written, after some algebraic manipulation, as

$$\left(\frac{\hat{\theta}}{\hat{\theta} + (1-\hat{\theta})F_u}, \frac{\hat{\theta}}{\hat{\theta} + (1-\hat{\theta})F_l} \right)$$

When complete samples are available,

$S_x = \sum_{i=2}^n (X_i - X_1)$, $S_y = \sum_{j=2}^m (Y_j - Y_1)$ one

of which is slightly different from those used in Bai and Hong (1992). They used

$\sum_{i=1}^n (X_i - \min(X_1, Y_1))$ $\sum_{j=1}^m (Y_j - \min(X_1, Y_1))$

instead of S_x , S_y respectively and obtained approximate confidence interval based on a mixed beta distribution.

Applications to Related Distributions

Suppose X and Y are independent two-parameter exponential random variables and φ is a monotonic function with inverse φ^{-1} . Because

$$P(Y < X) = P(\varphi(Y) < \varphi(X))$$

the tests and confidence bounds developed in the previous sections are also applicable to the variables $\varphi(X)$ and $\varphi(Y)$; the results are to be applied after making the transformation, φ , to the observations. The results are applicable to the Rayleigh distribution with $\varphi(X) = \sqrt{2X}$, $\varphi^{-1}(X) = X^2 / 2$ and the Pareto distribution with $\varphi(X) = \exp(X)$, $\varphi^{-1}(X) = \ln(X)$.

Numerical example

Suppose a system has two main parts, Y and X, whose lifetimes are exponentially distributed. Suppose $m=n=15$ component parts are put on test simultaneously and the failure times are {106, 108, 109, 113, 116, 126, 127, 132, 138, 141, 147, 164, 185, 202, 285} and {79, 82, 88, 89, 91, 107, 112, 118, 133, 149, 165, 167, 170, 202, 222} for Y and X respectively. Then $l=q=1$, $c_i, d_j=1$ for $i=j=1,2,...,15$, $t=0.0193$, $s_y=609$, $s_x=789$, and $v_x=v_y=14$. To test whether system failure may be equally likely due to either part, the test of $H_0 : \lambda \geq 0$ ($R \geq 0.5$) against $H_1 : \lambda < 0$ yields a P-value of 0.0004 which is sufficient evidence that X is more likely to fail before Y. If instead one is interested to test, say, $H_0 : R = \frac{1}{2} e^{-\lambda} \geq 0.4$ against $H_1 : R < 0.4$ then the P-value is 0.011. There is sufficient evidence to reject H_0 ; the probability that system failure will be due to Y is less than 0.4. If, for example, the values 108

and 109 for Y are missing, then one would set $d_2 = d_3 = 0$ and the recalculated values for the test of $H_0 : R \geq 0.4$ are $t = 0.0199$, $s_y = 568$, and $v_y = 12$ with a P-value equal 0.016.

Conclusion

Tests of hypotheses and confidence bounds for R have been developed for the two-parameter exponential distribution in two cases, namely one involving a common scale parameter and the other a common location parameter. Exact tests for the two cases are derived for situations in which data may be missing or incomplete. Exact confidence bounds for R in the common location case are also proposed and they provide an alternative to the approximate bounds that have been considered in a complete sample situation. Furthermore, these results are applicable to a larger class of distributions which includes the Raleigh and the Pareto distributions.

References

- Ali, M. M., Woo, J. & Pal, M. (2004). Inference on reliability $P(Y < X)$ in two-parameter exponential distribution. *International Journal of Statistical Sciences*, 3, 119-125.
- Baklizi, A. (2003). Confidence intervals for $P(X > Y)$ in the exponential case with common location parameter. *Journal of Modern Applied Statistical Methods*, 2, 341-349.
- Bai, D. S. & Hong (1992). Estimation of $P(X < Y)$ in the exponential case with common location parameter. *Communications in Statistics-Theory & Method*, 21, 269-282.
- Chao, A. (1982). On comparing estimators of $\Pr\{Y < X\}$ in the exponential case. *IEEE Transaction on Reliability*, R31, 389-392.
- Enis, P. & Geisser, S. (1971). Estimation of the probability that $Y < X$. *Journal American Statistical Association*, 66, 162-168.
- Gupta, R. D. & Gupta, R. C. (1988). Estimation of $P(Y_p > \max(Y_1, Y_2, \dots, Y_{p-1}))$ in the exponential case. *Communication in Statistics-Theory & Methods*, 17, 911-924.
- Likes, J. (1974). Prediction of sth ordered observation for the two-parameter exponential distribution. *Technometrics*, 16, 241-244.
- Ranganathan, J. & Kale, B. K. (1979). Tests of the hypotheses for reliability functions in two-parameter exponential models. *Canadian Journal of Statistics*, 7, 117-184.
- Tanis, E. A. (1964). Linear forms in the order statistics from an exponential distribution. *Annals Mathematical Statistics*, 35, 270-276.
- Tong, H. (1977). On the estimation of $P(Y < X)$ for the exponential families. *IEEE Transactions on Reliability*, R26, 54-56.

An Alternative to Warner's Randomized Response Model

Sat Gupta
University of North Carolina at Greensboro

Javid Shabbir
Quaid-I-Azam University

A modification to Warner's (1965) Randomized Response Model is suggested. The suggested model is more efficient than the original model.

Key words: randomized response model, survey bias

Introduction

Warner (1965) suggested an Indirect Questioning method to circumvent social desirability response bias in surveys involving sensitive questions. If π is the proportion of subjects in a population who have a sensitive characteristic A, then Warner's method recommends using a randomization device, such as a deck of cards, to scramble the true response. A known proportion (p) of the cards in the deck have the statement "I have characteristic A" and the remaining cards in the deck have the statement "I do not have characteristic A". A participant in the survey draws a card randomly from the deck and reports his/her agreement/disagreement with the statement on the card. Thus, a respondent who actually has the characteristic A, but draws a "I do not have characteristic A" card, will give a response "no" indicating lack of agreement with the statement on the card. Probability of a "yes" response (p_y) is given by

$$p_y = p\pi + (1-p)(1-\pi). \quad (1)$$

Sat Gupta is Professor of statistics. He received Ph. D. degree in Mathematics from University of Delhi and Ph. D. degree in Statistics from Colorado State University. Javid Shabbir is Associate Professor of Statistics. Dr Shabbir received a Ph. D. degree in Statistics from Kent University at Canterbury, UK.

Equation (1) can be rewritten as

$$\pi = \frac{p_y - (1-p)}{2p-1}. \quad (2)$$

Equation (2) suggests estimating π by $\hat{\pi}_w$ where

$$\hat{\pi}_w = \frac{\frac{n_1}{n} - (1-p)}{2p-1}, \quad p \neq .5 \quad (3)$$

where n_1 is proportion of "yes" responses in a simple random sample with replacement of size n . The fact that n_1 has a binomial distribution with parameters (n, p_y) can be used to prove that $\hat{\pi}_w$ is a maximum likelihood estimator of π . Its variance is given by

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(1-2p)^2}. \quad (4)$$

The second term in the above expression is the penalty due to indirect responding. Note that the penalty is smallest when p is closest to zero or one.

Several variations of Warner's model have been proposed in the literature. These include models by Greenberg et. al (1969), Mangat and Siingh (1990) and Christofides (2002). Gupta and Thornton (2002) have attempted to validate some of these models with actual survey data.

Proposed Alternative Strategy

It is clear from (4) that Warner's model works best when p is very close to zero or to one. But, both of these cases make the scrambling deck look very suspicious because almost all of the cards will be of the same type. Using two decks of the type described above is proposed, one with a low value of p (say p_1) and the other with a high value of p (say p_2). This will increase cooperation because the respondent is less suspicious in using decks of both kinds – one with a high value of p and one with a low value of p .

A simple random sample with replacement of size n is selected and each respondent is asked to give a response using each of the two decks. Let $(Z_{1i}, Z_{2i}), i = 1, 2, \dots, n$ be the responses where $Z_{ki} = 1$ if the response using the k^{th} deck ($k = 1, 2$) is "yes" and $Z_{ki} = 0$ if the response is "no". Let $n_{1i}, i = 1, 2$, be the number of "yes" responses from the two decks. Then one can construct two estimators of the type (3). These are given by

$$\hat{\pi}_{w1} = \frac{\frac{n_{11}}{n} - (1 - p_1)}{2p_1 - 1}, \quad p_1 \neq .5, \quad (5)$$

$$\hat{\pi}_{w2} = \frac{\frac{n_{12}}{n} - (1 - p_2)}{2p_2 - 1}, \quad p_2 \neq .5 \quad (6)$$

It is easy to note from (4) that both of these estimators have the same variance if p_1 and p_2 are symmetric about .5.

We now propose the estimator

$$\hat{\pi}_p = k_1 \hat{\pi}_{w1} + k_2 \hat{\pi}_{w2}, \quad k_1 + k_2 = 1. \quad (7)$$

Obviously $\hat{\pi}_p$ is unbiased because both $\hat{\pi}_{w1}$

and $\hat{\pi}_{w2}$ are unbiased. Also, variance of $\hat{\pi}_p$ is given by

$$V(\hat{\pi}_p) = k_1^2 V(\hat{\pi}_{w1}) + k_2^2 V(\hat{\pi}_{w2}) + 2k_1 k_2 \text{Cov}(\hat{\pi}_{w1}, \hat{\pi}_{w2}). \quad (8)$$

The following lemma is proven before exploring this variance further.

Lemma 1:

$$\text{Cov}(\hat{\pi}_{w1}, \hat{\pi}_{w2}) = \frac{\pi(1-\pi)}{n} \quad (9)$$

Proof:

$$\text{Note that } n_{1i} = \sum_i Z_{1i}.$$

Hence,

$$\begin{aligned} \text{Cov}(n_{11}, n_{12}) &= \text{Cov}\left(\sum_i Z_{1i}, \sum_i Z_{2i}\right) \\ &= \sum_i \text{Cov}(Z_{1i}, Z_{2i}) \end{aligned}$$

In the above block in the middle equation, please change the second summation to $\sum_j Z_{2j}$

because Z_{1i} and Z_{2j} are independent for $i \neq j$.

Hence,

$$\begin{aligned} \text{Cov}(n_{11}, n_{12}) &= n \text{Cov}(Z_{11}, Z_{12}) = n \{E(Z_{11} Z_{12}) - E(Z_{11})E(Z_{12})\} \\ &= n \{p_{yy} - p_{y1} p_{y2}\}, \end{aligned} \quad (10)$$

where p_{yy} is the probability of a "yes" response with both decks, p_{y1} is the probability of a "yes" response with Deck 1 and p_{y2} is the probability of a "yes" response with Deck 2. The following is provided as in (1).

$$p_{yi} = p_i \pi + (1 - p_i)(1 - \pi), \quad i = 1, 2 \quad (11)$$

and

$$p_{yy} = \pi p_1 p_2 + (1 - \pi)(1 - p_1)(1 - p_2). \quad (12)$$

Substituting (11) and (12) in (10), one can easily obtain

$$\text{Cov}(n_{11}, n_{12}) = n\pi(1 - \pi)(2p_1 - 1)(2p_2 - 1) \quad (13)$$

The lemma follows easily from (5), (6) and (13). Also, it is easy to verify that when $p_1 + p_2 = 1$, $p_1 \neq p_2$, the optimum values of (k_1, k_2) in (8) are $(.5, .5)$. This is because $V(\hat{\pi}_{w1}) = V(\hat{\pi}_{w2})$ if $p_2 = 1 - p_1$. With these choices for (k_1, k_2) , our proposed estimator becomes

$$\hat{\pi}_p = \frac{\hat{\pi}_{w1} + \hat{\pi}_{w2}}{2}. \quad (14)$$

As remarked earlier, $\hat{\pi}_p$ is unbiased because both $\hat{\pi}_{w1}$ and $\hat{\pi}_{w2}$ are unbiased.

Theorem 1:

When $p_1 + p_2 = 1$ and $p_1 \neq p_2$, estimator $\hat{\pi}_p$ is more efficient than Warner's estimators $\hat{\pi}_{w1}$ and $\hat{\pi}_{w2}$.

Proof:

Note that

$$\begin{aligned} V(\hat{\pi}_p) &= \frac{1}{4} \{V(\hat{\pi}_{w1}) + V(\hat{\pi}_{w2}) + 2 \text{Cov}(\hat{\pi}_{w1}, \hat{\pi}_{w2})\} \\ &= \frac{1}{2} V(\hat{\pi}_{w1}) + \frac{1}{2} \frac{\pi(1 - \pi)}{n}, \end{aligned}$$

because $V(\hat{\pi}_{w1}) = V(\hat{\pi}_{w2})$.

$$< V(\hat{\pi}_{w1})$$

because

$$\frac{\pi(1 - \pi)}{n} < V(\hat{\pi}_{w1}), \text{ from (4).}$$

Numerical Examples

In this section, the efficiency of the proposed estimator is compared with Warner's estimator for various choices of π , p_1 and p_2 ($p_2 = 1 - p_1$). Note that the proposed estimator is more efficient than Warner's estimator, as expected, for all choices of the parameters.

Table1: Efficiency of the proposed estimator compared to Warner's estimator

π	$p_1 = 0.1$	$p_1 = 0.2$	$p_1 = 0.3$	$p_1 = 0.4$
0.1	143.860	171.171	187.940	197.087
0.2	130.529	158.140	180.398	194.937
0.3	125.084	151.414	175.758	193.458
0.4	122.659	148.077	173.222	192.593
0.5	121.951	147.059	172.414	192.308
0.6	122.659	148.077	173.222	192.593
0.7	125.084	151.414	175.758	193.458
0.8	130.529	158.140	180.398	194.937
0.9	143.860	171.171	187.940	197.087

Conclusion

The proposed strategy is likely to induce greater cooperation from the survey participants because it provides greater diversity in the scrambling process. Moreover, the proposed strategy is clearly more efficient than Warner's model, particularly for higher values of p_1 .

References

Christofides, T. C. (2002). A generalized randomized response technique, *Metrika*, 57, 195-200

Greenberg, B. G, Abul-ela, A. L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical frame work. *Journal of American Statistical Association*, 64, 520-539.

Gupta, S. & Thornton, B. (2002). Circumventing social desirability response bias in personal interview surveys. *American Journal of Mathematical and Management Sciences*, 22, 369-383.

Mangat, N. S. & Singh, R. (1990), An alternative randomized response procedure. *Biometrika*, 77, 439-442.

Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *Journal of American Statistical Association*, 60, 63-69.

The Influence of Reliability on Four Rules for Determining the Number of Components to Retain

Gibbs Y. Kanyongo
Duquesne University

Imperfectly reliable scores impact the performance of factor analytic procedures. A series of Monte Carlo studies was conducted to generate scores with known component structure from population matrices with varying levels of reliability. The scores were submitted to four procedures: Kaiser rule, scree plot, parallel analysis, and modified Horn's parallel analysis to find if each procedure accurately determines the number of components at the different reliability levels. The performance of each procedure was judged by the percentage of the number of times that the procedure was correct and the mean components that each procedure extracted in each cell. Generally, the results show that when component loading was high, an increase in reliability resulted in an improvement of the accuracy of parallel analysis and modified horn's parallel analysis.

Key words: Monte Carlo, principal components analysis, factor analysis, parallel analysis, modified Horn's parallel analysis, scree plot, Kaiser rule.

Introduction

When dealing with a large number of variables, it is possible that some of those variables are highly correlated with one another such that they account for the same variance in the dependent variable. In such cases it may be possible to combine several variables into one factor without any substantial loss of information. This reduces the number of variables thereby facilitating the interpretation of the data and is achieved through several factor analytic procedures.

Factor analysis is a term used to refer to statistical procedures used in summarizing relationships among variables in a parsimonious but accurate manner. It is a generic term that includes several types of analyses, including

(a) common factor analysis, (b) principal component analysis (PCA), and (c) confirmatory factor analysis (CFA). Common factor analysis may be used when a primary goal of the research is to investigate how well a new set of data fits a particular well-established model (Merenda, 1997). Principal component analysis is usually used to identify the factor structure or model for a set of variables (Stevens, 2002). In contrast, confirmatory factor analysis is based on a strong theoretical foundation that allows the researcher to specify an exact model in advance. In this study, PCA will be of primary interest. Under PCA, the focus is going to be on procedures by which the number of components is determined. Specifically, two forms of parallel analysis, Horn's (1965) parallel analysis (HPA) and modified Horn's parallel analysis (MHPA), are going to be compared to the Kaiser (1960) rule and scree plot procedures under systematically varied conditions of reliability, component loading and variable-to-component loading.

Users of PCA are required to make decisions on a number of technical issues, including the number of components to retain, method of extraction and rotation techniques. Perhaps, the most important decision is that of determining how many components to retain

Gibbs Y. Kanyongo is Assistant Professor in the School of Education, Department of Foundations and Leadership at Duquesne University. He teaches courses in Educational Statistics and Research Methods.

(Merenda, 1997). Problems emerge when non-optimal numbers of components are extracted. Under-extraction compresses variables into a smaller number of components than what actually exists in the data, resulting in loss of important information, a neglect of potentially important components, a distorted fusing of two or more components, and an increase in error loadings (O'Connor, 2000). Over-extraction diffuses variables into a larger number of components than what actually exists in the data, potentially resulting in components with few components loadings and researchers' attributing excessive substantive importance to trivial components (O'Connor). Fava and Velicer (1992) emphasize that researchers should employ the most accurate procedures for determining the correct number of components in order to minimize such problems.

Methods for Determining the Number of Components

The four methods covered in this study are Kaiser rule, scree plot, Horn's parallel analysis procedure and modified Horn's parallel analysis procedure. Each of these methods is covered in detail below.

Kaiser rule.

The easiest and most commonly used method is to retain all components with eigenvalues greater than 1.0 procedure, which is known as the Kaiser rule. This method only provides a rough estimate of the optimal number of components that can be used to describe the data (Hutcheson & Sofroniou, 1999). The suggestion by Kaiser (1960) was based on a commonly used formula for the reliability of a total score (Cliff, 1988). This formula was first suggested by Kuder and Richardson (1937), and is called Kuder-Richardson formula 20 (K-R 20). The formula estimates the parallel form reliability of a total score from the internal consistency among subscores or items. The Kaiser rule uses the rationale that there are as many reliable components as there are eigenvalues greater than one. The reasoning is that an eigenvalue less than one implies that the scores on the component would have negative reliability.

Cliff (1988) argued that Kaiser's rationale for relating the reliability of components to the number of eigenvalues greater than one was based on a misapplication of a common formula for the reliability of a composite. The reliability of a principal component depends on the reliability of the measures. He pointed out that reliability of components cannot be deduced from the size of the eigenvalues, and that the Kaiser rationale for retaining as many components as there are eigenvalues greater than one does not have any logical basis.

The Kaiser rule continues to be one of the most widely employed, largely because the most widely used statistical packages (i.e., SPSS, SAS) continue to include this method as the default option. While this method continues to be widely used, several studies have shown it to be highly inaccurate and to tend to overestimate the number of components to retain (Velicer et al., 2000; Zwick & Velicer 1986). Linn (1968) performed a Monte Carlo study of the Kaiser rule and the degree of overestimation observed was extreme. On average the Kaiser rule overestimated the correct number of components by 66 percent.

In a simulation study conducted by Zwick and Velicer (1986) to compare different component extraction procedures, the Kaiser rule showed the poorest performance, indicating the correct number of components only 22 percent of the time. In some cases the Kaiser rule may sometimes lead to the extraction of fewer components than should have been extracted (Guttman 1954).

Cattell and Jaspers (1967) conducted a study to evaluate the accuracy of the Kaiser rule. In that study, the Kaiser rule was accurate when the number of variables was small (10 to 15) or moderate (20 to 30) and the communalities are high (.80). The communality of a variable is the amount of variance on a variable accounted for by the set of factors. However, the rule showed overestimation with a large number of variables (40) and low communalities (about .40). Stevens (2002) recommended that the Kaiser rule be used only when the number of variables is less than 30, sample size is more than 250, and when the mean communality is about .60. In summary, the Kaiser rule, although commonly used, is

believed by many researchers to overestimate (e.g., Velicer et al., 2000; Zwick & Velicer 1986) and by others to sometimes underestimate (e.g., Guttman 1954).

Cattell's scree plot

This procedure was proposed by Cattell (1966). With this procedure eigenvalues are plotted against their ordinal numbers and one examines to find where the break or a leveling of the slope of the plotted line occurs. Tabachnick and Fidell (2001) referred to the break point as the point where a line drawn through the points changes direction. The number of components is indicated by the number of eigenvalues above the point of the break. The eigenvalues below the break indicate error variance (Velicer et al., 2000). The graph for scree plot is available as an option in SPSS, and most other statistical programs.

Many studies have found this method to be reasonably effective in suggesting the correct number of components to retain (Catell & Jaspers, 1967; Catell & Vogelmann, 1977; Cliff, 1970; Linn, 1968; Tucker, Koopman & Linn, 1969; Velicer et al., 2000; Zwick & Velicer, 1982). Hakstian, Rogers and Cattell (1982) noted that when sample size is more than 250 and mean communality is .60, the scree plot extracts the correct number of components. However, Hakstian et al. found that the scree plot is less accurate with low communality data that resulted in the overestimation of the number of components to retain.

Zwick and Velicer (1982) found that the scree plot was especially effective with large sample sizes and with strong components. Stevens (2002) suggested that with this procedure, sample sizes greater than 200 are reasonable provided most of the communalities are large. Cliff (1970) and Linn (1968) found the method to be less accurate with small sample sizes.

Zwick and Velicer (1986) conducted a later study in which they found that the scree test was less accurate than several other methods they investigated. The visual rationale has potential to avoid some of the over-extraction problems because the trivial factors will not be visually compelling (Velicer et al., 2000). The scree procedure is recommended for use with

other procedures, not as a stand-alone procedure. In their studies Velicer et al. evaluated this procedure to be easy to implement because computer programs typically produce the eigenvalues. On accuracy of the procedure, they found mixed results; that is, some results indicated that the procedure extracted the correct number of components and others showed the procedure to over-extract.

Horn's parallel analysis.

Horn (1965) introduced HPA method for determining the number of components as an alternative to the Kaiser procedure. Horn's parallel analysis procedure involves the generation of a set of random data correlation matrices with the same number of variables and participants as the observed data (Velicer et al., 2000). Horn (1965) proposed that a number, say 50, of correlation matrices of p uncorrelated random normal variables and a sample size equal to n , where p and n are the same as the corresponding entries in the data set under study, be constructed and their mean eigenvalues across all the replications be determined. The eigenvalues of the observed data are then compared to the mean eigenvalues of the random data across the replications. The components with eigenvalues of the observed data that exceed the mean eigenvalues of the corresponding components of random data are retained because the observed eigenvalue is considered a real effect that is not likely to be due to chance sampling variability. Actual eigenvalues less than or equal to the mean random eigenvalues would be considered as due to random sampling variability (Glorfeld, 1995). Horn's parallel analysis can be taken to be a way of simulating the scree plot, while on the other hand the Kaiser rule is the theoretical scree plot.

Previous simulation studies have shown that Horn's parallel analysis procedure is accurate in determining the number of components to retain in PCA (Glorfeld, 1995; O'Connor, 2000; Velicer et al. 2000; Zwick & Velicer, 1986). Since this procedure makes use of mean eigenvalues, it requires a comparison of observed eigenvalues and mean eigenvalues of random data (Horn, 1965).

Modified Horn's parallel analysis

Glorfeld (1995) suggested a modification of Horn's parallel analysis procedure to come up with another procedure; modified Horn's parallel analysis which allows identification of any desired upper $(1-\alpha)$ percentile, such as the 95th percentile of the set of distribution (Glorfeld, 1995). This percentile is then used to determine whether the eigenvalue obtained is larger than what could be expected by chance.

Glorfeld (1995) argued that the use of mean eigenvalues is equivalent to conducting the analysis at the 50 percent significance level in conventional hypothesis testing. At this significance level, Horn's parallel analysis tends to over-extract by including minor components in the extraction. He noted that although Horn's parallel analysis is relatively accurate; it still tends to extract one or two more factors than is actually warranted and therefore sometimes retains poorly defined components. The modified Horn's parallel analysis compares real data eigenvalues to the eigenvalues at the desired percentile (typically the 95th) of the distribution of random data eigenvalues (Cota, Longman, Holden, Fekken & Xinaris, 1993; Glorfeld, 1995).

Buja and Eyuboglu (1992) advanced the same argument as Glorfeld (1995) when they pointed out, "the use of null averages as thresholds implies that the classical version of parallel analysis performs approximately at the .5 significance level" (p. 511). They suggested that a higher percentile be used rather than the mean. Buja and Eyuboglu's (1992) suggestion was also in line with the findings by Harshman and Reddon (1983), who pointed out that using the mean eigenvalue represented a potential flaw in the HPA procedure.

The Effects of Extracting the Wrong Number of Components

Over-extraction

Over-extraction is a situation where more components are extracted than the actual number of components in the data. Fava and Velicer (1992) noted, "A primary example of misspecification occurs as a result of employing Kaiser rule, it is likely to over-extract components ..." (p. 388). Moiser (1939)

conducted a simulation study to investigate the impact of over-extraction on the overall component structure. His findings provided some empirical evidence that supports the notion that the error of over-extraction is less harmful than the error of under-extraction (Gorsuch, 1983, Fava & Velicer, 1992).

There are two theoretical justifications that support the idea that over-extraction may not be as serious problem as under-extraction (Fava & Velicer, 1992). The first concerns the fact that the amount of variance explained by a component decreases for each succeeding component extracted (Cattell, 1958, Comrey, 1973), for example, the fifth extracted component will account for less variance than the fourth. This means if a true solution were a four component solution, it would be a more serious error to only extract three components than to over-extract five, because the information lost by the fourth component will be greater than the error added by the fifth component (Fava & Velicer, 1992). The second theoretical justification concerns the idea that, upon rotation, it is relatively easy to discard extra components that have been retained as trivial components without changing the substantive components (Cattell, 1958, Comrey, 1973).

The work by Fava and Velicer (1992) indicated that over-extraction does have negative effect, especially if component loading is low or sample size is low. They found that the worst effects occurred for cases of combined low component loading and low sample size during maximal over-extraction. In their results, they also noted, "there was also strong support for the hypothesis that over-extraction of a few (one or two) components will not cause major negative effects" (p. 413).

Under-extraction

Retaining too few components is another form of misspecification that occurs in PCA. Researchers point that logical arguments generally support the idea that under-extraction is a more serious problem than over-extraction. One argument is that the amount of explained variance decreases for each succeeding component extracted (Cattell, 1958, Comrey, 1973). This suggests that increasing the

dimensionality of a correct solution by one would result in less error variance relative to the amount of true variance that would be removed by decreasing the dimensionality by one (Fava & Velicer, 1996). It appears the general notion is that under-extraction is a more serious problem than over-extraction (Cattell, 1958, Comrey, 1973, Fava & Velicer, 1996). Although under-extraction is generally considered a worse problem than over-extraction, it has been studied less than over-extraction (Fava & Velicer, 1996). This means there is not much evidence to support the opinions that under-extraction is more serious.

Reliability

The decision to include reliability in this study is mainly because previous Monte Carlo studies to compare PCA procedures did not take reliability into account (e.g., Glorfeld, 1995, Velicer et al., 2000, Zwick & Velicer, 1986). These Monte Carlo studies generated the scores with the implicit assumption that the scores had perfect reliability. However, in this study, the scores were generated at varying levels of reliability to see whether the PCA procedures performed differently at each level of reliability.

Reliability is one of the most important considerations when selecting variables for analyses (Tabachnick & Fidell, 2001). Tabachnick and Fidell pointed out that unreliable variables degrade an analysis while reliable variables enhance the analysis. If variables are unreliable, the entire solution may not be trusted as it may contain a lot of measurement error. Gorsuch (1983) noted that if the variables have low reliabilities, then many more individuals would be needed for PCA. It is important that reliability be considered in studies dealing with PCA because in real world phenomenon perfectly reliable data rarely exist.

Reliability is important in principal components analysis because PCA studies correlations among variables and these correlations are impacted by reliability. When reliability is low, these correlations are reduced and consequently the magnitude of the components is also diminished. Also, PCA assumes perfect reliability, since it has 1.0 on the diagonal of the population correlation matrix. On the other hand, principal axis

factoring accommodates reliability and does not get impacted that much by a change in reliability.

McMillan and Schumacher (2001) advised that in any study, the reliability of scores should be established before the research is undertaken, and the type of reliability should be consistent with the use of the results. Reliability is a function of the trait being measured. This means some variables, such as most measures of achievement, provide highly reliable scores; whereas scores from measures of personality traits have lower reliability. It is common to have reliability values of .80 or above for achievement tests, whereas values of .70 may be acceptable for measuring personality traits (McMillan & Schumacher, 2001). What this means is that an achievement test with a reliability of .70 may be seen to be weak whereas a personality instrument with a reliability coefficient of .90 is considered excellent.

Tuckman (1999) provided the following factors that affect reliability of a measurement instrument, (a) familiarity with the particular measurement instrument, (b) participant fatigue, (c) emotional strain, (d) physical conditions of the room in which the test is administered, (e) participant's health, (f) participant's practice or experience in the specific skill being measured, and (g) specific knowledge gained outside the experience being evaluated by the measurement instrument. McMillan and Schumacher (2001) noted that to enhance reliability, it is best to establish standard conditions of data collection. For example, all participants should be given the same directions and have the same time frame in which to answer questions at the same time during the day. Error is often increased if different people administer the instrument.

Methodology

Monte Carlo Procedure

Monte Carlo simulations perform functions empirically through the analysis of random samples from populations whose characteristics are known to the researcher (Brooks, Barcikowski & Robey, 1999). That is, Monte Carlo methods use computer assisted simulations to provide evidence for problems

that cannot be solved mathematically, such as when the sampling distribution is unknown or hypothesis is not true.

The principle behind Monte Carlo simulation is that the behavior of a statistic in a random sample can be assessed by the empirical process of actually drawing many random samples and observing this behavior (Mooney, 1997). The idea is to create a pseudo-population through mathematical procedures for generating sets of numbers that resemble samples of data drawn from the population.

Mooney (1997) noted that other difficult aspects of the Monte Carlo design are writing the computer code to simulate the desired data conditions and interpreting the estimated sampling plan, data collection, and data analysis. An important point to note is that a Monte Carlo design takes the same format as a standard research design. This was noted by Brooks, Barcikowski, and Robey, (1999) when they wrote "It should be noted that Monte Carlo design is not very different from more standard research design, which typically includes identification of the population, description of the sampling plan, data collection and data analysis" (p. 3).

Like any methodology, Monte Carlo studies are not without disadvantages. Their usefulness depends in large part on the realism of the conditions that are modeled (Hutchinson & Bandalos, 1997). This means in setting up a Monte Carlo design, conditions set should resemble those found in practice otherwise results obtained will be of less utility. In this study, the conditions that are modeled are those that are encountered in real life situations (Zwick & Velicer, 1986). Another factor influencing results in a Monte Carlo study is the number of replications. This means in order to have a good model of a sampling distribution several replications should be performed.

Pseudo-Population

The conditions in the pseudo-population that were manipulated in this study are sample size, reliability, number of components, and variable-to-component ratio. These independent variables were selected on the basis of their importance in applied research and their ability to distinguish the accuracy of the PCA

procedures (Velicer et al., 2000, Zwick & Velicer, 1986). The conditions used in this study and the levels within them are given below:

Conditions Investigated

Component loading (a_{ij}).

The component loading is the Pearson correlation between a component and a variable. It is a measure of the degree of generalizability found between each variable and each component. This condition had two levels; a moderate coefficient of .50 and a very strong coefficient of .80 to represent small and large component loading respectively. These values were used by Zwick and Velicer (1986) because they generally bridge the range found in applied research situations and were shown to differentially affect the accuracy of the component extraction procedures. Component loading has been found to be one of the factors having the greatest effect on the accuracy of PCA procedures (Guadagnoli & Velicer, 1988, Velicer et al., 2000, Zwick & Velicer, 1986).

Reliability (ρ_{xx}).

Three levels of this variable were included. These levels are .60, .80 and 1.0. The value of .60 was included because it represents the near minimum acceptable reliability in applied research. The value of .80 was included because a reliability of .80 is generally considered good for most measures (McMillan & Schumacher, 2001). Previous Monte Carlo studies on PCA have assumed a reliability of 1.0. So this value was included in this study for comparison purposes with earlier studies. In this study each of the 24 variables had uniform reliability. That is they all had either reliability of .6, .8 or 1.0.

Variable-to-Component Ratio ($p:m$).

The number of components chosen was three and six, but because the number of variables was chosen to be constant at 24, this leads directly to the variable-to-component ratios of 8:1 and 4:1, respectively. These values were chosen to reflect those reported in literature (Velicer, Eaton, & Fava 2000; Zwick & Velicer, 1986). The number of variables was taken as a constant at a value of $p = 24$ across all the conditions. This value represents a moderately

sized data set (Stevens, 2002). It should be noted that other studies on principal component analysis (PCA) regard a value of $p = 36$ to be a small- to moderate-sized data set (e.g., Fava & Velicer, 1996, Zwick & Velicer, 1982, 1986).

Generation of Population Correlation Matrices

The underlying population correlation matrices were generated for each possible p , $p:m$ and a_{ij} combination in the following manner using RANCORR program by Hong (1999):

1. The component pattern matrix was specified based on the combination of values for $p:m$ and a_{ij} .
2. The population correlation matrix was produced from the pattern matrix.
3. The program was executed four times to yield four different population correlation matrices, one correlation matrix for each combination of conditions.

Generation of samples

After the population correlation matrices were generated as described in the above section, the MNDG program (Brooks, 2002) was then used to generate samples from the population correlation matrices. This program generated multivariate normally distributed data, and reliability was systematically varied in the program to create unreliable scores based on the classical test theory. That is, reliability is defined as the proportion of raw score variance explained by true score variance, σ_T^2/σ_X^2 , or equivalently $1 - \sigma_E^2/\sigma_X^2$, with each raw score generated taken to be a total score.

In the program, the reliability estimate was entered and was set to be the same for all the variables. For example, the value for reliability for .80 was provided as input to the program and was set to be the same for all the 24 variables. The population mean and standard deviation were set to be equal to 0.0 and 1.0 respectively. The data sets were generated for each of the combinations of the $3 \times 2 \times 2$ design representing three levels of reliability, two levels of component loading and two levels of variable-to-component ratio yielding 12 different cells.

Data Analysis

One of the decisions in Monte Carlo studies is determining the number of replications to be run to obtain individual samples from the pseudo-population. There seem to be no clear guidelines on how to select the optimum number of replications. The specific number depends on the type of phenomenon being studied, the extent to which the steps of the simulation can be automated, as well as available computer resources (Hutchinson & Bandalos, 1997). Guadagnoli and Velicer (1991) recommended that five replications per cell are adequate for PCA studies. This was also supported by Velicer et al. (2000) who used five replications in their study to compare various PCA procedures. However, this study used 10 replications (samples) per cell since this was feasible given the computer capabilities available. This yielded a total of 120 samples for the 12 cells.

Each of the samples was submitted to the four procedures to determine the number of components extracted. The number of times that each procedure was correct in determining the number of components was recorded. The accuracy was then measured by the percentage of the correct components extracted by each procedure. This was done by counting the number of times each procedure extracted the exact number of components in each cell, and then expressing it as a percentage of the total number of samples in that cell. The mean component extracted in each cell was calculated by summing the number of components across the 10 samples and then divide by 10.

Results

The issue in this study was whether reliability influences the accuracy of the Kaiser rule, scree plot, Horn's parallel analysis and modified Horn's parallel analysis procedures. This issue was investigated by calculating the percentage of times that each procedure was correct in determining the correct number of components. It became apparent in this study that reliability has an influence in determining the number of components to retain under certain conditions.

However, under other conditions, the influence of reliability on the performance of the procedures is not very apparent.

Figure 1 shows the performance of each procedure in terms of percentages at each level of reliability for variable-to-component ratio of 4:1 when the cells were collapsed across sample size and component loading. At the reliability of .60, all the procedures were less than 60 percent accurate. As reliability increased, so was the accuracy of the procedures. At lower reliability levels, Horn's parallel analysis procedure performed slightly better than the all the other procedures. However, at a reliability of 1.0, modified Horn's parallel analysis was the best procedure.

Figure 1. The overall percentage of the correct number of components extracted by each procedure collapsed across component loading for 4:1 variable-to-component ratio.

When the information is presented collapsed across sample size only (as in Figure 2) and when component loading was high, (.80), and variable-to-component ratio was 4:1, the scree plot, HPA, and MHPA were at least 70 percent correct even at the reliability of .60. However, the picture is the same as in Figure 1 that as reliability increased, so does the performance of the procedures as measured by the percentage correct. The performance of the

Kaiser rule is something that stood out in these results especially considering that in most previous studies; the Kaiser rule demonstrated poor performance. In this study, generally the procedure was poor as well, but a look at Figure 2 shows that when variable-to-component ratio was 4:1, and component loading was .80, the Kaiser rule did very well especially at reliability of 1.00. At this reliability level, the Kaiser rule was about 90 percent correct in determining the number of components.

Figure 2. Percentage of the correct number of components extracted by each procedure when the variable-to-component ratio is 4:1, component loading is .80.

Figure 3 shows that when variable-to-component ratio was 8:1; and component loading was .80, reliability did not have any influence on the accuracy of the scree plot, HPA, and MHPA. These three procedures were 100 percent accurate at all the reliability levels. However, reliability had a slight influence on the performance of the Kaiser rule. Under these conditions (with few variables per component), the Kaiser was very poor even though the other procedures were very accurate at all the levels of the reliability.

Figure 3. Percentage of the correct number of components extracted by each procedure when the variable-to-component ratio is 8:1, component loading is .80.

Figure 4. The mean number of components extracted by each procedure for variable-to-component ratio of 4:1 and component loading of .50.

Figure 4 shows the relationship between the mean number of components extracted by each procedure and reliability. The exact number of components to be extracted here is six, and is indicated on the graph by a horizontal line. Under conditions of component loading of .50, only the Kaiser rule appeared to be responsive to a change in reliability. It is clear from Figure 4 that most of the procedures did not extract the correct number of components under these conditions. Only the scree plot was accurate at reliability of .60, but as reliability increased, the scree plot displayed a tendency to over-extract. On the other hand, MHPA under-extracted at lower reliability levels, but as reliability increased, it displayed a tendency to over-extract. The bottom line is, when component loading is low, the pattern of the performance of the procedures is not clearly defined.

In Figure 5, the actual number of components was three and that is shown by a horizontal line at 3. The graph shows an almost similar picture as that in Figure 4 where the performance of the Kaiser rule improved with an increase in reliability. However, the pattern of the performance for the scree plot, HPA, and MHPA as reliability increased is not clearly defined. These results are consistent with previous research that found that when component loading is low, the component structure is more diffuse, and the performance of the different procedures cannot be clearly distinguished under these conditions.

Figure 5. The mean number of components extracted by each procedure for variable-to-component ratio of 8:1 and component loading of .50.

Practical implications

It is important for practitioners to know that imperfect scores impact the performance of the procedures they use to determine the number of components to retain. A stronger reliability is mostly desirable if the results will be used to make decisions about individuals. Although most practitioners and researchers use instruments with high reliability (at least reliability of .80); sometimes instruments with low reliability are unavoidable. For example, studies of groups can tolerate a lower reliability, sometimes as low as .50 in exploratory research (McMillan & Schumacher, 2001). The same authors also noted that measures of young children are usually less reliable than those of older participants. This could be because of problems in the reading and language level, as well as lack of clear instructions. Under these conditions of low reliability, it becomes important for practitioners to understand how that might impact their decisions on retaining the correct number of components.

What also became clear in this study is that practitioners should be cautious in determining the number of components to retain in PCA. This is especially critical when the data have low component loading. Under these conditions, almost all procedures are inaccurate

in determining the correct number of components. It would be best for practitioners to use more than one procedure and then compare the results. The scree plot can be a very useful procedure to use as an adjunct, but probably not as a stand-alone procedure.

It is important to mention that the best way to make a decision about how many components to extract should be based on the knowledge that the practitioner or researcher has about the data. These various PCA procedures we use are just tools that help us extract components but should never substitute the knowledge that the researcher has about the data. In other words, components that are extracted should have some practical significance, they should have meaning to the researcher and the tools should not be used blindly to determine the number of components to extract.

Conclusion

Although this study provided some important insights into how reliability influences the number of components to extract especially for high component loading, what seemed unclear is whether the differences between the different procedures are large enough to recommend one procedure over the other at a given reliability level. For example, can one say at low reliability level, the scree plot is a better procedure to use than HPA or MHPA? Such a recommendation is not possible in this study. Instead, researchers are recommended to take into account the reliability of the scores when they interpret the number of components extracted. This is supported by Henson (2001) who pointed, "... for tests that consist of scales measuring different constructs, internal consistency should be assessed separately for each scale" (p. 181).

References

- Brooks, G. P. (2002). MNDG. <http://oak.cats.ohiou.edu.edu/~brooksg/mndg.htm>.

- Brooks, G. P., Barcikowski, R. S., & Robey, R. R. (1999). *Monte Carlo simulation for perusal and practice*. A paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED449178).
- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, 27, 509-540.
- Cattell, R. B. (1958). Extracting the correct number of factors in factor analysis. *Educational and Psychological Measurement*, 18, 791-838.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cattell, R. B., & Jaspers, J. (1967). A general plasmode for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, 3, 1-212.
- Cattell, R. B., & Volgemann, S. A. (1977). A comprehensive trial of the scree and Kaiser-Guttman criteria for determining the number of factors. *Multivariate Behavioral Research*, 12, 289-325.
- Cliff, N. (1970). The relation between sample and population characteristics vectors. *Psychometrika*, 35, 163-178.
- Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, 103, 276-279.
- Comrey, A.L. (1973). *A first course in factor analysis*. New York: Plenum Press.
- Cota, A. A., Longman, R. S., Holden, R. R., & Fekken, G. C. (1993). Comparing different methods for implementing parallel analysis: A practical index of accuracy. *Educational and Psychological Measurement*, 53, 865-877.
- Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, 27, 387-415.
- Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement*, 56, 907-29.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55, 377-393.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103, 265-275.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149-162.
- Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, 17, 193-219.
- Harshman, R. A., & Reddon, J. R. (1983). Determining the number of factors by comparing real with random data: A serious flaw and some possible corrections. *Applied Psychological Measurement*, 9, 14-15.
- Henson, R.K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34, 177-189.
- Hong, S. (1999). Generating correlation matrices with model error for simulation studies in factor analysis: A combination of the Tucker-Koopman-Linn model and Wijsman's algorithm. *Behavior Research Methods, Instruments & Computers*, 31, 727-730.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
- Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist*. Thousand Oaks, CA: Sage Publications.
- Hutchinson, S. R., & Bandalos, D. L. (1997). A guide to Monte Carlo simulation research for applied researchers. *Journal of Vocational Education Research*, 22, 233-245.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika*, 2, 151-160.

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika*, 33, 37-71.

McMillan, J. H., & Schumacher, S. (2001). *Research in Education: A conceptual introduction*. (5th ed.). San Francisco, CA: Addison Wesley Longman.

Merenda, F. P. (1997). A Guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in Counselling and Development*, 30, 156-164.

Moiser, C. I. (1939). Influence of chance error on simple structure: An empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis. *Psychometrika*, 4, 33-44.

Mooney, C. Z. (1997). *Monte Carlo simulation* (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-116). Thousand Oaks, CA: Sage.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*, 32, 396-402.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights., MA: Pearson.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34, 421-459.

Tuckman, B. W. (1999). *Conducting educational research* (5th ed.). Orlando, FL: Harcourt Brace College.

Velicer, F. W., Eaton, C. A., & Fava, J. L. (2000). Construct Explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin, & E. Helmes (Eds.), *Problems and Solutions in Human Assessment* (pp. 42 -71). Boston: Kluwer Academic Publishers.

Zwick, R. W., & Velicer, W. F. (1982). Factors influencing four rules of determining the number of components to retain. *Multivariate Behavioral Research*, 17, 253-269.

Zwick, R. W., & Velicer, F. V. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

Determining Parallel Analysis Criteria

Marley W. Watkins
The Pennsylvania State University

Determining the number of factors to extract is a critical decision in exploratory factor analysis. Simulation studies have found the Parallel Analysis criterion to be accurate, but it is computationally intensive. Two freeware programs that implement Parallel Analysis on Macintosh and Windows operating systems are presented.

Key words: Exploratory factor analysis, parallel analysis, monte carlo, software

Introduction

Exploratory factor analysis is an important analytic tool for investigating test validity. Of all the decisions made in exploratory factor analysis, determining the number of factors to extract is perhaps the most critical because incorrect specification will obscure the factor structure (Cattell, 1978; Glorfeld, 1995; Goodwin & Goodwin, 1999). Although over-extraction might be somewhat less serious than under-extraction (Wood, Tataryn, & Gorsuch, 1996), it has been empirically demonstrated that both have deleterious effects (Fava & Velicer, 1992, 1996).

Many criteria for determining the number of factors to extract have been proposed (Benson & Nasser, 1998). Unfortunately, most are inaccurate guides to practice (Kanyongo, 2005; Zwick & Velicer, 1986). Based upon current simulation research (Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986), only two methods have consistently emerged as accurate: the Parallel Analysis (PA) method of Horn (1965) and the Minimum Average Partial (MAP) method of Velicer (1976).

The MAP procedure entails partialling each successive factor out of a correlation matrix to create a partial correlation matrix. The average of the squared correlations of the off-diagonal partial correlation matrix is then computed. This average should decrease as long as shared variance is being extracted, but begin to increase when error variance predominates. In contrast, PA requires that a set of random correlation matrices be generated based upon the same number of variables and participants as the experimental data. These random correlation matrices are then subjected to principal components analysis and the average of their eigenvalues is computed and compared to the eigenvalues produced by the experimental data. The criterion for factor extraction is where the eigenvalues generated by random data exceed the eigenvalues produced by the experimental data.

Thompson and Daniel (1996) explicitly recommended PA procedures for determination of the number of factors to extract. Unfortunately, neither MAP nor PA is included in common statistical software packages (Fabrigar, Wegener, MacCallum, & Strahan, 1999) and both methods are computationally intensive. Consequently, many published factor analytic studies have relied on inaccurate methods to determine the number of factors to retain (Fabrigar, et al. 1999).

The MAP method remains relatively inaccessible, but simpler alternatives have been developed for PA. Of these, regression equations to predict PA criteria have predominated.

Marley W. Watkins is Professor of school psychology at The Pennsylvania State University and a Diplomat of the American Board of Professional Psychology. Correspondence concerning this article should be addressed to Marley Watkins at mww10@psu.edu.

However, simulation studies indicated that they are inaccurate (Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986). Tables of random eigenvalues generated by Monte Carlo simulations have been found to be accurate and make PA criteria accessible (Lautenschlager, 1989; Velicer, et al., 2000). Unfortunately, they do not cover all possible variable-participant combinations and interpolation of tabled values is tedious and may be error-prone. Additionally, current tables only allow comparison of obtained eigenvalues with the mean of a sample of random eigenvalues. Several researchers have suggested that mean comparisons may retain more factors than is warranted and recommended that the 90th or 95th percentile be used instead (Cota, Longman, Holden, & Fekken, 1993; Glorfeld, 1995).

O'Connor (2000) provided a tutorial for using PA and MAP with existing general purpose statistical software, but use requires expensive software (i.e., SPSS or SAS) and manipulation of complex syntax code. In contrast, Kaufman and Dunlap (2000) published a standalone FORTRAN program to calculate PA criteria. Unfortunately, it only operates on the Windows platform (excluding Macintosh users) and does not accommodate problems with more than 50 variables or 1,000 subjects. The present paper presents two computer programs which make PA criteria more flexible and readily available.

Programs

MacParallel

Data were generated in a set of Monte Carlo simulations in which the number of variables (V) ranged from 5 to 100 in steps of 5 and sample sizes (N) were 100, 150, 200, 300, 400, 500, 750, 1000, 1500, 2000, and 2500. Using SPSS for the Macintosh, Version 6.1 (Norusis, 1994), random normal data were generated for each of the 220 combinations of variables and subjects and subsequently subjected to principal components analysis (PCA). Resulting eigenvalues were saved and each V by N simulation was replicated 100 times. The final output from these 22,000 simulations was the mean first through V th eigenvalues and associated standard errors.

When compared to the 3,950 overlapping values presented by Lautenschlager (1989), the largest difference was .036 and the average difference was .000051. Thus, these results appear to be consistent with previous simulations which were found to be accurate in determining the number of components to retain (Velicer, et al., 2000; Zwick & Velicer, 1986).

MacParallel is a standalone RealBASIC program which provides an electronic look-up table of these random data eigenvalues and standard errors with integral linear interpolation. To increase accessibility, identical versions are available for Macintosh and Windows operating systems.

Monte Carlo PCA for Parallel Analysis

Although MacParallel is quick and accurate, it does not directly calculate all possible variable-participant combinations. In contrast, Monte Carlo PCA for Parallel Analysis is a standalone RealBASIC program which allows specification of 3-300 variables, 100-2,500 participants, and 1-1,000 replications. The program: (a) generates random normal data for the quantity of variables and participants selected; (b) computes the correlation matrix; (c) calculates eigenvalues for those variables via a Jacobi routine; (d) repeats the process as many times as specified in the replications field; and (e) calculates the average and standard deviation of the eigenvalues across all replications. Identical versions are available for Macintosh and Windows operating systems.

Computation of random eigenvalues with Monte Carlo PCA for Parallel Analysis is dependent upon processor speed as well as the number of variables, participants, and replications requested. Results from 25 variables, 500 participants, and 100 replications were produced in 18 seconds by a Macintosh iMac G5 operating under System X.4.1. An identical analysis took 32 seconds under Windows NT on a 733 MHz Intel Pentium III processor.

Availability

Freeware versions of MacParallel and Monte Carlo PCA for Parallel Analysis are available for Macintosh and Windows operating

systems at:

<http://www.personal.psu.edu/mww10>

References

Benson, J., & Nasser, F. (1998). On the use of factor analysis as a research tool. *Journal of Vocational Education Research*, 23, 13-33.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum Press.

Cota, A. A., Longman, R. S., Holden, R. R., & Fekken, G. C. (1993). Comparing different methods for implementing parallel analysis: A practical index of accuracy. *Educational and Psychological Measurement*, 53, 865-876.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, 27, 387-415.

Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement*, 56, 907-929.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55, 377-393.

Goodwin, L. D., & Goodwin, W. L. (1999). Measurement myths and misconceptions. *School Psychology Quarterly*, 14, 408-427.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

Kanyongo, G. Y. (2005). Determining the correct number of components to extract from a principal components analysis: A Monte Carlo study of the accuracy of the scree plot. *Journal of Modern Applied Statistical Methods*, 4, 120-133.

Kaufman, J. D., & Dunlap, W. P. (2000). Determining the number of factors to retain: A Windows-based FORTRAN-IMSL program for parallel analysis. *Behavior Research Methods, Instruments, & Computers*, 32, 389-395.

Lautenschlager, G. J. (1989). A comparison of alternatives to conducting monte carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, 24, 365-395.

Norusis, M. J. (1994). *SPSS professional statistics 6.1, Macintosh version*. Chicago: SPSS, Inc.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32, 396-402.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56, 197-208.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 31, 321-327.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41-71). Boston: Kluwer Academic Publishers.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, 1, 254-365.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

Change Point Estimation of Bilevel Functions

Leming Qu
Department of Mathematics
Boise State University

Yi-Cheng Tu
Department of Computer Science
Purdue University

Reconstruction of a bilevel function such as a bar code signal in a partially blind deconvolution problem is an important task in industrial processes. Existing methods are based on either the local approach or the regularization approach with a total variation penalty. This article reformulated the problem explicitly in terms of change points of the 0-1 step function. The bilevel function is then reconstructed by solving the nonlinear least squares problem subject to linear inequality constraints, with starting values provided by the local extremas of the derivative of the convolved signal from discrete noisy data. Simulation results show a considerable improvement of the quality of the bilevel function using the proposed hybrid approach over the local approach. The hybrid approach extends the workable range of the standard deviation of the Gaussian kernel significantly.

Key words: Bar code, 0-1 step function, nonlinear least squares, constrained optimization

Introduction

Reconstruction of bilevel functions has important practical applications in modern life. A particular example of bilevel functions is the bar code signal. The ubiquitous alternating black and white strips (the bar code) are now widely used in every day life and industrial processes.

The problem of recovering a bar code signal $f(t)$ from the noisy signal $y(t)$ detected by a bar code scanner (Esedoglu, 2004, Shellhammer, Goren, & Pavlidis, 1999) is to construct a one-dimensional 0-1 step function $f(t)$ given the samples $y_i = y(t_i)$, $i = 1, \dots, M$, of the continuous-time observation

$$y(t) = \alpha \cdot G * f(t) + \varepsilon(t), t \in T = [0,1], \quad (1)$$

Leming Qu is an Assistant Professor of Statistics. His research interests include Wavelets in statistics, time series, Bayesian Analysis, statistical and computational inverse problems, nonparametric and semiparametric regression. Email him at qu@math.boisestate.edu. Yi-Cheng Tu is a doctoral student in Computer Science. Email at tuyc@cs.purdue.edu.

where $\alpha > 0$ is the unknown amplitude, the $\varepsilon(t)$ is the additive unobservable noise process and

$$G * f(t) = \int_T G(t-x)f(x)dx,$$

and $G(t)$ is a Gaussian kernel of the convolution:

$$G(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right),$$

where $\sigma > 0$ is the unknown standard deviation which increases as the scanner moves away from the bar code. According to Wittman, (2004), the $\sqrt{2}\sigma$ is assumed to be smaller than the width of the thinnest (WTB) bar. The approach proposed in this article can extend this range. The simulation result in section V shows that $f(t)$ can still be reconstructed even $\sqrt{2}\sigma$ is as large as 1.4 times the WTB at high signal to noise ratio level. Figure 1 illustrates the results of the convolution with additive noise for a UPC bar code (Wittman, 2004) encoding 0123456789.

Without using the special characteristics of a bar code signal, $f(t)$ in model (1) is treated as a bilevel function with levels 0 or 1. Thus, the goal is to reconstruct the bilevel function $f(t)$

from the noisy discrete data $y = (y_1, \dots, y_M)^T$. This problem differs slightly from standard restoration problems of image processing in that the convolution kernel contains unknown quantities. It is a partially blind deconvolution problem somewhat closer to blind deconvolution.

Previous Work and Current Approach

Previous work on the bar code reconstruction problem (Wittman, 2004) is based on the following:

- (a) local approach: finding local minima and maxima in the derivative of

$$s(t) = \alpha \cdot G * f(t). \quad (2)$$

Note the above is the noise-free version of model (1);

- (b) global approach: regularization methods for ill-posed inverse problems such as total variation based restoration (Esedoglu, 2004).

The approach (a) tries to relate the local minima and maxima in $s'(t)$ to the edges of bars which are the change points in $f(t)$. Locating these local extremas is sensitive to noise $\varepsilon(t)$. Furthermore, these local extrema are difficult to relate to the true change points of $f(t)$ due to convolution distortion (Shellhammer, Goren, & Pavlidis, 1999). Such techniques use only local information and would have difficulty to detect thin bars closely adjacent to each other. For example, in Figure 1, the $s(t)$ is near flat around location 0.55 even though three adjacent thin bars stand there.

To overcome these shortcomings, approach (b) in Esedoglu (2004) tried to recover $f(t)$ by regularization using the total variation penalty, a technique commonly used in image restoration literature. It models systematically the interaction of neighboring bars in $f(t)$ under convolution with $G(t)$, as well as the estimation of α and σ from global information contained in the $y(t)$. It is proved in Esedoglu (2004) that under certain regularity conditions, the infimum of the total-variation energy is attained.

Numerical results show that bar codes from highly degraded signals can be recovered.

The regularization approach in inverse problems must deal with the choice of the regularization parameter, a difficult problem itself. In Esedoglu (2004), there are two regularization parameters which need to be chosen. In the numerical results of Esedoglu (2004), the regularization parameters are preselected and kept fixed.

It is felt that all the existing works did not fully utilize the information about $f(t)$: a bilevel function with levels 0 or 1. To recover $f(t)$ is to recover the change points of $f(t)$ for $t \in T$. The number of change points in $f(t)$ is twice the number of bars in the bar code. Recovering the $f(t)$, $t \in T$ is usually an ill-posed problem, although recovering the change points is a well-posed problem if the number of observations exceed the number of unknown parameters. The well-posed problem is essentially to estimate the nonlinear parameters in finite dimension with linear inequality constraints. The formulation of the bilevel function deconvolution was not found in terms of the change points explicitly in the existing literature. Therefore, a nonlinear least squares solution is proposed to the change points of $f(t)$, σ and α with the constraints of the ordered change points. The local approach is used to provide the starting values for the global minimization problem. This method is a hybrid of local and global approaches in spirit.

Change Point Estimation

Assume the total number of bars of $f(t)$ is the known integer K . For example, $K = 22$ for the UPC bar code in the test problem. Denote ξ_{2j-1} and ξ_{2j} as the beginning and ending location of the j th bar for $j = 1, \dots, K$ of the bilevel function $f(t)$.

Then $f(t)$ can be defined explicitly as:

$$f(t) = I(\xi_{2j-1} < t \leq \xi_{2j}), t \in T, j = 1, \dots, K.$$

where $I()$ is the usual indicator function and

$$0 < \xi_1 < \xi_2 < \dots < \xi_{2K-1} < \xi_{2K} < 1$$

are the ordered change points.

The goal of the bilevel function reconstruction is to recover the change points $\xi = (\xi_1, \xi_2, \dots, \xi_{2K-1}, \xi_{2K})^T$ from the observed data $y = (y_1, \dots, y_M)^T$ at $t = (t_1, \dots, t_M)^T$, without any knowledge of σ and α .

With the special structure of $f(t)$, the convolution $G * f(t)$ can be explicitly expressed as a function of ξ and σ . Thus,

$$s(t) = \alpha \int_T G(t-x)f(x)dx = \alpha \sum_{j=1}^K \int_{\xi_{2j-1}}^{\xi_{2j}} G(t-x)dx.$$

Let

$$r_i = s_i - y_i = \alpha \sum_{j=1}^K \int_{\xi_{2j-1}}^{\xi_{2j}} G(t_i - x)dx - y_i$$

be the i th residual. Denote $r = (r_1, \dots, r_M)^T$ the residual vector and

$$h(\xi, \sigma, \alpha) = \frac{1}{2} \sum_{i=1}^M r_i^2 = \frac{1}{2} r^T r$$

the residual sums of squares. The least squares solution of ξ , σ and α is sought. That is to find $\hat{\xi}$, $\hat{\sigma}$, and $\hat{\alpha}$ which minimizes the merit function $h(\xi, \sigma, \alpha)$ subject to the required conditions.

More explicitly, the constrained nonlinear least squares problem is

$$\min_{\xi, \sigma, \alpha} h(\xi, \sigma, \alpha) \quad (3)$$

such that

$$0 < \xi_1 < \xi_2 < \dots < \xi_{2K-1} < \xi_{2K} < 1, \sigma > 0, \alpha > 0.$$

These constraints are simply linear inequality constraints $A[\xi^T, \sigma, \alpha]^T < u$ with a sparse matrix A whose elements are

$$\begin{aligned} A[1,1] &= -1, \\ A[i, i-1] &= 1, \\ A[i, i] &= 1 \quad \text{for } i = 2, \dots, 2K, \\ A[2K+1, 2K] &= 1, \\ A[2K+2, 2K+1] &= -1, \\ A[2K+3, 2K+2] &= -1; \end{aligned}$$

and $u = (0, \dots, 0, 1, 0, 0)^T$ is a $(2K+3)$ column vector.

The recast of the bilevel function reconstruction into a constrained nonlinear least squares problem enables us to utilize the existing techniques for solving nonlinear least square problem subject to linear inequality constraints in the statistical and numerical analysis literature.

The Fletcher-Xu hybrid Gauss-Newton and BFGS method (Fletcher & Xu, 1987) for nonlinear least squares problem is super linearly convergent. This method along with other five methods for constrained nonlinear least squares problems is implemented in the *cls Solve* solver of the TOMLAB 4.7 optimization environment [9].

The Gauss-Newton method needs the gradient of the merit function $h(\xi, \sigma, \alpha)$, which is the product of the Jacobian matrix of r and r . The Jacobian matrix of r is easily obtained by:

$$\begin{aligned} \frac{\partial r_i}{\partial \xi_{2j-1}} &= -\alpha \cdot G(t_i - \xi_{2j-1}), \\ \frac{\partial r_i}{\partial \xi_{2j}} &= \alpha \cdot G(t_i - \xi_{2j}), \\ \frac{\partial r_i}{\partial \sigma} &= \frac{\alpha}{\sigma} \sum_{j=1}^K \int_{\xi_{2j-1}}^{\xi_{2j}} \left[-1 + \frac{(t_i - x)^2}{\sigma^2} \right] G(t_i - x) dx, \\ \frac{\partial r_i}{\partial \alpha} &= \sum_{j=1}^K \int_{\xi_{2j-1}}^{\xi_{2j}} G(t_i - x) dx. \end{aligned}$$

The success of the $(2K+2)$ dimensional global minimization problem (3) heavily depends on good starting values. Our numerical experiments indicated that simple starting values such as equally spaced grids on T for ξ did not give satisfactory solutions. Next, the initial

parameter estimation based on the local approach is discussed.

Initial Estimation

The local extremas of the derivative of $s(t)$ are close to ξ , the edges of the bars. Then the initial estimation of ξ in terms of the local extremas of the derivative signal is the following problem: given the noisy discrete observations of $s(t)$:

$$y_i = s(t_i) + \varepsilon_i, i = 1, \dots, M,$$

finding the local extremas of $s'(t)$.

There are approaches of estimating $s'(t)$ based on different smoothing or denoising methods. Many of these try first to find $\hat{s}(t)$, the estimate of $s(t)$, using the chosen smoothing or denoising method; then estimate $s'(t)$ based on $\hat{s}(t)$. See, for example, the spline regression based method in Zhou and Wolfe (2000) or the wavelet denoising based method in Cai (2002). For equally spaced $\{t_i\}$ and when M is a power of 2, there exists a fast algorithm with complexity $O(M)$ to carry out discrete wavelet transform (DWT). In this simulation, the wavelet thresholding method is used to estimate $s(t_i)$ first, then estimate $s'(t_i)$ based on $\hat{s}(t_i)$ using a first derivative filter.

After obtaining the initial estimate ξ_0 of ξ by the K pairs of local maxima and minima of $\hat{s}'(t_i)$, σ_0 is estimated, the initial σ , by techniques suggested in Joseph and Pavlidis (1993), Joseph and Pavlidis (1994), and Esedoglu (2004). Proposition 1 of Joseph and Pavlidis (1993) suggested approximating σ by the distance from the last local maxima of $\hat{s}(t_i)$ to the last local minima of $\hat{s}'(t_i)$. Proposition 2 of Joseph and Pavlidis (1994) suggested approximating σ by $|\hat{s}'(t^*) / \hat{s}^{(3)}(t^*)|$ where t^* is a point such that $|\hat{s}'(t^*)| \geq |\hat{s}'(t_i)|$ for $i = 1, \dots, M$. The smaller value of $\hat{\sigma}$ based on the two propositions is used first. If it is outside the reasonable range $[0.001, 0.02]$, then the value 0.0079 is used as suggested in Esedoglu (2004) for the true σ ranging from 0.011 to 0.013. In contrast, the thinnest bar has width 0.0132 in

this simulation study. The initial value of α , α_0 , is simply the ordinary least squares estimate given the ξ_0 and σ_0 .

Simulation Results

In the experiment, a clean bar code $f(t)$ encoding 0123456789 was blurred by convolving it with $G(t)$ of known σ , amplified by the amplitude $\alpha=1$, sampled at $t_i = i/M$, for $i = 1, \dots, M$, followed by the addition of white Gaussian noise $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. The amount of the added noise makes the signal-to-noise ratio $SNR = 20 \log_{10} (std(s) / \sigma_\varepsilon)$ at the specified level.

Estimation of $s(t_i)$ is carried out by the soft Wavelet thresholding technique implemented in the Wavelet Toolbox in MATLAB. The thresholds are chosen by a heuristic variant of Stein's Unbiased Risk Estimate with multiplicative threshold rescaling using a single estimation of level noise based on the finest level wavelet coefficients (Donoho & Johnstone, 1995). The wavelet filter used is db6: the Daubechies wavelet with 6 vanishing moments.

The first derivative filter for estimating $s'(t_i)$ from $\hat{s}(t_i)$ is

$$d1 = [-0.015964, -0.121482, -0.193357, 0.00, 0.193357, 0.121482, 0.015964];$$

as recently suggested in Farid and Simoncelli (2004). This filter is significantly more accurate than those commonly used in the image and signal processing literature.

Table 1 shows the Monte-Carlo approximations to $MSE = E\left(\|\hat{\xi} - \xi\|_2^2 / (2K)\right)$ of this method based on 100 independent simulations as the σ is varied at 0.011, 0.012 and 0.0125, the sample size M is varied dynamically from $M = 256$ through 1024, and SNR is varied from high to moderate levels. The range of the σ is in contrast with the WTB equal to 0.0132.

Table 2 shows the Monte-Carlo approximations to $MSE = E(\|\xi_0 - \xi\|_2^2 / (2K))$. The results show a considerable reduction of MSE for $\hat{\xi}$ over ξ_0 in most cases. The most significant improvement occurs for the three tested values of σ with high SNR .

Table 1. Monte Carlo Approximations To $MSE = E\left(\|\hat{\xi} - \xi\|_2^2 / (2K)\right)$

		Means for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.0110	256	3.402e-09	3.358e-08	1.006e-04
0.0110	512	1.565e-09	1.628e-08	5.159e-05
0.0110	1024	7.769e-10	7.703e-09	1.471e-05
0.0120	256	3.685e-09	9.822e-05	2.992e-04
0.0120	512	1.939e-09	7.846e-06	3.084e-04
0.0120	1024	9.744e-10	1.220e-05	2.450e-04
0.0125	256	6.153e-05	2.020e-04	4.457e-04
0.0125	512	2.063e-09	5.569e-05	4.502e-04
0.0125	1024	1.055e-09	6.317e-05	3.847e-04

Table 2. Monte Carlo Approximations To $MSE = E(\|\zeta_0 - \xi\|_2^2 / (2K))$

		Means for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.0110	256	1.512e-05	1.657e-05	1.399e-04
0.0110	512	1.052e-05	1.128e-05	6.394e-05
0.0110	1024	9.366e-06	1.002e-05	3.060e-05
0.0120	256	2.050e-05	1.324e-04	3.486e-04
0.0120	512	1.277e-05	2.471e-05	3.496e-04
0.0120	1024	1.129e-05	2.670e-05	2.918e-04
0.0125	256	8.834e-05	2.485e-04	5.345e-04
0.0125	512	1.478e-05	8.177e-05	5.068e-04
0.0125	1024	1.264e-05	8.423e-05	4.555e-04

Figures 1 and 2 present results from two of these experiments. Figure 1 is an example of completely successful reconstruction while the Figure 2 an example that the estimation fails

when the noise level or the blur factor σ gets too high. It seems that when $\|\hat{\xi} - \xi\|_2^2 / 44$ is less than the scale of $1.0e-5$, a successful reconstruction is obtained.

Figure 1. Top to bottom: the bar code, the reconstructed bar code, the corresponding clean convolved signal, the noisy convolved signal. The true parameter used to generate the corrupted signal: $M = 512$, $\sigma = 0.0118$, $\alpha = 1$, $SNR = 18$. The estimated parameter: $\hat{\sigma} = 0.0119$, $\hat{\alpha} = 1.0036$. The square error $\|\hat{\xi} - \xi\|_2^2 / 44 = 9.6e-8$. CPU time: 51 seconds.

Figure 2. Top to bottom: the bar code, the failed reconstructed bar code, the corresponding clean convolved signal, the noisy convolved signal. The true parameter used to generate the corrupted signal: $M = 512$, $\sigma = 0.0129$, $\alpha = 1$, $SNR = 16$. The estimated parameter: $\hat{\sigma} = 0.0136$, $\hat{\alpha} = 1.0287$. The square error $\|\hat{\xi} - \xi\|_2^2 / 44 = 2.1e-4$. CPU time: 66 seconds.

Setting the cut-off $\|\hat{\xi} - \xi\|_2^2 / 44$ at $1.0e-5$ for a successful reconstruction, Table 3 displays the percentage of successful reconstructions of ξ . Table 4 displays the percentage of successful reconstructions of ξ_0 . The successful rate of our hybrid approach is 100% in most cases with high SNR. The local approach alone will not work.

The simulation result also indicates that $\hat{\sigma}$ gives much better solution than the initial estimate σ_0 in terms of the reduction of MSE . The same is true for $\hat{\alpha}$.

The computational time for finding the solution is relatively fast. For example, for the

most time consuming scenario $\sigma = 0.125$, $M = 1024$, $SNR = 21$, the average CPU time is 86 seconds, contrary to the reported 6 minutes using the regularization approach in Wittman (2004).

Most current bar code decoding works only in the range when σ is up to 0.7 times the WTB. The result in Table 3 indicates that this approach can attain 100% successful rate even when σ is up to 0.95 times the WTB at the $SNR = 38$ level. In contrast, the local approach alone can attain a successful rate of 94% when σ is up to 0.83 times the WTB at the $SNR = 38$ level as indicated in Table 4. This approach seems to break down at $\sigma = 0.0120$ and $SNR = 21$.

Table 3. Percentage Of Successful Reconstruction Of ξ

		Percentages for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.011000	256	100	100	85
0.011000	512	100	100	90
0.011000	1024	100	100	98
0.012000	256	100	81	54
0.012000	512	100	97	52
0.012000	1024	100	98	51
0.012500	256	88	63	40
0.012500	512	100	89	37
0.012500	1024	100	87	33

Table 4. Percentage Of Successful Reconstruction Of ξ_0

		Percentages for the following levels of SNR		
σ	M	$SNR = 38$	$SNR = 28$	$SNR = 21$
0.011000	256	0	0	0
0.011000	512	19	13	1
0.011000	1024	94	54	8
0.012000	256	0	0	0
0.012000	512	0	0	0
0.012000	1024	0	0	0
0.012500	256	0	0	0
0.012500	512	0	0	0
0.012500	1024	0	0	0

Conclusion

A nonlinear least squares estimation for change points of a bilevel function is proposed. The local information contained in the derivative of the convolved signal is used to provide starting values for the global optimization solution. This hybrid approach uses all available information for parameter estimation to the full extent. Monte Carlo simulation results show the good performance of the hybrid approach over the local approach. It also indicates that the hybrid approach extends the workable range of the standard deviation of the Gaussian kernel significantly.

If extra information such as the WTB or the width of thickest bar or space is available, they can be easily incorporated into the linear inequality constraints. Actually, in UPC barcode system, the WTB is always known. A barcode is decoded based on the bar widths relative to the WTB. Each should be 1, 2, 3, or 4 times the WTB. This knowledge could lead to an interesting combinatorial problem for bar code deconvolution and an additional constraint: choosing the change points to reflect the fact that the widths are 1 through 4 times the WTB.

Currently, the value K of the number of bars is assumed to be known in advance. In real bar code, the range of possible K values will be fairly small. A future research effort is to estimate the bilevel function without this assumption. Then model selection methods are needed for such problem.

References

- Cai, T. (2002). On adaptive wavelet estimation of a derivative and other related linear inverse problems. *Journal of Statistical Planning and Inference*, 108, 329-349
- Donoho, D., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association*, 90, 1200-1224
- Esedoglu, S. (2004). Blind deconvolution of bar code signals. *Inverse Problems*, 20, 121-135
- Farid, H. & Simoncelli, E. P. (2004). Differentiation of discrete multidimensional signals. *IEEE Transactions on Image Processing*, 13(4), 496-508.
- Fletcher, R. & Xu, C. (1987). Hybrid methods for nonlinear least squares. *IMA Journal of Numerical Analysis*, 7, 371-389.

Joseph, E. & Pavlidis, T. (1993). Deblurring of bilevel waveforms. *IEEE Transactions on Image Processing*, 2, 223-235.

Joseph, E. & Pavlidis, T. (1994). Bar code waveform recognition using peak locations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 630-640.

Shellhammer, S. J., Goren, D. P., & Pavlidis, T. (March, 1999). Novel signal-processing techniques in barcode scanning. *IEEE Robotics & Automation Magazine*, 57-65

Tomlab Optimization Inc., (2005). User's Guide For TOMLAB 4.7. <http://tomlab.biz>.

Wittman, T. (September, 2004). Lost in the supermarket: Decoding blurry barcodes. *SIAM News*.

Zhou, S. & Wolfe, D. A., (2000). On derivative estimation in spline regression. *Statistica Sinica*, 10, 93-108.

Large Sample and Bootstrap Intervals for the Gamma Scale Parameter Based on Grouped Data

Ayman Baklizi Amjad AL-Nasser
Department of Statistics
Yarmouk University

Interval estimation of the scale parameter of the gamma distribution using grouped data is considered in this article. Exact intervals do not exist and approximate intervals are needed. Recently, Chen and Mi (2001) proposed alternative approximate intervals. In this article, some bootstrap and jackknife type intervals are proposed. The performance of these intervals is investigated and compared. The results show that some of the suggested intervals have a satisfactory statistical performance in situations where the sample size is small with heavy proportion of censoring.

Key words: bootstrap, gamma distribution, grouped data, interval estimation.

Introduction

In many practical studies, the collected data may not be complete observations, they may be in a form of counts of observations in certain intervals; such data is often called grouped data. Grouped data arise frequently in life testing experiments when inspecting the test units intermittently for failure, this procedure is frequently used because it requires less testing effort than continuous inspection. The data obtained from intermittent inspection consists only of the number of failures in each inspection interval. Other examples of natural occurrences of grouped data are given in Pettitt and Stephens (1977).

Ayman Baklizi is an Associate Professor of statistics, and currently is at the department of mathematics and physics at Qatar University. He is an elected-member at the ISI. Amjad AL-Nasser is an Assistant Professor of statistics at Yarmouk University, Jordan.

In a recent article, Chen and Mi (2001) provided a general method for constructing intervals for the unknown parameters in the distribution using grouped data. Assume that the data are grouped in the classes $[0, t_1), [t_1, t_2), \dots, [t_{k-1}, t_k), [t_k, \infty)$. The i -th interval is $[t_{i-1}, t_i)$. Assume that $t_0 = 0$ and $t_{k+1} = \infty$. Let r_i be the number of failures in the i -th interval. Define the random variable $\zeta_n = \sum_{i=1}^k r_i t_i + r_{k+1} t_k$. It follows that $\frac{\zeta_n - ng(\lambda)}{\sqrt{ns_n}} \rightarrow N(0,1)$ in law as $n \rightarrow \infty$,

where $g(\lambda) = \sum_{i=1}^k t_i p_i + t_k p_{k+1}$ and

$$s_n = \left(\sum_{i=1}^k t_i^2 \hat{p}_i + t_k^2 \hat{p}_{k+1} \right) - \left(\sum_{i=1}^k t_i \hat{p}_i + t_k \hat{p}_{k+1} \right)^2.$$

It follows that, asymptotically,
$$P\left(\frac{\zeta_n}{n} - z_{\alpha/2} \frac{s_n}{\sqrt{n}} < g(\lambda) < \frac{\zeta_n}{n} + z_{\alpha/2} \frac{s_n}{\sqrt{n}} \right) = 1 - \alpha$$

When the function $g(\lambda)$ is monotone, an approximate $(1 - \alpha)\%$ confidence interval for λ , call it the CM interval, can be obtained as

$$\left[g^{-1}\left(\frac{\zeta_n}{n} - z_{\alpha/2} \frac{s_n}{\sqrt{n}}\right), g^{-1}\left(\frac{\zeta_n}{n} + z_{\alpha/2} \frac{s_n}{\sqrt{n}}\right) \right].$$

However, the above interval possesses exact coverage probabilities and symmetry probabilities only for sufficiently large sample sizes. In this article, the properties of these intervals are investigated and some bootstrap based intervals that use the result of Chen and Mi (2001) are proposed. A similar problem has been investigated for the Burr type X distribution by Al-Nasser and Baklizi (2004).

Bootstrap Intervals

Let $x_1, \dots, x_n$ be a random sample from the gamma distribution whose probability density function is given by

$$f(x, \lambda, c) = \frac{\lambda^c}{\Gamma(c)} x^{c-1} e^{-\lambda x}, \quad x > 0.$$

Let r_i be the number of observations falling in the i -th interval (t_{i-1}, t_i) , $i = 1, \dots, k+1$. The joint probability function of $r_1, \dots, r_{k+1}$ is multinomial with parameters n and $p_1, \dots, p_{k+1}$. The following confidence intervals are based on the Bootstrap approach (Efron & Tibshirani, 1993). There are several Bootstrap based intervals discussed in the literature, the most common ones are the bootstrap -t interval, the percentile interval and the bias corrected and accelerated (BC_a) interval.

The Bootstrap - t Interval (BTS Intervals)

Let ζ_n be the random variable defined as $\zeta_n = \sum_{i=1}^k r_i t_i + r_{k+1} t_k$ calculated from the original data and let ζ_n^* be calculated from the bootstrap sample. Let z_α^* be the α quantile of the bootstrap distribution of $Z^* = \frac{(\zeta_n^* - \zeta_n)}{s_n^*}$,

where s_n^* is estimated variance of ζ_n calculated from the bootstrap sample. The bootstrap-t interval for λ is given by $\left[g^{-1}(\zeta_n - z_{\alpha/2}^* s_n^*), g^{-1}(\zeta_n + z_{\alpha/2}^* s_n^*) \right]$ where z_α^* is determined by simulation

The Percentile Interval (PRC Interval)

Here, the bootstrap distribution of ζ_n^* are simulated by resampling repeatedly from the parametric model of the original data and calculating $\zeta_{n,i}^*, i = 1, \dots, B$ where B is the number of bootstrap samples. Let $\hat{G}$ be the cumulative distribution function of ζ_n^* , then the $1 - \alpha$ interval is given by

$$\left[\hat{G}^{-1}\left(\frac{\alpha}{2}\right), \hat{G}^{-1}\left(1 - \frac{\alpha}{2}\right) \right].$$

The Bias Corrected Interval (BC Interval)

The bias corrected interval (Efron, 1982) is calculated using the percentiles of the bootstrap distribution of ζ_n^* . The determination of the appropriate percentiles depends on a number ($\hat{z}_0$) called the bias correction. The $1 - \alpha$ interval is given by

$$[\hat{G}^{-1}(\alpha_1), \hat{G}^{-1}(\alpha_2)]$$

where

$$\begin{aligned} \alpha_1 &= \Phi(2\hat{z}_0 + z_{\alpha/2}), \\ \alpha_2 &= \Phi(2\hat{z}_0 + z_{1-\alpha/2}), \\ &\Phi(\cdot) \end{aligned}$$

is the standard normal cumulative distribution function, z_α is the α quantile of the standard normal distribution. The value of $\hat{z}_0$ are calculated as $\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\zeta_n^* < \zeta_n\}}{B}\right)$.

The Bias Corrected and Accelerated Interval (BCa Interval)

The bias corrected and accelerated interval is calculated also using the percentiles of the bootstrap distribution of. The percentiles depend on two numbers $\hat{a}$ and $\hat{z}_0$ called the acceleration and the bias correction. The $1 - \alpha$ interval is given by

$$(\hat{G}^{-1}(\alpha_1), \hat{G}^{-1}(\alpha_2))$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right),$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right),$$

$$\Phi(\cdot)$$

is the standard normal cumulative distribution function, z_α is the α quantile of the standard normal distribution. The values of $\hat{a}$ and $\hat{z}_0$ are calculated as follows;

$$\hat{a} = \frac{\sum_{i=1}^n (\zeta_n(\cdot) - \zeta_n(i))^3}{6 \left\{ \sum_{i=1}^n (\zeta_n(\cdot) - \zeta_n(i))^2 \right\}^{3/2}}$$

where $\zeta_n(i)$ is calculated using the original data excluding the i -th observation and

$$\zeta_n(\cdot) = \frac{\sum_{i=1}^n \zeta_n(i)}{n}.$$

The value of $\hat{z}_0$ is given, as before, by

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\zeta_n^* < \zeta_n\}}{B}\right).$$

Jackknife Intervals (JAC Intervals)

An interval based on the jackknife (Efron & Tibshirani, 1993) can be constructed as follows;

$$\zeta_n(\cdot) \pm z_{\alpha/2} s\hat{e},$$

where

$$s\hat{e}^2 = \frac{n-1}{n} \sum_{i=1}^n (\zeta_n(\cdot) - \zeta_n(i))^2$$

is the jackknife estimate of the variance of ζ_n .

Intervals Based on the Bootstrap Standard Deviation (BSD Intervals)

An interval similar in form to the based on the jackknife can be constructed as follows;

$$\zeta_n \pm z_{\alpha/2} s\tilde{e},$$

where

$$s\tilde{e}^2 = \frac{1}{B-1} \sum_{i=1}^B (\zeta_{i,n}^* - \bar{\zeta}_n^*)^2,$$

$$\bar{\zeta}_n^* = \frac{1}{B} \sum_{i=1}^B \zeta_{i,n}^*$$

is the bootstrap estimate of the variance of ζ_n .

Small Sample Performance of the Intervals

For the confidence intervals with nominal confidence coefficient $(1 - \alpha)$, the criterion of attainment of lower and upper error probabilities (Jennings, 1987) is used, which are

both taken equal to $\frac{\alpha}{2}$. A simulation study is

conducted to investigate the performance of the intervals. The sample sizes chosen are $n = 20, 30, 50, 100$. The number of groups $k+1$ is taken as 3, 5 and 9. The censoring proportion (cp) is taken as 0.6, 0.4, and 0.2. The confidence coefficient is taken as 95%, and the shape parameter r is taken as 0.4, 0.8, 1.2, 1.6, and 2. For each combination of the simulation indices 2000 samples were generated from the

gamma distribution with $\lambda = 1$. The intervals are calculated, $B = 2000$ was used for bootstrap calculations. The following quantities are simulated for each interval using the results of the 2000 samples;

1. Lower error rates (L): The fraction of intervals that fall entirely above the true parameter.
2. Upper error rates (U): The fraction of intervals that fall entirely below the true parameter.
3. Total error rates (T): The fraction of intervals that did not contain the true parameter value.

The results are given in Tables 1-3.

Results

From the simulation results, it appears that for $k = 2$, small sample size ($n = 20, 30$) and heavy censoring ($cp = 0.4, 0.6$), the CM intervals tend to be anti-conservative. This is also true for PRC, BC and BCa intervals. On the other hand, the BTS, BSD, and JAC intervals tend to attain

the nominal sizes. As the censoring proportion is light to moderate, the PRC and the BTS intervals tend to be highly conservative while the BC and BCa intervals tend to be grossly anti-conservative. For larger sample sizes ($n = 50, 100$) all intervals attain their nominal sizes except for the BC and BCa intervals where they remain anticonservative. In situations where $k = 2$ and small sample size, all intervals are asymmetric. As k increases, the intervals tend generally to be more symmetric. The performance of the BC and BCa intervals improves considerably for larger values of k . Also their performance improves for higher values of r , that is, the more symmetric the parent gamma distribution, the more symmetric the BC and BCa intervals tend to be.

Conclusion

It appears that the intervals proposed by Chen and Mi (2001) have a good performance except for situations of small sample size and heavy censoring. In this case the BTS, JAC, and especially BSD intervals provide better alternatives.

Table 1. Results for K=2

		n		20			30			50			100	
R	CP	M	L1	U1	T1	L2	U2	T2	L3	U3	T3	L4	U4	
0.4	0.6	CM	0.045	0.038	0.083	0.033	0.039	0.071	0.017	0.025	0.042	0.023	0.027	0.049
		BTS	0.043	0.012	0.055	0.032	0.016	0.047	0.018	0.018	0.036	0.018	0.010	0.028
		PRC	0.013	0.012	0.025	0.010	0.016	0.026	0.007	0.018	0.025	0.009	0.026	0.035
		BC	0.027	0.104	0.130	0.022	0.079	0.100	0.017	0.079	0.096	0.017	0.059	0.076
		BCA	0.015	0.104	0.119	0.021	0.082	0.103	0.012	0.083	0.095	0.013	0.060	0.073
		JAC	0.045	0.038	0.083	0.033	0.039	0.071	0.017	0.025	0.042	0.023	0.027	0.049
		BSD	0.013	0.038	0.051	0.017	0.039	0.055	0.017	0.027	0.044	0.023	0.027	0.050
	0.4	CM	0.019	0.050	0.069	0.027	0.028	0.055	0.014	0.034	0.048	0.020	0.035	0.055
		BTS	0.019	0.017	0.036	0.026	0.009	0.035	0.018	0.014	0.032	0.030	0.014	0.043
		PRC	0.018	0.004	0.022	0.026	0.002	0.028	0.018	0.014	0.032	0.030	0.009	0.038
		BC	0.116	0.173	0.289	0.011	0.101	0.112	0.007	0.132	0.139	0.015	0.079	0.093
		BCA	0.042	0.160	0.202	0.009	0.100	0.109	0.007	0.124	0.131	0.014	0.079	0.092
		JAC	0.019	0.050	0.069	0.027	0.028	0.055	0.030	0.034	0.064	0.020	0.035	0.055
		BSD	0.019	0.050	0.069	0.027	0.028	0.055	0.017	0.034	0.051	0.020	0.035	0.055
	0.2	CM	0.007	0.067	0.074	0.023	0.085	0.108	0.011	0.055	0.066	0.015	0.034	0.048
		BTS	0.007	0.020	0.027	0.032	0.017	0.049	0.026	0.012	0.038	0.049	0.003	0.052
		PRC	0.037	0.000	0.037	0.032	0.005	0.037	0.026	0.004	0.030	0.049	0.002	0.051
		BC	0.006	0.224	0.230	0.005	0.238	0.242	0.005	0.234	0.239	0.005	0.173	0.178
		BCA	0.005	0.184	0.188	0.005	0.234	0.239	0.005	0.230	0.235	0.005	0.172	0.177
		JAC	0.007	0.067	0.074	0.023	0.085	0.108	0.011	0.055	0.066	0.015	0.034	0.048
		BSD	0.007	0.064	0.070	0.023	0.038	0.060	0.012	0.041	0.053	0.017	0.034	0.051
0.8	0.6	CM	0.033	0.037	0.070	0.013	0.037	0.050	0.030	0.022	0.051	0.035	0.016	0.051
		BTS	0.033	0.013	0.045	0.013	0.003	0.016	0.013	0.011	0.023	0.029	0.016	0.045
		PRC	0.000	0.042	0.042	0.003	0.020	0.023	0.006	0.024	0.030	0.009	0.029	0.038
		BC	0.005	0.178	0.182	0.013	0.080	0.093	0.028	0.047	0.075	0.027	0.035	0.061
		BCA	0.001	0.178	0.178	0.008	0.087	0.095	0.012	0.050	0.061	0.020	0.040	0.059
		JAC	0.033	0.037	0.070	0.013	0.037	0.050	0.030	0.022	0.051	0.035	0.029	0.063
		BSD	0.005	0.037	0.042	0.013	0.037	0.050	0.030	0.022	0.052	0.031	0.017	0.048
	0.4	CM	0.029	0.035	0.064	0.026	0.032	0.057	0.019	0.030	0.049	0.022	0.025	0.047
		BTS	0.025	0.024	0.049	0.024	0.022	0.045	0.021	0.015	0.036	0.030	0.014	0.044
		PRC	0.005	0.024	0.029	0.011	0.022	0.032	0.011	0.015	0.026	0.017	0.014	0.031
		BC	0.010	0.168	0.178	0.011	0.137	0.147	0.014	0.078	0.092	0.014	0.079	0.092
		BCA	0.005	0.168	0.173	0.005	0.137	0.141	0.013	0.078	0.091	0.011	0.079	0.089
		JAC	0.029	0.035	0.064	0.026	0.032	0.057	0.019	0.030	0.049	0.022	0.025	0.047
		BSD	0.029	0.035	0.064	0.026	0.032	0.057	0.019	0.030	0.049	0.022	0.026	0.048
	0.2	CM	0.013	0.056	0.069	0.025	0.046	0.071	0.021	0.041	0.062	0.022	0.030	0.051
		BTS	0.013	0.019	0.032	0.025	0.014	0.039	0.034	0.010	0.044	0.047	0.009	0.056
		PRC	0.013	0.007	0.020	0.025	0.006	0.031	0.034	0.003	0.036	0.047	0.004	0.051
		BC	0.002	0.214	0.216	0.010	0.183	0.193	0.008	0.105	0.113	0.009	0.098	0.106
		BCA	0.002	0.214	0.216	0.008	0.173	0.180	0.008	0.100	0.108	0.008	0.097	0.105
		JAC	0.013	0.056	0.069	0.025	0.046	0.071	0.021	0.041	0.062	0.022	0.030	0.051
		BSD	0.013	0.056	0.069	0.025	0.046	0.071	0.021	0.041	0.062	0.022	0.030	0.051
1.2	0.6	CM	0.010	0.038	0.048	0.015	0.031	0.046	0.017	0.024	0.041	0.017	0.024	0.041
		BTS	0.010	0.014	0.024	0.015	0.024	0.039	0.017	0.004	0.021	0.016	0.020	0.036
		PRC	0.000	0.014	0.014	0.002	0.027	0.029	0.006	0.018	0.024	0.003	0.039	0.041
		BC	0.010	0.089	0.099	0.017	0.063	0.080	0.018	0.050	0.068	0.023	0.024	0.047
		BCA	0.000	0.089	0.089	0.015	0.063	0.078	0.017	0.056	0.073	0.018	0.027	0.044
		JAC	0.010	0.038	0.048	0.015	0.031	0.046	0.017	0.024	0.041	0.017	0.024	0.041
		BSD	0.010	0.038	0.048	0.015	0.031	0.046	0.017	0.021	0.038	0.017	0.024	0.041
	0.4	CM	0.020	0.033	0.053	0.013	0.050	0.062	0.024	0.028	0.051	0.019	0.026	0.045
		BTS	0.020	0.032	0.052	0.032	0.017	0.049	0.024	0.014	0.037	0.022	0.021	0.042
		PRC	0.004	0.032	0.036	0.012	0.017	0.029	0.011	0.014	0.025	0.014	0.021	0.035
		BC	0.004	0.163	0.166	0.013	0.099	0.111	0.016	0.094	0.109	0.019	0.065	0.084
		BCA	0.003	0.163	0.166	0.011	0.100	0.111	0.012	0.094	0.105	0.016	0.066	0.082
		JAC	0.020	0.033	0.053	0.033	0.050	0.083	0.024	0.028	0.051	0.019	0.041	0.059
		BSD	0.020	0.033	0.053	0.013	0.050	0.062	0.024	0.034	0.058	0.019	0.034	0.053

Table 1. Continued

		n		20			30			50			100	
R	CP	M	L1	U1	T1	L2	U2	T2	L3	U3	T3	L4	U4	
	0.2	CM	0.016	0.050	0.066	0.018	0.059	0.077	0.024	0.051	0.075	0.020	0.026	0.045
		BTS	0.016	0.017	0.033	0.018	0.024	0.041	0.035	0.013	0.048	0.028	0.009	0.037
		PRC	0.016	0.006	0.022	0.017	0.011	0.028	0.035	0.008	0.043	0.028	0.009	0.037
		BC	0.003	0.167	0.169	0.008	0.188	0.195	0.012	0.145	0.157	0.015	0.113	0.127
		BCA	0.002	0.165	0.167	0.004	0.188	0.191	0.011	0.143	0.154	0.014	0.111	0.125
		JAC	0.016	0.050	0.066	0.018	0.059	0.077	0.024	0.051	0.075	0.020	0.026	0.045
		BSD	0.016	0.050	0.066	0.018	0.047	0.065	0.024	0.036	0.060	0.023	0.027	0.049
1.6	0.6	CM	0.034	0.046	0.080	0.023	0.021	0.044	0.017	0.018	0.035	0.025	0.024	0.049
		BTS	0.034	0.008	0.041	0.023	0.004	0.026	0.017	0.017	0.034	0.012	0.028	0.039
		PRC	0.000	0.044	0.044	0.000	0.021	0.021	0.001	0.042	0.043	0.005	0.050	0.054
		BC	0.016	0.066	0.081	0.019	0.052	0.071	0.029	0.054	0.083	0.030	0.038	0.067
		BCA	0.009	0.068	0.076	0.005	0.061	0.066	0.016	0.057	0.073	0.026	0.043	0.068
		JAC	0.034	0.046	0.080	0.023	0.021	0.044	0.017	0.018	0.035	0.025	0.024	0.049
		BSD	0.000	0.046	0.046	0.006	0.021	0.027	0.017	0.019	0.036	0.022	0.026	0.047
	0.4	CM	0.012	0.034	0.045	0.037	0.024	0.060	0.019	0.038	0.056	0.030	0.036	0.066
		BTS	0.048	0.010	0.058	0.036	0.009	0.045	0.019	0.017	0.036	0.025	0.020	0.045
		PRC	0.012	0.010	0.022	0.015	0.009	0.024	0.010	0.017	0.027	0.014	0.020	0.034
		BC	0.013	0.073	0.086	0.015	0.051	0.066	0.018	0.051	0.069	0.024	0.049	0.073
		BCA	0.010	0.073	0.083	0.015	0.051	0.066	0.016	0.052	0.068	0.019	0.050	0.068
		JAC	0.012	0.034	0.045	0.037	0.024	0.060	0.019	0.038	0.056	0.030	0.036	0.066
		BSD	0.012	0.034	0.045	0.015	0.024	0.038	0.019	0.031	0.049	0.018	0.034	0.052
	0.2	CM	0.015	0.063	0.078	0.018	0.043	0.061	0.023	0.048	0.071	0.017	0.038	0.055
		BTS	0.015	0.021	0.036	0.018	0.018	0.035	0.031	0.016	0.047	0.033	0.013	0.045
		PRC	0.014	0.008	0.022	0.018	0.007	0.024	0.031	0.008	0.039	0.033	0.007	0.040
		BC	0.005	0.225	0.230	0.010	0.130	0.140	0.009	0.108	0.117	0.015	0.068	0.083
		BCA	0.005	0.211	0.215	0.007	0.124	0.130	0.008	0.103	0.111	0.014	0.067	0.081
		JAC	0.015	0.063	0.078	0.018	0.043	0.061	0.023	0.048	0.071	0.017	0.038	0.055
		BSD	0.015	0.063	0.078	0.018	0.043	0.061	0.023	0.046	0.069	0.017	0.037	0.054
2.0	0.6	CM	0.043	0.019	0.062	0.007	0.021	0.028	0.015	0.023	0.038	0.023	0.021	0.044
		BTS	0.043	0.014	0.057	0.007	0.010	0.017	0.015	0.016	0.030	0.010	0.022	0.032
		PRC	0.000	0.018	0.018	0.000	0.036	0.036	0.000	0.042	0.042	0.001	0.041	0.042
		BC	0.000	0.038	0.038	0.007	0.055	0.061	0.015	0.045	0.059	0.035	0.023	0.057
		BCA	0.000	0.043	0.043	0.005	0.091	0.096	0.015	0.045	0.060	0.025	0.025	0.050
		JAC	0.043	0.019	0.062	0.007	0.021	0.028	0.015	0.023	0.038	0.023	0.021	0.044
		BSD	0.000	0.019	0.019	0.007	0.021	0.028	0.015	0.022	0.037	0.023	0.021	0.044
	0.4	CM	0.025	0.049	0.074	0.029	0.023	0.052	0.036	0.021	0.057	0.022	0.027	0.048
		BTS	0.025	0.013	0.038	0.029	0.012	0.041	0.018	0.009	0.027	0.013	0.019	0.031
		PRC	0.005	0.013	0.017	0.009	0.012	0.020	0.009	0.021	0.030	0.006	0.036	0.042
		BC	0.008	0.095	0.103	0.024	0.052	0.076	0.019	0.079	0.098	0.013	0.062	0.075
		BCA	0.005	0.097	0.102	0.014	0.052	0.066	0.015	0.083	0.097	0.012	0.067	0.079
		JAC	0.025	0.049	0.074	0.029	0.023	0.052	0.036	0.021	0.057	0.022	0.027	0.048
		BSD	0.025	0.049	0.074	0.027	0.023	0.050	0.032	0.022	0.053	0.015	0.027	0.041
	0.2	CM	0.033	0.035	0.068	0.024	0.047	0.070	0.014	0.048	0.062	0.018	0.039	0.057
		BTS	0.031	0.015	0.046	0.021	0.020	0.040	0.016	0.022	0.038	0.031	0.009	0.040
		PRC	0.009	0.015	0.023	0.021	0.008	0.028	0.016	0.007	0.023	0.031	0.009	0.040
		BC	0.005	0.197	0.201	0.009	0.157	0.165	0.011	0.099	0.109	0.014	0.071	0.084
		BCA	0.003	0.188	0.190	0.008	0.153	0.161	0.010	0.098	0.107	0.011	0.071	0.082
		JAC	0.033	0.081	0.114	0.024	0.047	0.070	0.014	0.048	0.062	0.018	0.039	0.057
		BSD	0.016	0.036	0.052	0.024	0.047	0.070	0.014	0.041	0.055	0.019	0.033	0.051

Table 2: Results for K=4

		n		20			30			50			100	
R	CP	M	L1	U1	T1	L2	U2	T2	L3	U3	T3	L4	U4	
0.4	0.6	CM	0.017	0.035	0.052	0.021	0.032	0.053	0.016	0.030	0.046	0.028	0.028	0.056
		BTS	0.013	0.012	0.025	0.014	0.015	0.028	0.009	0.021	0.029	0.011	0.025	0.036
		PRC	0.002	0.030	0.031	0.003	0.033	0.036	0.002	0.034	0.036	0.007	0.036	0.043
		BC	0.027	0.077	0.104	0.028	0.062	0.090	0.026	0.048	0.074	0.040	0.037	0.077
		BCA	0.020	0.084	0.104	0.022	0.063	0.085	0.021	0.052	0.073	0.037	0.038	0.074
		JAC	0.018	0.040	0.058	0.023	0.033	0.056	0.016	0.030	0.046	0.028	0.029	0.057
		BSD	0.014	0.039	0.053	0.018	0.035	0.052	0.014	0.030	0.044	0.026	0.029	0.055
	0.4	CM	0.015	0.048	0.063	0.022	0.046	0.068	0.019	0.035	0.054	0.019	0.040	0.059
		BTS	0.011	0.022	0.032	0.018	0.022	0.040	0.017	0.017	0.034	0.019	0.019	0.037
		PRC	0.007	0.015	0.022	0.015	0.019	0.034	0.013	0.017	0.030	0.017	0.021	0.037
		BC	0.027	0.113	0.140	0.022	0.097	0.119	0.018	0.081	0.099	0.016	0.073	0.088
		BCA	0.013	0.111	0.124	0.019	0.090	0.108	0.017	0.079	0.096	0.015	0.073	0.088
		JAC	0.015	0.055	0.070	0.022	0.049	0.071	0.020	0.036	0.055	0.019	0.040	0.059
		BSD	0.016	0.049	0.065	0.021	0.047	0.068	0.021	0.037	0.057	0.018	0.040	0.058
	0.2	CM	0.015	0.068	0.083	0.017	0.054	0.071	0.016	0.045	0.061	0.018	0.038	0.055
		BTS	0.014	0.034	0.048	0.020	0.019	0.039	0.024	0.011	0.034	0.036	0.010	0.046
		PRC	0.023	0.007	0.030	0.030	0.006	0.035	0.026	0.006	0.032	0.039	0.006	0.045
		BC	0.012	0.182	0.194	0.008	0.152	0.160	0.009	0.141	0.150	0.006	0.130	0.136
		BCA	0.010	0.170	0.180	0.008	0.142	0.149	0.009	0.134	0.142	0.006	0.126	0.132
		JAC	0.018	0.072	0.090	0.017	0.058	0.075	0.017	0.046	0.062	0.018	0.038	0.055
		BSD	0.017	0.053	0.069	0.017	0.041	0.058	0.017	0.037	0.054	0.018	0.035	0.052
0.8	0.6	CM	0.018	0.045	0.063	0.020	0.032	0.052	0.019	0.025	0.044	0.020	0.030	0.049
		BTS	0.008	0.018	0.025	0.013	0.016	0.029	0.009	0.019	0.027	0.012	0.037	0.048
		PRC	0.000	0.047	0.047	0.001	0.046	0.047	0.002	0.037	0.038	0.004	0.055	0.058
		BC	0.024	0.070	0.094	0.032	0.054	0.086	0.034	0.029	0.062	0.050	0.029	0.078
		BCA	0.015	0.076	0.091	0.025	0.061	0.086	0.030	0.033	0.062	0.044	0.033	0.077
		JAC	0.018	0.048	0.066	0.021	0.033	0.054	0.019	0.026	0.045	0.020	0.031	0.051
		BSD	0.008	0.047	0.055	0.015	0.035	0.050	0.014	0.026	0.040	0.018	0.030	0.048
	0.4	CM	0.018	0.050	0.067	0.017	0.043	0.060	0.015	0.038	0.053	0.024	0.028	0.052
		BTS	0.009	0.019	0.028	0.010	0.020	0.029	0.012	0.021	0.033	0.020	0.015	0.035
		PRC	0.003	0.018	0.021	0.005	0.024	0.029	0.006	0.028	0.034	0.015	0.018	0.033
		BC	0.022	0.087	0.109	0.026	0.090	0.115	0.022	0.064	0.086	0.029	0.043	0.072
		BCA	0.016	0.085	0.101	0.020	0.089	0.109	0.018	0.064	0.082	0.027	0.044	0.071
		JAC	0.021	0.055	0.075	0.023	0.045	0.068	0.016	0.038	0.054	0.024	0.030	0.054
		BSD	0.013	0.051	0.064	0.018	0.045	0.062	0.015	0.039	0.054	0.024	0.028	0.052
	0.2	CM	0.014	0.056	0.070	0.019	0.053	0.072	0.013	0.044	0.056	0.014	0.037	0.051
		BTS	0.008	0.019	0.027	0.017	0.020	0.036	0.015	0.012	0.027	0.023	0.010	0.033
		PRC	0.009	0.004	0.013	0.020	0.007	0.027	0.015	0.009	0.023	0.026	0.007	0.032
		BC	0.009	0.142	0.151	0.014	0.133	0.147	0.006	0.105	0.111	0.009	0.103	0.111
		BCA	0.006	0.135	0.141	0.013	0.128	0.140	0.005	0.101	0.106	0.009	0.100	0.108
		JAC	0.014	0.063	0.076	0.020	0.057	0.076	0.013	0.044	0.056	0.015	0.038	0.052
		BSD	0.013	0.044	0.056	0.019	0.045	0.063	0.012	0.037	0.049	0.015	0.035	0.050
1.2	0.6	CM	0.014	0.031	0.045	0.018	0.028	0.046	0.013	0.028	0.041	0.018	0.029	0.047
		BTS	0.007	0.014	0.021	0.006	0.016	0.022	0.005	0.027	0.032	0.008	0.044	0.052
		PRC	0.000	0.040	0.040	0.000	0.049	0.049	0.001	0.065	0.065	0.002	0.065	0.067
		BC	0.023	0.052	0.075	0.034	0.030	0.064	0.035	0.031	0.066	0.050	0.021	0.071
		BCA	0.011	0.058	0.069	0.028	0.039	0.067	0.030	0.039	0.069	0.049	0.026	0.074
		JAC	0.014	0.035	0.049	0.021	0.029	0.050	0.014	0.029	0.043	0.018	0.029	0.047
		BSD	0.004	0.034	0.038	0.007	0.028	0.035	0.009	0.027	0.036	0.014	0.029	0.042
	0.4	CM	0.010	0.047	0.057	0.015	0.034	0.049	0.019	0.039	0.058	0.020	0.030	0.050
		BTS	0.008	0.016	0.024	0.010	0.020	0.030	0.011	0.020	0.031	0.012	0.025	0.036
		PRC	0.004	0.022	0.025	0.003	0.024	0.026	0.006	0.029	0.035	0.006	0.030	0.036
		BC	0.013	0.074	0.087	0.024	0.054	0.078	0.029	0.055	0.084	0.035	0.038	0.073
		BCA	0.011	0.074	0.085	0.018	0.055	0.072	0.025	0.057	0.082	0.031	0.040	0.070
		JAC	0.011	0.054	0.065	0.015	0.035	0.050	0.021	0.040	0.061	0.020	0.030	0.050
		BSD	0.009	0.048	0.056	0.013	0.035	0.048	0.018	0.039	0.056	0.019	0.031	0.050

Table 2. Continued

R	CP	n	M	L1	20 U1	T1	L2	30 U2	T2	L3	50 U3	T3	L4	100 U4
	0.2		CM	0.020	0.058	0.078	0.017	0.053	0.070	0.019	0.044	0.063	0.020	0.038
			BTS	0.015	0.019	0.033	0.016	0.021	0.036	0.019	0.020	0.038	0.025	0.018
			PRC	0.014	0.006	0.020	0.015	0.009	0.024	0.018	0.017	0.035	0.024	0.017
			BC	0.018	0.128	0.146	0.014	0.113	0.127	0.019	0.096	0.115	0.014	0.081
			BCA	0.014	0.122	0.136	0.013	0.110	0.123	0.019	0.092	0.111	0.014	0.079
			JAC	0.022	0.070	0.092	0.018	0.056	0.074	0.019	0.046	0.065	0.022	0.038
			BSD	0.019	0.052	0.071	0.016	0.049	0.065	0.019	0.042	0.061	0.022	0.037
1.6	0.6		CM	0.005	0.030	0.034	0.014	0.024	0.037	0.018	0.029	0.046	0.024	0.029
			BTS	0.002	0.012	0.014	0.009	0.014	0.023	0.006	0.035	0.040	0.007	0.053
			PRC	0.000	0.063	0.063	0.000	0.043	0.043	0.000	0.075	0.075	0.002	0.078
			BC	0.015	0.047	0.062	0.032	0.026	0.057	0.048	0.029	0.076	0.066	0.021
			BCA	0.009	0.057	0.065	0.025	0.029	0.053	0.041	0.034	0.074	0.061	0.024
			JAC	0.005	0.034	0.039	0.014	0.027	0.040	0.018	0.030	0.047	0.024	0.029
			BSD	0.002	0.030	0.032	0.007	0.025	0.031	0.009	0.027	0.036	0.018	0.026
	0.4		CM	0.013	0.040	0.052	0.019	0.042	0.060	0.019	0.026	0.045	0.020	0.031
			BTS	0.007	0.016	0.023	0.012	0.023	0.034	0.011	0.020	0.031	0.011	0.030
			PRC	0.000	0.035	0.035	0.001	0.048	0.049	0.002	0.028	0.030	0.007	0.039
			BC	0.013	0.071	0.083	0.021	0.073	0.094	0.030	0.040	0.069	0.032	0.036
			BCA	0.008	0.078	0.086	0.018	0.076	0.094	0.027	0.043	0.070	0.030	0.037
			JAC	0.013	0.043	0.056	0.020	0.045	0.064	0.020	0.028	0.048	0.020	0.031
			BSD	0.009	0.040	0.049	0.015	0.043	0.058	0.014	0.027	0.041	0.018	0.032
	0.2		CM	0.013	0.052	0.064	0.017	0.049	0.066	0.022	0.037	0.059	0.015	0.035
			BTS	0.011	0.017	0.028	0.011	0.015	0.026	0.018	0.015	0.033	0.014	0.015
			PRC	0.007	0.013	0.020	0.010	0.011	0.021	0.013	0.015	0.028	0.011	0.015
			BC	0.012	0.125	0.137	0.016	0.106	0.121	0.018	0.097	0.115	0.013	0.065
			BCA	0.011	0.121	0.132	0.013	0.103	0.116	0.018	0.096	0.114	0.012	0.064
			JAC	0.014	0.058	0.072	0.017	0.050	0.067	0.022	0.040	0.062	0.015	0.036
			BSD	0.013	0.043	0.055	0.015	0.043	0.058	0.022	0.035	0.057	0.015	0.032
2	0.6		CM	0.002	0.041	0.043	0.011	0.036	0.047	0.020	0.027	0.047	0.022	0.027
			BTS	0.000	0.019	0.019	0.004	0.031	0.035	0.008	0.046	0.054	0.007	0.070
			PRC	0.000	0.081	0.081	0.000	0.079	0.079	0.000	0.108	0.108	0.001	0.101
			BC	0.013	0.055	0.068	0.025	0.035	0.060	0.057	0.025	0.082	0.074	0.015
			BCA	0.002	0.070	0.072	0.020	0.042	0.062	0.049	0.033	0.082	0.069	0.019
			JAC	0.002	0.043	0.045	0.011	0.038	0.049	0.020	0.031	0.051	0.022	0.027
			BSD	0.000	0.038	0.038	0.002	0.035	0.036	0.009	0.025	0.034	0.013	0.024
	0.4		CM	0.014	0.040	0.054	0.011	0.039	0.049	0.011	0.042	0.053	0.020	0.034
			BTS	0.005	0.013	0.018	0.005	0.021	0.026	0.006	0.034	0.039	0.006	0.038
			PRC	0.000	0.026	0.026	0.001	0.037	0.038	0.003	0.051	0.053	0.003	0.063
			BC	0.017	0.057	0.074	0.020	0.054	0.074	0.024	0.049	0.073	0.039	0.034
			BCA	0.013	0.061	0.074	0.013	0.058	0.071	0.021	0.052	0.073	0.033	0.037
			JAC	0.014	0.042	0.056	0.011	0.044	0.055	0.013	0.044	0.056	0.020	0.036
			BSD	0.006	0.040	0.046	0.007	0.039	0.046	0.009	0.042	0.051	0.016	0.032
	0.2		CM	0.008	0.068	0.076	0.012	0.048	0.060	0.015	0.046	0.061	0.017	0.042
			BTS	0.008	0.025	0.033	0.008	0.017	0.024	0.012	0.021	0.033	0.015	0.021
			PRC	0.003	0.018	0.020	0.006	0.013	0.019	0.007	0.021	0.027	0.014	0.017
			BC	0.008	0.137	0.144	0.013	0.094	0.106	0.014	0.090	0.104	0.018	0.062
			BCA	0.006	0.134	0.140	0.011	0.092	0.103	0.012	0.088	0.100	0.017	0.062
			JAC	0.009	0.075	0.083	0.015	0.049	0.064	0.016	0.047	0.063	0.017	0.042
			BSD	0.008	0.059	0.066	0.012	0.043	0.055	0.015	0.041	0.056	0.016	0.039

Table 3: Results for K=8

		N		20			30			50			100	
R	CP	M	L1	U1	T1	L2	U2	T2	L3	U3	T3	L4	U4	
0.4	0.6	CM	0.022	0.048	0.069	0.019	0.039	0.057	0.021	0.033	0.053	0.021	0.021	0.042
		BTS	0.010	0.024	0.033	0.009	0.023	0.032	0.008	0.028	0.036	0.010	0.025	0.035
		PRC	0.001	0.046	0.047	0.001	0.047	0.048	0.003	0.042	0.044	0.005	0.040	0.045
		BC	0.042	0.073	0.114	0.045	0.054	0.099	0.049	0.036	0.085	0.052	0.022	0.074
		BCA	0.034	0.074	0.108	0.036	0.056	0.092	0.044	0.038	0.081	0.048	0.024	0.072
		JAC	0.022	0.051	0.073	0.022	0.040	0.062	0.021	0.033	0.054	0.022	0.022	0.043
	0.4	BSD	0.014	0.053	0.066	0.014	0.040	0.054	0.020	0.037	0.056	0.020	0.023	0.043
		CM	0.012	0.046	0.058	0.019	0.041	0.060	0.021	0.036	0.057	0.020	0.024	0.044
		BTS	0.010	0.024	0.033	0.011	0.022	0.032	0.012	0.020	0.032	0.012	0.015	0.027
		PRC	0.007	0.022	0.028	0.008	0.020	0.028	0.007	0.023	0.029	0.010	0.015	0.025
		BC	0.027	0.087	0.114	0.027	0.081	0.108	0.026	0.065	0.091	0.025	0.044	0.068
		BCA	0.016	0.080	0.096	0.023	0.077	0.099	0.024	0.065	0.088	0.024	0.044	0.068
	0.2	JAC	0.013	0.049	0.062	0.021	0.044	0.064	0.023	0.038	0.060	0.021	0.024	0.045
		BSD	0.012	0.047	0.059	0.020	0.044	0.064	0.022	0.039	0.061	0.021	0.027	0.048
		CM	0.011	0.075	0.086	0.014	0.059	0.072	0.011	0.045	0.056	0.016	0.040	0.055
		BTS	0.007	0.031	0.038	0.013	0.025	0.038	0.015	0.018	0.033	0.028	0.013	0.041
		PRC	0.012	0.005	0.016	0.018	0.007	0.025	0.024	0.005	0.028	0.036	0.008	0.043
		BC	0.011	0.156	0.166	0.011	0.149	0.159	0.009	0.124	0.132	0.010	0.111	0.121
	0.6	BCA	0.009	0.144	0.153	0.011	0.139	0.149	0.009	0.117	0.125	0.010	0.105	0.115
		JAC	0.013	0.082	0.095	0.015	0.062	0.077	0.012	0.046	0.058	0.016	0.040	0.055
		BSD	0.013	0.059	0.071	0.015	0.050	0.065	0.014	0.043	0.057	0.017	0.038	0.055
		CM	0.016	0.040	0.056	0.021	0.033	0.054	0.023	0.032	0.055	0.020	0.039	0.059
		BTS	0.010	0.024	0.034	0.008	0.025	0.033	0.007	0.041	0.047	0.007	0.060	0.067
		PRC	0.000	0.063	0.063	0.001	0.053	0.054	0.001	0.073	0.074	0.001	0.076	0.077
0.8	0.6	BC	0.038	0.053	0.090	0.042	0.034	0.076	0.059	0.032	0.090	0.067	0.031	0.097
		BCA	0.030	0.057	0.087	0.037	0.037	0.074	0.055	0.034	0.089	0.063	0.032	0.095
		JAC	0.017	0.043	0.060	0.023	0.038	0.060	0.024	0.034	0.058	0.020	0.041	0.061
		BSD	0.011	0.044	0.055	0.015	0.038	0.053	0.017	0.034	0.051	0.016	0.040	0.056
	0.4	CM	0.015	0.050	0.064	0.017	0.037	0.054	0.021	0.038	0.058	0.018	0.036	0.054
		BTS	0.005	0.027	0.032	0.005	0.022	0.027	0.012	0.028	0.040	0.009	0.032	0.041
		PRC	0.001	0.032	0.033	0.002	0.028	0.030	0.006	0.036	0.042	0.006	0.038	0.044
		BC	0.022	0.082	0.104	0.032	0.055	0.087	0.033	0.054	0.086	0.030	0.043	0.073
		BCA	0.019	0.080	0.099	0.027	0.054	0.081	0.031	0.054	0.085	0.028	0.044	0.071
		JAC	0.015	0.058	0.072	0.018	0.039	0.057	0.021	0.039	0.060	0.018	0.037	0.055
	0.2	BSD	0.011	0.055	0.066	0.016	0.040	0.055	0.021	0.041	0.062	0.018	0.038	0.056
		CM	0.012	0.062	0.074	0.013	0.052	0.065	0.013	0.038	0.050	0.017	0.034	0.050
		BTS	0.007	0.021	0.027	0.010	0.016	0.025	0.011	0.018	0.029	0.019	0.011	0.030
		PRC	0.007	0.007	0.014	0.011	0.007	0.018	0.012	0.015	0.026	0.022	0.011	0.033
		BC	0.014	0.121	0.135	0.014	0.112	0.126	0.012	0.103	0.115	0.014	0.082	0.096
		BCA	0.014	0.107	0.121	0.014	0.104	0.117	0.011	0.099	0.110	0.014	0.079	0.092
	0.6	JAC	0.015	0.065	0.080	0.013	0.055	0.068	0.013	0.038	0.051	0.017	0.034	0.050
		BSD	0.012	0.053	0.065	0.014	0.047	0.060	0.013	0.036	0.049	0.018	0.033	0.051
		CM	0.009	0.038	0.047	0.013	0.034	0.047	0.013	0.031	0.044	0.024	0.028	0.052
		BTS	0.003	0.024	0.027	0.006	0.032	0.037	0.006	0.052	0.057	0.007	0.062	0.069
		PRC	0.000	0.072	0.072	0.000	0.091	0.091	0.000	0.097	0.097	0.002	0.094	0.096
		BC	0.023	0.042	0.065	0.035	0.031	0.066	0.059	0.020	0.079	0.082	0.012	0.094
1.2	0.6	BCA	0.018	0.049	0.067	0.030	0.035	0.065	0.053	0.025	0.077	0.077	0.016	0.092
		JAC	0.009	0.042	0.051	0.014	0.037	0.051	0.013	0.034	0.047	0.024	0.028	0.052
		BSD	0.003	0.040	0.043	0.008	0.037	0.045	0.011	0.033	0.044	0.021	0.026	0.047

Table 3. Continued

R	CP	n		20			30			50			100	
		M	L1	U1	T1	L2	U2	T2	L3	U3	T3	L4	U4	
	0.4	CM	0.012	0.056	0.068	0.011	0.039	0.049	0.020	0.034	0.054	0.020	0.030	0.050
		BTS	0.005	0.029	0.033	0.005	0.027	0.032	0.008	0.029	0.037	0.010	0.035	0.044
		PRC	0.001	0.045	0.046	0.001	0.038	0.039	0.002	0.043	0.045	0.005	0.043	0.048
		BC	0.024	0.075	0.099	0.029	0.052	0.081	0.050	0.041	0.090	0.045	0.033	0.078
		BCA	0.018	0.073	0.090	0.026	0.052	0.078	0.045	0.042	0.086	0.042	0.033	0.075
		JAC	0.012	0.063	0.075	0.012	0.042	0.054	0.021	0.036	0.057	0.021	0.030	0.051
		BSD	0.009	0.059	0.067	0.009	0.042	0.050	0.018	0.036	0.054	0.019	0.030	0.049
	0.2	CM	0.011	0.059	0.070	0.017	0.043	0.059	0.016	0.046	0.062	0.015	0.030	0.045
		BTS	0.006	0.024	0.030	0.011	0.021	0.031	0.010	0.022	0.032	0.013	0.019	0.032
		PRC	0.004	0.014	0.018	0.011	0.015	0.026	0.011	0.019	0.030	0.015	0.018	0.033
		BC	0.016	0.115	0.131	0.021	0.075	0.095	0.017	0.082	0.099	0.017	0.059	0.076
		BCA	0.011	0.107	0.118	0.019	0.072	0.090	0.017	0.079	0.096	0.017	0.058	0.074
		JAC	0.014	0.062	0.076	0.017	0.044	0.061	0.016	0.048	0.064	0.016	0.031	0.046
		BSD	0.012	0.052	0.064	0.016	0.039	0.055	0.015	0.044	0.059	0.016	0.030	0.045
1.6	0.6	CM	0.005	0.035	0.040	0.011	0.031	0.042	0.011	0.033	0.044	0.020	0.031	0.051
		BTS	0.001	0.026	0.027	0.002	0.038	0.040	0.002	0.070	0.072	0.004	0.088	0.091
		PRC	0.000	0.080	0.080	0.000	0.112	0.112	0.000	0.136	0.136	0.000	0.121	0.121
		BC	0.029	0.033	0.062	0.036	0.026	0.062	0.051	0.019	0.070	0.095	0.009	0.104
		BCA	0.023	0.038	0.060	0.032	0.032	0.064	0.046	0.023	0.069	0.089	0.014	0.103
		JAC	0.007	0.039	0.046	0.011	0.033	0.044	0.011	0.034	0.045	0.021	0.031	0.052
		BSD	0.000	0.036	0.036	0.002	0.032	0.034	0.004	0.031	0.034	0.011	0.029	0.040
	0.4	CM	0.011	0.046	0.056	0.014	0.041	0.055	0.012	0.036	0.047	0.017	0.033	0.050
		BTS	0.006	0.029	0.035	0.004	0.029	0.033	0.005	0.038	0.043	0.006	0.047	0.053
		PRC	0.000	0.048	0.048	0.001	0.053	0.054	0.002	0.055	0.056	0.002	0.057	0.059
		BC	0.025	0.056	0.081	0.032	0.045	0.076	0.032	0.033	0.064	0.042	0.027	0.069
		BCA	0.018	0.056	0.074	0.028	0.046	0.074	0.028	0.035	0.063	0.040	0.028	0.068
		JAC	0.011	0.050	0.060	0.015	0.045	0.059	0.012	0.037	0.048	0.018	0.033	0.051
		BSD	0.009	0.049	0.058	0.009	0.041	0.050	0.009	0.035	0.044	0.014	0.033	0.047
	0.2	CM	0.013	0.055	0.068	0.010	0.047	0.057	0.017	0.044	0.061	0.014	0.035	0.048
		BTS	0.007	0.024	0.031	0.006	0.021	0.027	0.006	0.026	0.032	0.008	0.022	0.030
		PRC	0.005	0.018	0.022	0.004	0.016	0.020	0.005	0.022	0.027	0.008	0.021	0.029
		BC	0.020	0.091	0.111	0.014	0.081	0.095	0.021	0.068	0.089	0.019	0.055	0.074
		BCA	0.015	0.083	0.098	0.012	0.076	0.088	0.020	0.063	0.083	0.018	0.052	0.070
		JAC	0.014	0.062	0.076	0.011	0.049	0.060	0.017	0.045	0.062	0.014	0.035	0.049
		BSD	0.012	0.050	0.061	0.010	0.045	0.055	0.015	0.040	0.055	0.013	0.034	0.047
2	0.6	CM	0.002	0.037	0.039	0.010	0.032	0.041	0.013	0.029	0.042	0.019	0.029	0.048
		BTS	0.000	0.030	0.030	0.004	0.053	0.057	0.004	0.071	0.075	0.004	0.118	0.122
		PRC	0.000	0.112	0.112	0.000	0.108	0.108	0.000	0.135	0.135	0.000	0.156	0.156
		BC	0.019	0.036	0.054	0.046	0.017	0.063	0.066	0.016	0.082	0.102	0.013	0.115
		BCA	0.013	0.049	0.062	0.040	0.025	0.064	0.065	0.019	0.084	0.098	0.015	0.113
		JAC	0.002	0.046	0.048	0.011	0.037	0.048	0.014	0.030	0.044	0.019	0.030	0.048
		BSD	0.000	0.036	0.036	0.003	0.027	0.030	0.006	0.024	0.030	0.011	0.024	0.035
	0.4	CM	0.006	0.040	0.046	0.006	0.052	0.057	0.018	0.034	0.052	0.018	0.030	0.047
		BTS	0.001	0.027	0.028	0.002	0.045	0.047	0.008	0.045	0.053	0.007	0.049	0.056
		PRC	0.000	0.059	0.059	0.000	0.078	0.078	0.001	0.067	0.067	0.002	0.073	0.075
		BC	0.018	0.053	0.071	0.021	0.052	0.072	0.047	0.030	0.076	0.049	0.020	0.069
		BCA	0.015	0.056	0.070	0.017	0.056	0.073	0.040	0.034	0.074	0.044	0.022	0.066
		JAC	0.007	0.045	0.052	0.006	0.056	0.062	0.019	0.035	0.053	0.018	0.031	0.049
		BSD	0.003	0.040	0.043	0.004	0.051	0.055	0.013	0.034	0.047	0.015	0.028	0.043

Table 3. Continued

		n		20			30			50			100	
R	CP	M	L1	U1	T1	L2	U2	T2	L3	U3	T3	L4	U4	
	0.2	CM	0.014	0.065	0.078	0.010	0.048	0.057	0.014	0.048	0.062	0.018	0.037	0.055
		BTS	0.006	0.032	0.038	0.003	0.028	0.031	0.007	0.028	0.034	0.009	0.027	0.036
		PRC	0.001	0.029	0.030	0.001	0.029	0.030	0.004	0.027	0.031	0.007	0.031	0.038
		BC	0.018	0.100	0.117	0.015	0.086	0.101	0.022	0.071	0.093	0.023	0.049	0.071
		BCA	0.015	0.091	0.105	0.011	0.083	0.094	0.021	0.067	0.088	0.022	0.048	0.070
		JAC	0.015	0.069	0.084	0.010	0.051	0.061	0.014	0.050	0.064	0.018	0.038	0.056
		BSD	0.011	0.061	0.072	0.008	0.048	0.056	0.012	0.048	0.059	0.017	0.038	0.055

References

- Al-Nasser, A., & Baklizi, A. (2004). Large sample and bootstrap intervals for the scale of burr type x distribution based on grouped data. *Journal of Modern Applied Statistical Methods*, 3(2), 386 – 392.
- Chen, Z., & Mi, J. (2001). An approximate confidence interval for the scale parameter of the gamma distribution based on grouped data. *Statistical Paper*, 4, 285-299.
- Cohen, A., & Whitten, B. (1988). *Parameter estimation in reliability and life span models*. New York, N.Y: Marcel Dekker.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia, PA: SIAM.
- Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, N.Y.: Chapman and Hall.
- Jennings, D. (1987). How do we judge confidence intervals adequacy? *The American Statistician*, 41(4), 335-337.
- Kulldorff, G. (1961). *Estimation from grouped and partially grouped samples*. New York, N.Y.: Wiley.
- Meeker, Jr, W. (1986). Planning Life tests in which units are inspected for failure. *IEEE Trans. on Reliability R-35*, 571-578.
- Pettitt, A. N., & Stephens, M. A. (1977). The Kolmogrov-Smirnov goodness-of-fit statistic with discrete and grouped data. *Technometrics*, 19, 205 – 210.

Applications of Some Improved Estimators in Linear Regression

B. M. Golam Kibria
Department of Statistics
Florida International University

The problem of estimation of the regression coefficients under multicollinearity situation for the restricted linear model is discussed. Some improved estimators are considered, including the unrestricted ridge regression estimator (URRE), restricted ridge regression estimator (RRRE), shrinkage restricted ridge regression estimator (SRRRE), preliminary test ridge regression estimator (PTRRE), and restricted Liu estimator (RLIUE). They were compared based on the sampling variance-covariance criterion. The RRRE dominates other ridge estimators when the restriction does or does not hold. A numerical example was provided. The RRRE performed equivalently or better than the RLIUE in the sense of having smaller sampling variance.

Key words: Bias; non-Central F, preliminary test, restricted Liu estimator, ridge regression, shrinkage estimation, variance matrix.

Introduction

Multiple linear regression model plays an important role in statistical inference and is used extensively in business, environmental, industrial, and social sciences. In linear regression model, one usually assumes that the explanatory variables are independent. However, in practice, there may be strong or near to strong linear relationships among the explanatory variables. In that case, the independence assumptions are no longer valid, which causes the problem of multicollinearity. In the presence of multicollinearity, it is difficult to estimate the unique effects of individual variables in the

regression equation. Moreover, the estimated regression coefficient will have large sampling variance which affects both inference and prediction. In the literature, there are various methods existing to solve this problem. Among them, ridge regression is the most popular one which has much usefulness in real life. To describe the ridge regression, the following multiple linear regression model was considered:

$$y = X\beta + e, \quad (1)$$

where y is a random vector of length n , β is a $p \times 1$ vector of fixed but unknown regression coefficients, X is a $n \times p$ known design matrix of rank p and e is a $n \times 1$ vector random variable, which is distributed as multivariate normal with mean vector 0 and covariance matrix $\sigma^2 \times I_n$, I_n is an identity matrix of order n , and $\sigma^2 > 0$ is unknown error variance. Of primary interest is an estimation of the regression coefficient β , when it is *a priori* suspected that β may be restricted to the subspace

$$H\beta = h, \quad (2)$$

B. M. Golam Kibria is Assistant Professor. He is a fellow of the Royal Statistical Society and the recipient of the 2005 Chapter Service Recognition Award. He is the Associate Editor and member of the editorial board of three international statistical journals. He has authored/co-authored about 40 articles in statistical journals.

where H is a $q \times p$ known matrix of rank $q (< p)$ and h is an $q \times 1$ vector of known constants. The usual least squares estimate (ULSE) or the maximum likelihood estimate (MLE) of β is given by $\tilde{\beta} = C^{-1}X'y$, which will heavily depend on the characteristics of the matrix $C = X'X$. If the C matrix is ill conditioned ($\det(X'X \approx 0)$), the LSE are sensitive to a number of errors, for example, some of the regression coefficients may be statistically insignificant with wrong sign and meaningful statistical inference become difficult for the practitioners. To solve this problem, Hoerl and Kennard (1970) suggested use of $C(k) = C + kI_p, (k \geq 0)$, instead of C for estimating β . The resulting estimators are given as:

$$\tilde{\beta}(k) = (C + kI_p)^{-1}X'y, \quad (3)$$

which are known as unrestricted ridge regression estimators (URRE). The constant, $k > 0$ is known as shrinkage or biasing or ridge parameter. Though these estimators result in biasing for $k \geq 0$, they yield minimum mean square error (MSE) compared to the unrestricted least squares estimator (URLSE) (for examples, see Hoerl & Kennard, 1970, Saleh & Kibria, 1993). The objective of this article is to compare the performance of some ridge regression estimators under the sampling variance-covariance criterion. The findings will be illustrated with a numerical example.

Proposed Estimators, Biases and Variances

In this section, some ridge type estimators are considered for β , when $H\beta = h$ is *a priori* suspected in the case of multicollinearity. Biases and variance-covariance matrices of the estimators were also provided.

Unrestricted Ridge Regression Estimator (URRE):

The URRE in (3), can be re-written as follows:

$$\tilde{\beta}(k) = W\tilde{\beta}, \quad W = [I_p + kC^{-1}]^{-1}; \quad k \geq 0, \quad (4)$$

where $\tilde{\beta}$ is the unrestricted least square estimator (URLSE) of β . The bias and the variance matrix of URRE are as follows:

$$\begin{aligned} \text{Bias} = B_1(k) &= E(\tilde{\beta}(k)) - \beta = -kC^{-1}(k)\beta; \\ C(k) &= [C + kI_p]; \\ V(\tilde{\beta}(k)) = V_1(k) &= \sigma^2(WC^{-1}W'). \end{aligned} \quad (5)$$

Restricted Ridge Regression Estimator (RRRE):
Sarkar (1992) proposed the following RRRE:

$$\hat{\beta}(k) = W\hat{\beta}, \quad (6)$$

where $\hat{\beta} = \tilde{\beta} - C^{-1}H'(HC^{-1}H')^{-1}(H\tilde{\beta} - h)$ is the restricted least squares estimator (RLSE) of β . The bias and the variance matrix of RRRE are as follows:

$$\begin{aligned} B_2(k) &= -W\eta - kC^{-1}(k)\beta; \\ V_2(k) &= \sigma^2(WC^{-1}W') - \sigma^2(WAW'), \end{aligned} \quad (7)$$

where

$$A = C^{-1}H'(HC^{-1}H')^{-1}HC^{-1},$$

$$\eta = C^{-1}H'(HC^{-1}H')^{-1}\delta$$

and

$$\delta = H\beta - h.$$

Shrinkage Restricted Ridge Regression Estimator (SRRRE):

Haq and Kibria (1996) proposed the following SRRRE,

$$\hat{\beta}^{SR}(k) = W\hat{\beta}^{SR} \quad (8)$$

where

$$\hat{\beta}^{SR} = d\tilde{\beta} + (1-d)\hat{\beta}, \quad 0 \leq d \leq 1,$$

is the shrinkage restricted least squares estimator (SRLSE) of β and the coefficient d is known as the coefficient of distrust and may completely be determined by the user, depending on the degree of belief about the null hypothesis. For $d=1$, $\tilde{\beta}$ is obtained and for $d=0$, $\hat{\beta}$ is obtained. Therefore, $\hat{\beta}^{SR}(k)$ is a convex combination of $\tilde{\beta}$ and $\hat{\beta}$. The bias and variance matrix of SRRRE are given as follows:

$$\begin{aligned} B_3(k, d) &= -(1-d)W\eta - kC^{-1}(k)\beta; \\ V_3(k, d) &= \sigma^2(WC^{-1}W) - (1-d^2)\sigma^2(WAW). \end{aligned} \quad (9)$$

Preliminary Test Ridge Regression Estimator (PTRRE):

Saleh and Kibria (1993), first introduced the preliminary test ridge regression estimator,

$$\hat{\beta}^{PT}(k) = W\hat{\beta}^{PT} = \begin{cases} W\tilde{\beta} & \text{if } H_0 \text{ is false} \\ W\hat{\beta} & \text{if } H_0 \text{ is true,} \end{cases} \quad (10)$$

where $\hat{\beta}^{PT} = \tilde{\beta}I(Ln \geq F_\alpha) + \hat{\beta}I(Ln < F_\alpha)$ is the preliminary test least squares estimator (PTLSE) of β , which is introduced by Bancraft (1944) and $I(X)$ is the indicator function of X . Here Ln is the test-statistic for testing the null-hypothesis $H_0: H\beta = h$ against $H_1: H\beta \neq h$ and is given by

$$Ln = \frac{(H\tilde{\beta}_n - h)'(HC^{-1}H')^{-1}(H\tilde{\beta}_n - h)}{q\tilde{\sigma}_e^2}, \quad (11)$$

where

$$\tilde{\sigma}_e^2 = (n-p)^{-1}(Y - X\tilde{\beta}_n)'(Y - X\tilde{\beta}_n)$$

is the unrestricted least squares estimator of σ^2 . The corresponding restricted least squares estimator of σ^2 is given by

$$\hat{\sigma}_e^2 = (n-p+q)^{-1}(Y - X\hat{\beta}_n)'(Y - X\hat{\beta}_n).$$

Under H_0 , the test-statistic Ln follows a central F -distribution with $(q, n-p)$ degrees of freedom and F_α is the upper α -level critical value from this distribution. However, when H_0 does not hold, Giles (1991) had shown that the probability density function of Ln is non-central F -distribution with non-centrality parameter $\frac{\Delta}{2}$, where

$$\Delta = \frac{(H\beta - h)'(HC^{-1}H')^{-1}(H\beta - h)}{\sigma^2} = \frac{\eta' C \eta}{\sigma^2}. \quad (12)$$

is called the departure parameter. For the power of the F -test, readers are referred to Sutradhar (1988) and Giles (1991) among others. Preliminary test and ridge regression approaches have been studied by Bancroft (1944, 1964), Bock and Yancey (1973), Giles (1991), Han and Bancroft (1968), Judge and Bock (1978), Kibria and Saleh (1993), Kibria (2003), Kibria (2004a,b), Kibria and Saleh (2004, 2005a, 2005b, 2006), Obenchain et al. (1975) and Saleh and Han (1990) to mention a few. The bias and variance matrix of the PTRRE are as follows:

$$\begin{aligned} B_4(k, \Delta, \alpha) &= -W\eta G_{q+2, n-p}(I_\alpha^*; \Delta) \\ &\quad - kC^{-1}(k)\beta; \\ V_4(k, \Delta, \alpha) &= \sigma^2(WC^{-1}W) \\ &\quad - \sigma^2(WAW)G_{q+2, n-p}(I_\alpha^*; \Delta) \\ &\quad + (W\eta\eta'W)G(\Delta, \alpha), \end{aligned} \quad (13)$$

where

$$G(\Delta, \alpha) = \begin{bmatrix} 2G_{q+2, n-p}(l_{\alpha}^*; \Delta) \\ -G_{q+4, n-p}(l_{\alpha}^{**}; \Delta) \\ -G_{q+2, n-p}^2(l_{\alpha}^*; \Delta) \end{bmatrix} \quad (14)$$

and $G_{m,n}(\cdot; \Delta)$ is the cumulative density function (*cdf*) of a non-central F-distribution with (m, n) degrees of freedom and the non-centrality parameter Δ . Also $l_{\alpha}^* = \frac{q}{q+2}F_{\alpha}$ and $l_{\alpha}^{**} = \frac{q}{q+4}F_{\alpha}$. When $\alpha = 1$, then $F_{\alpha} = 0$ and $G_{m,n}(\cdot; \Delta) = 0$, consequently, the H_0 will be rejected and thus PTRRE becomes URRE. On the other hand, when $\alpha = 0$, then $F_{\alpha} = \infty$ and $G_{m,n}(\cdot; \Delta) = 1$, consequently, the H_0 will not be rejected and thus PTRRE becomes RRRE.

Restricted Liu Estimator (RLIUE):

Kaciranlar et al. (1999) proposed the following restricted Liu estimator (RLIUE):

$$\hat{\beta}^{RL}(b) = F_b \hat{\beta}, \quad (15)$$

based on the corresponding unrestricted Liu estimator (ULIUE) (Liu 1993), $\hat{\beta}^{UL}(b) = F_b \tilde{\beta}$, where $F_b = (C + I)^{-1}(C + bI)$ and b is the constant to be determined. Note that for $b = 1$, the ULIUE and RLIUE become corresponding URLSE and RLSE respectively. It also noted that for $k = 1$ and $b = 0$ both RRRE and RLIUE become the same. The bias and the variance matrix of the RLIUE are as follows:

$$\begin{aligned} B_5(b, \Delta) &= (b-1)(C + I)^{-1} \beta + F_b \eta; \\ V_5(b) &= \sigma^2 \{F_b (C^{-1} - A) F_b'\}. \end{aligned} \quad (16)$$

In the following section, the performance of the proposed ridge type

estimators will be proposed under the restriction (2) and using the sampling variance criterion. This is because the value of the test statistics depends on the sampling variance than the mean squares error (MSE). Because the detailed analysis about ULIUE and RLIUE are available in Kaciranlar et al. (1999), for convenience, the comparison of ridge regression estimators with that of RLIUE would be provided only by an example.

Performance of the estimators under H_0

Comparison between $\tilde{\beta}$ and $\tilde{\beta}(k)$

The difference between the sampling variance of URLSE and URRE is

$$\begin{aligned} Var(\tilde{\beta}) - Var(\tilde{\beta}(k)) &= \sigma^2 [C^{-1} - C^{-1}(k)W'] \\ &= \sigma^2 W[W^{-1}C^{-1}W'^{-1} - C^{-1}]W' \\ &= \sigma^2 W[(I_p + kC^{-1})C^{-1}(I_p + kC^{-1})' - C^{-1}]W' \\ &= \sigma^2 W[2kC^{-2} + k^2C^{-3}]W' \\ &= k\sigma^2 C^{-1}(k)EC^{-1}(k)', \end{aligned} \quad (17)$$

where $C(k) = [C + kI]$ and $E = (2I + kC^{-1})$ are the positive definite (p.d.) matrices. Thus, the difference in (17) is non-negative definite (n.n.d) for $k > 0$ and therefore, $\tilde{\beta}(k)$ is superior to $\tilde{\beta}$ with respect to sampling variance. It further notes that the variance of $\tilde{\beta}$ will be equal to that of $\tilde{\beta}(k)$ iff $k = 0$.

Comparison between $\hat{\beta}$ and $\hat{\beta}(k)$

The difference between the sampling variance of RLSE and RRRE is

$$\begin{aligned} V_2(0) - V_2(k) &= \sigma^2 [(C^{-1} - A) - W(C^{-1} - A)W'] \\ &= k\sigma^2 C^{-1}(k)[E - F_0]C^{-1}(k)', \end{aligned} \quad (18)$$

where $F_0 = (CA + AC + kA)$ is a symmetric matrix. Then, following Anderson (1984)

(Theorem A.2.2, p.589), there existed a nonsingular matrix P such that $P'EP = I$ and $P'F_0P = \Lambda$, where Λ is a diagonal matrix and its elements are roots of the polynomial equation $|E - \lambda F_0| = 0$. Because

$(P'EP - P'F_0P) = P'(E - F_0)P = I - \Lambda$ is a non-negative definite matrix, then $(1 - \lambda_i) \geq 0$. That means, $(P'EP - P'F_0P)$ to be the n.n.d. matrix, it is necessary that $\lambda_{\max}(E^{-1}F_0) \leq 1$.

Let $\lambda_{\max}(E^{-1}F_0) \leq 1$. Because E is positive definite matrix and F_0 is symmetric matrix, then $\lambda_p \leq \frac{x'F_0x}{x'Ex} \leq \lambda_1$, where

$\lambda_1(E^{-1}F_0) \geq \lambda_2(E^{-1}F_0) \leq \dots \leq \lambda_p(E^{-1}F_0)$ are the roots of $|E - \lambda F_0| = 0$ (see Anderson, 1984, Theorem A.2.4, p. 590). Then, it is obvious that $V_2(0) - V_2(k)$ is non-negative definite for $k \geq 0$ if and only if $\lambda_{\max}(E^{-1}F_0) \leq 1$. Therefore, sampling variance of RRRE is always less than or equal to that of RLSE if and only if $\lambda_{\max}(E^{-1}F_0) \leq 1$.

If the sampling variance of URRE is compared with that of RRRE, then

$$V_1(k) - V_2(k) = \sigma^2(WAW'); k \geq 0,$$

which is a positive semidefinite matrix. Therefore, RRRE is always superior to that of URRE, when the sampling variance criterion is considered.

Comparison between $\hat{\beta}^{SR}$ and $\hat{\beta}^{SR}(k)$

The difference between sampling variance of SRLSE and SRRRE is

$$\begin{aligned} & V_3(0, d) - V_3(k, d) \\ &= \sigma^2[(C^{-1} - (1 - d^2)A) - W(C^{-1} - (1 - d^2)A)W'] \\ &= k\sigma^2 C^{-1}(k)[E - F_1]C^{-1}(k)', \end{aligned} \quad (19)$$

where $E = 2I + kC^{-1}$ and $F_1 = (1 - d^2)[CA + AC + kA]$. Then, it is

obvious that $V_3(0, d) - V_3(k, d)$ is non-negative definite for $k \geq 0$ and $0 \leq d \leq 1$ if and only if $\lambda_{\max}(E^{-1}F_1) \leq 1$.

If the sampling variance of SRRRE is compared with that of URRE, then

$$V_1(k) - V_3(k, d) = \sigma^2(1 - d^2)WC^{-1}W',$$

which is always positive semi definite matrix for all $k \geq 0$ and $0 \leq d \leq 1$. Thus, SRRRE is superior to URRE. If the sampling variance of SRRRE is compared with that of RRRE, then

$$V_2(k) - V_3(k, d) = -\sigma^2 d^2 WC^{-1}W',$$

which is negative semidefinite. Therefore, RRRE is superior to SRRRE for all d and k when the restriction holds.

Comparison between $\hat{\beta}^{PT}$ and $\hat{\beta}^{PT}(k)$

The difference between the sampling variance of PTLSE and PTRRE is

$$\begin{aligned} & V_4(0, 0, \alpha) - V_4(k, 0, \alpha) \\ &= k\sigma^2 C^{-1}(k)[E - F_2]C^{-1}(k)', \end{aligned} \quad (20)$$

where $F_2 = [CA + AC + kA]G_{q+2, n-p}(l_\alpha^*; 0)$.

Then, the sampling variance of $\hat{\beta}^{PT}(k)$ is always less than or equal to that of $\hat{\beta}^{PT}$ if and only if $\lambda_{\max}(E^{-1}F_2) \leq 1$ for all α . If the sampling variance of PTRRE is compared with that of URRE, then

$$\begin{aligned} & V_1(k) - V_4(k, 0, \alpha) \\ &= \sigma^2(WAW')G_{q+2, n-p}(l_\alpha^*; 0), \end{aligned}$$

which is always positive semi definite matrix for all $k \geq 0$. Thus, PTRRE is superior to URRE. If PTRRE is compared with that of RRRE, then

$$\begin{aligned} & V_2(k) - V_4(k, 0, \alpha) \\ &= \sigma^2(WAW')(G_{q+2, n-p}(l_\alpha^*; 0) - 1), \end{aligned}$$

which is negative semidefinite for all $\alpha \in (0, 1)$. Therefore, RRRE is superior to PTRRE for all α and k when the restriction hold. If the variance of PTRRE is compared with that of SRRRE, then

$$V_3(k, d) - V_4(k, 0, \alpha) = \sigma^2(WAW') \begin{Bmatrix} G_{q+2, n-p}(I_\alpha^*; 0) \\ -(1-d^2) \end{Bmatrix}.$$

Thus, the PTRRE will dominates SRRRE if

$$F_\alpha > \left(\frac{q+2}{q} \right) G_{q+2, n-p}^{-1}(1-d^2), \quad (21)$$

otherwise, SRRRE will dominate PTRRE.

It is observed from equations (5), (7), (9), (13) and (16), that under the restrictions, all the proposed ridge estimators are biased and the amount of biases are same for the ridge regression estimators. For $k=0$, all proposed ridge estimators become corresponding usual least squares estimator and under restrictions they are unbiased. It is also observed that the RRRE dominates all other estimators under the variance criterion and when the restriction holds. In the following section, the performance of the estimators will be compared when the restriction does not hold.

Performance of the estimators under H_a

Comparison between $\hat{\beta}^{PT}$ and $\hat{\beta}^{PT}(k)$

The difference between the sampling variance of PTLSE and PTRRE is

$$\begin{aligned} V_4(0, \Delta, \alpha) - V_4(k, \Delta, \alpha) \\ = k\sigma^2 C^{-1}(k)[E - F_3]C^{-1}(k)', \end{aligned} \quad (22)$$

where $E = (2I + kC^{-1})$ is positive definite matrix and

$$\begin{aligned} F_3 &= (CA + AC + kA)G_{q+2, n-p}(I_\alpha^*; \Delta) \\ &\quad - \frac{1}{\sigma^2}(\eta\eta'C + C\eta\eta' + k\eta\eta')G(\Delta, \alpha) \end{aligned}$$

is a symmetric matrix and $G(\Delta, \alpha)$ is greater than 0 for all α and Δ . Then, following Anderson (1974, Theorems A.2.2 and A.2.4), the difference in (22) will be non-negative definite if and only if $\lambda_{\max}(E^{-1}F_3) \leq 1$. Therefore, under the alternative hypothesis, sampling variance of $\hat{\beta}^{PT}(k)$ ($k \geq 0$) is always less than or equal to that of $\hat{\beta}^{PT}$ if and only if $\lambda_{\max}(E^{-1}F_3) \leq 1$.

Comparison between PTRRE and URRE

$$\begin{aligned} V_1(k) - V_4(k, \Delta, \alpha) \\ = \sigma^2(WAW')G_{q+2, n-p}(I_\alpha^*; \Delta) - W\eta\eta'W'G(\alpha, \Delta), \end{aligned} \quad (23)$$

where $G(\alpha, \Delta)$ is defined in (14).

Let P be the orthogonal matrix with eigen vectors of C so that

$$P'CP = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p).$$

Because (WAW') and $(W\eta\eta'W')$, are the $p \times p$ symmetric matrices, the following may be written

$$WAW' = P[\Lambda + kI]^{-1}\Lambda A^*\Lambda[\Lambda + kI]^{-1}P'$$

where $P'AP = A^*$. Then, the i^{th} diagonal element of the matrix WAW' is $\frac{a_i^* \lambda_i^2}{(\lambda_i + k)^2}$.

Similarly, the i^{th} diagonal element of the matrix

$W\eta\eta'W'$ is $\frac{\eta_i^{*2} \lambda_i^2}{(\lambda_i + k)^2}$, where η_i^* is the i^{th}

element if the vector $\eta^* = \eta'P$. Then, the difference in (23) is positive semidefinite if,

$$a_i^* \geq \left[\frac{2G_{q+2, n-p}(I_\alpha^*; \Delta) - G_{q+4, n-p}(I_\alpha^{**}; \Delta) - G_{q+2, n-p}^2(I_\alpha^*; \Delta)}{G_{q+2, n-p}(\alpha, \Delta)} \right] \eta_i^{*2}. \quad (24)$$

Thus, PTRRE will dominate URRE when (24) holds.

Special Cases on Δ

Consider $k = 0$ in (23), and then following Judge and Bock (1978, p.115-117), the sampling variance of PTLSE is obtained and is smaller than that of the URLSE when,

$$\Delta \leq \frac{G_{q+2,n-p}(l_{\alpha}^*; \Delta)}{2[2G_{q+2,n-p}(l_{\alpha}^*; \Delta) - G_{q+4,n-p}(l_{\alpha}^{**}; \Delta) - G_{q+2,n-p}^2(l_{\alpha}^*; \Delta)]}$$

This result coincides with that of Judge and Bock (1978, p.116).

Comparison between PTRRE and RRRE

$$V_2(k) - V_5(k, d, \Delta, \alpha) = - \left\{ \sigma^2(WAW') [1 - G_{q+2,n-p}(l_{\alpha}^*; \Delta)] + [(W\eta\eta'W)G(\alpha, \Delta)] \right\}$$

It is observed that the above difference is always negative semi-definite for all $\alpha \in (0, 1)$, $\Delta \in (0, \infty)$ and $d \in (0, 1)$. Therefore, the RRRE dominates PTRRE, SRRRE and URRE whether the restriction does or does not holds.

Comparison between PTRRE and SRRRE

$$V_3(k, d) - V_4(k, \Delta, \alpha) = - \left\{ \sigma^2(WAW') \{ (1 - d^2 - G_{q+2,n-p}(l_{\alpha}^*; \Delta)) \} + W\eta\eta'W'G(\alpha, \Delta) \right\} \quad (25)$$

The difference in (25) will be negative semidefinite and therefore, the SRRRE will dominates PTRRE if

$$d \leq \sqrt{1 - G_{q+2,n-p}(l_{\alpha}^*; \Delta)},$$

otherwise, PTRRE will dominates SRRRE.

Numerical Example

An example is considered that demonstrates the performance of the proposed estimators. Accordingly, the dataset on Portland cement, originally due to Woods et al. (1932), was analyzed. This dataset has widely been analyzed by Hald (1952, p.635-652), Gorman and Torman (1966, p.735), Nomura (1988, p.735), Piepel and Redgate (1998) and recently Kaciranlar et al. (1999) among many researchers. This data came from an experimental investigation of the heat evolved during the setting and hardening of the Portland cements of varied composition as a function of the percentage of four compounds in the clinkers from which the cement was produced. Woods et al. (1932) considered the following four compounds: Tricalcium aluminate ($3CaO.Al_2O_3$), Tricalcium silicate ($3CaO.SiO_2$), Tetracalcium aluminoferrite ($4CaO.Al_2O_3.Fe_2O_3$), and β -dicalcium silicate ($2CaO.SiO_2$). The heat evolved after 180 days of curing is measured in calories per gram of cement. For details about the dataset, please see Woods et al. (1932). Consider the following linear model to analyze the cement dataset.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e_i; \quad i = 1, 2, \dots, 13, \quad (26)$$

where y_i denotes the i^{th} measurement for the heat evolved in calories per gram of cement. x_1 = Tricalcium aluminate ($3CaO.Al_2O_3$), x_2 = Tricalcium silicate ($3CaO.SiO_2$), x_3 = Tetracalcium aluminoferrite ($4CaO.Al_2O_3.Fe_2O_3$), x_4 = β -dicalcium silicate ($2CaO.SiO_2$) and e is a random error which is distributed as $N(0, \sigma^2)$. The model (26) is known as *inhomogeneous* linear regression model and will be used to analyze the cement dataset to compare the performance of the proposed estimators and compared them with that of Kaciranlar et al. (1999). The least square estimates and their corresponding

estimated standard error (SE)s, together with associated t_8 -statistic and two-sided P-values are given in Table 1.

Table 1: The Regression Analysis

Coefs	SE	t -value	P-value
62.4054	70.0710	0.8906	0.3991
1.5511	0.7450	2.0821	0.0709
0.5102	0.7239	0.7048	0.5010
0.1019	0.7550	0.1350	0.8960
-0.1441	0.7092	-0.2031	0.8441

Note that S-plus software was used to analyze the dataset. It was found that $R^2 = 0.9824$, which implies that 98.24% of

Moreover, the singular values for the design matrix are: 211.36746660, 77.23614495, 28.45965700, 10.26736008, and 0.03490017 and the condition number of X is $211.36746660/0.03490017 = 6056.3744$. So, the X matrix is severely ill-conditioned (Kaciranlar, 1999), which probably make the regression coefficients insignificant and therefore, it is hard to make a valid inference or prediction using usual least squares method. To make a valid inference and to compare the performance of the proposed estimators, the cement dataset would be the most appropriate one to analyze.

Table 2: The correlation matrix

	x_1	x_2	x_3	x_4	y
x_1	1.0000	0.2286	-0.8241	-0.2454	0.7307
x_2	0.2286	1.0000	-0.1392	-0.9730	0.8163
x_3	-0.8241	-0.1392	1.0000	0.0295	-0.5347
x_4	-0.2454	-0.9730	0.0295	1.0000	-0.8213
y	0.7307	0.8163	-0.5347	-0.8213	1.0000

the total variation has been explained by the regressors, $F_{0.05,4,8} = 111.48$, which is highly significant. However, none of the regression coefficients is statistically significant at 5% level. This could possibly be due to the high correlation among the explanatory variables, which can be observed from the correlation matrix provided in Table 2.

To compare the performance of the proposed ridge estimators with that of ULIUE and RLIUE, consider the following parametric restriction: $\beta_1 - \beta_2 + \beta_3 = 0$ as that of Kaciranlar (1999), which can be expressed as

$$\begin{aligned} H_0 &: H\beta = 0 \\ H_a &: H\beta \neq 0, \end{aligned} \quad (27)$$

where $H = (0, 1, -1, 1, 0)$ is an 1×5 known matrix. In the notation, $q = 1$ and $p = 5$. To test the hypotheses in (27), use (11) and observe that $L_n = 1.92$. Because $F_{0.05, 1, 8} = 5.32$, one would not reject H_0 at 5% significance level. It is noted that the proposed ridge regression estimators depend on the value of k . Many different techniques for estimating k have been proposed or suggested by different researchers. Among them, Hoerl et al (1975), Lawless and Wang (1976), are notable.

For detailed about the estimation technique of k , see Kibria (2003) and references therein. The estimated values of k by using, Hoerl et al. (1975, p.107), and Lawless and Wang (1976, p.311) are 0.0077 and 0.00025 respectively. The estimated σ^2 by using URLSE and RLSE are 5.982955, 6.595198 respectively. The sampling variance of the proposed estimators depends on the value of σ^2 .

Therefore, 5.982955 was used to estimate the sampling variance of URLSE, URRE and URLIUE and 6.595198 for ridge regression estimators and RLIUE. The estimated standard error (SE) has been provided for all proposed estimators along with unrestricted Liu Estimator (ULIUE) and restricted Liu estimator (RLIUE) in Table 3. It may be observed from Table 3

that all the proposed ridge regression estimators dominate corresponding usual least squares estimators in the sense of having smaller sampling variances. Assume $d = 0.5$; the natural choice for the analysis.

It was also found from Table 3 that for any value of $k = 0, 0.00025$ and 0.0077 ,

the RRRE produces the smallest sampling variances compared to other ridge estimators. Specifically, for $k = 0.0077$, all ridge estimators dominate the restricted Liu estimator. Thus, using ridge regression technique, one would be able to improve the estimators. From Table 3, it was also observed that the sampling variance of RLIUE is smaller than that of ULIUE, which is the consequences of Theorem 3.1.1. (Kaciranlar 1999, p.448). Finally, using $k = 0.0077$, the data was analyzed and URRRE along with corresponding least squares estimators (in the parentheses) are provided in Table 4. Table 5 provides the regression analysis for RRRE and the restricted Liu estimator (in the parentheses).

It may be seen from Table 4, that under the OLS method, the regression coefficients are insignificant, yet under the ridge regression approach, all the regression coefficients are highly significant. It was also observed that the sign of the estimated regression coefficients has been changed, which is expected for highly correlated data (see Hoerl & Kennard, 1970). When the RLIUE was compared with that of RRRE, it was seen from Table 5 that all the regression coefficients, with the exception of β_2 under RLIUE, is not significant.

Next, the following parametric restrictions were considered: $\beta_1 - \beta_2 + \beta_3 = 0$, $\beta_2 - \beta_3 + \beta_4 = 0$, $\beta_1 - \beta_2 - \beta_4 = 0$, which yield the following hypotheses,

$$\begin{aligned} H_0 &: H\beta = 0 \\ H_a &: H\beta \neq 0, \end{aligned} \quad (28)$$

Table 3: Estimated standard error of the proposed estimators

URLSE	RLSE	SRLSE	PTLSE	ULIUE
$\alpha = 0.05 \quad d = 0.5 \quad k=0$				
70.0710	31.6114	44.4629	43.5948	55.7471
0.7450	0.1789	0.4025	0.3899	0.6032
0.7239	0.3674	0.4817	0.4733	0.5766
0.7550	0.2074	0.4183	0.4062	0.6083
0.7092	0.3464	0.4648	0.4561	0.5647
URRE	RRRE	SRRRE	PTRRE	RLIUE
$\alpha = 0.05 \quad d=0.5 \quad k=0.00025$				
58.1381	26.2285	36.8912	36.1710	31.6215
0.6269	0.1265	0.3317	0.3209	0.1721
0.6008	0.3114	0.4037	0.3975	0.3690
0.6325	0.1549	0.3435	0.3332	0.2031
0.5891	0.2933	0.3886	0.3821	0.3481
URRE	RRRE	SRRRE	PTRRE	RLIUE
$\alpha = 0.05 \quad d=0.5 \quad k=0.0077$				
9.5703	4.3195	6.0738	5.9553	31.6215
0.2121	0.1304	0.1549	0.1517	0.1721
0.1095	0.0894	0.0949	0.0949	0.3690
0.1897	0.1049	0.1342	0.1304	0.2031
0.1049	0.0775	0.0837	0.0837	0.3481

Table 4: Ridge regression (Usual Least Squares) analysis without restriction

Coefficients	SE	<i>t</i> -value	P-value
8.5642 (62.4054)	9.5703 (70.0710)	0.8949 (0.8906)	0.3970 (0.3991)
2.1048 (1.5511)	0.2121 (0.7450)	9.9238 (2.0820)	0.0000 (0.0708)
1.0651 (0.5102)	0.1095 (0.7239)	9.7268 (0.7047)	0.0000 (0.5009)
0.6683 (0.1019)	0.1897 (0.7550)	3.5231 (0.1350)	0.0078 (0.8959)
0.3998 (-0.1441)	0.1049 (0.7092)	3.8115 (-0.2031)	0.0052 (0.8441)

Table 5: Ridge regression (Restricted Liu) analysis under the restriction

Coefficients	SE	<i>t</i> -value	P-value
20.4006 (142.0379)	4.3195 (31.6215)	3.3528 (4.4918)	0.0100 (0.0020)
1.8724 (0.6207)	0.1304 (0.1721)	15.2503 (3.6065)	0.0000 (0.0069)
0.9716 (-0.2819)	0.0894 (0.3690)	11.3907 (-0.7640)	0.0000 (0.4668)
0.4504 (-0.8298)	0.1049 (0.2031)	5.3322 (-4.0857)	0.0007 (0.0035)
0.2985 (-0.9302)	0.0775 (0.3481)	4.5051 (-2.6721)	0.0020 (0.0283)

where H matrix is defined as

$$\begin{vmatrix} 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 1 & -1 & 0 & -1 \end{vmatrix}$$

To test the null hypothesis in (28), use (11) and observe that $L_n = 133.522$. The upper percentage point of the non-central F distribution with 3 and 8 degrees of freedom and non-centrality parameter 10.78 is 16, so, the null hypothesis is rejected at the 5% significance level. Now, the analysis will be carried out under the alternative hypothesis. The unrestricted and restricted estimated values of σ^2 are found to be 5.982955 and 222.221 respectively. It is noticed that the estimation of σ^2 in a restriction model very much depends of the restriction on the parameters. Using (12),

the non-centrality parameter Δ is estimated by replacing β by $\tilde{\beta}$ and σ^2 by $\tilde{\sigma}^2$ as 10.78.

The estimated SE for the proposed estimators along with ULIUE and RLIUE are presented in Tables 6 and 7, where σ^2 is estimated by $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ respectively. It was found that the estimated SE in Table 7 is 6 times bigger than the SE in Table 6, as $\hat{\sigma}^2$ is bigger than the $\tilde{\sigma}^2$. From both Tables 6 and 7, it was observed that the proposed ridge regression estimators have smaller sampling variances compared to usual least squares estimator for any values of k . It was also noticed that the RRRE performed better than the URRRE, SRRRE and PTRRE in the sense of having smaller as well as stable sampling variance. For the restriction (27), the PTRRE performs better than URRRE and SRRRE.

Table 6: Estimated SE of the proposed estimators using σ^2 estimated by $\tilde{\sigma}^2$

URLSE	RLSE	SRLSE	PTLSE	URLIE
$\alpha = 0.05 \quad d = 0.5 \quad k=0$				
70.0710	1.1688	35.0501	32.9698	70.0464
0.7450	0.0316	0.3742	1.0232	0.7445
0.7239	0.0000	0.3619	0.3795	0.7235
0.7550	0.0316	0.3782	0.2811	0.7545
0.7092	0.0316	0.3564	0.2608	0.7088
URRE	RRRE	SRRRE	PTRRE	RLIUE
$\alpha = 0.05 \quad d=0.5 \quad k=0.00025$				
58.1381	0.9701	29.0812	27.3548	1.1632
0.6269	0.0316	0.3146	0.9762	0.0370
0.6008	0.0000	0.3000	0.3225	0.0001
0.6325	0.0447	0.3178	0.2569	0.0371
0.5891	0.0316	0.2966	0.2121	0.0370
URRE	RRRE	SRRRE	PTRRE	RLIUE
$\alpha = 0.05 \quad d=0.5 \quad k=0.0077$				
9.5703	0.1612	4.7872	4.5010	1.1632
0.2121	0.0316	0.1095	0.8000	0.0370
0.1095	0.0000	0.0548	0.0949	0.0001
0.1897	0.0447	0.1049	0.2881	0.0371
0.1049	0.0316	0.0548	0.1378	0.0370

Table 7: Estimated SE of the proposed estimators when σ^2 is estimated $\hat{\sigma}^2$

URRE	RRRE	SRRRE	PtrRE	URLIE
$\alpha = 0.05 \quad d = 0.5$				
427.0441	7.1237	213.6112	152.6839	426.8942
4.5389	0.2258	2.2779	1.8218	4.5375
4.4111	0.0000	2.2054	1.5818	4.4095
4.5996	0.2258	2.3080	1.6462	4.5980
4.3212	0.2258	2.1696	1.5479	4.3198
URRE	RRRE	SRRRE	PtrRE	RLIUE
$\alpha = 0.05 \quad d=0.5 \quad k=0.00025$				
354.3197	5.9108	177.2338	126.6822	7.0890
3.8196	0.2168	1.9189	1.5837	0.2255
3.6636	0.0000	1.8319	1.3153	0.0004
3.8555	0.2366	1.9386	1.3889	0.2261
3.5882	0.2168	1.8039	1.2888	0.2255
URRE	RRRE	SRRRE	PtrRE	RLIUE
$\alpha = 0.05 \quad d=0.5 \quad k=0.0077$				
58.3256	0.9742	29.1750	20.8533	7.0890
1.2900	0.1789	0.6633	0.8246	0.2255
0.6701	0.0632	0.3391	0.2550	0.0004
1.1589	0.2811	0.6285	0.5404	0.2261
0.6419	0.1789	0.3564	0.3033	0.2255

The estimated optimum value of b using equations (3.10) and (3.11) of Kaciranlar (1999, p. 450) are found to be $b_{OLS} = 0.9996484$ and $b_{RLE} = 0.995122$ respectively. It was noticed that both of the estimated b are approximately equal to 1. One would therefore expect that ULIUE will give similar result like that of URLSE and RLIUE will produce similar result like that of RLSE. These results can be evident from Tables 6 and 7. It is noted from numerical analysis that the RRRE performed better than URLSE, URRE, RLSE, SRLSE, SRRRE, PTLSE, and PTRRE in the sense of having smaller sampling variance. From the analysis, it is also noted that the RRRE

performed equivalently or better than RLIUE in the sense of having smaller sampling variance.

Using $k = 0.0077$, the regression analysis for the RRRE and the RLIUE (in the parenthesis) have been presented in Tables 8 and 9 where σ^2 is estimated by $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ respectively. It may be found from Table 8 that all the regression coefficients are statistically significant under both RRRE and RLIUE. However, from Table 9, it was found that all the regression coefficients are statistically significant under the restricted ridge regression model and only β_2 is significant under the RLIUE model.

Table 8: Ridge regression (Restricted Liu) analysis, when σ^2 is estimated by $\tilde{\sigma}^2$

Coefficients	SE	t-value	P-value
13.9436 (101.3373)	0.1612 (1.1632)	69.8131 (87.1194)	0.0000 (0.0000)
0.6545 (-0.2444)	0.0316 (0.0370)	43.6612 (-6.6063)	0.0000 (0.0002)
0.9058 (0.0051)	0.0000 (0.0001)	Inf (51.2085)	0.0000 (0.0000)
1.1742 (0.2547)	0.0447 (0.0371)	20.6098 (6.8648)	0.0000 (0.0001)
0.6383 (-0.2445)	0.0316 (0.0370)	16.4263 (-6.6076)	0.0000 (0.0002)

Table 9: Ridge regression (Restricted Liu) analysis, when σ^2 is estimated by $\hat{\sigma}^2$

Coefficients	SE	t-value	P-value
13.9436 (101.3373)	0.9742 (7.0890)	11.5519 (14.2950)	0.0000 (0.0000)
0.6545 (-0.2444)	0.1789 (0.2255)	7.7121 (-1.0840)	0.0001 (0.3100)
0.9058 (0.0051)	0.0632 (0.0004)	15.5926 (12.8021)	0.0000 (0.0000)
1.1742 (0.2547)	0.2811 (0.2261)	3.2773 (1.1264)	0.0112 (0.2926)
0.6383 (-0.2445)	0.1789 (0.2255)	2.9015 (-1.0842)	0.0198 (0.3099)

Conclusion

In this article, some ridge type estimators were considered for estimating the regression parameters under a parametric restriction. The proposed estimators have been studied by using the variance criterion. It was found that a sufficient and necessary condition for all the ridge estimators to have smaller sampling variance than their corresponding least squares estimators. It was found that, under both H_0 and H_a , the RRRE dominates all other ridge estimators and corresponding least squares estimators. As Δ moves away from 0, the conditions of superiority of the PTRRE and SRRRE over the URRE and URLSE are determined. The findings were illustrated by analyzing the Portland Cement dataset. From the analysis, it is evident that RRRE is superior to other proposed ridge regression estimators and as good as restricted Liu estimator.

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. (2nd Ed.). New York, N.Y.: John Wiley.

Bancroft, T. A. (1944). On biases in estimation due to use of preliminary tests of significance. *Annals of Mathematics and Statistics*, 15, 190-204.

Bancroft, T. A. (1964). Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance. *Biometrics*, 20, 427-442.

Bock, M. E., Yancey, T. A. & Judge, G. G. (1973). The statistical consequences of preliminary test estimators in regression. *Journal of the American Statistical Association*, 68, 109-116.

Giles, A. J. (1991). Pre-testing for linear restrictions in a regression model with spherically symmetric disturbances. *Journal of Econometrics*, 50, 377-398.

Gorman, J. W., & Toman, R. J. (1966). Selection of variables for fitting equations to data. *Technometrics*, 8, 27-51.

Hald, A. (1952). *Statistical theory with Engineering Applications*. New York, N.Y.: Wiley.

Han, C-P, & Bancroft, T. A. (1968). On pooling mean when variance is unknown. *Journal of the American Statistical Association*, 63, 1333-1342.

Haq, M. S. & Kibria, B. M. G. (1996). A shrinkage estimator for the restricted linear regression model: Ridge regression approach. *Journal of Applied Statistical Science*, 3(4), 301-316.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55-67.

Hoerl, A. E., Kennard, R. W. & Baldwin, K. F. (1975). Ridge regression: Some simulation. *Communications in Statistics*, A 4, 105-123.

Judge, G. G. & Bock, M. E. (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics*. Amsterdam: North-Holland Publishing Company.

Kaciranlar, S., Akdeniz, S. S. F., Styan, G. P. & Werner, H. J. (1999). A new biased estimators in linear regression and detailed analysis of the widely-analysed dataset on portland cement. *Sankhya*, Series B, 443-459.

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 32, 419-435.

Kibria, B. M. G. (2004a). Performance of the shrinkage preliminary test ridge regression estimators based on the conflicting of W, LR and LM tests. *Journal of Statistical Computation and Simulation*, 74(11), 793 – 810.

Kibria, B. M. G. (2004b). On some ridge regression estimators under possible stochastic constraint. *Pakistan Journal of Statistics* -Special issue in the honor of Professor Dr. Mir Masoom Ali, 20(1), 1-23.

Kibria, B. M. G. and Saleh, A. K. Md. E. (1993). Performance of shrinkage preliminary test estimator in regression analysis. *Jahangirnagar Review*, A 17, 133-148.

Kibria, B. M. G. & Saleh, A. K. Md. E. (2004). Performance of positive rule estimator in the ill-conditioned Gaussian regression model. *Calcutta Statistical Association Bulletin*, 55 (219-220), 211-241.

Kibria, B. M. G. & Saleh, A. K. Md. E. (2005a). Pooling multivariate data under W, LR, and LM tests. *Statistical Papers*, 47, 49-68.

Kibria, B. M. G. & Saleh, A. K. Md. E. (2005b). Comparison between Han-Bancroft and Brook methods to determine the optimal significance level for pre-test Estimator. *Journal of Probability and Statistical Science*, 3, 293-303.

Kibria, B. M. G. & Saleh, A. K. Md. E. (2006). Optimum critical value for pre-test estimators. *Communications in Statistics-Simulation and Computation*, 35 (2), 309-319.

Lawless, J. F. & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics*, A 5, 307-323.

Liu, K. (1993). A new class of biased estimator in linear regression. *Communications in Statistics*, A, 22, 393-402.

Nomura, M. (1988). On the almost unbiased ridge regression estimator. *Communications in Statistics*, B, 17, 729-743.

Obenchain, R. L. (1975). Ridge analysis following a preliminary test of the shrunken hypothesis. *Technometrics*, 17, 431-441.

Piepel, G. & Redgate, T. (1998). A mixture experimnt analysis of the Hald cement data. *The American Statistician*, 52, 23-30.

Saleh, A. K. Md. E. & Han, C-P. (1990). Shrinkage estimation in regression analysis. *Estadistica*, 42-40.63.

Saleh, A. K. Md. & Kibria, B. M. G. (1993). Performances of some new preliminary test ridge regression estimators & their properties. *Communications in Statistics-Theory and Methods*, 22, 2747-2764.

Sarkar, N. (1992). A new estimator combining the ridge regression and the restricted least squares method of estimation. *Communications in Statistics*, A 21, 1987-2000.

Sutradhar, B. C. (1990). A note on the power of the F-test for testing linear hypothesis with elliptic t errors. . *Communications in Statistics*, A 19, 863--867.

Woods, H., Steinour, H. H., & Starke, H. R. (1932). Effect of composition of Portland cement on heat evolved during hardening. *Industrial and Engineering Chemistry*, 24, 1207-1214.

Bootstrap Intervals of the Parameters of Lognormal Distribution Using Power Rule Model and Accelerated Life Tests

Mohammed Al-Haj Ebrahim
Department of Statistics
Yarmouk University

Assumed that the distribution of the lifetime of any unit follows a lognormal distribution with parameters μ and σ . Also, assume that the relationship between μ and the stress level V is given by the power rule model. Several types of bootstrap intervals of the parameters were studied and their performance was studied using simulations and compared in term of attainment of the nominal confidence level, symmetry of lower and upper error rates and the expected width. Conclusions and recommendations are given.

Key words: Power rule model, lognormal distribution, bootstrap intervals, accelerated life test.

Introduction

The lognormal distribution has many special features that allowed it to be used as a model in various real life applications. In particular, it is used in analyzing biological data (Koch, 1966), and for analyzing data in workplace exposure to contaminants (Lyles & Kupper, 1997). It is also of importance in modeling lifetimes of products and individuals (Lawless, 1982). Various other motivations and applications of the lognormal distribution may also be found (see Johnson et. al., 1994, Schneider, 1986).

In a life testing experiment, the problem is that most units have a very long life under the normal conditions. Therefore, by the time the experiment is completed and an estimate of the reliability is obtained, the results will be outdated. To overcome this delay, accelerated life testing was introduced (Mann. et. al., 1974).

In an accelerated life testing experiment a certain number of units are subjected to a stress that is higher than the normal stress. The

experiment is repeated under different values of stress. In order to do so, some relationship between the parameters of the time to failure distribution of the unit and the corresponding stress level must be postulated.

It is assumed that density function of the time to failure of a unit depends on one parameter say θ , and the environment depends on one stress V and that the relationship between θ and V is given by $\theta = \frac{C}{V^P}$ where C and P are positive constants. This relationship is known as the power rule model.

Consider the interval estimation for the parameters of the lognormal distribution after reparametrizing the location parameter μ as a function of the stress V using power rule model. The performance of the bootstrap and Jackknife intervals (Efron & Tibshirani, 1993) in term of attainment of the nominal confidence level, symmetry of lower and upper error rates and the expected width of the intervals will be compared.

Mohammed Al-Haj Ebrahim is Assistant Professor in the Department of Statistics at Yarmouk University, Irbid-Jordan. His research interest is in reliability, accelerated life test and non-parametric regression models. E-mail: m_hassanb@hotmail.com.

The Model and The Maximum likelihood Estimation

It is assumed that the lifetime (T) of any unit follows a lognormal distribution with location parameter μ and scale parameter σ . The probability density function of T is given by (Lawless, 1982):

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln t - \mu)^2}{2\sigma^2}\right), 0 < t < \infty. \quad (1)$$

The location parameter μ was reparameterized as a function of the stress V using the power rule model $\mu = \frac{C}{V^{P_0}}$, therefore c and σ are the new parameters of the model. The unknown parameters c and σ were estimated using k complete samples. The j -th sample is obtained by using n_j units and the value V_j for the stress, $j = 1, 2, \dots, k$.

The likelihood function of the k complete samples is given by:

$$L(\mu, \sigma) = \frac{e^{-\frac{1}{2\sigma^2} \sum_{j=1}^k \sum_{i=1}^{n_j} (\ln t_{ij} - \mu_j)^2}}{\sigma^{\sum_{j=1}^k n_j} (2\pi)^{\frac{\sum_{j=1}^k n_j}{2}} \prod_{j=1}^k \prod_{i=1}^{n_j} t_{ij}} \quad (2)$$

Using the power rule model $\mu_j = \frac{C}{V_j^{P_0}}$ $j = 1, 2, \dots, k$, the likelihood function is given by:

$$L(C, \sigma) = \frac{e^{-\frac{1}{2\sigma^2} \sum_{j=1}^k \sum_{i=1}^{n_j} \left(\ln t_{ij} - \frac{c}{V_j^{P_0}}\right)^2}}{\sigma^{\sum_{j=1}^k n_j} (2\pi)^{\frac{\sum_{j=1}^k n_j}{2}} \prod_{j=1}^k \prod_{i=1}^{n_j} t_{ij}} \quad (3)$$

It is easy to show that the Maximum likelihood estimators of C and σ are given by:

$$\hat{C} = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} \ln t_{ij} / v_j^{P_0}}{\sum_{j=1}^k n_j / v_j^{2P_0}} \quad (4)$$

and

$$\hat{\sigma} = \sqrt{\frac{\sum_{j=1}^k \sum_{i=1}^{n_j} \left(\ln t_{ij} - \frac{\hat{C}}{v_j^{P_0}}\right)^2}{\sum_{j=1}^k n_j}} \quad (5)$$

It is obvious that $\hat{C}$ is an unbiased estimator of C while $\hat{\sigma}$ is a biased estimator of σ .

The Percentile Interval

The methods of deriving confidence intervals presented in this section and section 4 are based on the parametric bootstrap approach (Efron & Tibshirani, 1993); they are constructed by resampling from the estimated parametric distribution. To construct the percentile interval, a simulation of the bootstrap distribution of $\hat{C}$ and $\hat{\sigma}$ is done by resampling from the parametric model of the original data. That is, a B bootstrap sample is generated and for each sample $\hat{C}^*$ and $\hat{\sigma}^*$ are calculated using equation (4) and (5) respectively. The calculated values are denoted by $\hat{C}^*$ and $\hat{\sigma}^*$.

Let $\hat{G}_1$ denotes the cumulative distribution of $\hat{C}^*$, the $(1-\alpha)\%$ percentile interval of C is $\left(\hat{G}_1^{-1}\left(\frac{\alpha}{2}\right), \hat{G}_1^{-1}\left(1-\frac{\alpha}{2}\right)\right)$, similarly let $\hat{G}_2$ denotes the cumulative distribution of $\hat{\sigma}^*$, the $(1-\alpha)\%$ percentile interval of σ is $\left(\hat{G}_2^{-1}\left(\frac{\alpha}{2}\right), \hat{G}_2^{-1}\left(1-\frac{\alpha}{2}\right)\right)$.

The Bias Corrected and Accelerated Interval (BCa Interval)

The bias corrected and accelerated interval is constructed by calculating two numbers $\hat{a}$ and $\hat{z}_0$ called the accelerated and the bias correction factor respectively, they are calculated using the following formulas

$$\hat{a} = \frac{\sum_{i=1}^n (\hat{C}(\cdot) - \hat{C}(i))^3}{6 \left(\sum_{i=1}^n (\hat{C}(\cdot) - \hat{C}(i))^2 \right)^{3/2}} \quad (6)$$

where $\hat{C}(i)$ is the maximum likelihood estimator of C using the original data excluding the i-th

observation and $\hat{C}(\cdot) = \frac{\sum_{i=1}^n \hat{C}(i)}{n}$, $n = \sum_{j=1}^k n_j$.

The value of $\hat{z}_0$ is given by

$$\hat{z}_0 = \Phi^{-1} \left(\frac{\#(\hat{C}^* < \hat{C})}{B} \right) \quad (7)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The $(1-\alpha)\%$ BCa interval of C is $(\hat{G}_1^{-1}(\alpha_1), \hat{G}_1^{-1}(\alpha_2))$ where

$$\alpha_1 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})} \right)$$

and

$$\alpha_2 = \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})} \right) \quad (8)$$

where z_α is the α quantile of the standard normal distribution. In the same way, the $(1-\alpha)\%$ BCa interval of σ can be constructed.

Jackknife Interval

A $(1-\alpha)\%$ Jackknife interval of C (Efron and Tibshirani, 1993) is constructed as follows:

$$\hat{C}(\cdot) \pm Z_{(\alpha/2)} \hat{S}_{Jack},$$

where

$$\hat{S}_{Jack}^2 = \frac{n-1}{n} \sum_{i=1}^n (\hat{C}(\cdot) - \hat{C}(i))^2, \quad \hat{C}(\cdot), \hat{C}(i)$$

and n were defined in section 4. Similarly, the $(1-\alpha)\%$ Jackknife interval of σ by replacing C by σ in the above interval.

Simulation Study

A simulation study is conducted to investigate the performance of the intervals discussed in sections 3, 4 and 5 above. The indices of the simulation study are:

k: The number of lognormal populations, in this study $k = 2$.

n_1 : Sample size from the first lognormal population, in this study $n_1 = 5, 10, 30$.

n_2 : Sample size from the second lognormal population, in this study $n_2 = 5, 10, 30$.

C: Parameter of the power rule model, in this study $C = 3$.

P_0 : In this study $P_0 = 0.3$.

V_1 : The value of stress for the first lognormal population, in this study $V_1 = 100$.

V_2 : The value of stress for the second lognormal population, in this study $V_2 = 200$.

σ : In this study $\sigma = 1$.

B: The number of bootstrap samples, in this study $B = 1000$.

For each combination of n_1 and n_2 2000 samples are generated and a $(1-\alpha)\%$ Percentile interval is constructed, BCa interval and Jackknife interval for C and σ . Two values are considered for α , 0.05 and 0.10. The following were obtained for each interval:

- 1- The expected width (IW): the average of widths of the 2000 intervals.
- 2- Lower error rate (LER): the fraction of intervals that fall entirely above the true parameter.

- 3- Upper error rate (UER): the fraction of intervals that fall entirely below the true parameter.
- 4- Total error rate (TER): the fraction of intervals that did not contain the true parameter value.

Results and Conclusions

The results are given in tables 1 – 12. Table 1 has simulation results of the percentile interval of the parameter C using $\alpha = 0.05$. Table 2 has simulation results of the BCa interval of the parameter C using $\alpha = 0.05$. Table 3 has simulation results of the Jackknife interval of the parameter C using $\alpha = 0.05$. Table 4 has simulation results of the percentile interval of the parameter C using $\alpha = 0.1$. Table 5 has simulation results of the BCa interval of the parameter C using $\alpha = 0.1$. Table 6 has simulation results of the Jackknife interval of the parameter C using $\alpha = 0.1$. Table 7 has simulation results of the percentile interval of the parameter σ using $\alpha = 0.05$. Table 8 has simulation results of the BCa interval of the parameter σ using

$\alpha = 0.05$. Table 9 has simulation results of the Jackknife interval of the parameter σ using $\alpha = 0.05$. Table 10 has simulation results of the percentile interval of the parameter σ using $\alpha = 0.1$. Table 11 has simulation results of the BCa interval of the parameter σ using $\alpha = 0.1$. Table 12 has simulation results of the Jackknife interval of the parameter σ using $\alpha = 0.1$. From these results the following can be concluded:

For the parameter C , the three intervals have almost the same expected width, and the expected width decreases as the sample sizes increases. In term of attainment of coverage probability and symmetry of lower and upper rates, the three intervals behave in the same way. It is recommended that the Jackknife interval be used because its calculation is simpler than the BCa and the percentile intervals.

For the parameter σ , the expected width for the percentile interval is nearly smaller than the other two intervals. On the other hand, in term of attainment of coverage probability and symmetry of lower and upper rates, the BCa interval behaves the best. It is therefore recommended that the BCa interval be used in this case.

Table 1. Percentile Interval of the parameter C with $\alpha = 0.05$

n_1	n_2	IW	LER	UER	TER
5	5	4.983	0.048	0.054	0.102
5	10	4.341	0.051	0.031	0.081
5	30	3.083	0.061	0.016	0.077
10	5	4.067	0.053	0.029	0.082
10	10	3.654	0.051	0.025	0.075
10	30	2.811	0.047	0.016	0.063
30	5	2.649	0.051	0.013	0.064
30	10	2.540	0.050	0.012	0.062
30	30	2.180	0.047	0.017	0.063

Table 2. BCa Interval of the parameter C with $\alpha = 0.05$

n_1	n_2	IW	LER	UER	TER
5	5	5.007	0.048	0.053	0.101
5	10	4.362	0.056	0.026	0.082
5	30	3.118	0.071	0.011	0.082
10	5	4.085	0.054	0.026	0.079
10	10	3.684	0.057	0.020	0.077
10	30	2.841	0.054	0.008	0.062
30	5	2.688	0.061	0.009	0.069
30	10	2.577	0.058	0.008	0.065
30	30	2.205	0.052	0.010	0.062

Table 3. Jackknife Interval of the parameter C with $\alpha = 0.05$

n_1	n_2	IW	LER	UER	TER
5	5	5.266	0.043	0.047	0.089
5	10	4.495	0.043	0.033	0.076
5	30	3.120	0.045	0.020	0.065
10	5	4.211	0.043	0.030	0.073
10	10	3.762	0.042	0.024	0.066
10	30	2.842	0.037	0.021	0.058
30	5	2.685	0.041	0.016	0.056
30	10	2.570	0.038	0.016	0.053
30	30	2.197	0.038	0.025	0.063

Table 4. Percentile Interval of the parameter C with $\alpha = 0.1$

n_1	n_2	IW	LER	UER	TER
5	5	4.182	0.074	0.081	0.154
5	10	3.641	0.083	0.056	0.139
5	30	2.571	0.087	0.036	0.123
10	5	3.410	0.077	0.065	0.142
10	10	3.062	0.084	0.049	0.133
10	30	2.346	0.072	0.042	0.114
30	5	2.211	0.072	0.034	0.105
30	10	2.119	0.075	0.038	0.112
30	30	1.820	0.070	0.046	0.116

Table 5. BCa Interval of the parameter C with $\alpha = 0.1$

n_1	n_2	IW	LER	UER	TER
5	5	4.194	0.074	0.077	0.151
5	10	3.658	0.090	0.052	0.142
5	30	2.603	0.097	0.030	0.126
10	5	3.423	0.082	0.058	0.139
10	10	3.086	0.093	0.042	0.135
10	30	2.372	0.085	0.033	0.118
30	5	2.243	0.083	0.024	0.107
30	10	2.149	0.088	0.025	0.113
30	30	1.840	0.084	0.036	0.119

Table 6. Jackknife Interval of the parameter C with $\alpha = 0.1$

n_1	n_2	IW	LER	UER	TER
5	5	4.419	0.065	0.073	0.138
5	10	3.772	0.067	0.057	0.124
5	30	2.619	0.076	0.040	0.116
10	5	3.534	0.067	0.061	0.128
10	10	3.157	0.071	0.049	0.119
10	30	2.385	0.063	0.045	0.108
30	5	2.254	0.065	0.039	0.104
30	10	2.157	0.065	0.040	0.105
30	30	1.844	0.061	0.053	0.113

Table 7. Percentile Interval of the parameter σ with $\alpha = 0.05$

n_1	n_2	IW	LER	UER	TER
5	5	0.790	0.000	0.198	0.198
5	10	0.667	0.000	0.145	0.145
5	30	0.456	0.003	0.081	0.084
10	5	0.669	0.001	0.149	0.150
10	10	0.584	0.002	0.128	0.130
10	30	0.427	0.003	0.090	0.093
30	5	0.455	0.004	0.094	0.098
30	10	0.428	0.006	0.069	0.075
30	30	0.350	0.007	0.067	0.074

Table 8. BCa Interval of the parameter σ with $\alpha = 0.05$

n_1	n_2	IW	LER	UER	TER
5	5	0.904	0.017	0.067	0.084
5	10	0.760	0.021	0.041	0.062
5	30	0.499	0.029	0.028	0.057
10	5	0.763	0.021	0.044	0.065
10	10	0.660	0.019	0.031	0.050
10	30	0.463	0.028	0.026	0.054
30	5	0.497	0.029	0.031	0.060
30	10	0.464	0.026	0.021	0.047
30	30	0.371	0.026	0.025	0.051

Table 9. Jackknife Interval of the parameter σ with $\alpha = 0.05$

n_1	n_2	IW	LER	UER	TER
5	5	0.847	0.005	0.165	0.170
5	10	0.702	0.006	0.119	0.125
5	30	0.467	0.008	0.076	0.084
10	5	0.702	0.004	0.124	0.128
10	10	0.608	0.008	0.109	0.117
10	30	0.435	0.009	0.082	0.091
30	5	0.462	0.009	0.087	0.096
30	10	0.432	0.013	0.072	0.085
30	30	0.354	0.009	0.066	0.075

Table 10. Percentile Interval of the parameter σ with $\alpha = 0.1$

n_1	n_2	IW	LER	UER	TER
5	5	0.666	0.000	0.253	0.253
5	10	0.561	0.003	0.206	0.209
5	30	0.383	0.012	0.134	0.146
10	5	0.563	0.002	0.215	0.217
10	10	0.490	0.006	0.199	0.204
10	30	0.358	0.012	0.133	0.144
30	5	0.381	0.012	0.139	0.151
30	10	0.359	0.015	0.116	0.131
30	30	0.294	0.016	0.113	0.129

Table 11. BCa Interval of the parameter σ with $\alpha = 0.1$

n_1	n_2	IW	LER	UER	TER
5	5	0.770	0.042	0.098	0.140
5	10	0.636	0.042	0.065	0.106
5	30	0.414	0.060	0.049	0.109
10	5	0.639	0.047	0.069	0.115
10	10	0.547	0.045	0.058	0.103
10	30	0.385	0.054	0.060	0.114
30	5	0.412	0.059	0.059	0.117
30	10	0.386	0.052	0.046	0.098
30	30	0.310	0.051	0.047	0.098

Table 12. Jackknife Interval of the parameter σ with $\alpha = 0.1$

n_1	n_2	IW	LER	UER	TER
5	5	0.711	0.012	0.202	0.214
5	10	0.590	0.013	0.171	0.184
5	30	0.392	0.026	0.113	0.139
10	5	0.590	0.013	0.160	0.172
10	10	0.511	0.015	0.160	0.175
10	30	0.366	0.023	0.117	0.139
30	5	0.388	0.023	0.128	0.151
30	10	0.363	0.024	0.101	0.125
30	30	0.298	0.026	0.099	0.125

References

- Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Johnson, N. L., Kotz, S., & Balakrishnan. (1994). *Continuous univariate distributions: vol 1*. New York: Wiley
- Koch, A.L. (1966). The logarithm in biology. *Journal of Theoretical Biology*, 12, 276 – 290.
- Lawless, J.F. (1982). *Statistical models and methods for lifetime data*. New York: Wiley.
- Lyles, R. H., Kupper, L. L., & Rappaport, S. M. (1997). Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA model. *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 64 – 86.
- Mann, N. R., Schafer, R. E., & Singpurwalla, N. D. (1974). *Methods for statistical analysis of reliability and life data*. New York: John Wiley and Sons Inc.
- Schneider, H. (1986). *Truncated and censored samples from normal populations*. New York: Marcel Dekker.
- Efron, B. & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, N.Y.: Chapman and Hall.
- Jennings, D. (1987). How do we judge confidence intervals adequacy? *The American Statistician*, 41(4), 335-337.
- Kulldorff, G. (1961). *Estimation from grouped and partially grouped samples*. New York, N.Y.: Wiley.
- Meeker, Jr, W. (1986). Planning Life tests in which units are inspected for failure. *IEEE Trans. on Reliability R-35*, 571-578.
- Pettitt, A. N., & Stephens, M. A. (1977). The Kolmogorov-Smirnov goodness-of-fit statistic with discrete and grouped data. *Technometrics*, 19, 205 – 210.

Determination of Optimal Block Designs With Pre-assigned Variance For Elementary Contrasts

Seemon Thomas Alex Thannippara

St.Thomas College, Pala

Department of Statistics

S.C. Bagui

University of West Florida

Department of Mathematics and Statistics

D.K. Ghosh

Saurashtra University

Department of Mathematics Statistics

A method for obtaining optimal designs from the class of variance balanced and connected designs was developed for comparing treatment effects with a pre-assigned variance. The properties of the C-matrix of a block design are employed in developing this method. Some new results concerning the design parameters and the non-zero characteristic root of the C-matrix are also presented.

Key words: Contrasts, C-matrix, connected design, balanced incomplete block design, optimal design.

Introduction

Block designs have a wide range of applications in agriculture, engineering, medicine, social and behavioral sciences. In experimental designs for social and behavioral sciences problems, the

choice of appropriate design with a specified variance for the elementary (estimable) contrasts is very useful for inference purposes. Similar issues also arise in the design of industrial experiments for statistical quality control problems. There are situations where the value of σ^2 is known, but there is no procedure available for determining the parameters of a block design in order to compare treatment effects with a pre-determined variance for the elementary (estimable) contrasts under consideration.

Let there be v treatments arranged in b blocks, such that the j -th block contains k_j experimental units and the i -th treatment appears r_i times in the entire design, $i = 1, 2, \dots, v$; $j = 1, 2, \dots, b$. The $v \times b$ matrix $N = (n_{ij})$, where n_{ij} is the number of times i -th treatment occurs in the j -th block, is the incidence matrix of the block design. If $K = \text{diag}(k_1, \dots, k_b)$ and $R = \text{diag}(r_1, \dots, r_v)$, then the C-matrix of the design is defined as

$$C = R - NK^{-1}N' \quad (1)$$

Seemon Thomas is Lecturer of Statistics. His areas of research are construction of optimal designs and statistical modeling. email: ktm_stpala@sancharnet.in. Alex

Thannippara is Lecturer of Statistics. His areas of research are construction of combinatorial and optimal designs. email: ktm_stpala@sancharnet.in. Subhash

Bagui is Professor of Statistics. His areas of research are statistical classification and pattern recognition, bio-statistics, construction of designs, tolerance regions, and statistical computing. email: sbagui@uwf.edu. Dilip

Ghosh is Professor and Head of the Department of Mathematics and Statistics. His areas of research are combinatorial designs, construction of designs and its analysis, optimal designs, biostatistics, graph theory, and statistical analysis of cancer data. email: ghosh_dkg@rediffmail.com.

The following discussion is based on Gini (1986).

Definition 1.1

A contrast $\sum_j C_j t_j = C't$ with $C = (C_1, \dots, C_v)'$ in the treatment effects $t = (t_1, \dots, t_v)'$ is called an elementary contrast if C has only two non-zero elements 1 and -1. It is often desirable to design experiments where all elementary contrasts $t_i - t_j$, $i \neq j$ are estimable.

Definition 1.2

A design where all elementary contrasts are estimable is called a connected design.

Definition 1.3

A connected design is said to be balanced if all elementary contrasts in the treatment effects can be estimated with the same precision.

Definition 1.4

A balanced incomplete block design (BIBD) is an arrangement of v treatments in b blocks each containing k experimental units ($k < v$) such that

- (i) every treatment occurs at most once in each block,
- (ii) every treatment occurs exactly in r blocks
- (iii) every pair of treatments occur together in exactly λ blocks.

For any estimable contrast $\psi = \sum_j C_j t_j$ corresponding to a BIBD, the variance of its unbiased estimator $\hat{\psi} = \sum_j C_j \hat{t}_j$ is given by

$$\text{Var}(\hat{\psi}) = (r\varepsilon)^{-1} \sigma^2 \sum_j C_j^2$$

Where $\varepsilon = \frac{v(k-1)}{k(v-1)}$ is called the efficiency factor of BIBD. If it is a randomized block design (RBD), then

$$\text{Var}(\hat{\psi}) = \frac{\sigma^2}{r} \sum_j C_j^2.$$

Definition 1.5

If the elements of the incidence matrix take only values 0 or 1, then the incomplete block design is called a binary design.

Definition 1.6

An incomplete block design with $k_1 = \dots = k_b = k$ is called a proper design.

Some Properties of C -matrix

The C -matrix of block designs is very useful in the analysis and construction of various block designs. The following results on C -matrix are well known and more details can be seen from Dey(1986) and Raghavarao(1971).

1. For a balanced design all the non-zero characteristic roots of its C -matrix are equal.
2. For a balanced design each of the elementary contrasts is estimated with a variance $\frac{2\sigma^2}{\theta}$, where θ is the non zero characteristic root of the C -matrix.
3. For a balanced design the C -matrix is of the form $C = \theta(I_v - v^{-1}JJ')$ where $J = (1, \dots, 1)'$.
4. A block design is connected if the rank of its C -matrix is $v-1$.
5. In a connected design the diagonal elements of the C -matrix are all positive. Further, the principal minors of all orders $(1, 2, \dots, v-1)$ of C are all positive.
6. A connected block design is variance balanced if its C -matrix has all its diagonal elements equal and all its off diagonal elements equal. That is, C is given by $C = (a-b)I + bJJ'$ where a and b are some scalars.

7. For an equireplicate, proper block design,

$$C = rI - \frac{NN'}{k}$$

and if the design is binary,

$$\text{tr}(C) = vr - \frac{vr}{k} = vr - b.$$

If it is variance balanced, then

$$\text{tr}(C) = \theta(v-1).$$

If all the above conditions hold, then

$$\theta = \frac{(vr - b)}{(v-1)}.$$

8. For any general incomplete block design with $\sum_{ij} n_{ij} = bk$,

$$\text{tr}(C) = bk - k^{-1} \sum_{ij} n_{ij}^2$$

is maximized by the BIBD whenever it exists.

Methodology

A method for identifying optimal designs for comparing treatment effects of equireplicate, proper, variance balanced, and connected block designs is introduced. The procedure can be extended to other cases also.

Determination of Optimal Designs

Suppose that the process variability σ^2 is known. Then, following are the steps involved in determining the parameters of the corresponding design.

1. Decide upon the precision of the elementary (estimable) contrast to be estimated.

2. Choose an integer value for θ so that the variance of the elementary contrasts $\frac{2\sigma^2}{\theta}$ will meet the specification in step 1.

3. Select an appropriate C -matrix of dimension $v \times v$ such that the non-zero characteristic root is θ which is determined in step 2. The value of v is known from the objective of the underlying experiment.

The exploration of a C -matrix corresponding to a variance balanced and connected design is not tedious as it has the following format.

$$(a_{ij}) = \begin{bmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \cdots & \vdots \\ b & b & \cdots & a \end{bmatrix}_{v \times v}$$

The characteristic roots of the above matrix are $a + (v-1)b$ and $(a-b)$ of multiplicity $v-1$.

4. Determine all possible combinations of r and b satisfying the condition $\theta(v-1) = vr - b$.
5. Since θ and v are known quantities, one can consider those combinations of r and b satisfying the above condition. Among these combinations of b and r , those values satisfying the condition $vr = bk$, where k must be an integer such that $k \leq v$ are selected. The existence of an incidence matrix corresponding to the resulting design must be verified. If it is a proper block design, then NN' will be such that

$$NN' = k(rI - C).$$

If it is a BIBD, then

$$NN' = (r - \lambda)I + \lambda JJ'$$

where λ is such that $\lambda(v-1) = r(k-1)$. Hence, the following theorem.

Theorem 2.1

The above process will lead to a set of optimal designs, which include one RBD and BIBDs provided the latter should exist.

The resulting designs were called optimal designs because Kiefer (1958) proved that RBD and BIBD are A , D , E and L optimum. Further, Roy (1958) proved that in the case of proper binary incomplete block designs, a most efficient design if it exists, is necessarily a BIBD.

Corollary 2.1

Corresponding to a given C -matrix there can be more than one design with the same variance for all the elementary contrasts.

Theorem 2.2

If θ is the non-zero characteristic root of the C -matrix corresponding to an RBD, then $\theta = b = r$.

Proof. The C -matrix corresponding to an RBD satisfies the condition

$$C = rI - \frac{NN'}{k}$$

and hence

$$\text{tr}(C) = vr - \frac{vr}{k} = vr - b.$$

Because $vr = bk$ it can be rewritten as

$$\text{tr}(C) = b(k-1) \quad (2)$$

Further the design is variance balanced and hence all the non-zero characteristic roots of C -matrix are equal to θ . Hence

$$\text{tr}(C) = \theta(v-1) \quad (3)$$

From (2) and (3) it follows that

$$\theta = \frac{b(k-1)}{(v-1)}.$$

Now v and k are equal in the case of RBD and hence $\theta = b$. Because $vr = bk$, $v = k$ implies $r = b$.

The above result is justified because the variance of each elementary contrast of a RBD is $\frac{2\sigma^2}{r}$. Therefore, an RBD with a suitable r will have a required precision for its estimable contrasts. Table 2.1 gives the manner in which θ changes with respect to the other parameters of an RBD having 5 treatments.

Table 2.1

r	2	3	4	5	6	7	8
b	2	3	4	5	6	7	8
k	5	5	5	5	5	5	5
θ	2	3	4	5	6	7	8

Suppose one wishes to compare the effects of v treatments with the variance of the elementary contrast being $\frac{2\sigma^2}{\theta}$, where θ is the non-zero

characteristic root of the C -matrix of a variance balanced design. The following example illustrates the method of determining the values of b , r , and k corresponding to an optimal design.

Example

Let $v = 5$ and $\theta = 5$. Then the C -matrix of a block design with non-zero characteristic root as 5 can be given as

$$C = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & -1 & 4 \end{bmatrix}$$

values of b , r , and k are to be determined such that there exists integers satisfying the

conditions $vr = bk$ and $\theta(v-1) = vr - b$. One can easily verify that the following three designs which satisfy all the requirements given in steps 1 to 4.

1. RBD with parameters $b = r = k = 5$
2. BIBD with parameters $b = 10, r = 6, k = 3, \lambda = 3$.
3. BIBD with parameters $b = 10, r = 8, k = 2, \lambda = 2$.

More examples are presented in table 2.2.

From the above examples it can be seen that corresponding to a positive integer θ there always exists an RBD described in Theorem 2 but BIBDs need not exist for a positive integer θ . Note that if BIBDs exist then the parameters will satisfy the relation $\theta = \frac{\lambda v}{k}$.

References

- Dey, A. (1986). *Theory of block designs*. Wiley Eastern Limited.
- Giri, N. (1986). *Analysis of variance*. New Delhi: South Asian Publishers.
- Kiefer, J. (1958). On the non-randomized optimality and randomized nonoptimality of symmetrical designs. *The Annals of Mathematical Statistics* 29, 675-699.
- Raghavarao, D. (1971). *Constructions and combinatorial problems in design of experiments*. New York, N.Y.: Wiley.
- Roy, J. (1958). On the efficiency factor of block designs. *Sankhya* 19, 181-188.

Table 2.2

v	θ	C-matrix of order v	RBD Parameters	RBD Parameters
3	3	$c_{ii} = 2, c_{ij} = -1, i \neq j$	$r = b = k = 3$	$b = 6, r = 4, k = 2, \lambda = 2$
5	10	$c_{ij} = 8, c_{ij} = -2, i \neq j$	$r = b = 10, k = 5$	$b = 20, r = 12, k = 3, \lambda = 6$
6	6	$c_{ii} = 5, c_{ij} = -1, i \neq j$	$r = b = k = 6$	$b = 30, r = 10, k = 2, \lambda = 2$
6	12	$c_{ii} = 10, c_{ij} = -2, i \neq j$	$r = b = 12, k = 6$	$b = 24, r = 8, k = 3, \lambda = 2$
9	6	$c_{ii} = \frac{16}{3}, c_{ij} = -\frac{2}{3}, i \neq j$	$r = b = 6, k = 9$	$b = 24, r = 8, k = 3, \lambda = 2$
7	3	$c_{ii} = \frac{5}{2}, c_{ij} = -\frac{1}{2}, i \neq j$	$r = b = 3, k = 7$	Does not exist
8	16	$c_{ii} = 14, c_{ij} = -2, i \neq j$	$r = b = 16, k = 8$	Does not exist

Interval Estimation of Risk Difference in Simple Compliance Randomized Trials

Kung-Jong Lui
San Diego State University

Consider the simple compliance randomized trial, in which patients randomly assigned to the experimental treatment may switch to receive the standard treatment, while patients randomly assigned to the standard treatment are all assumed to receive their assigned treatment. Six asymptotic interval estimators for the risk difference in probabilities of response among patients who would accept the experimental treatment were developed. Monte Carlo methods were employed to evaluate and compare the finite-sample performance of these estimators. An example studying the effect of vitamin A supplementation on reducing mortality in preschool children was included to illustrate their practical use.

Key words: interval estimation; coverage probability; average length; efficiency; simple compliance trial.

Introduction

In randomized clinical trials, because the characteristics of the experimental treatment effect are often not completely known, some patients randomly assigned to this treatment may not comply with their assigned treatment. For example, consider the study of vitamin A supplementation to reduce mortality among preschool children in rural Indonesia (Sommer & Zeger, 1991, Sommer, Tarwotjo, Djunaedi, et al., 1986). Children who resided in 225 randomly selected villages out of 450 villages were assigned to receive a large oral dose of vitamin A two to three months following baseline enumeration and again six months later. Children in the remaining 225 villages would receive no vitamin A supplementation. The number of deaths in both comparison groups was ascertained in a second population census

12 months following the baseline census. Nearly 20 percent of children assigned to the experimental group failed to receive vitamin A supplementation. The mortality rates between the groups were then compared. Because the data structure for this simple compliance study is essentially the same as that of the single consent randomized design (Zelen, 1979, 1990, Anbar, 1983, Brunner & Neumann, 1985, Ellenberg, 1984, Bernhard & Compagnone, 1989, McHugh, 1984, Matts & McHugh, 1987, 1993, Lui & Lin, 2003), all statistical methods developed here for the simple compliance trial are applicable to the latter (Sommer & Zeger, 1991, Zelen, 1986) as well.

In this article, six asymptotic interval estimators of the risk difference between two treatments among patients who would accept the experimental treatment were developed. These included the interval estimator using Wald's statistic (Casella & Berger, 1990, Sommer & Zeger, 1991), the interval estimator using $\tanh^{-1}(x)$ transformation (Edwardes, 1995, Lui, 2002) of the maximum likelihood estimator (MLE), the interval estimator derived from a quadratic equation based on the MLE and its asymptotic properties, the interval estimator using an idea similar to that of Fieller's Theorem (Casella & Berger, 1990), and the interval estimator using a randomization-based approach with and without the continuity correction (Mark & Robins, 1993, Sato, 2000). To evaluate and compare the finite sample performance of these

Kung-Jong Lui is Professor of Mathematics and Statistics at San Diego State University. His current research interests are in statistical methods in epidemiology and randomized clinical trials, categorical data analysis, and sample surveys. He is a fellow of the American Statistical Association, and is serving as an Associate Editor for *Biometrical Journal* as well as an Editorial Advisor for *Journal of Probability and Statistical Science*.

estimators, Monte Carlo simulation was employed. Finally, an example studying the effect of vitamin A supplementation on reducing mortality in preschool children (Sommer, Tarwotjo, Djunaedi, et al., 1986) was included to illustrate their practical use.

Methods and Notations

Consider comparing an experimental treatment with a standard treatment in a simple compliance randomized trial. Patients randomly assigned to the experimental treatment are allowed to switch to receive the standard treatment, while patients randomly assigned to the standard treatment are assumed to all receive their assigned treatment. For clarity, the probabilities p_{ij} of response for the experimental treatment are summarized in the following table. For example, the parameter p_{11} denotes probability of obtaining a patient who would accept the experimental treatment and have a positive response. Define $p_{i+} = p_{i1} + p_{i0}$ and $p_{+j} = p_{1j} + p_{0j}$ for $i = 1, 0$ and $j = 1, 0$ (see Table 1).

Similarly, the parameter p_{ij}^* ($i = 1, 0$ and $j = 1, 0$) denotes the corresponding cell probability of response in the following table for the standard treatment. Define $p_{i+}^* = p_{i1}^* + p_{i0}^*$ and $p_{+j}^* = p_{1j}^* + p_{0j}^*$ for $i = 1, 0$ and $j = 1, 0$ (see Table 2).

Because a patient assigned to the experimental treatment will receive the standard

treatment if he/she declines to receive the experimental treatment and because patients are randomly assigned to either treatment, the equality $p_{i0} = p_{i0}^*$ (for $i = 1, 0$) can be reasonably assumed to hold (Sommer & Zeger, 1991). These imply that the proportions of patients who would consent to accept the experimental treatment between the two treatment groups are equal (i.e., $p_{+1} = p_{+1}^*$).

Suppose that there are n and m patients independently randomly assigned to receive the experimental and standard treatments, respectively. Let n_{ij} (and m_{ij}) denote the observed frequencies corresponding to the cell probabilities p_{ij} (and p_{ij}^*) in the experimental (and the standard) treatment. Then the random vector $n = (n_{11}, n_{10}, n_{01}, n_{00})'$ follows the multinomial distribution with parameters n and $(p_{11}, p_{10}, p_{01}, p_{00})'$. Because patients assigned to the standard treatment were not asked whether they would accept the experimental treatment, only the marginal total number of responses $m_{1+} (= m_{11} + m_{10})$, following the binomial distribution with parameters m and $p_{1+}^* (= p_{11}^* + p_{10}^*)$, was observed.

In this article, searching for a good interval estimator for the risk difference $\Delta (= (p_{11} / p_{+1}) - (p_{11}^* / p_{+1}^*) = (p_{1+} - p_{1+}^*) / p_{+1})$ between two treatments among patients who would accept the experimental treatment is the main focus of interest here. Note that the range for Δ is, by definition, $-1 < \Delta < 1$.

Table 1. Experimental Treatment: Consent to Receive the Experimental Treatment

Response	Yes	No	Total
Positive	p_{11}	p_{10}	p_{1+}
Negative	p_{01}	p_{00}	p_{0+}
Total	p_{+1}	p_{+0}	1

Table 2. Standard Treatment: Consent to Receive the Experimental Treatment

	Yes	No	Total
Positive	p_{11}^*	p_{10}^*	p_{1+}^*
Negative	p_{01}^*	p_{00}^*	p_{0+}^*
Total	p_{+1}^*	p_{+0}^*	1

Based on the intent-to-treat analysis (Zelen, 1979), patients are compared according to the treatments to which they are originally designated despite whether patients comply with their regimen. The maximum likelihood estimator (MLE) for Δ is given by

$$\hat{\Delta} = (\hat{p}_{1+} - \hat{p}_{1+}^*) / \hat{p}_{+1} \quad (1)$$

where $\hat{p}_{1+} = \hat{p}_{11} + \hat{p}_{10}$, $\hat{p}_{+1} = \hat{p}_{11} + \hat{p}_{01}$, $\hat{p}_{ij} = n_{ij} / n$, and $\hat{p}_{1+}^* = m_{1+} / m$. Furthermore, on the basis of the delta method (Appendix), an estimated asymptotic variance of $\hat{\Delta}$ is obtained as

$$\begin{aligned} & \hat{Var}(\hat{\Delta}) \\ &= [\hat{p}_{1+}(\hat{p}_{10} + \hat{p}_{01}) - \hat{p}_{1+}^*(2\hat{p}_{10} - \hat{p}_{1+}^*\hat{p}_{+0})] / (n\hat{p}_{+1}^3) \\ &+ \hat{p}_{1+}^*(1 - \hat{p}_{1+}^*) / (m\hat{p}_{+1}^2) \end{aligned} \quad (2)$$

Based on (1) and (2), an asymptotic $100(1-\alpha)$ percent confidence interval using Wald's statistic for Δ is given by

$$\begin{aligned} & [\max\{\hat{\Delta} - Z_{\alpha/2} (\hat{Var}(\hat{\Delta}))^{1/2}, -1\}, \\ & \min\{\hat{\Delta} + Z_{\alpha/2} (\hat{Var}(\hat{\Delta}))^{1/2}, 1\}] \end{aligned} \quad (3)$$

where $Z_{\alpha/2}$ is the upper $100(\alpha/2)$ th percentile of the standard normal distribution.

Because the sampling distribution of $\hat{\Delta}$ can be skewed, interval estimator (3) may not perform well when the number of patients assigned to either treatment is not large. On the basis of some empirical results that the transformation $\tanh^{-1}(x) = \frac{1}{2} \log((1+x)/(1-x))$ has been successfully applied to improve the performance of statistics relevant to the difference in proportions under other situations (Edwardes, 1995, Lui, 2002), this transformation is considered in this article as well. Based on the delta method again, an estimated asymptotic variance can be shown to be given by $\hat{Var}(\tanh^{-1}(\hat{\Delta})) = \hat{Var}(\hat{\Delta}) / (1 - \hat{\Delta}^2)^2$. This leads to produce an asymptotic $100(1-\alpha)$ percent confidence interval for Δ to be

$$\begin{aligned} & [\tanh(\tanh^{-1}(\hat{\Delta}) - Z_{\alpha/2} (\hat{Var}(\tanh^{-1}(\hat{\Delta})))^{1/2}), \\ & \tanh(\tanh^{-1}(\hat{\Delta}) + Z_{\alpha/2} (\hat{Var}(\tanh^{-1}(\hat{\Delta})))^{1/2})] \end{aligned} \quad (4)$$

Note that when both n and m are large, the probability $P((\hat{\Delta} - \Delta)^2 / \hat{Var}(\hat{\Delta}) \leq Z_{\alpha/2}^2) \approx 1 - \alpha$, where $\hat{Var}(\hat{\Delta})$ is, as shown in Appendix, given by

$$\begin{aligned} & \Delta[(p_{1+} - p_{1+}^*)(1 - p_{+1}) - 2(p_{11} - p_{1+}p_{+1})] / (np_{+1}^2) + \\ & p_{1+}(1 - p_{1+}) / (np_{+1}^2) + p_{1+}^*(1 - p_{1+}^*) / (mp_{+1}^2). \end{aligned}$$

This suggests the following quadratic inequality:

$$\Delta^2 - 2B\Delta + C \leq 0, \quad (5)$$

where

$$\begin{aligned} A &= 1 \\ B &= \hat{\Delta} + Z_{\alpha/2}^2 [(\hat{p}_{1+} - \hat{p}_{1+}^*)(1 - \hat{p}_{1+}) \\ &\quad - 2(\hat{p}_{11} - \hat{p}_{1+}\hat{p}_{+1})] / (2n\hat{p}_{+1}^2) \\ C &= \hat{\Delta}^2 - Z_{\alpha/2}^2 [\hat{p}_{1+}(1 - \hat{p}_{1+}) / (n\hat{p}_{+1}^2) \\ &\quad + \hat{p}_{1+}^*(1 - \hat{p}_{1+}^*) / (m\hat{p}_{+1}^2)] \end{aligned}$$

Thus, if $B^2 - AC > 0$, then an asymptotic $100(1-\alpha)$ percent confidence interval for Δ would be given by

$$\begin{aligned} &[\max\{(B - \sqrt{B^2 - AC}) / A, -1\}, \\ &\min\{(B + \sqrt{B^2 - AC}) / A, 1\}] \end{aligned} \quad (6)$$

To alleviate the concern that the sampling distribution of the MLE $\hat{\Delta} = (\hat{p}_{1+} - \hat{p}_{1+}^*) / \hat{p}_{+1}$ can be, as noted previously, skewed when n or m is small, the idea of Fieller's Theorem (Casella & Berger, 1990) may be considered. Define $Z^* = (\hat{p}_{1+} - \hat{p}_{1+}^*) - \Delta\hat{p}_{+1}$. First, note that the expectation $E(Z^*)$ is equal to 0. The variance of Z^* is

$$\begin{aligned} Var(Z^*) &= p_{1+}(1 - p_{1+}) / n + p_{1+}^*(1 - p_{1+}^*) / m + \\ &\Delta^2 p_{+1}(1 - p_{+1}) / n - 2\Delta(p_{11} - p_{1+}p_{+1}) / n \end{aligned} \quad (7)$$

Thus, when both n and m are large, the probability $P((Z^*)^2 / Var(Z^*) \leq Z_{\alpha/2}^2) \approx 1 - \alpha$. This leads to considering the following quadratic equation:

$$A^* \Delta^2 - 2B^* \Delta + C^* \leq 0 \quad (8)$$

where

$$\begin{aligned} A^* &= \hat{p}_{+1}^2 - Z_{\alpha/2}^2 \hat{p}_{+1}(1 - \hat{p}_{+1}) / n \\ B^* &= (\hat{p}_{1+} - \hat{p}_{1+}^*) \hat{p}_{+1} - Z_{\alpha/2}^2 (\hat{p}_{11} - \hat{p}_{1+}\hat{p}_{+1}) / n \\ C^* &= (\hat{p}_{1+} - \hat{p}_{1+}^*)^2 - \\ &\quad Z_{\alpha/2}^2 \left(\frac{\hat{p}_{1+}(1 - \hat{p}_{1+})}{n} + \frac{\hat{p}_{1+}^*(1 - \hat{p}_{1+}^*)}{m} \right) \end{aligned}$$

If $(B^*)^2 - A^*C^* > 0$, then an asymptotic $100(1-\alpha)$ percent confidence interval for Δ would be given by

$$\begin{aligned} &[\max\{(B^* - \sqrt{(B^*)^2 - A^*C^*}) / A^*, -1\}, \\ &\min\{(B^* + \sqrt{(B^*)^2 - A^*C^*}) / A^*, 1\}] \end{aligned} \quad (9)$$

Following Mark and Robins (1993), Sato (1995, 2000) discussed sample size calculation using a randomization-based approach in binary outcome data. Based on the same arguments as those given by (Sato, 2000, pp. 2691-2692), an asymptotic $100(1-\alpha)$ percent confidence interval for Δ is obtained as

$$\begin{aligned} &[\max\{(B_{-}^{**} - \sqrt{(B_{-}^{**})^2 - A^{**}C_{-}^{**}}) / A^{**}, -1\}, \\ &\min\{(B_{+}^{**} + \sqrt{(B_{+}^{**})^2 - A^{**}C_{+}^{**}}) / A^{**}, 1\}] \end{aligned} \quad (10)$$

where

$$\begin{aligned} A^{**} &= n_{+1}^2 [m^2 + Z_{\alpha/2}^2 (nm) / N] \\ B_{\pm}^{**} &= mn_{+1} (mn_{1+} - nm_{1+} \pm c) \\ &\quad - Z_{\alpha/2}^2 nm n_{+1} (N - 2(n_{1+} + m_{1+})) / (2N) \\ C_{\pm}^{**} &= (mn_{1+} - nm_{1+} \pm c)^2 \\ &\quad - Z_{\alpha/2}^2 nm (n_{1+} + m_{1+}) (N - (n_{1+} + m_{1+})) / N \end{aligned}$$

$N = n + m$, $n_{1+} = n_{11} + n_{10}$, $n_{+1} = n_{11} + n_{01}$, and $c = N/2$ when one wishes to employ the continuity correction; and $c = 0$, otherwise.

Monte Carlo Simulation

To evaluate and compare the finite sample performance of interval estimators (3), (4), (6), (9), and (10), Monte Carlo simulation is employed. Given the population proportion of a randomly selected patient who would consent to accept the experimental treatment $p_{+1} (= p_{+1}^*)$ and the probability of positive response $p_{r|c}^* (= p_{11}^* / p_{+1}^*)$ among patients who would consent to accept the experimental treatment (if he/she had been assigned to the experimental treatment) in the standard treatment, the cell probability $p_{11}^* = p_{r|c}^* p_{+1}^* = p_{r|c}^* p_{+1}$ can then be uniquely determined. For given a value Δ , the cell probability: $p_{11} = p_{11}^* + \Delta p_{+1}$.

Similarly, the cell probability $p_{10}^* (= p_{10})$ can be determined by $p_{r|\bar{c}}^* (1 - p_{+1}^*)$ when the probability $p_{r|\bar{c}}^*$ of positive response among patients who would decline to receive the experimental treatment in the standard treatment is given. In the simulation, $p_{r|\bar{c}}^*$ is arbitrarily set equal to $p_{r|c}^* / 3$. For simplicity, the case of equal sample allocation (i.e., $n = m$) is focused here.

This article considers the situations, in which the probability of a randomly selected patient who would consent to accept the experimental treatment $p_{+1} = 0.30, 0.50, 0.80$; the underlying difference between two treatments among patients who would consent to accept the experimental treatment $\Delta = 0.0, 0.10, 0.20$; the conditional probability of positive responses among patients who would consent to accept the experimental treatment in the standard treatment group $p_{r|c}^* = 0.20, 0.50$; and the number of patients assigned to either treatment $n (= m) = 30, 50, 100$.

For each configuration determined by a combination of these parameters, SAS (1990) is applied to generate 10000 repeated samples of

observations following the desired multinomial distributions to estimate the coverage probability and the average length. Note that if a sample led to $\hat{p}_{+1} = 0$ or an estimate $\hat{\Delta}$ was out of the range $-1 < \Delta < 1$, any interval estimator discussed here would be inapplicable.

Furthermore, if the two distinct real roots of a quadratic equation did not exist, the corresponding interval estimator could not be employed either. The estimated coverage probability and average length are calculated over those samples for which the corresponding interval estimator exists. For completeness, the probability of failing to produce a confidence interval for each interval estimator in all the above situations is also calculated.

Results

Table 3 summarizes the estimated coverage probability and average length of 95% confidence interval using interval estimators (3), (4), (6), (9), as well as the interval estimator (10) with and without the continuity correction. Note that the coverage probability of interval estimator (10) using the randomization approach without the continuity correction can be frequently less than 95% by more than 1% when the underlying difference Δ is not equal to 0. By contrast, the interval estimator (10) with the continuity correction tends to be conservative and hence lose efficiency with respect to the average length as compared with the other estimators considered here (Table 3). Note further that interval estimators (3) and (6) are essentially equivalent and both are preferable to interval estimator (9) with respect to efficiency.

Finally, note that interval estimator (4) is consistently the most efficient with respect to the average length among all interval estimators with the coverage probability larger than or equal to the desired 95% confidence level considered here. In fact, interval estimator (4) is the only one estimator that has the estimated coverage probability larger than the 95% confidence level in all the situations discussed in Table 3.

Table 3. The estimated coverage probability and average length (in parenthesis) of 95% confidence interval using interval estimators (3), (4), (6), (9), as well as the interval estimator (10) with and without continuity correction. Each entry is calculated on the basis of 10000 repeated samples.

p_{+1}	Δ	$p_{r c}^*$	n	(3)	(4)	(6)	(9)	(10)	
0.30	0.00	0.20	30	0.965 (1.060)	0.992✓ (1.004)	0.949 (1.044)	0.946 (1.166)	0.995 ^a (1.204)	0.950 ^b (1.016)
			50	0.956 (0.831)	0.984✓ (0.792)	0.937 (0.829)	0.937 (0.915)	0.980 (0.941)	0.944 (0.815)
			100	0.956 (0.582)	0.970✓ (0.565)	0.950 (0.583)	0.950 (0.613)	0.976 (0.646)	0.951 (0.580)
			30	0.972 (1.395)	0.998✓ (1.328)	0.964 (1.352)	0.962 (1.453)	0.988 (1.483)	0.972 (1.341)
			50	0.965 (1.161)	0.995✓ (1.096)	0.950 (1.145)	0.950 (1.231)	0.975 (1.237)	0.956 (1.125)
			100	0.954 (0.835)	0.983✓ (0.795)	0.946 (0.836)	0.946 (0.877)	0.966 (0.886)	0.949 (0.821)
	0.10	0.20	30	0.965 (1.093)	0.994✓ (1.034)	0.949 (1.084)	0.948 (1.210)	0.987 (1.229)	0.944 (1.036)
			50	0.958 (0.868)	0.987✓ (0.825)	0.945 (0.868)	0.946 (0.959)	0.976 (0.956)	0.938✖ (0.828)
			100	0.953 (0.605)	0.968✓ (0.587)	0.947 (0.606)	0.946 (0.638)	0.964 (0.647)	0.939✖ (0.581)
			30	0.969 (1.391)	0.999✓ (1.325)	0.960 (1.362)	0.958 (1.461)	0.987 (1.499)	0.967 (1.354)
			50	0.961 (1.158)	0.998✓ (1.091)	0.949 (1.148)	0.950 (1.236)	0.975 (1.244)	0.955 (1.130)
			100	0.951 (0.837)	0.986✓ (0.795)	0.943 (0.838)	0.943 (0.879)	0.964 (0.891)	0.945 (0.825)
	0.20	0.20	30	0.960 (1.107)	0.994✓ (1.061)	0.944 (1.106)	0.944 (1.227)	0.980 (1.247)	0.932✖ (1.047)
			50	0.956 (0.882)	0.989✓ (0.841)	0.944 (0.886)	0.946 (0.975)	0.970 (0.961)	0.931✖ (0.831)
			100	0.952 (0.617)	0.972✓ (0.599)	0.946 (0.619)	0.946 (0.650)	0.960 (0.647)	0.933✖ (0.580)
			30	0.973 (1.372)	1.000✓ (1.328)	0.964 (1.362)	0.962 (1.440)	0.991 (1.504)	0.975 (1.361)
			50	0.959 (1.142)	0.998✓ (1.088)	0.946 (1.143)	0.946 (1.229)	0.978 (1.248)	0.955 (1.134)
			100	0.958 (0.824)	0.991✓ (0.785)	0.952 (0.828)	0.952 (0.868)	0.972 (0.889)	0.957 (0.823)

^a: This column is calculated with the continuity correction.

^b: This column is calculated without the continuity correction.

✕: means that the estimated coverage probability is less than the desired 95% confidence level by more than 1%.

✓: indicates that the interval estimator has the shortest average length among interval estimators subject to its estimated coverage probability less than the desired 95% confidence level by no more than 1%.

Table 3 (Continued)

p_{+1}	Δ	$p_{r c}^*$	n	(3)	(4)	(6)	(9)	(10)	
0.50	0.00	0.20	30	0.949 (0.698)	0.971✓ (0.670)	0.941 (0.699)	0.941 (0.755)	0.987 (0.819)	0.944 (0.688)
			50	0.955 (0.538)	0.967✓ (0.524)	0.947 (0.538)	0.947 (0.563)	0.979 (0.612)	0.949 (0.533)
			100	0.952 (0.379)	0.957✓ (0.374)	0.950 (0.379)	0.950 (0.387)	0.973 (0.417)	0.950 (0.377)
			30	0.950 (0.973)	0.985✓ (0.919)	0.945 (0.968)	0.945 (1.033)	0.976 (1.057)	0.953 (0.935)
			50	0.951 (0.755)	0.976✓ (0.723)	0.947 (0.756)	0.947 (0.790)	0.971 (0.814)	0.952 (0.736)
			100	0.945 (0.528)	0.957✓ (0.516)	0.940 (0.529)	0.940 (0.540)	0.961 (0.561)	0.942 (0.521)
	0.10	0.20	30	0.948 (0.734)	0.976✓ (0.703)	0.943 (0.735)	0.943 (0.793)	0.979 (0.821)	0.934✗ (0.689)
			50	0.947 (0.564)	0.965✓ (0.549)	0.944 (0.565)	0.944 (0.590)	0.969 (0.610)	0.933✗ (0.532)
			100	0.951 (0.399)	0.957✓ (0.393)	0.948 (0.399)	0.948 (0.407)	0.961 (0.417)	0.937✗ (0.378)
			30	0.951 (0.968)	0.990✓ (0.912)	0.944 (0.967)	0.944 (1.033)	0.978 (1.060)	0.953 (0.937)
			50	0.949 (0.746)	0.977✓ (0.714)	0.944 (0.748)	0.944 (0.781)	0.971 (0.810)	0.952 (0.732)
			100	0.951 (0.525)	0.963✓ (0.513)	0.948 (0.526)	0.948 (0.536)	0.967 (0.561)	0.952 (0.521)
0.20	0.20	30	0.943 (0.753)	0.975✓ (0.723)	0.938 (0.757)	0.939 (0.815)	0.969 (0.824)	0.919✗ (0.692)	
		50	0.951 (0.581)	0.966✓ (0.565)	0.946 (0.582)	0.947 (0.608)	0.963 (0.612)	0.928✗ (0.534)	
		100	0.948 (0.411)	0.956✓ (0.405)	0.947 (0.411)	0.946 (0.419)	0.953 (0.418)	0.927✗ (0.378)	
		30	0.950 (0.953)	0.992✓ (0.901)	0.942 (0.957)	0.942 (1.022)	0.978 (1.062)	0.955 (0.938)	
		50	0.949 (0.735)	0.977✓ (0.706)	0.943 (0.738)	0.942 (0.770)	0.971 (0.810)	0.952 (0.732)	
		100	0.951 (0.516)	0.962✓ (0.505)	0.947 (0.517)	0.948 (0.527)	0.968 (0.560)	0.955 (0.521)	

Table 3 (continued)

p_{+1}	Δ	$p_{r c}^*$	n	(3)	(4)	(6)	(9)	(10)	
0.80	0.00	0.20	30	0.945	0.953✓	0.941	0.941	0.982	0.951
				(0.473)	(0.464)	(0.473)	(0.482)	(0.546)	(0.467)
			50	0.945	0.950✓	0.942	0.942	0.974	0.945
				(0.368)	(0.364)	(0.368)	(0.372)	(0.414)	(0.365)
			100	0.949	0.952✓	0.950	0.950	0.968	0.950
				(0.261)	(0.259)	(0.261)	(0.262)	(0.284)	(0.260)
		0.50	30	0.948	0.964	0.951	0.951	0.977	0.952✓
				(0.624)	(0.604)	(0.624)	(0.636)	(0.682)	(0.603)
			50	0.946	0.953	0.944	0.944	0.966	0.944✓
				(0.484)	(0.475)	(0.484)	(0.489)	(0.522)	(0.474)
			100	0.947	0.952	0.945	0.945	0.959	0.945✓
				(0.343)	(0.340)	(0.343)	(0.345)	(0.364)	(0.339)
	0.10	0.20	30	0.944	0.953✓	0.941	0.942	0.969	0.930✖
				(0.505)	(0.494)	(0.505)	(0.514)	(0.545)	(0.466)
			50	0.950	0.956✓	0.950	0.950	0.965	0.934✖
				(0.393)	(0.388)	(0.393)	(0.397)	(0.414)	(0.365)
			100	0.947	0.950✓	0.946	0.946	0.956	0.932✖
				(0.279)	(0.277)	(0.279)	(0.280)	(0.284)	(0.260)
		0.50	30	0.943	0.960✓	0.937	0.939	0.973	0.950
				(0.619)	(0.600)	(0.620)	(0.631)	(0.682)	(0.603)
			50	0.949	0.958✓	0.947	0.947	0.972	0.952
				(0.481)	(0.472)	(0.481)	(0.486)	(0.523)	(0.475)
			100	0.945	0.950✓	0.944	0.944	0.963	0.948
				(0.340)	(0.337)	(0.340)	(0.342)	(0.364)	(0.339)
0.20	0.20	30	0.944	0.956✓	0.943	0.943	0.961	0.918✖	
			(0.524)	(0.513)	(0.524)	(0.534)	(0.546)	(0.467)	
		50	0.943	0.951✓	0.943	0.943	0.951	0.917✖	
			(0.408)	(0.402)	(0.408)	(0.412)	(0.413)	(0.365)	
		100	0.946	0.952	0.947	0.947	0.946✓	0.918✖	
			(0.289)	(0.287)	(0.289)	(0.291)	(0.285)	(0.260)	
	0.50	30	0.949	0.964✓	0.947	0.947	0.979	0.960	
			(0.605)	(0.588)	(0.606)	(0.617)	(0.683)	(0.605)	
		50	0.946	0.957✓	0.945	0.945	0.974	0.957	
			(0.468)	(0.460)	(0.469)	(0.473)	(0.523)	(0.474)	
		100	0.950	0.953✓	0.948	0.949	0.969	0.956	
			(0.332)	(0.329)	(0.332)	(0.334)	(0.364)	(0.340)	

Finally, Table 4 shows that except for the extreme cases where both the probability of consent $\hat{p}_{+1}$ and the number of patients assigned to each treatment n are small (i.e., $\hat{p}_{+1} = 0.30$ and $n = 30$), the probability of failing to produce a 95% confidence interval by using interval estimators discussed here is generally negligible (≤ 0.002) in all the situations considered in Table 3.

An Example

Consider the randomized trial of studying vitamin A supplementation to reduce the mortality among preschool children in rural Indonesia (Sommer, Tarwotjo, Djunaedi, et al., 1986). As described previously, children were randomly assigned to either the treatment group of receiving a large oral dose of vitamin A for two to three months following baseline evaluation and again six months later, or to the control group without receiving any vitamin A supplementation. In the control group, children were precluded from receiving a placebo for ethical reasons and therefore, only the total number of survival children without the information on children who would fail to take vitamin A if they were assigned to receive this vitamin was observed. The results on mortality from month 4 (following completion of the first distribution cycle) to month 12 were compared.

As shown elsewhere (Sommer & Zeger, 1991), the number of survival children in the control group, ($m_{1+} =$) was 11,514 out of ($m =$) 11,588 children. Furthermore, the frequencies: $n_{11} = 9,663$, $n_{10} = 2,385$, $n_{01} = 12$, $n_{00} = 34$, were obtained for the total number 12,094 of children assigned to the treatment of receiving vitamin A. Suppose that one is interested in estimation of vitamin A effect on the non-death rate between the group receiving vitamin A and the group without receiving vitamin A supplement among children who would consent to accept vitamin A.

Given the above data, the MLE $\hat{\Delta}$ is 0.0032. Applying interval estimators (3), (4), (6), (9), as well as (10) with and without the continuity correction leads to produce 95% confidence intervals to be: [0.0010, 0.0055], [0.0010, 0.0055], [0.0010, 0.0055], [0.0010,

0.0055], [0.0008, 0.0061], and [0.0009, 0.0060], respectively. Because the total number of children in this randomized trial is quite large, the resulting 95% confidence interval using interval estimators (3), (4), (6), and (9) are essentially identical. Note that the 95% confidence interval using (10) with the continuity correction tends to have the length larger than the others. This is actually consistent with the previous findings obtained in simulations. Because all the above lower limits fall above 0, there is a significant evidence to support that taking vitamin A can increase the survival rate for preschool children at 5%-level.

Conclusion

It was found that the interval estimator (10) using the randomization-based approach with the continuity correction can lose efficiency, while this estimator without the continuity correction can be slightly liberal when the underlying difference Δ is not equal to 0. This article also finds that the interval estimator (4) using the $\tanh^{-1}(\hat{\Delta})$ transformation is probably preferable to all the other estimators discussed here with respect to the coverage probability and the average length. Thus, interval estimator (4) is recommended for general use.

Following Brunner and Neumann (1985) as well as Bernhard and Compagnone (1989), one can also discuss interval estimation of the selection effect, defined as $p_{11}^* / p_{+1}^* - p_{10}^* / p_{+0}^*$ which is the difference in probabilities of positive response for the standard treatment between patients who would agree and patients who would decline to receive the experimental treatment. By the functional invariance property, the MLE for this selection effect is simply equal to $(\hat{p}_{1+}^* - \hat{p}_{10}^* / \hat{p}_{+0}^*) / \hat{p}_{+1}^*$ (Appendix). Thus, it is straightforward to extend the above discussion to account for interval estimation of this selection effect. However, the detailed derivation and discussion on the selection effect are beyond the scope of this article and can be a future possible research topic.

Table 4. The estimated probability of failing to apply interval estimators (3), (4), (6), (9), as well as the interval estimator (10) with and without the continuity correction.

p_{+1}	Δ	$p_{r c}^*$	n	(3)	(4)	(6)	(9)	(10)	
0.30	0.00	0.20	30	0.012	0.012	0.013	0.017	0.047 ^a	0.012 ^b
			50	0.001	0.001	0.001	0.001	0.006	0.001
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.041	0.041	0.041	0.043	0.041	0.041
			50	0.010	0.010	0.010	0.010	0.010	0.010
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.10	30	0.010	0.010	0.010	0.014	0.041	0.010
			50	0.002	0.002	0.002	0.002	0.004	0.002
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.038	0.038	0.038	0.041	0.038	0.038
			50	0.008	0.008	0.008	0.008	0.008	0.008
			100	0.000	0.000	0.000	0.000	0.000	0.000
	0.20	0.20	30	0.016	0.016	0.016	0.020	0.048	0.016
			50	0.002	0.002	0.002	0.002	0.005	0.002
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.040	0.040	0.040	0.043	0.040	0.040
			50	0.009	0.009	0.009	0.009	0.009	0.009
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.10	30	0.000	0.000	0.000	0.000	0.015	0.000
			50	0.000	0.000	0.000	0.000	0.001	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.002	0.002	0.002	0.002	0.002	0.002
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
0.50	0.00	0.20	30	0.000	0.000	0.000	0.000	0.015	0.000
			50	0.000	0.000	0.000	0.000	0.001	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.002	0.002	0.002	0.002	0.002	0.002
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.10	30	0.000	0.000	0.000	0.000	0.015	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.001	0.001	0.001	0.001	0.001	0.001
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
	0.20	0.20	30	0.000	0.000	0.000	0.000	0.014	0.000
			50	0.000	0.000	0.000	0.000	0.001	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.002	0.002	0.002	0.002	0.002	0.002
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.10	30	0.000	0.000	0.000	0.000	0.015	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000

^a : This column is calculated with the continuity correction.^b : This column is calculated without the continuity correction

Table 4 (continued)

0.80	0.00	0.20	30	0.000	0.000	0.000	0.000	0.004	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.000	0.000	0.000	0.000	0.000	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
	0.10	0.20	30	0.000	0.000	0.000	0.000	0.003	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.000	0.000	0.000	0.000	0.000	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
	0.20	0.20	30	0.000	0.000	0.000	0.000	0.003	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000
		0.50	30	0.000	0.000	0.000	0.000	0.000	0.000
			50	0.000	0.000	0.000	0.000	0.000	0.000
			100	0.000	0.000	0.000	0.000	0.000	0.000

In summary, six interval estimators for estimating the risk difference in a simple compliance randomized trial have been developed. The evaluation and comparison of the finite-sample performance of these interval estimators have been carried out in a variety of situations. The interval estimator using the $\tanh^{-1}(\hat{\Delta})$ transformation of the MLE has been shown to be the best among all interval estimators for the risk difference considered here. The results and the discussion presented in this article should have use for biostatisticians and clinicians when they encounter data under a simple compliance randomized trial.

References

- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444-455.
- Anbar, D. (1983). The relative efficiency of Zelen's prerandomization design for clinical trials. *Biometrics*, 39, 711-718.
- Bernhard, G. & Compagnone, D. (1989). Binary data in prerandomized designs. *Biometrical Journal*, 31, 29-33.
- Brunner, E. & Neumann, N. (1985). On the mathematical basis of Zelen's prerandomized designs. *Methods of Information in Medicine*, 24, 120-130.
- Casella, G. & Berger, R. L. (1990). *Statistical Inference*. Belmont, CA: Duxbury.
- Edwardes, M. D. (1995). A confidence interval for $\Pr(X < Y) - \Pr(X > Y)$ estimated from simple cluster samples. *Biometrics*, 51, 571-578.
- Ellenberg, S. S. (1984). Randomization designs in comparative clinical trials. *The New England Journal of Medicine*, 310, 1404-1408.
- Lui, K.-J. (2002). Notes on estimation of the general odds ratio and the general risk difference for paired-sample data. *Biometrical Journal*, 44, 957-968.
- Lui, K.-J. & Lin, C.-D. (2003). Interval estimation of treatment effects in double consent randomized design. *Statistica Sinica*, 13, 179-187.

Mark, S. D. & Robins, J. M. (1993). A method for the analysis of randomized trials with compliance information: an application to the multiple risk factor intervention trial. *Controlled Clinical Trials*, 14, 79-97.

Matts, J. & McHugh, R. (1987). Randomization and efficiency in Zelen's single consent design. *Biometrics*, 43, 885-894.

Matts, J. P. and McHugh, R. B. (1993). Precision estimation in Zelen's single-consent design. *Biometrical Journal*, 35, 65-72.

McHugh, R. (1984). Validity and treatment dilution in Zelen's single consent design. *Statistics in Medicine*, 3, 215-218.

SAS Institute Inc. (1990). *SAS Language, Reference Version 6*, 1st edition. Cary, North Carolina: SAS Institute.

Sato, T. (2000). Sample size calculations with compliance information. *Statistics in Medicine*, 19, 2689-2697.

Sommer, A., Tarwotjo, I., Djunaedi, E., West, K. P., Loedin, A. A., Tilden, R., & Mele, L. (1986). Impact of vitamin A Supplementation on childhood mortality: A randomized controlled community trial. *Lancet* i, 1169-1173.

Sommer, A. & Zeger, S. L. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine*, 10, 45-52.

Zelen, M. (1979). A new design for randomized clinical trials. *The New England Journal of Medicine*, 300, 1242-1245.

Zelen, M. (1986). Response. *Journal of Chronic Disease*, 39, 247-249.

Zelen, M. (1990). Randomized consent designs for clinical trials: an update. *Statistics in Medicine*, 9, 645-656.

Appendix

Let R denote the response random variable: $R = 1$ if the underlying response is positive and $= 0$, otherwise. Furthermore, let T and C denote the random variables for the treatment assignment and the acceptance of the experimental treatment, respectively: $T = 1$ for the experimental treatment, and $T = 0$ for the standard treatment; $C = 1$ for accepting the experimental treatment and $C = 0$, otherwise. Let $P(R = 1|T = i)$ denote the probability of response for treatment i ($i=1$ and 0). Because patients were randomly assigned to the

treatment, it may reasonably be assumed that the proportion of patients who would consent to accept the experimental between the experimental and standard treatment groups are equal

$$(i.e., P(C = 1|T = 1) = P(C = 1|T = 0) = P(C = 1))$$

Note that

$$\begin{aligned} P(R = 1|T = 1) - P(R = 1|T = 0) = \\ [P(R = 1|T = 1, C = 1) - \\ P(R = 1|T = 0, C = 1)] P(C = 1|T = 1) + \\ [P(R = 1|T = 1, C = 0) - \\ P(R = 1|T = 0, C = 0)] P(C = 0|T = 1). \end{aligned} \quad (A.1)$$

Because a patient assigned to the experimental treatment who does not consent to accept his/her assigned treatment will receive the standard treatment (i.e., $T = 0$), by the exclusion restriction assumption (Angrist, Imbens, & Rubin, 1996), the equality $P(R = 1|T = 1, C = 0) = P(R = 1|T = 0, C = 0)$ holds. Thus, the second component of (A.1) is zero. These suggests that equation (A.1) can be expressed in terms of notations p_{ij} and p_{ij}^* as

$$p_{1+} - p_{1+}^* = \Delta p_{+1} \quad (A.2)$$

where $\Delta = (p_{11} / p_{+1}) - (p_{11}^* / p_{+1}^*)$ is the difference in probabilities of positive response among patients who would consent to accept the experimental treatment. Say, there are n and m patients independently randomly assigned to receive the experimental and standard treatments, respectively. Let n_{ij} (and m_{ij}) denote the observed frequency corresponding to the cell probability p_{ij} (and p_{ij}^*) in the experimental (and the standard) treatment. Then, the random vector $n = (n_{11}, n_{10}, n_{01}, n_{00})'$ follows the multinomial distribution with parameters n and $(p_{11}, p_{10}, p_{01}, p_{00})'$. Furthermore, the marginal number $m_{1+} (= m_{11} + m_{10})$ of patients with positive response in the standard treatment

follows the binomial distribution with parameters m and $p_{1+}^* (= p_{11}^* + p_{10}^*)$. Therefore, by the functional invariance property (Casella & Berger, 1990) of the MLE, the MLE for Δ is

$$\hat{\Delta} = (\hat{p}_{1+} - \hat{p}_{1+}^*) / \hat{p}_{+1} \quad (\text{A.3})$$

where $\hat{p}_{1+} = \hat{p}_{11} + \hat{p}_{10}$, $\hat{p}_{+1} = \hat{p}_{11} + \hat{p}_{01}$, and $\hat{p}_{1+}^* = \hat{p}_{11}^* + \hat{p}_{10}^*$, where $\hat{p}_{ij} = n_{ij} / n$ and $\hat{p}_{ij}^* = m_{ij} / m$. By the Central Limit Theorem (Casella & Berger, 1990), as both n and m are large, the random vector $(\hat{p}_{1+}, \hat{p}_{+1}, \hat{p}_{1+}^*)'$ asymptotically follows the multivariate normal distribution with mean vector $(p_{1+}, p_{+1}, p_{1+}^*)'$ and covariance matrix with diagonal terms given by

$$\begin{aligned} \text{Var}(\hat{p}_{1+}) &= p_{1+}(1 - p_{1+}) / n, \\ \text{Var}(\hat{p}_{+1}) &= p_{+1}(1 - p_{+1}) / n, \\ \text{Var}(\hat{p}_{1+}^*) &= p_{1+}^*(1 - p_{1+}^*) / m, \end{aligned}$$

and off-diagonal terms given by

$$\text{Cov}(\hat{p}_{1+}, \hat{p}_{+1}) = (p_{11} - p_{1+}p_{+1}) / n,$$

and

$$\text{Cov}(\hat{p}_{1+}, \hat{p}_{1+}^*) = \text{Cov}(\hat{p}_{+1}, \hat{p}_{1+}^*) = 0.$$

On the basis of the delta method, an asymptotic variance of $\hat{\Delta}$ is given by

$$\begin{aligned} \text{Var}(\hat{\Delta}) &= \Delta[(p_{1+} - p_{1+}^*)(1 - p_{+1}) \\ &\quad - 2(p_{11} - p_{1+}p_{+1})] / (np_{+1}^2) + \end{aligned}$$

$$p_{1+}(1 - p_{1+}) / (np_{+1}^2) + p_{1+}^*(1 - p_{1+}^*) / (mp_{+1}^2). \quad (\text{A.4})$$

From (A.4), the following variance estimator

$$\begin{aligned} \hat{\text{Var}}(\hat{\Delta}) &= [\hat{p}_{1+}(\hat{p}_{10} + \hat{p}_{01}) \\ &\quad - \hat{p}_{1+}^*(2\hat{p}_{10} - \hat{p}_{1+}^*(1 - \hat{p}_{+1}))] / (n\hat{p}_{+1}^3) \\ &\quad + \hat{p}_{1+}^*(1 - \hat{p}_{1+}^*) / (m\hat{p}_{+1}^2) \end{aligned} \quad (\text{A.5})$$

is obtained by simply substituting $\hat{\Delta}$ for Δ , $\hat{p}_{1+}$ for p_{1+} , $\hat{p}_{+1}$ for p_{+1} , $\hat{p}_{1+}^*$ for p_{1+}^* , and $\hat{p}_{11}$ for p_{11} . For assessing the selection effect (Brunner & Neumann, 1985, Bernhard & Compagnone, 1989), consider the following difference:

$$\begin{aligned} &P(R = 1 | T = 0) - P(R = 1 | T = 1, C = 0) \\ &= [P(R = 1 | T = 0, C = 1) \\ &\quad - P(R = 1 | T = 0, C = 0)]P(C = 1 | T = 1) \end{aligned} \quad (\text{A.6})$$

In terms of notations p_{ij} and p_{ij}^* , formula (A.6) may be rewritten as

$$p_{1+}^* - p_{10} / p_{+0} = (p_{11}^* / p_{+1}^* - p_{10}^* / p_{+0}^*) p_{+1}. \quad (\text{A.7})$$

Thus, the selection effect, defined as $p_{11}^* / p_{+1}^* - p_{10}^* / p_{+0}^*$ which is simply the difference in probabilities of positive response for the standard treatment between patients who would accept and patients who would decline to the experimental treatment, can be estimated by using the MLE:

$$(\hat{p}_{1+}^* - \hat{p}_{10} / \hat{p}_{+0}) / \hat{p}_{+1}. \quad (\text{A.8})$$

Simulation of Non-normal Autocorrelated Variables

H.E.T. Holgersson

Department of Economics and Statistics
Jönköping International Business School

All statistical methods rely on assumptions to some extent. Two assumptions frequently met in statistical analyses are those of normal distribution and independence. When examining robustness properties of such assumptions by Monte Carlo simulations it is therefore crucial that the possible effects of autocorrelation and non-normality are not confounded so that their separate effects may be investigated. This article presents a number of non-normal variables with non-confounded autocorrelation, thus allowing the analyst to specify autocorrelation or shape properties while keeping the other effect fixed.

Key words: Autocorrelation, non-normality, confounding, robustness.

Introduction

All statistical methods rely on assumptions to some extent. These assumptions, for example, may be that some moments are finite or that the variance is homogenous at all data points. Other assumptions involve normal distribution or independence. If some of the assumptions are violated then the expected properties of the method may no longer hold. For example, a statistical hypothesis test that requires independence of the data may seriously over reject under the null hypothesis if the data possess autocorrelation.

It is therefore important to investigate the robustness properties of statistical methods before they are applied to real data. Although modern computers are developing at a rapid pace, it has become increasingly popular to

perform robustness studies of such assumptions by Monte Carlo simulations. However, when examining robustness to autocorrelation and non-normality, some technical problems arise.

Because autocorrelation usually is generated by a recursive sequence of random numbers, the central limit theorem will force the autocorrelated variable to be more normal when compared to the variable used to generate the sequence. For example, imagine the problem of investigating the robustness of a non-normality test to autocorrelation. If a skewed variable is used to recursively generate a skewed and autocorrelated variable, then this new variable will be more symmetric than the original one and will be more symmetric the larger the autocorrelation is. Thus, the simulation study will not reveal the separate effects of autocorrelation and non-normality as was intended.

Several such examples are to be found in the literature. For example, Shukur (2000) examined the robustness of an autocorrelation test to non-normality by generating a first order autoregressive process with non-normal disturbances and Bai and Ng (2005) applied the first order autoregressive process with non-normal disturbances to investigate the robustness of a non-normality test to autocorrelation. Such effects of confounding can be avoided by using alternative methods of generating the variables.

H. E. T. Holgersson is Associate Professor at Jönköping International Business School with research and teaching. His research has mainly been focused on assessing distributional properties. Research also includes multivariate analysis and financial statistics.

A number of variables are proposed that are non-normal but autocorrelated and allow for separate control of the shape properties (skewness/kurtosis) and autocorrelation. All proposed variables are easy to generate in standard software packages.

Non-Normal And Autocorrelated Variables

The question of how to simulate random variables with given distributional properties have been given a great deal of attention (see Johnson (1987) for a general description). One of the most frequently used methods is the so-called inverse method. Assume the problem consists of generating a variate X whose distribution is specified by F and that F is strictly increasing with inverse function F^{-1} . Then, the inverse method consists of first generating a uniformly distributed variate U and then calculating the variable of interest by $X = F^{-1}(U)$. This method is fast and simple because all statistical software provides facilities for generating uniformly distributed variates.

Unfortunately, the issue becomes much more complicated when generating autocorrelated variables with given distribution because the mapping $U \mapsto X$ usually will change the autocorrelation pattern of U drastically. Furthermore, it is not an easy task to generate autocorrelated variates which are uniformly distributed. This suggests that the inverse method is not very useful for the problem of generating autocorrelated variates with given marginal distribution, and so other methods are usually applied for that purpose. In particular, such variates are frequently generated by finite order ARMA models, often the AR(1) process, with given distribution of the disturbances. Unfortunately, this method will not result in the distribution aimed at. This can be seen from the following: Consider the linear process defined by

$$Y_t - \mu_Y = \sum_{i=0}^{\infty} \psi_i \delta_{t-i} \quad (2.1)$$

where δ_t are some zero mean independently identically distributed variables and the ψ_i 's are

constants such that $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Without loss of generality, assume that $\mu_Y = 0$ and $\text{Var}(\delta_t) = 0 < \sigma_\delta^2 < \infty$. The sequence (2.1) is known as the infinite order moving average process, also referred to as a linear process, and encompasses all stationary variables according to Wold's decomposition theorem. It is seen directly from (2.1) that because the right hand is the sum of a (possible infinite) number of variables, The left hand side will in general be at least as normally distributed as the δ_t due to the central limit theorem. For example, the AR(1) process defined by

$$Y_t = \phi Y_{t-1} + \delta_t \quad (2.2)$$

is a special case of (2.1) where $\psi_i = \phi^i$. Then it follows that the distribution of Y_t will be more normal when compared to δ_t for $\phi \neq 0$. Now, it is not generally true that $|\psi_i| > 0 \forall_i$. An example is the finite order MA(q) model.

However, it is readily seen that already the skewness of an MA(1) process is closer to that of a normal distribution when compared to the skewness of δ . In fact, it may be shown that for a process defined by

$$Y_t = \delta_t + \theta \delta_{t-1} \quad (2.3)$$

the skewness of Y_t is given by

$$\begin{aligned} \beta_{1,Y} &:= E[Y - E[Y]]^3 / \left(E[Y - E[Y]]^2 \right)^{3/2} \\ &= \beta_{1,\delta} \frac{(1 + \theta^3)}{(1 + \theta^2)^{3/2}} \end{aligned}$$

where $\beta_{1,\delta}$ is the skewness coefficient of δ (see Appendix). Hence, if for example $\beta_{1,\delta} = 5$ and $\theta = -0.7$ the skewness becomes

$\beta_{1,Y} = 5(1-0.7^3) / (1+0.7^2)^{3/2} \approx 1.8$, which is less than half that of $\beta_{1,\delta}$. Thus, if one wishes to investigate the power properties of an autocorrelation test when applied to non-normal data, and applies model (2.3) for different values of θ , then the effect of non-normality will be confounded with the power properties because the skewness is a direct function of the autocorrelation.

In light of the above discussion, one may wonder how autocorrelated and non-normal variables should be generated then. The fact that autocorrelation in general will smooth out non-normality suggests that autocorrelated normally distributed variables should be generated at a first step, and the non-normality should be imposed in the second step. Furthermore, the transformation to non-normality should result in a simple relation between the autocorrelation of the original variable and that of the transformed, thus allowing for total control of the autocorrelation pattern. In the following, that principle will be applied in a series of theorems that describe the generation of the variables and its properties.

Theorem 1.

Let $Y_t = \phi Y_{t-1} + \delta_t$ where $\delta_t \stackrel{iid}{\sim} N(0,1)$, $|\phi| < 1$, and define $\tilde{Y}_t := Y_t / \sqrt{\sigma_Y^2}$ where $\sigma_Y^2 = V(Y) = 1/(1-\phi^2)$. Then $\tilde{Y}_t^2 \sim \chi_{(1)}^2$ independently of the value of ϕ , and the autocorrelations of $\tilde{Y}_t^2$ is given by $\rho_{\tilde{Y}^2}(k) = \phi^{2k}$.

Proof of Theorem 1:

The variance of the AR(1) process is well known to be given by $\sigma_Y^2 = 1/(1-\phi^2)$ and hence $\tilde{Y}_t \sim N(0,1)$ and the chi-square distribution of $\tilde{Y}_t^2$ follows directly. The autocovariances $\gamma_{\tilde{Y}^2}(k)$ of $\tilde{Y}^2$ are given by:

$$\begin{aligned} \gamma_{\tilde{Y}^2}(k) &= E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] - E[\tilde{Y}_t^2] E[\tilde{Y}_{t-k}^2] \\ &= \sigma_Y^{-2,2} E[Y_t^2 Y_{t-k}^2] - 1 \cdot 1 \\ &= \sigma_Y^{-2,2} E[(\phi Y_{t-1} + \varepsilon_t)^2 Y_{t-k}^2] - 1 \\ &= \sigma_Y^{-2,2} E[(\phi^2 Y_{t-1}^2 + 2\phi Y_{t-1} \varepsilon_t + \varepsilon_t^2) Y_{t-k}^2] - 1 \\ &= E \left[\frac{\phi^2 \sigma_Y^{-2} Y_{t-1}^2 \sigma_Y^{-2} Y_{t-k}^2}{+ 2\phi \sigma_Y^{-2} Y_{t-1} \varepsilon_t \sigma_Y^{-2} Y_{t-k}^2 + \sigma_Y^{-2} \varepsilon_t^2 \sigma_Y^{-2} Y_{t-k}^2} \right] - 1 \\ &= \phi^2 E[\tilde{Y}_{t-1}^2 \tilde{Y}_{t-k}^2] \\ &\quad + \left\{ -\phi^2 E[\tilde{Y}_{t-1}^2] E[\tilde{Y}_{t-k}^2] + \phi^2 E[\tilde{Y}_{t-1}^2] E[\tilde{Y}_{t-k}^2] \right\} \\ &\quad + 0 + \sigma_Y^{-2} - 1 \\ &= \phi^2 (E[\tilde{Y}_{t-1}^2 \tilde{Y}_{t-k}^2] - E[\tilde{Y}_{t-1}^2] E[\tilde{Y}_{t-k}^2]) \\ &\quad + \phi^2 E[\tilde{Y}_{t-1}^2] E[\tilde{Y}_{t-k}^2] + \sigma_Y^{-2} - 1 \\ &= \phi^2 \gamma_{\tilde{Y}^2}(k-1) + \phi^2 \cdot 1 \cdot 1 + \frac{1}{(1/(1-\phi^2))} - 1 \\ &= \phi^2 \gamma_{\tilde{Y}^2}(k-1). \end{aligned}$$

Furthermore,

$$\begin{aligned} \gamma_{\tilde{Y}^2}(0) &= E[\tilde{Y}_t^2 \tilde{Y}_t^2] - E[\tilde{Y}_t^2] E[\tilde{Y}_t^2] \\ &= 3 - 1 \cdot 1 \\ &= 2 \end{aligned}$$

By using recursion, it follows that that $\gamma_{\tilde{Y}^2}(k) = 2\phi^{2k}$. The autocorrelation function of $\tilde{Y}_t^2$ is thus determined by $\rho_{\tilde{Y}^2}(k) = \gamma_{\tilde{Y}^2}(k) / \gamma_{\tilde{Y}^2}(0) = \phi^{2k}$ and Theorem 1 follows. In other words, the autocorrelation behaves like that of an AR(1) process with autoregressive parameter ϕ^2 while the distribution is $\chi_{(1)}^2$. Also note that the shape property is independent of the autoregressive parameter.

Now, this variable is highly skewed and may not be appropriate in situations where near-normal distributions are required. The next

theorem proposes a general $\chi_{(r)}^2$ distributed variable (which limits a normal distribution as r increases) with non-confounded autocorrelation:

Theorem 2.

Let $Z_t := \sum_{j=1}^r \tilde{Y}_{jt}^2$ where $\tilde{Y}_{jt}$, $j=1,2,\dots,r$ are mutually independent variables defined as in Theorem 1 with common autocorrelation parameter ϕ . Then the skewness and kurtosis of Z_t are given by $\beta_{1,Z} = \sqrt{8/r}$ and $\beta_{2,Z} = 3 + 12/r$ respectively, independently of ϕ , and the autocorrelation of Z_t is given by $\rho_Z(k) = \phi^{2k}$, independently of r .

Proof of Theorem 2:

The chi-square distribution follows directly from the fact that a sum of r independent $\chi_{(1)}^2$ variates is distributed as $\chi_{(r)}^2$. The skewness and kurtosis of such variates are well known and can be found in Johnson et al (1994). The autocovariance is obtained by using the property $E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] - 1 = \gamma_{\tilde{Y}^2}(k)$ (given in the derivation of Theorem 1):

$$\begin{aligned} \gamma_Z(k) &:= E[Z_t Z_{t-k}] - E[Z_t] E[Z_{t-k}] \\ &= E\left[\sum_{i=1}^r \tilde{Y}_{i,t}^2 \sum_{j=1}^r \tilde{Y}_{j,t-k}^2\right] - E\left[\sum_{i=1}^r \tilde{Y}_{i,t}^2\right] E\left[\sum_{j=1}^r \tilde{Y}_{j,t-k}^2\right] \text{ given by:} \\ &= E\left[\sum_{i=1}^r \tilde{Y}_{i,t}^2 \tilde{Y}_{i,t-k}^2 + \sum_{i \neq j}^r \tilde{Y}_{i,t}^2 \tilde{Y}_{j,t-k}^2\right] - r \cdot r \\ &= \sum_{i=1}^r E[\tilde{Y}_{i,t}^2 \tilde{Y}_{i,t-k}^2] + \sum_{i \neq j}^r E[\tilde{Y}_{i,t}^2] E[\tilde{Y}_{j,t-k}^2] \\ &= r E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] + r(r-1) - r^2 \\ &= r(E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] - 1) \\ &= r \gamma_{\tilde{Y}^2}(k) \\ &= r 2 \phi^{2k}. \end{aligned}$$

In particular, $\gamma_Z(0) = r 2 \phi^0$ and so $\rho_Z(k) = \phi^{2k}$ as was to be shown. In other words, the shape-part of the distribution of Z_t is

$\chi_{(r)}^2$ and the autocorrelation-part of the distribution is $\rho_Z(k) = \phi^{2k}$, and none of the effects is confounded to the other. Thus, the skewness and kurtosis can be determined over an arbitrary (though discrete) range of values independently of the autocorrelation. The next theorem proposes a symmetric non-normal variable with non-confounded autocorrelation:

Theorem 3.

Let Z_t be defined as in Theorem 2. Also, let $c_t = \phi c_{t-1} + \varepsilon_t$ where $\varepsilon_t \stackrel{iid}{\sim} N(0,1)$, and define $\tilde{c}_t := c_t / \sqrt{1/(1-\phi^2)}$ and $W_t := \tilde{c}_t (Z_t - r)$. Then the skewness and kurtosis of W_t are given by $\beta_{1,W} = 0$ and $\beta_{2,W} = 3(3 + 12/r)$ independently of ϕ while the autocorrelations are given by ϕ^{3k} , independently of r .

Proof of Theorem 3:

Firstly, note that $\tilde{c}_t$ and $(Z_t - r)$ are mutually independent and that $E[W_t] = 0$. Hence, the skewness and kurtosis of W_t are

$$\begin{aligned} \beta_{1,W} &= E[W_t^3] / E[W_t]^{3/2} \\ &= E[\tilde{c}_t^3] E[(Z - r)^3] / E[W_t]^{3/2} \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} \beta_{2,W} &= E[W_t^4] / (E[W_t^2])^2 \\ &= E[\tilde{c}_t^4] E[Z_t^4] / (E[\tilde{c}_t^2] E[Z_t^2])^2 \\ &= \{E[\tilde{c}_t^4] / (E[\tilde{c}_t^2])^2\} \{E[Z_t^4] / (E[Z_t^2])^2\} \\ &= 3\{3 + 12/r\}. \end{aligned}$$

Hence, the W variable is symmetric with kurtosis determined by r with range $9 < \beta_{2,W} \leq 45$. Furthermore, by using the results of the proofs of Theorem 2 and Theorem 3, the autocovariances may be obtained:

$$\begin{aligned}
 \gamma_W(k) &= E[W_t W_{t-k}] - 0 \\
 &= E[\tilde{c}_t Z_t \tilde{c}_{t-k} Z_{t-k}] \\
 &= E[\tilde{c}_t \tilde{c}_{t-k}] E[(Z_t - r)(Z_{t-k} - r)] \\
 &= \phi^k E\{(Z_t Z_{t-k}) - r Z_t - r Z_{t-k} + r^2\} \\
 &= \phi^k \{E(Z_t Z_{t-k}) - r^2\} \\
 &= \phi^k \{r E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] + r(r-1) - r^2\} \\
 &= \phi^k r \{E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] - 1\} \\
 &= \phi^k r \gamma_{\tilde{Y}^2}(k) \\
 &= \phi^k 2r \phi^{2k}.
 \end{aligned}$$

Hence, $\gamma_W(k) = 2r\phi^{3k}$ and in particular $\gamma_W(0) = 2r$. Thus, the autocorrelations are given by $\rho_W(k) = 2r\phi^{3k}/2r = \phi^{3k}$ independently of r as was to be shown. Thus, Theorem 3 provides a symmetric but non-normal variable where the autocorrelations can be identical to those of an AR(1) process by putting the autocorrelation parameter of the original variable equal to $\phi^{1/3}$.

In general, the variables presented in Theorems 1-3 share the property that they all have autocorrelations that decay slowly in the sense that they are non zero at all lags if $\phi \neq 0$. In many instances it is of interest to generate autocorrelations that are zero above a certain lag. Therefore, some short memory processes of MA(1) type will also be proposed. These are presented below:

Theorem 4.

Let $Y_t = \delta_t - \theta \delta_{t-1}$ where $\delta_t \stackrel{iid}{\sim} N(0,1)$ and define $\tilde{Y}_t := Y_t / \sqrt{\sigma_Y^2}$ where $\sigma_Y^2 = 1 + \theta^2$.

Then, $\tilde{Y}_t^2 \sim \chi_{(1)}^2$ independently of ϕ and the autocorrelations are given by $\rho_{\tilde{Y}^2}(1) = \theta^2 / (1 + \theta^2)^2$ $k=1$ and $\rho_{\tilde{Y}^2}(k) = 0$ $k > 1$ independently of r .

Proof of Theorem 4.

The chi-square distribution follows trivially as $\tilde{Y}$ is a standard normally distributed variate. The second order moments are given by

$$\begin{aligned}
 \gamma_{\tilde{Y}^2}(0) &= V(\tilde{Y}_t^2) = 2. \\
 \gamma_{\tilde{Y}^2}(1) &= E[\tilde{Y}_t^2 \tilde{Y}_{t-1}^2] - E[\tilde{Y}_t^2] E[\tilde{Y}_{t-1}^2] \\
 &= \left(1/(1 + \theta^2)^2\right) E[Y_t^2 Y_{t-1}^2] - 1 \cdot 1 \\
 &= \left(1/(1 + \theta^2)^2\right) E[(\delta_t - \theta \delta_{t-1})^2 (\delta_{t-1} - \theta \delta_{t-2})^2] - 1 \\
 &= \left(1/(1 + \theta^2)^2\right) \left\{ E[\delta_t^2 \delta_{t-1}^2] + E[\delta_t^2 \theta^2 \delta_{t-2}^2] \right. \\
 &\quad \left. + E[\theta^2 \delta_{t-1}^4] + E[\theta^4 \delta_{t-1}^2 \delta_{t-2}^2] \right\} - 1 \\
 &= (1 + 4\theta^2 + \theta^4) / (1 + \theta^2)^2 - 1 \\
 &= 2\theta^2 / (1 + \theta^2)^2.
 \end{aligned}$$

An analogous proof reveals that $\gamma_{\tilde{Y}^2}(k) = 0$ if $k > 1$. Hence the

autocorrelations of $\tilde{Y}_t^2$ are given by $\rho_{\tilde{Y}^2}(1) = 2\theta^2 / 2(1 + \theta^2)^2 = \theta^2 / (1 + \theta^2)^2$ $k=1$ and $\rho_{\tilde{Y}^2}(k) = 0$ $k > 1$ as was to be shown.

Thus, the autocorrelations of $\tilde{Y}_t^2$ are those of an MA(1) process where the autocorrelation at lag 1 equals the root of $\theta^2 / (1 + \theta^2)^2$ which, in turn, is bounded between 0 and 0.25 (the maximum being reached at $\theta = 1$). This variable may also be extended to an arbitrary $\chi_{(r)}^2$ distribution:

Theorem 5.

Let $Z_t := \sum_{j=1}^r \tilde{Y}_{jt}^2$ where $\tilde{Y}_{jt}, j=1,2,\dots,r$ are mutually independent variables defined as in Theorem 4 with common autoregressive parameter θ . Then the skewness and kurtosis of Z_t are given by $\beta_{1,Z} = \sqrt{8/r}$ and $\beta_{2,Z} = 3 + 12/r$ independently of θ , and the autocorrelations of Z are given by $\rho_{\tilde{Y}_t^2}(1) = \theta^2 / (1 + \theta^2)^2$ $k=1$ and $\rho_{\tilde{Y}_t^2}(k) = 0$ $k > 1$, independently of r .

Proof of Theorem 5.

The skewness and kurtosis are motivated in Theorem 2. Analogous to the proof of Theorem 2, $\gamma_Z(k) = r(E[\tilde{Y}_t^2 \tilde{Y}_{t-k}^2] - 1)$, and as the variance is given by $\gamma_{Z_t}(0) = \text{Var}[Z_t] = 2r$ it follows that the autocorrelations are given by

$$\begin{aligned} \rho_{\tilde{Y}_t^2}(1) &= r \left\{ 2\theta^2 / (1 + \theta^2)^2 \right\} / 2r \\ &= \theta^2 / (1 + \theta^2)^2 \quad k=1 \end{aligned}$$

and $\rho_{\tilde{Y}_t^2}(k) = 0$ $k > 1$ as was to be shown. The Z variable may also be symmetrized according to the following:

Theorem 6.

Let Z_t be defined as in Theorem 5. Also, let $c_t = \varepsilon_t - \theta \varepsilon_{t-1}$ where $\varepsilon_t \sim N(0,1)$, and define $\tilde{c}_t := c_t / \sqrt{1 + \theta^2}$ and $W_t := \tilde{c}_t (Z_t - r)$. Then, the skewness and kurtosis of W_t are given by $\beta_{1,W} = 0$ and $\beta_{2,W} = 3(3 + 12/r)$ independently of θ and the autocorrelations of W_t are given by $\rho_W(1) = -\theta^3 / (1 + \theta^2)^2$, $\rho_W(k) = 0$ $k > 1$ independently of r .

Proof of Theorem 6.

The skewness and kurtosis are given in the proof of Theorem 3. The autocovariance of W_t is

$$\begin{aligned} \gamma_W(1) &= E[W_t W_{t-1}] \\ &= E[\tilde{c}_t \tilde{c}_{t-1}] E[(Z_t - r)(Z_{t-1} - r)] \\ &= -\theta [E(Z_t Z_{t-1}) - r^2] \\ &= -\theta ((\text{Cov}[Z_t Z_{t-1}] + r^2) - r^2) \\ &= -\theta r (E[\tilde{Y}_t^2 \tilde{Y}_{t-1}^2] - 1) \\ &= -\theta r \gamma_{\tilde{Y}_t^2}(k). \end{aligned}$$

Hence,

$$\gamma_W(0) = 2r, \quad \gamma_W(1) = -r\theta \left\{ 2\theta^2 / (1 + \theta^2)^2 \right\},$$

$\gamma_W(k) = 0$ $k > 1$ and the theorem follows.

Hence, the autocorrelations behave like those of an MA(1) process with MA parameter determined by the root of $-\theta^3 / (1 + \theta^2)^2$ which is bounded in the interval $(-0.32, 0.32)$ with maximum at $\theta = \pm 1.73$.

The variables proposed above are all univariate. It will sometimes be of interest to generate multivariate variables with cross correlation between pairwise marginal variables. When imposing such cross correlation one wish to do that in a manner that does not alter the marginal distributions. This can be achieved by letting one or several variables used to form the marginal variables be identical (fixed) in all marginal variables. The next theorem describes such variables and its main properties:

Theorem 7.

Let $\tilde{Y}_{ijt}, i=1,2,\dots,P, j=1,2,\dots,r$ be mutually independent variables defined as in Theorem 1 or Theorem 4, depending on whether the AR(1) or MA(1) process have been used to generate the autocorrelation, with common autocorrelation parameter ϕ (or θ). Then

define $Z_{i,t} := \sum_{j=1}^h \tilde{Y}_{1,j,t}^2 + \sum_{j=h+1}^r \tilde{Y}_{i,j,t}^2$,
 $i = 1, 2, \dots, P$ and let $\mathbf{Z}_t = [Z_{1,t}, \dots, Z_{P,t}]$ be a
 random vector of the marginal variables $Z_{i,t}$.
 Then, the cross correlations between two
 marginal variables of $\mathbf{Z}_t$ are given by

$$\text{Corr}(Z_{i,t}, Z_{i',t}) = h/r, \quad i \neq i'.$$

Proof of Theorem 7.

Assume for the moment that $h=1$.
 Then, on observing that each $Z_{i,t}$ is a chi square
 variate, the covariance becomes:

$$\begin{aligned} & \text{Cov}(Z_{i,t}, Z_{i',t}) \\ &= E[Z_{i,t} Z_{i',t}] - E[Z_{i,t}] E[Z_{i',t}] \\ &= E\left[\left(\tilde{Y}_{1,1,t}^2 + \sum_{j=2}^r \tilde{Y}_{i,j,t}^2\right)\left(\tilde{Y}_{1,1,t}^2 + \sum_{j=2}^r \tilde{Y}_{i',j,t}^2\right)\right] - r^2 \\ &= E\left[\begin{aligned} & \left(\tilde{Y}_{1,1,t}^2\right)^2 + \tilde{Y}_{1,1,t}^2 \sum_{j=2}^r \tilde{Y}_{i',j,t}^2 \\ & + \left(\tilde{Y}_{i,2,t}^2 \tilde{Y}_{1,1,t}^2 + \tilde{Y}_{i,2,t}^2 \sum_{j=2}^r \tilde{Y}_{i',j,t}^2\right) + \\ & \dots + \left(\tilde{Y}_{i,r,t}^2 \tilde{Y}_{1,1,t}^2 + \tilde{Y}_{i,r,t}^2 \sum_{j=2}^r \tilde{Y}_{i',j,t}^2\right) \end{aligned}\right] - r^2 \\ &= 3 + 1 \cdot (r-1) + (1 \cdot 1 + 1(r-1)) + \\ & \dots + (1 \cdot 1 + 1(r-1)) - r^2 \\ &= 3 + (r-1) + r + \dots + r - r^2 \\ &= 2. \end{aligned}$$

Because the marginal $Z_{i,t}$ variables are
 distributed as $\chi_{(r)}^2$ it follows that the correlation
 is given by $\text{Corr}(Z_{i,t}, Z_{i',t}) = 2/2r = 1/r$. By
 applying an analogous proof for a general
 $0 \leq h \leq r$ it is seen that $\text{Cov}(Z_{i,t}, Z_{i',t}) = 2h$
 and hence the cross correlation is given by
 $\text{Corr}(Z_{i,t}, Z_{i',t}) = 2h/2r = h/r$ which
 completes the proof.

In other words, Theorem 7 proposes a
 random vector with marginal variables of the

kind described in Theorem 2 (or Theorem 5)
 though with cross correlations given by h/r . It
 is also possible to generate symmetric
 multivariate variables as shown in the next
 theorem:

Theorem 8.

Let Z_t be defined as in Theorem 7 and
 let $W_t := \tilde{c}_t(Z_t - r)$ where $\tilde{c}_t$ is defined as in
 Theorem 3 (or Theorem 6). Then the cross
 correlation between $W_{i,t}$ and $W_{i',t}$ is given by
 $\text{Corr}(W_{i,t}, W_{i',t}) = h/r$ independently of ϕ (or
 θ).

Proof of Theorem 8.

The cross covariance is given by

$$\begin{aligned} & \text{Cov}(W_{i,t}, W_{i',t}) \\ &= E(W_{i,t}, W_{i',t}) - 0 \\ &= E(\tilde{c}_t(Z_{i,t} - r) \tilde{c}_t(Z_{i',t} - r)) \\ &= E[\tilde{c}_t^2] E((Z_{i,t} - r)(Z_{i',t} - r)) \\ &= E(Z_{i,t} Z_{i',t}) - r^2 \\ &= (2h + r^2) - r^2 \\ &= 2h. \end{aligned}$$

Finally,

$$\begin{aligned} & \text{Var}(W_{i,t}) \\ &= E(W_{i,t}^2) - E^2(W_{i,t}) \\ &= E(\tilde{c}_t^2 (Z_{i,t} - r)^2) - 0 \\ &= E(\tilde{c}_t^2) E(Z_{i,t} - r)^2 \\ &= 1 \cdot \text{Var}(Z_{i,t}) \\ &= 2r. \end{aligned}$$

This completes the proof.

The variables proposed above allow for
 separate control of shape properties and
 autocorrelations. In particular, the variables
 formed by transformations of AR(1) processes

(those of Theorems 1-3) behave like AR(1) processes and the domain of the autoregressive parameter remains $-1 < \phi < 1$. In other words these variables provide a tool for generating AR(1) processes with non-normal distributions. On the other hand, the domains of the autoregressive parameter (which is equal to the whole real line though usually kept in the span $-1 < \theta < 1$) of the MA(1) type variables (those of Theorems 4-6) do not remain when transformed to non-normality.

This might be a drawback in some instances though they do provide a tool for investigating the effect of non-normal short memory processes. In general, the proposed variables should cover most robustness problems met in practice. Further research in the matter could involve the development of generating non-normal processes with long memory of ARFIMA type with autocorrelations non-confounded with shape properties. Other relevant issues involve non-stationary processes with given shape properties. This, however, is beyond the scope of this article and is left for future studies.

Conclusion

In this article, it is argued that care must be taken when simulating autocorrelated and non-normal variables so that the autocorrelation is not a function of the shape property and vice versa. Furthermore, a number of random variables specially designed for simulation studies concerning shape properties and autocorrelation are proposed. The variables involve univariate or multivariate distributions, which are symmetric or skewed and have short memory or long memory. Thus, they cover a fairly wide range of applications. Furthermore, all variables are easily generated in any standard statistical software packages that have facilities to generate AR or MA processes.

References

- Bai, J. & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1), 49-60.
- Johnson, M. E. (1987). *Multivariate statistical simulation*. Wiley.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions*, Vol. 1. Wiley.
- Shukur, G. (2000). The robustness of the systemwise Breach-Godfrey auto-correlation test for non-normal error terms. *Communications in Statistics, Simulation, & Computation*, 29(2), 419-448.

Appendix

Let $\mu_{X,h} := E[X_t^h]$ $\mu_{\varepsilon,h} := E[\varepsilon_t^h]$ for $h = 1, 2, 3$. Then, if ε_t is a sequence of iid (zero mean) uniformly integrable variables, the following holds:

$$\begin{aligned}\mu_{X,2} &= E\left[\left(\sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}\right)^2\right] \\ &= E\left[\sum_{i=0}^{\infty} \theta_i^2 \varepsilon_{t-i}^2 + \sum \sum_{i \neq j} \theta_i \varepsilon_{t-i} \theta_j \varepsilon_{t-j}\right] \\ &= \sum_{i=0}^{\infty} \theta_i^2 E[\varepsilon_{t-i}^2] \\ &= \mu_{\varepsilon,2} \sum_{i=0}^{\infty} \theta_i^2.\end{aligned}$$

$$\begin{aligned}\mu_{X,3} &= E\left[\left(\sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i}\right)^3\right] \\ &= E\left[\sum_{i=0}^{\infty} \theta_i^3 \varepsilon_{t-i}^3 + 3 \sum \sum_{i \neq j} \theta_i^2 \varepsilon_{t-i}^2 \theta_j \varepsilon_{t-j} + \sum \sum \sum_{i \neq j \neq k} \theta_i \varepsilon_{t-i} \theta_j \varepsilon_{t-j} \theta_k \varepsilon_{t-k}\right] \\ &= \sum_{i=0}^{\infty} \theta_i^3 E[\varepsilon_{t-i}^3] \\ &= \mu_{\varepsilon,3} \sum_{i=0}^{\infty} \theta_i^3.\end{aligned}$$

$$\begin{aligned}
\mu_{X,4} &= E \left[\left(\sum_{i=0}^{\infty} \theta_i \varepsilon_{t-i} \right)^4 \right] \\
&= E \left[\sum_{i=0}^{\infty} \theta_i^4 \varepsilon_{t-i}^4 + 4 \sum_{i \neq j} \sum \theta_i^3 \varepsilon_{t-i}^3 \theta_j \varepsilon_{t-j} + \right. \\
&\quad \left. 3 \sum_{i \neq j} \sum \theta_i^2 \varepsilon_{t-i}^2 \theta_j^2 \varepsilon_{t-j}^2 + 6 \sum_{i \neq j \neq k} \sum \theta_i^2 \varepsilon_{t-i}^2 \theta_j \varepsilon_{t-j} \theta_k \varepsilon_{t-k} \right. \\
&\quad \left. + 6 \sum_{i \neq j \neq k \neq l} \sum \theta_i^2 \varepsilon_{t-i}^2 \theta_j \varepsilon_{t-j} \theta_k \varepsilon_{t-k} \theta_l \varepsilon_{t-l} \right] \\
&= \sum_{i=0}^{\infty} \theta_i^4 E[\varepsilon_{t-i}^4] \\
&\quad + 3 \sum_{i \neq j} \sum \theta_i^2 E[\varepsilon_{t-i}^2] \theta_j^2 E[\varepsilon_{t-j}^2] \\
&= \mu_{\varepsilon,4} \sum_{i=0}^{\infty} \theta_i^4 + 3 \mu_{\varepsilon,2}^2 \sum_{i \neq j} \sum \theta_i^2 \theta_j^2.
\end{aligned}$$

Hence, if an MA(1) process is determined by

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad (\text{A1})$$

it follows that the skewness is given by

$$\begin{aligned}
\beta_{1,X} &= \frac{\mu_{X,3}}{\mu_{X,2}^{3/2}} \\
&= \frac{\mu_{\varepsilon,3} \sum_{i=0}^{\infty} \theta_i^3}{\left(\mu_{\varepsilon,2} \sum_{i=0}^{\infty} \theta_i^2 \right)^{3/2}} \\
&= \frac{\mu_{\varepsilon,3} (1^3 + \theta^3)}{\mu_{\varepsilon,2}^{3/2} (1^2 + \theta^2)^{3/2}} \\
&= \beta_{1,\varepsilon} \frac{(1 + \theta^3)}{(1 + \theta^2)^{3/2}}
\end{aligned} \quad (\text{A2})$$

Because $\frac{(1 + \theta^3)}{(1 + \theta^2)^{3/2}}$ is bounded in the span $(-1, 1)$ it follows that $|\beta_{1,X}| \leq |\beta_{1,\varepsilon}|$.

Furthermore, the kurtosis of the MA(1) process is given by

$$\begin{aligned}
\beta_{2,X} &= \frac{\mu_{X,4}}{\mu_{X,2}^2} \\
&= \frac{\mu_{\varepsilon,4} \sum_{i=0}^{\infty} \theta_i^4 + 3 \mu_{\varepsilon,2}^2 \sum_{i \neq j} \sum \theta_i^2 \theta_j^2}{\left(\mu_{\varepsilon,2} \sum_{i=0}^{\infty} \theta_i^2 \right)^2} \\
&= \frac{\mu_{\varepsilon,4} \sum_{i=0}^{\infty} \theta_i^4}{\left(\mu_{\varepsilon,2} \sum_{i=0}^{\infty} \theta_i^2 \right)^2} + \frac{3 \mu_{\varepsilon,2}^2 \sum_{i \neq j} \sum \theta_i^2 \theta_j^2}{\left(\mu_{\varepsilon,2} \sum_{i=0}^{\infty} \theta_i^2 \right)^2} \\
&= \frac{\mu_{\varepsilon,4} (1 + \theta^4)}{\left(\mu_{\varepsilon,2} (1 + \theta^2) \right)^2} + 3 \mu_{\varepsilon,2}^2 \frac{(1 \cdot \theta^2 + \theta^2 \cdot 1)}{\left(\mu_{\varepsilon,2} (1 + \theta^2) \right)^2} \\
&= \frac{\mu_{\varepsilon,4} (1 + \theta^4)}{\mu_{\varepsilon,2}^2 (1 + \theta^2)^2} + 3 \mu_{\varepsilon,2}^2 \frac{2\theta^2}{\mu_{\varepsilon,2}^2 (1 + \theta^2)^2} \\
&= \beta_{2,\varepsilon} \frac{(1 + \theta^4)}{(1 + \theta^2)^2} + 3 \frac{2\theta^2}{(1 + \theta^2)^2}.
\end{aligned} \quad (\text{A3})$$

Note that (A3) is 3 if $\beta_{2,\varepsilon} = 3$. Furthermore, it may be shown that the kurtosis of $\beta_{2,X}$ is always closer to the kurtosis of the normal distribution when compared to $\beta_{2,\varepsilon}$, i.e. $|\beta_{2,X} - 3| \leq |\beta_{2,\varepsilon} - 3|$.

Restricted Quasi-Independent Model Resolves Paradoxical Behaviors of Cohen's Kappa

Vicki Stover Hertzberg Frank Xu Michael Haber
Department of Biostatistics
Emory University

Cohen's kappa, an index of inter-rater agreement, behaves paradoxically in 2×2 tables. λ_A is derived, an index from the restricted quasi-independent model for 2×2 tables. Simulation studies are used to demonstrate λ_A has superior performance compared to Scott's pi. Moreover, λ_A does not show paradoxical behavior for 2×2 tables.

Keywords: Quasi-independent model; Cohen's kappa; Scott's pi; inter-rater agreement

Introduction

In clinical trials and epidemiology studies, agreement studies are often conducted in order to assess and characterize the extent to which two sets of measurements on the same unit of observation agree. Examples of such studies include when raters examine a group of subjects to determine the presence or absence of a trait, sort them into previously arranged categories, or rate them according to a previously defined scale. Examples of areas in which rater variability is of concern include the interpretations of image results in radiology, diagnoses made on the basis of laboratory measurements, or psychiatric evaluations. Data from a study in which raters A and B classify $N_{..}$ subjects into k categories are the counts $\{N_{ij}; i=1, \dots, k; j=1, \dots, k\}$ where N_{ij} is the number of

subjects that are simultaneously classified as category i by rater A and category j by rater B.

A variety of measures are available to assess the extent of agreement between ratings. Because some agreement can be expected merely due to chance, an important consideration in selecting such a measure is whether or not it is a chance-corrected index. The more popular indices that are chance-corrected include the S statistic (Bennett, et al., 1954), Scott's pi (Scott, 1955), and Cohen's kappa (Cohen, 1960). Among these measures, Cohen's kappa is a popular choice, due to its intuitive means for correcting for chance. The population value for Cohen's kappa can be written as

$$\kappa = \frac{\pi_0 - \pi_e}{1 - \pi_e} \quad (1)$$

where π_0 is the proportion of observed agreement and π_e is the proportion of agreement expected by chance alone. Cohen's kappa, κ , is estimated in the $k \times k$ table as

$$\hat{\kappa} = \frac{\sum_{i=1}^k \frac{N_{ii}}{N_{..}} - \sum_{i=1}^k \frac{N_{i.} N_{.i}}{N_{..}^2}}{1 - \sum_{i=1}^k \frac{N_{i.} N_{.i}}{N_{..}^2}} \quad (2)$$

Vicki Stover Hertzberg has worked in academic biostatistics at University of Cincinnati and Emory University in the Department of Biostatistics. Email: vhertz@sph.emory.edu. Frank Xu works as a statistician for Spectrx, a medical device manufacturing firm in the Atlanta metropolitan area. Email: fxu1967@hotmail.com. Michael Haber is in the Department of Biostatistics. Email: mhaber@sph.emory.edu. This work was supported in part by US Public Health Service grant 1 R01 MH070028-01A1.

where

$$N_{i.} = \sum_{j=1}^k N_{ij},$$

$$N_{.j} = \sum_{i=1}^k N_{ij},$$

and

$$N_{..} = \sum_{i=1}^k \sum_{j=1}^k N_{ij}.$$

There are a variety of generalizations of Cohen's kappa, such as versions that are weighted for ordinal scale assessments as well as versions for use in the assessment of multi-rater agreement. In this article, the discussion is confined to the assessment of the agreement and disagreement between two raters in a simple square contingency table.

Despite its popularity as an index of agreement, Cohen's kappa exhibits paradoxical behaviors in 2×2 tables (Feinstein & Cicchetti, 1990). For a given 2×2 table, the marginal probabilities are called symmetrical if either $(N_{1.}/N_{..} \geq 0.5 \text{ and } N_{.1}/N_{..} \geq 0.5)$ or $(N_{1.}/N_{..} \leq 0.5 \text{ and } N_{.1}/N_{..} \leq 0.5)$. The marginal probabilities are called balanced if both $N_{1.}/N_{..}$ and $N_{.1}/N_{..}$ are close to 0.5. One such paradox is that κ estimated for a table with symmetrical unbalanced marginal probabilities can be substantially less than κ estimated for a table with symmetrical balanced marginal probabilities although both tables have the same amount of observed agreement. In addition, a table with asymmetrical unbalanced marginal probabilities will have larger estimated κ than a table with symmetrical unbalanced marginal probabilities even though the observed agreement is the same, the second paradox.

Several authors (Brennan, et al., 1981, Cicchetti & Feinstein, 1990, Lantz & Nebenzahl, 1996, Byrt, et al., 1993) investigated this problematic behavior. They have suggested companion statistics to be reported along with Cohen's kappa; however these companion statistics are not model-based and are arbitrary in the treatment of the correction for chance.

Thus an alternative index which does not exhibit such paradoxical behaviors is desirable. The use of a measure of agreement is

explored; λ_A , derived from the quasi-independent (QI) model (Goodman, 1968). The QI model was developed for application to $k \times k$ tables, specifically for the analysis of truncated tables (i.e., tables with missing entries due to study design or data collection). One limitation of the QI model is that it is not directly applicable to 2×2 tables. This limitation is due to lack of degrees of freedom. In this article, a restricted QI model for interrater agreement that allows for rater bias in 2×2 tables is examined. The introduction of the restriction allows us to overcome the problem with degrees of freedom.

The notion of quasi-independence assumes that a sub-table, which is part of the whole table, is independent (Bishop, et al., 1975, Agresti, 1990). A two-dimensional table is said to be QI if for a subset of cells U there exist constants p_{ri} and p_{cj} such that the probability of cell (i,j) given it is in U equals $p_{ri}p_{cj}$. The remaining cells are in U^* .

Guggermoos-Holzman and Vonk (1998) showed that the QI model is related to latent class models. This relationship is exploited to apply the QI concept to the context of rater agreement studies. Suppose that there are two groups of subjects (latent classes) to be classified into k categories. Group 1 is systematically classified by all raters. If the raters agree on the classification then systematic agreement is said to have occurred; otherwise systematic disagreement has resulted due to the use of different classification rules by the raters.

For Group 1 subjects the classifications by the raters are not made independently, thus they contribute only to U^* , the set of cells with systematic agreement or disagreement. Group 2 comprises subjects for whom at least one rater randomly classifies according to a multinomial distribution, that is, the raters classify these subjects with independent marginal probabilities p_{ri} and p_{cj} respectively. Group 2 subjects contribute to the frequencies of all cells in the table. To illustrate this concept, consider Table 1. In this scenario, raters A and B classify 100 subjects into three categories. Unbeknownst to A and B there are 80 subjects in Group 1 and 20 in

Table 1. Illustrative Example of Quasi-independent Data.

Easy to Classify Subjects (Group 1)					+	Difficult to Classify Subjects (Group 2)				=	Whole Table			
Rater A category						Rater A category					Rater A category			
Rater B category	1	2	3	Total		1	2	3	Total		1	2	3	Total
1	25 ¹	5		30		4	4	2	10		29	9	2	40
2		30		30		2	2	1	5		2	32	1	35
3			20	20		2	2	1	5		2	2	21	25
Total	25	35	20	80	8	8	4	20	33	43	24	100		

¹Cells in set U* denoted with boldface.

Group 2. Group 2 classifications are made using independent marginal probabilities of (0.5, 0.25, 0.25) for categories 1, 2, and 3 respectively by rater A and (0.4, 0.4, and 0.2) by rater B. The set U* comprises cells (1,1), (1,2), (2,2), and (3,3). In U*, the cells (1,1), (2,2), and (3,3) represent systematic agreement, while the cell (1,2) represents systematic disagreement. Systematic disagreement may arise when the raters use slightly different rules for classification. In the case of Table 1, the rules used by rater A are such that s/he tends to over-read category 2 subjects versus category 1 in comparison to rater B.

Suppose that λ is the proportion of the population of subjects in Group 1 and $1-\lambda$ is the proportion in Group 2. Thus λ is the total proportion of systematic agreement and disagreement. If cell (i,j) is in U* then define $d_{ij} = 1$, and $d_{ij} = 0$ otherwise. When $i=j$, χ_{ij} is the proportion of systematic agreement and when $i \neq j$, χ_{ij} is the proportion of systematic disagreement, defined only for cells in U*. For

each cell (i,j) let $\frac{d_{ij}\chi_{ij}}{\lambda}$ be the conditional probability that a subject is classified into that cell given that it is in Group 1.

Thus $\lambda = \sum_{i=1}^k \sum_{j=1}^k d_{ij}\chi_{ij}$. By the total probability

theorem, the probability of cell (i,j) can be written as

$$\pi_{ij} = (1-\lambda)p_{r_i}p_{c_j} + d_{ij}\chi_{ij}. \quad (3)$$

One may solve for λ by multiplying both sides of equation (3) by d_{ij} and summing over all cells obtaining

$$\lambda = \frac{\sum_{i=1}^k \sum_{j=1}^k d_{ij}\pi_{ij} - \sum_{i=1}^k \sum_{j=1}^k d_{ij}p_{r_i}p_{c_j}}{1 - \sum_{i=1}^k \sum_{j=1}^k d_{ij}p_{r_i}p_{c_j}}. \quad (4)$$

Note the similarity of λ to the formulation of a chance-corrected agreement index. In this formulation the terms d_{ij} are terms that must be specified before any further calculations can be made. There are k^2-1 degrees of freedom available in the $k \times k$ table, of which $2(k-1)$ are the parameters for the marginal probabilities. Thus, at most $(k-1)^2$ parameters of the d_{ij} can be set to 1 in equation (4). As a result, the QI model can only be used in $k \times k$ tables where $k \geq 3$.

Furthermore, λ can be expressed as the sum of systematic agreement and disagreement as follows:

$$\lambda = \lambda_A + \lambda_D = \sum_{i=1}^k d_{ii} \chi_{ii} + \sum_{i=1}^k \sum_{j \neq i}^k d_{ij} \chi_{ij} . \quad (5)$$

The log-likelihood of the general QI model then is given by

$$\ln(l) = \sum_{i=1}^k \sum_{j=1}^k N_{ij} \ln[(1-\lambda)p_{r_i}p_{c_j} + d_{ij}\chi_{ij}]. \quad (6)$$

The unknown parameters are p_{ri} , p_{cj} , and χ_{ij} , with $\lambda = \sum \chi_{ij}$.

The following iterative procedure may be used to derive the maximum likelihood estimates for the model:

$$\begin{aligned} \hat{p}_{r_i}^{(l)} &= \frac{p_{i\cdot} - \sum_{j=1}^k d_{ij} \hat{\chi}_{ij}^{(l)}}{1 - \hat{\lambda}^{(l)}} \\ \hat{p}_{c_j}^{(l)} &= \frac{p_{\cdot j} - \sum_{i=1}^k d_{ij} \hat{\chi}_{ij}^{(l)}}{1 - \hat{\lambda}^{(l)}} \end{aligned} \quad (7)$$

where

$$\hat{\chi}_{ij}^{(l)} = d_{ij} [p_{ij} - (1 - \hat{\lambda}^{(l)}) \hat{p}_{r_i}^{(l)} \hat{p}_{c_j}^{(l)}],$$

and $p_{i\cdot}$ and $p_{\cdot j}$ are the observed marginal probabilities. Initial values are $\hat{\lambda}^{(0)} = 0$, and $\hat{p}_{r_i}^{(0)}$ and $\hat{p}_{c_j}^{(0)}$ are set to the observed marginal probabilities, $i, j = 1, \dots, k$.

To derive these estimates one must set $d_{ij} = 1$ or 0 on the basis of either *a priori* knowledge or using a data driven method. Agresti (1990) assumed $d_{ij} = 1$ for all diagonal cells while Bergan (1980) and Aickin (1990) used a trial and error method to determine d_{ij} from the data.

Some illustrations are explored. Returning to Table 1, it may be seen that $\lambda = 0.8$. The values of χ_{ij} are 0.25, 0.05, 0.30, and 0.20 for cells in U^* and not defined otherwise.

Next, turn to Table 2. If systematic agreement is assumed in every diagonal cell and no systematic disagreement, then $\hat{\lambda}(\text{se}) = \hat{\lambda}_A(\text{se}) = 0.554(0.008)$, compared to $\hat{\kappa}(\text{se}) = 0.493(0.057)$, where standard errors are obtained by bootstrap. A goodness of fit test for this table results in $\chi^2 = 11.7$ with 5 degrees of freedom, $p = 0.039$, giving an indication of lack of fit.

Table 2. Diagnosis of Carcinoma for Pathologists A and B

Classification of Pathologist A	Classification of Pathologist B				
	1	2	3	4 & 5	Total
1	22	2	2	0	26
2	5	7	14	0	26
3	0	2	36	0	38
4 & 5	0	1	17	10	28
Total	27	12	69	10	118

Source: Derived from Landis and Koch (1977) as described in Agresti (1990)

Note the relatively small amount of agreement in cell (2,2) (5.9% of the 118 observations versus 18.6%, 30.5%, and 8.5% in cells (1,1), (3,3), and (4,4) respectively) and the large error frequency in cell (4,3) (14.4%). Setting U^* to contain cells (1,1), (3,3), (4,4), and (4,3), the following are obtained $\hat{\lambda}(se) = 0.69(0.005)$, with $\hat{\lambda}_A(se) = 0.554(.005)$ and $\hat{\lambda}_D(se) = 0.136(.003)$. $\chi^2 = 2.18$ with 5 degrees of freedom, $p=0.82$, may be further derived from this model, indicating much better fit.

Methodology

For the case of 2×2 tables, it is assumed that $i=1$ or $j=1$ indicates that the prevalent condition is positive. Due to lack of degrees of freedom, it must also be assumed in this case that only agreement is systematic, i.e., there is no systematic disagreement. Thus U^* contains only the two diagonal cells. The model can now be rewritten as

$$\pi_{ij} = (1 - \lambda_A) p_{r_i} p_{c_j} + I_{(i=j)} \chi_{ij} \quad (8)$$

for $i = 1, 2$ and $j = 1, 2$, where $0 \leq p_{r_i} \leq 1$, $p_{r1} + p_{r2} = 1$, $0 \leq p_{c_j} \leq 1$, $p_{c1} + p_{c2} = 1$, $0 \leq \chi_{ii} \leq 1$, and $0 \leq \lambda = \chi_{11} + \chi_{22} \leq 1$.

As mentioned above, there are three degrees of freedom and four parameters: χ_{11} , χ_{22} , p_{r1} , and p_{c1} . If no restriction is placed on the independent marginal probabilities then a restriction must be placed on χ_{11} and χ_{22} .

The common correlation model for Scott's π_i , denoted κ_s , assumes 1) no rater bias and 2) the rater prevalence in Group 1 equals that in Group 2. The second assumption follows from 3) the underlying true prevalence in Group 1 equals that in Group 2 and 4) the common rater prevalence is an unbiased estimator of the true prevalence. Scott's π_i is limited by the assumption of no rater bias, in particular the assumption of no rater bias in Group 2. It is theorized that the two observers are likely to have different rater prevalence's in Group 2. In fact, many agreement studies show evidence of rater bias. Thus, to adequately apply the QI concept to rater agreement in 2×2 tables, one must assume (5) the true prevalence in Group 2

is equal to that in Group 1, and (6) the rater prevalence in Group 2 differs between raters, but the average is an unbiased estimator for the true prevalence.

Under assumption (6), χ_{11}/λ_A can be interpreted as the common rater prevalence in Group 1 since the two raters classify with certainty and agree. Thus, χ_{11}/λ_A is the best estimator of true prevalence. If then one allows for the prevalence for each rater for Group 2 subjects, then $(p_{r1} + p_{c1})/2$ is also an estimator of true prevalence. Under assumption (5) then one has $\chi_{11}/\lambda_A = (p_{r1} + p_{c1})/2$, giving $\chi_{ii} = (p_{r_i} + p_{c_i})\lambda_A/2$. Thus, under assumptions (5) and (6) one has the restricted QI model,

$$\pi_{ij} = (1 - \lambda_A) p_{r_i} p_{c_j} + I_{(i=j)} \frac{p_{r_i} + p_{c_j}}{2} \lambda_A \quad (9)$$

for $i = 1, 2$ and $j = 1, 2$ where $0 \leq p_{r_i} \leq 1$, $p_{r1} + p_{r2} = 1$, $0 \leq p_{c_j} \leq 1$, $p_{c1} + p_{c2} = 1$, $0 \leq \chi_{ii} \leq 1$, and $0 \leq \lambda = \chi_{11} + \chi_{22} \leq 1$. The log-likelihood function for this restricted model is

$$\ln(I) = \sum_{i=1}^2 \sum_{j=1}^2 N_{ij} \ln \left[(1 - \lambda_A) p_{r_i} p_{c_j} + I_{(i=j)} \frac{p_{r_i} + p_{c_j}}{2} \lambda_A \right] \quad (10)$$

The maximum likelihood estimators can be derived by setting the score equations with respect to the parameters equal to zero and solving for the unknown parameters. Alternatively, estimates can be obtained by solving the equations $E(N_{ij}) = N_{ij}$, $i = 1, 2$, $j = 1, 2$.

Thus, the following maximum likelihood estimates are obtained:

$$\begin{aligned} \hat{\lambda}_A = & \frac{N_{11}}{2N_{11} + N_{12} + N_{21}} \\ & + \frac{N_{22}}{2N_{22} + N_{12} + N_{21}} \\ & - \frac{T}{(2N_{11} + N_{12} + N_{21})(2N_{22} + N_{12} + N_{21})} \end{aligned} \quad (11)$$

where

$$T = \sqrt{\frac{(N_{11} - N_{22})^2 (N_{12} + N_{21})^2}{+4N_{21}N_{12}(2N_{11} + N_{12} + N_{21})}}$$

$$\hat{p}_{r1} = \frac{N_{11} + N_{12}}{N_{..}} - \frac{(N_{12} - N_{21})\hat{\lambda}_A}{2N_{..}(1 - \hat{\lambda}_A)}$$

and

$$\hat{p}_{c1} = \frac{N_{11} + N_{21}}{N_{..}} - \frac{(N_{12} - N_{21})\hat{\lambda}_A}{2N_{..}(1 - \hat{\lambda}_A)}$$

Note that when $N_{11}N_{22}=N_{12}N_{21}$ (independence), $\hat{\lambda}_A=0$ and when $N_{12}=N_{21}=0$ (total agreement), $\hat{\lambda}_A=1$. Moreover, the estimated independent marginal probabilities are the same as the observed marginal probabilities iff $\hat{\lambda}_A=0$ or $N_{12}=N_{21}$.

If both N_{12} and N_{21} are replaced by $(N_{12}+N_{21})/2$ on RHS of equation (11) the estimator for Scott's pi, κ_s is derived. When the assumption of no rater bias is made, the extended model (9) reduces to the common correlation model (Donner & Eliasziw, 1992). Because the extended model allows for rater bias, $\hat{\lambda}_A$ has several advantages over κ_s as follows:

1. The common correlation model does not fit the data if in fact there is substantial rater bias. Thus applications of the common correlation model of Scott's pi could be misleading in the absence of a test for rater bias, such as McNemar's test. However, even with such a test, the power may be insufficient to detect rater bias for a small sample size, and an improper application of Scott's pi may occur.

2. Both the common correlation model and the extended model assume only random agreement, i.e., no systematic disagreement. If the two raters have different independent marginal probabilities they are less likely to agree with each other by chance, and the observed agreement is more likely to have been achieved for reasons other than chance. Therefore, if there is rater bias, the estimated agreement index should increase with this bias given the same amount of observed agreement. Scott's pi does not change with observer bias while $\hat{\lambda}_A$ increases with increasing rater bias.

3. When the table is independent ($N_{11}N_{22}=N_{12}N_{21}$) one would expect the estimated systematic agreement to be zero, as is the case with $\hat{\lambda}_A$. However Scott's pi does not equal zero unless $N_{12}=N_{21}$ in addition.

It may be written

$$\kappa_s = \frac{(1 - \lambda_A) \sum_{i=1}^2 p_{ri} p_{ci} + \lambda_A - \sum_{i=1}^2 \left(\frac{p_{ri} + p_{ci}}{2} \right)^2}{1 - \sum_{i=1}^2 \left(\frac{p_{ri} + p_{ci}}{2} \right)^2} \quad (12)$$

and

$$error(\kappa_s, \lambda_A) = \kappa_s - \lambda_A = \frac{(1 - \lambda_A)(p_{r1} - p_{c1})^2}{2(1 - \sum_{i=1}^2 \left(\frac{p_{ri} + p_{ci}}{2} \right)^2)} \quad (13)$$

When there is no rater bias, $p_{r1}=p_{c1}$ and $\kappa_s=\lambda_A$. As $p_{r1}-p_{c1}$ increases, so does $error(\kappa_s, \lambda_A)$. Table 3 shows the values of $error(\kappa_s, \lambda_A)$ at different values of $(p_{r1}, p_{c1}, \lambda_A)$.

Table 3. Error of Scott's pi under the Restricted QI Model

p_{r1}	p_{c1}	$\lambda=0$	$\lambda=0.1$	$\lambda=0.3$	$\lambda=0.5$	$\lambda=0.7$	$\lambda=0.9$
0.9	0.1	0.640	0.576	0.448	0.320	0.192	0.064
0.8	0.2	0.360	0.324	0.252	0.180	0.108	0.036
0.7	0.3	0.160	0.144	0.112	0.080	0.048	0.016
0.6	0.4	0.040	0.036	0.028	0.020	0.012	0.004

Table 4. Diagnosis of Carcinoma for Pathologists A and B

Classification of Pathologist A	Classification of Pathologist B		Total
	Class 1 or 2	Class 3 or 4 or 5	
Class 1 or 2	36	16	52
Class 3 or 4 or 5	3	63	66
Total	39	79	118

Consider Table 4 which shows the collapsed version of Table 2. The p-value for McNemar's test for this table is 0.003, indicating substantial rater bias. For this table, $\hat{\kappa}_S=0.660$ and $\hat{\lambda}_A=0.703$ are obtained.

Results

It has been shown above that $\hat{\lambda}_A$ increases with increasing rater bias while Scott's pi does not change, and that $\hat{\lambda}_A = 0$ for an independent 2×2 table, whereas Scott's pi is only zero when the off-diagonal elements are equal to each other. Thus, λ_A is preferable to κ_S as a measure of agreement in the presence of rater bias. When there is no rater bias, is λ_A as good as κ_S ? Simulations were conducted to investigate the performance of these agreement indices.

Simulations were conducted for a total of 100 configurations: 4 different sample sizes, $N_{\cdot}=\{20,50,100,200\}$; 5 different nominal values of systematic agreement $\lambda_A = \{0,0.1,0.3,0.6,0.9\}$; and 5 combinations of independent marginal probabilities $(p_{r1}, p_{c1}) = \{(0.9,0.9), (0.8,0.8), (0.7,0.7), (0.6,0.6), (0.5,0.5)\}$. The four values for sample size range from small to moderate to sufficiently large. The five values of λ_A cover the whole range. The independent marginal probabilities all represent the case where there is no rater bias. Because of symmetry, only probabilities of 0.5 to 0.9 are investigated.

The probability of each cell is computed according to the extended model specified in (8) given the nominal values of systematic agreement and independent marginal probabilities. The frequency of each cell is generated as a multinomial random number

given the sample size, using the GENMUL routine (Brown and Lovato). For each configuration, 1000 tables were generated.

The efficiency of the two indices were compared in terms of empirical bias, empirical standard deviation (defined as the standard deviation of the estimated values from the 1000 tables) and empirical residual mean square error (RMSE) (defined as the square root of the mean square of differences between the estimated values of the index and the nominal values over the 100 tables). Figure1 displays a side-by-side comparison of the bias of $\hat{\lambda}_A$ compared to the bias of $\hat{\kappa}_S$ as a function of the nominal values of systematic agreement, independent marginal probability and sample size. Figures 2 and 3 give similar displays for the standard deviations and RMSEs.

From these figures, it is observed that $\hat{\kappa}_S$ is negatively biased, underestimating the true value of systematic agreement in all situations, whereas $\hat{\lambda}_A$ is either positively or negatively biased. There is an increasing trend in bias for $\hat{\lambda}_A$ as λ_A increases. Both indices are increasingly likely to underestimate λ_A as $p_{r1}=p_{c1}$ increases. The biases of both $\hat{\lambda}_A$ and $\hat{\kappa}_S$ decrease as sample size increases. There are no differences between the standard errors and RMSE values of $\hat{\lambda}_A$ and $\hat{\kappa}_S$ that are visually discernible. The standard error and RMSE of both indices tend to be smaller when λ_A is close to either 0 or 1, when the independent marginal probability is close to 0.5, or when the sample size is large.

Figure 1. Side-by-side comparison of bias of $\hat{\lambda}_A$ (Lambda) and $\hat{\kappa}_S$ (Scott's Pi) as a function of nominal values for systematic agreement, independent marginal probability, and sample size.

Figure 2 Side-by-side comparison of standard deviation (STD) of $\hat{\lambda}_A$ (Lambda) and $\hat{\kappa}_S$ (Scott's Pi) as a function of nominal values for systematic agreement, independent marginal probability, and sample size.

Figure 3. Side-by-side comparison of residual mean square error (RMSE) of $\hat{\lambda}_A$ (Lambda) and $\hat{\kappa}_S$ (Scott's Pi) as a function of nominal values for systematic agreement, independent marginal probability, and sample size.

The new index is now compared to Cohen's kappa. Cohen's kappa corrects for chance agreement using the assumption of independence between raters. However, the assumption of independence in agreement studies is not valid. Some degree of agreement is usually expected in most agreement studies. If there is systematic agreement present, the classifications of the raters cannot be independent since the raters are dealing with the same information (i.e., the same subject) on which to base each of their classifications. The assumption of blindness of ratings is reasonable, leading to an assumption of conditional independence. Thus the formula

$$\Pr(i, j) = p_{r_i} p_{c_j} \quad (14)$$

should not be used to estimate the expected agreement by chance.

Alternatively, one can use the QI concept to investigate agreement in this context. If it is assumed that no systematic disagreement is present, then equation (3) reduces to

$$\pi_{ij} = (1 - \lambda_A) p_{r_i} p_{c_j} + d_{ii} \chi_{ii} \quad (15)$$

Because $\sum_{i=1}^k \hat{\pi}_{ii} = p_0$ where p_0 is the observed agreement, an estimator is obtained

$$\hat{\lambda}_A = \frac{\left[p_0 - \sum_{i=1}^k \hat{p}_{r_i} \hat{p}_{c_j} \right]}{\left[1 - \sum_{i=1}^k \hat{p}_{r_i} \hat{p}_{c_j} \right]} \quad (16)$$

where $\hat{p}_{r_i}$ and $\hat{p}_{c_j}$ are estimates of the independent marginal probabilities, estimated only in the difficult to classify group (Group 2) where the observers classify independently. Note the similar formulation of $\hat{\lambda}_A$ and $\hat{\kappa}$ given in (2), with the difference being in the marginal probabilities used in calculating these two agreement indices.

If one assumes that there is either no systematic agreement or systematic disagreement, the quasi independent concept resolves the paradoxes posed by Cohen's kappa in 2x2 tables described earlier. Consider the following illustrations.

A. The case of the first paradox is illustrated by the two independent tables shown in Table 5. The values of lambda are 0 for both table 5.1 and table 5.2. All agreement is random agreement. In these cases the observed marginal probabilities are equal to the independent marginal probabilities. Table 5.1 has a set of symmetrical balanced independent marginal probabilities and table 5.2 has a set of symmetrical unbalanced marginal probabilities. Intuitively, the table with unbalanced independent marginal probabilities yields more agreement. A subject in table 5.1 has a 50% chance to be agreed upon by the two observers while the chance is 82% that a subject in table 5.2 will be agreed upon. The agreement is not systematic and can be considered as random agreement. Thus, a set of symmetrical unbalanced marginal probabilities yields more random agreement and less systematic agreement than a set of symmetrical balanced marginal probabilities.

Next consider Table 6. $\hat{\lambda}_A = 0.70$ and $\hat{\lambda}_A = 0.32$ are calculated for table 6.1 and table 6.2, respectively, which agree with the Cohen's kappa estimates. However, one is also able to use the QI concept to derive estimates of the independent marginal probabilities, obtaining $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.53, 0.42)$ and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.91, 0.84)$. Thus, table 6.1 yields a set of symmetrical balanced independent marginal probabilities while table 6.2 yields a set of symmetrical unbalanced independent marginal probabilities. Given the same observed agreement, the amount of systematic agreement (estimated by $\hat{\lambda}_A$) should be greater for the tables with symmetrical balanced independent marginal probabilities. Thus, by using the construct of a latent class of subjects who are systematically classified, one arrives at a resolution of the first paradox.

Table 5. Balanced and unbalanced independent marginal probabilities

Table 5.1: $\hat{P}_o = 0.50, \hat{\lambda}_A = 0$			
	Observer A		
Observer B	Yes	No	Total
Yes	25	25	50
No	25	25	50
Total	50	50	100

Table 5.2: $\hat{P}_o = 0.82, \hat{\lambda}_A = 0$			
	Observer A		
Observer B	Yes	No	Total
Yes	81	9	90
No	9	1	10
Total	90	10	100

Table 6. Balanced and Unbalanced Marginal Probabilities

Table 6.1: Balanced Marginal Probabilities			
	Observer A		
Observer B	Yes	No	Total
Yes	40	9	49
No	6	45	51
Total	46	54	100

Table 6.2: Unbalanced Marginal Probabilities			
	Observer A		
Observer B	Yes	No	Total
Yes	80	10	90
No	5	5	10
Total	85	15	100

For table 6.1 we obtain $\hat{P}_o = 0.85$, $\hat{\kappa} = 0.70$, $\hat{\lambda}_A = 0.70$, and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.53, 0.42)$, while for table 6.2 we obtain $\hat{P}_o = 0.85$, $\hat{\kappa} = 0.32$, $\hat{\lambda}_A = 0.32$, and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.91, 0.84)$.

Source: Derived from Feinstein and Cicchetti (1990)

B. Consider now the case of the second paradox illustrated by the two independent tables in Table 7. Table 7.1 has a set of symmetrical marginal probabilities while table 7.2 has a set of asymmetrical marginal probabilities. There is no systematic agreement in either table. The observed marginal probabilities are the independent marginal probabilities. Intuitively, the agreement achieved by the observers in Table 7.1 is much more than that in table 7.2. Thus a set of symmetrical independent marginal probabilities yields more agreement than a set of asymmetrical independent marginal probabilities.

Next consider Table 8. $\hat{\lambda}_A = 0.13$ and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.59, 0.71)$, $\hat{\lambda}_A = 0.33$ and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.67, 0.23)$ are estimated for tables 8.1 and 8.2 respectively. The independent marginal probabilities have more symmetry for Table 8.1, yielding more random agreement and less systematic agreement than Table 8.2. Given the same amount of observed agreement, there is less systematic agreement (estimated by $\hat{\lambda}_A$) for Table 8.1, thus resolving the second paradox.

Table 7. Symmetrical and asymmetrical independent marginal probabilities

Table 7.1: $\hat{P}_o = 0.82, \hat{\lambda}_A = 0$			
	Observer A		
Observer B	Yes	No	Total
Yes	81	9	90
No	9	1	10
Total	90	10	100

Table 7.2: $\hat{P}_o = 0.18, \hat{\lambda}_A = 0$			
	Observer A		
Observer B	Yes	No	Total
Yes	9	81	90
No	1	9	10
Total	10	90	100

Table 8. Symmetrical and Asymmetrical Unbalanced Marginal Probabilities

Table 8.1: Symmetrical Unbalanced Marginal Probabilities			
	Observer A		
Observer B	Yes	No	Total
Yes	45	15	60
No	25	15	40
Total	70	30	100

Table 8.2: Asymmetrical Unbalanced Marginal Probabilities			
	Observer A		
Observer B	Yes	No	Total
Yes	25	35	60
No	5	35	40
Total	30	70	100

For table 8.1 we obtain $\hat{P}_o = 0.60, \hat{\kappa} = 0.13, \hat{\lambda}_A = 0.13$ and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.59, 0.71)$, while for table 8.2 we obtain $\hat{P}_o = 0.60, \hat{\kappa} = 0.26, \hat{\lambda}_A = 0.33$ and $(\hat{p}_{r_1}, \hat{p}_{c_1}) = (0.67, 0.23)$.

Source: Derived from Feinstein and Cicchetti (1990)

Conclusion

It has been shown how the quasi-independent concept can be applied to studies of inter-rater agreement. When applied to 2×2 tables, the use of the QI concept results in a paradigm for agreement that resolves the paradoxical behavior of the popular measure, Cohen's kappa, although the resulting measure can only be derived after the user decides on cells representing systematic agreement or disagreement. This measure has other desirable properties; specifically it allows for assessment of the independent marginal probabilities, which can be reported as companion statistics. Unlike the other statistics that have been suggested for

reporting along with Cohen's kappa, these independent marginal probabilities are model-based. Thus, further use and study of the application of the QI concept in inter-rater agreement studies is warranted.

References

- Agresti, A. (1990). *Categorical Data Analysis*. New York, N.Y.: John Wiley & Sons.
- Aickin, M. (1990). Maximum likelihood estimation of agreement in the constant predictive probability model, and its relation to Cohen's kappa. *Biometrics*, 46, 293-302.
- Bennett, E. M., Alpert, R., & Goldstein, A. C. (1954). The measurement of observer disagreement in the recording of signs. *Journal of the Royal Statistical Society A*, 129, 98-109.

- Bergan, J. R. (1980). Measuring observer agreement using the quasi-independent concept. *Journal of Educational Measurement*, 17, 59-69.
- Bishop, Y. N., Feinberg, S. W., & Holland, P. W. (1975). *Discrete Multivariate Analysis*. Cambridge: MIT Press.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41, 687-698.
- Brown, B. W., & Lovato, J. GENMUL routine. Library of C Routines for Random Number Generation. Department of Biomathematics, Box 237, University of Texas, M. D. Anderson Cancer Center, Houston, TX 77030.
- Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence, and kappa. *Journal of Clinical Epidemiology*, 46, 423-429.
- Cicchetti, D. V. & Feinstein, A. (1990). High agreement by low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43, 551-558.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Donner, A., & Eliasziw, M. (1992). A goodness-of-fit approach to inference procedure for the kappa statistic: Confidence interval construction, significance testing and sample size estimation. *Statistics in Medicine*, 11, 1511-1519.
- Feinstein, A., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problem of two paradoxes. *Journal of Clinical Epidemiology*, 43, 543-549.
- Goodman, L. A. (1968). The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing data entries. *Journal of the American Statistical Association*, 63, 1091-1131.
- Guggenmoos-Holzmam, I., & Vonk, R. (1998). Kappa-like indices of observer agreement viewed from a latent class perspective. *Statistics in Medicine*, 17, 797-812.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-74.
- Lantz, C. A., & Nebenzahl, E. (1996). Behavior and interpretation of the κ statistic; resolution of the two paradoxes. *Journal of Clinical Epidemiology*, 49, 431-434.
- Scott, W. A. (1955). Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly*, 19, 321-325.

A Method for Analyzing Unreplicated Experiments Using Information on the Intraclass Correlation Coefficient

Jamis J. Perrett

Department of Applied Statistics and Research Methods
University of Northern Colorado

Many studies are performed on units that cannot be replicated; however, there is often an abundance of subsampling. By placing a reasonable upper bound on the intraclass correlation coefficient (ICC), it is possible to carry out classical tests of significance that have conservative levels of significance.

Key words: Intraclass correlation coefficient, unreplicated experiment, subsamples

Introduction

A researcher, wishing to compare two different teaching methods, teaches two classes: one with method 1 and one with method 2. The grade of each student in the two classes is recorded with the purpose of comparing the average grade for the students taught by method 1 to the average grade for the students taught by method 2. The within class variation is the variability from student to student. The between class variation is due to such factors as time of day, difference in classroom setting, etc. One would expect the variation from class to class to be small relative to the within class variation, regardless of whether the students are being taught mathematics, creative writing, etc. The majority of the total variability will be explained by the difference in performance of the students within a class, and that should be fairly similar for one

subject as it is another. Thus, the intraclass correlation coefficient (ICC) should be consistent in studies of similar types, and it will tend to be small (In an education example such as this, it would not be unusual for the ICC to be less than 0.1) in many situations.

An unreplicated experiment is one in which a treatment of interest is applied to only one unit. Some experiments logistically cannot be replicated. Circumstances that might prevent replication are cost in time or money or both, scarcity of experimental units, destructive experimentation, among other things. Some farmers just don't have an extra plot of land to experiment on. Consideration is given for what can be done in such cases.

The Model

Let y_{ij} be the measurement taken on the j^{th} student given the i^{th} treatment, μ_i is the fixed effect of treatment i , δ_i is the random effect of class i , and ε_{ij} is the random effect of student j given treatment i , $i = 1, 2, \dots, t$, $j = 1, 2, \dots, n_i$. Let $\delta_i \sim n(0, \sigma_\delta^2)$, where σ_δ^2 represents the between class variability; let $\varepsilon_{ij} \sim n(0, \sigma_\varepsilon^2)$, where σ_ε^2 represents the between student within class variability. It is assumed that δ_i and ε_{ij} are independent. A model is written for the experiment as follows:

Jamis Perrett is Assistant professor of Applied Statistics at the University of Northern Colorado. His research interests include the analysis of unreplicated experiments, computational statistics, and repeated measures analysis. E-mail: jamis.perrett@unco.edu

$$y_{ij} = \mu_i + \delta_i + \varepsilon_{ij} \quad (1)$$

This type of model is a single factor completely randomized design with subsampling, where classes are the experimental units for each treatment level, and the students within each class are the subsamples, or observational units. A researcher, who uses the students as the experimental units, ignores the variability that can exist between different classes receiving the same treatment. Such an assumption is to claim that $\sigma_\delta^2=0$. If the researcher correctly uses classes as the experimental units, there is only one unit per treatment level and zero degrees of freedom available for testing the difference between these treatment means (Barcikowski, 1981, Blair, 1986).

The Intraclass Correlation Coefficient and the Independence Assumption

The intraclass correlation coefficient (ICC) is defined as the correlation between y_{ij} and $y_{ij'}$ (two subsample units within one experimental unit). In this study, ρ refers to the true value of the ICC and ρ_0 refers to a best guess value, chosen by the researcher, to plug into formulas in place of the ICC in the analysis. The ICC for the model in Equation 1 can be obtained using the following formula:

$$\rho = \frac{\text{cov}(y_{ij}, y_{ij'})}{\sqrt{\text{var}(y_{ij})\text{var}(y_{ij'})}} = \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma_\varepsilon^2} \quad (2)$$

Thus, if $\sigma_\delta^2=0$, the ICC is also zero. The result is independent samples assuming normality of error terms. To illustrate the ideas, consider the two-treatment case in which $H_0: \mu_1 = \mu_2$ versus $H_A: \mu_1 \neq \mu_2$ is tested. The variance of the difference between the two sample means is given by

$$\text{var}(\bar{y}_1 - \bar{y}_2) = 2\sigma_\delta^2 + \sigma_\varepsilon^2 \left[\frac{n_1 + n_2}{n_1 n_2} \right]$$

$$= \sigma_\varepsilon^2 \left\{ 2 \left(\frac{\rho}{1-\rho} \right) + \left[\frac{n_1 + n_2}{n_1 n_2} \right] \right\} \quad (3)$$

using the substitution $\sigma_\delta^2 = \sigma_\varepsilon^2 \frac{\rho}{1-\rho}$.

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\sigma_\varepsilon^2 \left\{ 2 \left(\frac{\rho}{1-\rho} \right) + \left[\frac{n_1 + n_2}{n_1 n_2} \right] \right\}}} \quad (4)$$

where Z has a standard normal distribution. If ρ is incorrectly assumed to be zero, then

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\sigma_\varepsilon^2 \left(\frac{n_1 + n_2}{n_1 n_2} \right)}} \quad (5)$$

A test statistic for a hypothesis test based on the incorrect assumption that observations are independent will be too large and consequently inflate the associated Type 1 error.

Bounding the ICC

In practice, σ_ε^2 may be estimated from the pooled variance of observational units within the experimental unit. With many subsamples, σ_ε^2 can be estimated quite accurately. However, in an unreplicated experiment there is no way to estimate σ_δ^2 and consequently ρ .

Although it may not be possible to know the value of the ICC, in many cases it may be reasonable to make assumptions about its upper bound. To do so, one must consider the relative size of the between unit variability to the within unit variability. In the example considered in this study, the classes were similar, so it is reasonable to assume the component of the variance due to classes is relatively small. On the other hand, the component of variance due to differences among students within a class tends to be relatively large due to the inherent differences in students: maturity, study habits, initial understanding, intelligence, etc. Thus, it

would seem to be reasonable to place a bound on the ICC that is less than .5 and possibly quite a bit smaller than this. Data discussed in this study indicate that a bound of $\rho < .15$ is reasonable for this example.

Other examples of this also are common in agriculture. Consider for instance feeding treatments applied to pens with measurements made on individual animals within pens. For many measurements such as weight gain, body condition scores, and various blood parameters, the greatest source of variability is among the animals within the pens. The component of variance due to pens, while not negligible, is often just a fraction of the component of variance due to the animals. In such cases it is quite reasonable to assume that ρ is small. The upper bound will be denoted as $\rho_{\max}$.

The importance of a small value of ρ can be seen in Equation 3 which shows that the variance of the difference of sample means gets smaller as ρ gets smaller. In the limit, the variance is that of the difference of means of independent observations. Intuitively, the closer ρ is to zero, the more the observations behave as if they were independent. Moreover, the analysis using a known ICC becomes more powerful as ρ gets closer to zero with the limiting power being that obtained when the observations are independent.

Methodology

The Testing Strategy

Although ρ is not known, if prior experience allows for an upper bound to be placed on it, then it is possible to carry out statistical tests for unreplicated experiments. Consider a test of the hypotheses $H_0: \mu_1 = \mu_2$ vs. $H_A: \mu_1 \neq \mu_2$. Let ρ_0 denote a value of ρ that the researcher assumes to be reasonable based on prior experience. Let $\mu_{01} - \mu_{02}$ represent the hypothesized difference of the mean for treatment 1 and the mean for treatment 2 respectively. The test statistic is

$$T = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_{01} - \mu_{02})}{\sqrt{\hat{\sigma}_\varepsilon^2 \left\{ 2 \left(\frac{\rho_0}{1 - \rho_0} \right) + \left[\frac{n_1 + n_2}{n_1 n_2} \right] \right\}}} \quad (6)$$

where $\hat{\sigma}_\varepsilon^2$ represents a pooled estimate of the within class variance. Let

$$s_i^2 = \frac{\sum_{k=1}^{n_i} (y_{ik} - \bar{y}_i)^2}{n_i - 1}, \quad i=1, 2 \quad (7)$$

be the variance of the measurements under treatment i . Then

$$\hat{\sigma}_\varepsilon^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (8)$$

Let α_0 denote the nominal level of the test and α the true level of significance. Depending on assumptions, α_0 may or may not equal α .

Balancing simplicity and desirable properties, it is suggested that tests be based on p-values obtained when $\rho_0 = \rho_{\max}$. It is also recommended that confidence intervals and multiple comparisons be carried out using $\rho_0 = \rho_{\max}$.

Properties of the test statistic

For simplicity, assume the number of subsamples per class is the same for all classes, i.e. $n = n_1 = n_2$. The sample size is one-one class per treatment. Let $\mu_1 - \mu_2$ be the true value of the difference between the treatment means to be compared in the hypothesis test. Let $v=2(n-1)$, the degrees of freedom for the test. Let $t_{0.05,v}$ denote the upper tail 0.05 value of the t-distribution with v degrees of freedom. The probability of rejecting H_0 for an upper-tail test at $\alpha_0 = 0.05$ can be determined using the following steps:

$$P \left[\frac{(\bar{y}_1 - \bar{y}_2) - (\mu_{01} - \mu_{02})}{\sqrt{2\hat{\sigma}_\varepsilon^2 \left(\frac{\rho_0}{1-\rho_0} + \frac{1}{n} \right)}} > t_{0.05, v} \right] \quad (9)$$

If the following is defined as

$$Z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{2\sigma_\varepsilon^2 \left(\frac{\rho}{1-\rho} + \frac{1}{n} \right)}} \quad (10)$$

and

$$\lambda = \frac{(\mu_1 - \mu_2) - (\mu_{01} - \mu_{02})}{\sqrt{2\sigma_\varepsilon^2 \left(\frac{\rho}{1-\rho} + \frac{1}{n} \right)}} \quad (11)$$

then Equation 9 can be expressed as

$$P \left[\frac{Z + \lambda}{\sqrt{U/v}} > (t_{0.05, v}) \frac{\sqrt{\frac{\rho_0}{1-\rho_0} + \frac{1}{n}}}{\sqrt{\frac{\rho}{1-\rho} + \frac{1}{n}}} \right] \quad (12)$$

where $Z \sim N(0,1)$, λ is a constant, and $U \sim \chi_v^2$. To the left of the inequality is a random variable with a non-central t-distribution with non-centrality parameter λ and degrees of freedom, v .

Evaluation

The plug-in method involves choosing a value ρ_0 to plug into Equation 6. The method is evaluated by determining significance levels and power curves for the tests, for different values of ρ_0 and ρ . In particular, the method is considered useful if it can maintain a

Type 1 error level close to the nominal level (α_0) while providing power to detect differences in treatment means for a reasonable range of ρ_0 near ρ .

Equation 12 may be used to evaluate the probability of rejection for tests of the hypothesis that the two means are equal. These probabilities depend on the values for ρ , ρ_0 , n , and the non-centrality parameter, λ . In order to measure the deviation between the two means, a standardized difference is defined as $\text{StDiff} = \frac{\mu_1 - \mu_2}{\sigma_\varepsilon}$. Probabilities will depend on StDiff

through λ . Using Equation 12, probabilities were generated using the following values as indices:

$$\text{StDiff} = 0, 1$$

$$n = \infty$$

$$\rho = 0 \text{ through } 0.99 \text{ in increments of } 0.01$$

$$\rho_0 = 0 \text{ through } 0.99 \text{ in increments of } 0.01$$

Figure 1 is a result of the generated probabilities. The data used for this plot were created by evaluating Equation 12. For the plot, let $n \rightarrow \infty$ resulting in $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$ being used in

Equation 12. The nominal significance level $\alpha_0 = 0.05$ is used for the plot. Also, neither the power nor the significance level is defined at $\rho=1$ or $\rho_0=1$. This plot depicts the case with a theoretically infinite number of students per class. A similar graph results from using 10 students per class (omitted). This power plot is overlaid with the corresponding plot of two-tail significance levels. The red, green, and blue areas denote power in the ranges ≤ 0.2 , 0.2 to 0.5 , and 0.5 to 0.9 respectively. The lines represent the two-tail probabilities of rejecting the null hypothesis. If $\rho = \rho_0$,

Figure 1. Large (sub)sample power of detecting a difference of size 1 between two treatment means with no replication, overlaid with large sample significance (α) levels.

then $\alpha=0.05$. If the plug-in value ρ_0 is less than the true value ρ ($\rho \geq \rho_0$), the Type 1 error is inflated ($\alpha \geq 0.05$). For example, if a value of $\rho_0=0.2$ is used when $\rho=0.4$, alpha will be 0.2253. If the plug-in value ρ_0 is greater than the true value ρ ($\rho \leq \rho_0$), the probability of a Type 1 error is smaller than $\alpha_0=0.05$.

Because the p-value increases as ρ_0 increases, the p-value obtained when $\rho_0 = \rho_{\max}$ is greater than or equal to the true p-value, so tests using this methodology are conservative. On the other hand, if $\rho_{\max}$ is set too small by mistake, Type 1 error will be inflated. In terms

of power and length of confidence intervals, a smaller $\rho_{\max}$ is better than a larger one.

This study does not suggest that problems with lack of replication magically disappear with this methodology. Even if ρ is known, $\rho > 0$ presents problems. In the two sample case, for instance, the variance of the difference between two means when the number of subsamples n_1 and n_2 approaches infinity becomes

$$= 2\sigma_e^2 \frac{\rho}{1-\rho} \quad (13)$$

This would be the formula for the variance of the difference between two sample means if

each treatment had $n = \frac{1-\rho}{\rho}$ replications.

Thus, $\frac{1-\rho}{\rho}$ can be thought of as the number of pseudo replications. For instance if $\rho=0.2$, the number of pseudo replications is 4. This study simply suggests the proposed methodology gives researchers a way to analyze data when conventional analysis of variance could not be carried out due to lack of replication.

It is also useful to examine the maximum power attainable using a plug-in value for ρ in hypothesis testing. Figure 2 demonstrates the maximum attainable power under the most ideal conditions: namely $\rho = \rho_0$ and $n = \infty$.

As seen in Figure 2, if the true value ρ is 0.5, the power is low. For smaller values of ρ

the power increases considerably. For instance, for a standardized difference of 1.0 and $\rho=0.1$, the power is 0.564. Thus, using a plug-in value for ρ is only going to be effective for smaller values of ρ .

What the Researcher May Know About ρ

It is possible for the researcher to obtain information about ρ based on prior experiments of a similar nature, or from knowledge about the behavior of ρ for a current experiment.

Distributional Information

If a large amount of distributional information is available from prior studies of a nature similar to that of the current study, the researcher may be able to put a prior distribution or empirical distribution on ρ (not considered in this study).

Figure 2. Maximum attainable power for testing the difference between two treatment means using the plug-in method with different values of $\rho = \rho_0$ and infinite subsampling.

Point or Interval Information

The researcher may not have extensive distributional information about ρ , but may have an indication of a reasonable minimum or maximum value of ρ .

The Plug-in Value

In the case of no replication, zero degrees of freedom exist for conducting a hypothesis test comparing means. So, the test cannot be performed using traditional methods. With this procedure a value, chosen by the researcher, is used as if it were the true value, ρ . This value, ρ_0 , called a plug-in value, can be used in hypothesis testing and in producing confidence intervals of differences of treatment means. The strategies for choosing a value for ρ_0 given in this section are proposed to researchers who have an unreplicated experiment and have a reasonable idea of the actual value, ρ . Other methods for dealing with unreplicated experiments have been investigated with similar results to the plug-in methods.

Proposed Strategies

Two strategies that make use of prior information about ρ to test hypotheses about treatment means in an unreplicated, two-treatment experiment will be presented in this study.

Strategy 1: Plot of the Conditional P-value Given ρ_0

Description

For a given set of data, the p-value may be obtained for various assumed values of ρ . It is computed as the probability that a t-distributed random variable with $n_1 + n_2 - 2$ degrees of freedom is more extreme than the observed value of the statistic defined by Equation 10. A conditional p-value plot plots p-values for testing a certain hypothesis over the range of possible values of ρ .

Properties

The three situations a researcher may encounter with a conditional p-value plot are 1) all p-values are above α_0 for reasonable values of ρ , 2) some p-values are below α_0 for reasonable values of ρ and some are above α_0 for reasonable values of ρ , and 3) all p-values are below α_0 for reasonable values of ρ . When the first situation occurs, the result of the test is to fail to reject the null hypothesis at level α_0 . When the third situation occurs, the result of the test is to reject the null hypothesis at level α_0 . When the second situation occurs, it is less obvious what the results of the test of hypothesis should be. The researcher is advised to refrain from making a decision about the acceptance of the null hypothesis if the second situation is observed. The accuracy of the results will depend on the accuracy of the choice of the likely range of ρ .

Implementation

The researcher determines the range of reasonable values for the ICC based on information obtained from prior studies. Then the researcher creates the conditional p-value plot given the possible values of ρ . The decision to reject or fail to reject the null hypothesis, or abstain from judgment on the acceptance of the null hypothesis is made based on the observed p-values within the range of reasonable values of ρ .

Strategy 2: Maximum Rho

Description

The Maximum Rho procedure simply involves choosing $\rho_0 = \rho_{\max}$. That value, ρ_0 , is then incorporated into the test statistic (Equation 6). The test rejects the null hypothesis if the p-value is less than α_0 .

Properties

The Maximum Rho procedure assures the true significance level, α , is less than or equal to the nominal value α_0 , with equality when $\rho_0 = \rho$. The closer ρ_0 is to ρ , the higher the power of the test.

Implementation

To implement the Maximum Rho procedure, the researcher simply computes a p-value for the test based on using $\rho_0 = \rho_{\max}$ in Equation 6.

Example

The class data consists of final course grades for two classes of introductory statistics taught by two different methods. The sample means for the two classes were 2.83 and 3.37 with sample standard deviations of 1.04 and 0.84 respectively. A researcher would like to see if there is a difference between the average grade received by students taught by the two different methods. Only one class was observed for each of the two methods making this an unreplicated experiment with class as the unit of study, student as the subsampling unit. Let $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$. Let $\alpha_0 = 0.05$. It is assumed that replication would have yielded small class-to-class variability relative to the total variability. To determine if

that could in fact be the case, a study was conducted on prior undergraduate courses taught to multiple sections over multiple semester.

College Course Grades

The value ρ was estimated from a variety of courses offered at Kansas State University using the SAS® MIXED procedure. The components of variance consist of variability of scores due to section, σ_δ^2 ; and the variability of scores due to students within sections, σ_ϵ^2 . Fourteen different undergraduate courses were selected (CHM 111, 210, 230; CIS 101; ENGL 100, 125; MATH 010, 100; MUSIC 250, 255; PSYCH 110, 202, 350; SPAN 161) each with multiple sections, covering both Fall and Spring semesters over the years 2001-2003, for a total of 43 course-semester combinations. These values are graphed in a frequency histogram in Figure 3.

All 43 estimated ICC values are at or below 0.33. The majority, 95%, is at or below 0.2—90% are below 0.15. Only one value is at 0.33. That value is for a honors English course (ENGL 125). Other courses include undergraduate courses in chemistry, English, music, CIS, math, psychology, and Spanish. Based on this, it would be reasonable to use the range of ρ from 0.02 to 0.15, with an assumed maximum at $\rho_0 = \rho_{\max} = 0.15$ for the plug-in methods involving grades for this study.

Figure 3. Estimated ICC values for KSU grades for 43 courses with multiple sections (2001-2003).

Results

Plot of the Conditional P-value Given ρ_0

Figure 4 plots the p-value as a function of ρ for the class data. It can be seen that p-values are only significant if ρ is less than 0.013. Because the likely value of ρ is greater than 0.013, the result of the test is to fail to reject the null hypothesis at $\alpha_0 = 0.05$.

Maximum Rho

Using $\rho_0 = 0.15$ and Equation 6, the test statistic is found to be $t = -0.894$ with a p-value of 0.374. So the result of the test is to fail to reject the null hypothesis in favor of the alternative (at $\alpha_0 = 0.05$).

Based on both the conditional p-value plot and the results of the Maximum Rho procedure, the conclusion is that the difference between the mean grades for the two classes is not significant. Assuming $\rho_0 \geq \rho$, the probability of a Type 1 error for the test in this example is at most 0.05. If, in fact, $\rho = \rho_0 = 0.15$, the power for detecting a difference of one grade point is approx. 0.3373. However, if $\rho < 0.15$, the power will be less. A significant difference would have been detected using a classical t-test to test the same hypothesis under the assumption of independent samples. However, the probability of a type 1 error would be inflated under that assumption.

Figure 4. Conditional P-value Plot. P-value curve for different values of ρ_0 using the Class data.

Conclusion

If replication is feasible, a replicated experiment is always preferred over an unreplicated experiment. However, many studies are performed on units that cannot be replicated. The method described in this paper makes it possible to accurately analyze unreplicated experiments in which the intraclass correlation coefficient (ICC) is small and relatively stable. By placing a reasonable upper bound on the ICC, it is possible to carry out classical tests of significance that have conservative levels of significance. This methodology has wide applicability for analyzing unreplicated experiments in many fields of research and its simple computations will surely appeal to the applied researcher.

References

- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics*, 6, 267-285.
- Blair, R. C., et. al. (1983). An investigation of the robustness of the t-test to unit of analysis violations. *Educational and Psychological Measurement*, 43, 69-80.
- Blair, R. C., & Higgins, J. J. (1986). Comment on statistical power with group mean as the unit of analysis. *Journal of Educational Statistics Summer*, 2, 161-169.
- Casella, G., & Berger, R. L. (2002). *Statistical inference (2nd ed.)*. Pacific Grove, CA: Duxbury.
- Graybill, F. A. (1976). *Theory and application of the linear model*. Pacific Grove, CA: Wadsworth.
- Little, R. C., et. al. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute Inc.

Milliken, G. A., & Johnson, D. E. (1992). *Analysis of messy data, Volume 1: Experimental design*. London, UK: Chapman & Hall.

SAS Institute Inc. (1999). *SAS/STAT user's guide, version 8*. Cary, NC: Author.

SAS is a registered trademark of SAS Institute Inc., Cary, N.C.

A Comparison of the Spearman-Brown and Flanagan-Rulon Formulas for Split Half Reliability under Various Variance Parameter Conditions

David A. Walker
Northern Illinois University

Differences between the Spearman-Brown and Flanagan-Rulon formulas are examined when the variance parameters for two halves of a test had the following ratios: 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0 and also had a correlation between the two halves of a test at 1.00, .95, .90, .80, .70, .60, .50, .40, .30, .20, .10, .05. It was found that use of the Spearman-Brown formula to estimate the population ρ when the ratio between the standard deviations on two halves of a test is disparate, or beyond .9 to 1.1, was not warranted. Applied and theoretical examples are employed, as well as syntax for user application.

Key words: Split half reliability, Spearman-Brown, Flanagan-Rulon

Introduction

Examination of the difference between estimators of the population ρ for split-half reliability has been studied in the past (Charter, 1996, Cronbach, 1951, Feldt & Brennan, 1989, Kelley, 1942, Rulon, 1939, Stanley, 1971). To estimate the score reliability of a test split in half, the Spearman-Brown formula is the typical method used. Charter (1996) showed that:

$$r = 2r_{xy} / (1 + r_{xy}) \quad (1)$$

where, r_{xy} = the correlation between the two halves of a test.

One major assumption with this formula is that the two halves of a test have equal variance parameters. Rulon (1939, attributed to Flanagan) proposed a split-half formula for use when the variance parameters were unequal. Charter (1996) showed that:

$$r = (4r_{xy} SD_x SD_y) / (SD_x^2 + SD_y^2 + 2r_{xy} SD_x SD_y) \quad (2)$$

where, SD_x and SD_y are the standard deviations for the two halves of the test.

Cronbach (1951), and more recently Charter (1996), found that when the standard deviations of the two halves of a test are not equal, that is, the previously-noted major assumption guiding the Spearman-Brown formula is violated; its use will lead to an over-estimation of the reliability coefficient.

Purpose

The purpose of this research is to elaborate upon the work of Cronbach (1951) and Charter (1996) and strengthen the evidence in the literature that indicates that in most instances when unequal standard deviations for two halves of a test are present, regardless of the correlation between the two halves, the Flanagan-Rulon formula is the better estimator of ρ in a split-halves reliability situation. Thus, this research will build upon Cronbach's work and show, via various graphs and a detailed table, the differences between the Spearman-Brown formula and the Flanagan-Rulon formula when the variance parameters for two halves of a test have the following ratios (either greater or lesser): 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0 and also have a correlation between the two halves of a test at 1.00, .95, .90, .80, .70,

David Walker is Associate Professor of Educational Research and Assessment at Northern Illinois University. His research interests include structural equation modeling, effect sizes, factor analyses, predictive discriminant analysis, predictive validity, weighting, and bootstrapping.

.60, .50, .40, .30, .20, .10, .05. As well, SPSS (Statistical Package for the Social Sciences) syntax is provided in the Appendix section for users to create Tables 1 and 2 or calculate a Spearman-Brown or Flanagan-Rulon value given certain data.

Results

As was found by both Cronbach (1951) and Charter (1996), a large discrepancy between the variance parameters on the two halves of a test results in a substantial decrease in r . A small difference between standard deviations has the opposite effect, which is to be expected. Cronbach noted that when the ratio between the standard deviations of the two halves of a test are .9 to 1.1, the results from either the Spearman-Brown or the Flanagan-Rulon

formulas are virtually the same and thus, the former formula should be used.

For example from empirical data, Gordon (1970) used the Musical Aptitude Profile (i.e., $n = 190$, under normal distribution, and with non-missing data) and estimated the split half reliability via the Spearman-Brown method. These data yielded a range of ratios between variance parameters of .9 to 1.1 (i.e., .949 to 1.141), which produced Spearman-Brown and Flanagan-Rulon values that were nearly identical. That is, there was no discrepancy between the Spearman-Brown formula results and Flanagan-Rulon estimates (i.e., the ratio of Spearman-Brown to Flanagan-Rulon ranged from 1.000 to 1.005). Thus, in this instance, the Spearman-Brown formula did not over-estimate the population ρ and was the proper choice.

Table 1. Comparison of SB and FR Estimates Using the Musical Aptitude Profile

r_{xy}	SD ₁	SD ₂	SD Ratio	SB	FR	Ratio of SB to FR
.695	9.57	10.08	.949	.820	.819	1.001
.600	10.01	9.59	1.044	.750	.750	1.000
.818	9.17	8.93	1.027	.900	.900	1.000
.600	7.66	7.29	1.051	.750	.749	1.001
.575	9.33	8.69	1.074	.730	.729	1.002
.786	7.71	6.76	1.141	.880	.876	1.005
.563	9.28	9.03	1.028	.720	.720	1.000
.600	10.20	9.14	1.116	.750	.747	1.004
.538	8.96	8.28	1.082	.700	.698	1.002
.818	7.95	7.22	1.101	.900	.898	1.003
.905	7.06	6.56	1.076	.950	.949	1.001

r_{xy} = correlation between two halves of a test, SD₁ & SD₂ = standard deviations for two halves of a test, SD Ratio = SD₁ / SD₂, SB = Spearman-Brown, FR = Flanagan-Rulon, Ratio of SB to FR = SB / FR

Note. Table adapted from Gordon (1970), where SB was reported only and not r_{xy} . To calculate r_{xy} the inverse of the SB formula was applied, where $r_{xy} = (SB_r/2) / (1-r_{SB})$.

The current research extended Cronbach (1951) and Charter's (1996) work by creating a detailed table and figures which demonstrated that the range of the ratio between the standard deviations for two halves of a test could not be extended beyond .9 to 1.1. Table 2 shows that the deviation between the results yielded by the Spearman-Brown and Flanagan-Rulon formulas

when the variance parameter ratio was .9 to 1.1 had a decrease in r of $< 1\%$ and a Spearman-Brown to Flanagan-Rulon ratio range difference $< 1\%$. This was not the case, though, when the variance parameter ratio was .8 to 1.2. The Spearman-Brown to Flanagan-Rulon ratio was $\geq 1\%$ starting at $r_{xy} = .70$, which generated a ratio = 1.010, and ended at $r_{xy} = .05$ or a ratio = 1.016.

Table 2. Comparison of SB and FR Formulas under Various r_{xy} and Variance Parameter Conditions

Ratio of SDs for Two Halves				
r_{xy}	Two Halves	SB	Ratio of FR	SB to FR
1.00	2.000	1.000	.889	1.125
.95	2.000	.974	.864	1.128
.90	2.000	.947	.837	1.132
.80	2.000	.889	.780	1.139
.70	2.000	.824	.718	1.147
.60	2.000	.750	.649	1.156
.50	2.000	.667	.571	1.167
.40	2.000	.571	.485	1.179
.30	2.000	.462	.387	1.192
.20	2.000	.333	.276	1.208
.10	2.000	.182	.148	1.227
.05	2.000	.095	.077	1.238
1.00	1.900	1.000	.904	1.107
.95	1.900	.974	.878	1.109
.90	1.900	.947	.852	1.112
.80	1.900	.889	.795	1.118
.70	1.900	.824	.732	1.125
.60	1.900	.750	.662	1.133
.50	1.900	.667	.584	1.142
.40	1.900	.571	.496	1.152
.30	1.900	.462	.397	1.164
.20	1.900	.333	.283	1.178
.10	1.900	.182	.152	1.194
.05	1.900	.095	.079	1.203

Table 2. Continued

1.00	1.800	1.000	.918	1.089
.95	1.800	.974	.893	1.091
.90	1.800	.947	.866	1.094
.80	1.800	.889	.809	1.099
.70	1.800	.824	.746	1.105
.60	1.800	.750	.675	1.111
.50	1.800	.667	.596	1.119
.40	1.800	.571	.507	1.127
.30	1.800	.462	.406	1.137
.20	1.800	.333	.290	1.148
.10	1.800	.182	.157	1.162
.05	1.800	.095	.081	1.169
1.00	1.700	1.000	.933	1.072
.95	1.700	.974	.907	1.074
.90	1.700	.947	.881	1.076
.80	1.700	.889	.823	1.080
.70	1.700	.824	.759	1.085
.60	1.700	.750	.688	1.090
.50	1.700	.667	.608	1.096
.40	1.700	.571	.518	1.103
.30	1.700	.462	.415	1.111
.20	1.700	.333	.298	1.120
.10	1.700	.182	.161	1.131
.05	1.700	.095	.084	1.137
1.00	1.600	1.000	.947	1.056
.95	1.600	.974	.921	1.058
.90	1.600	.947	.894	1.059
.80	1.600	.889	.837	1.063
.70	1.600	.824	.772	1.066
.60	1.600	.750	.701	1.070
.50	1.600	.667	.620	1.075
.40	1.600	.571	.529	1.080
.30	1.600	.462	.425	1.087
.20	1.600	.333	.305	1.094
.10	1.600	.182	.165	1.102
.05	1.600	.095	.086	1.107

Table 2. Continued

1.00	1.500	1.000	.960	1.042
.95	1.500	.974	.934	1.043
.90	1.500	.947	.908	1.044
.80	1.500	.889	.850	1.046
.70	1.500	.824	.785	1.049
.60	1.500	.750	.713	1.052
.50	1.500	.667	.632	1.056
.40	1.500	.571	.539	1.060
.30	1.500	.462	.434	1.064
.20	1.500	.333	.312	1.069
.10	1.500	.182	.169	1.076
.05	1.500	.095	.088	1.079
1.00	1.400	1.000	.972	1.029
.95	1.400	.974	.947	1.029
.90	1.400	.947	.920	1.030
.80	1.400	.889	.862	1.032
.70	1.400	.824	.797	1.034
.60	1.400	.750	.724	1.036
.50	1.400	.667	.642	1.038
.40	1.400	.571	.549	1.041
.30	1.400	.462	.442	1.044
.20	1.400	.333	.318	1.048
.10	1.400	.182	.173	1.052
.05	1.400	.095	.090	1.054
1.00	1.300	1.000	.983	1.017
.95	1.300	.974	.957	1.018
.90	1.300	.947	.930	1.018
.80	1.300	.889	.872	1.019
.70	1.300	.824	.807	1.020
.60	1.300	.750	.734	1.022
.50	1.300	.667	.652	1.023
.40	1.300	.571	.558	1.025
.30	1.300	.462	.450	1.027
.20	1.300	.333	.324	1.029
.10	1.300	.182	.176	1.031
.05	1.300	.095	.092	1.033
1.00	1.200	1.000	.992	1.008
.95	1.200	.974	.966	1.009
.90	1.200	.947	.939	1.009
.80	1.200	.889	.881	1.009
.70	1.200	.824	.816	1.010
.60	1.200	.750	.742	1.010
.50	1.200	.667	.659	1.011
.40	1.200	.571	.565	1.012
.30	1.200	.462	.456	1.013
.20	1.200	.333	.329	1.014
.10	1.200	.182	.179	1.015
.05	1.200	.095	.094	1.016

Table 2. Continued

1.00	1.100	1.000	.998	1.002
.95	1.100	.974	.972	1.002
.90	1.100	.947	.945	1.002
.80	1.100	.889	.887	1.003
.70	1.100	.824	.821	1.003
.60	1.100	.750	.748	1.003
.50	1.100	.667	.665	1.003
.40	1.100	.571	.570	1.003
.30	1.100	.462	.460	1.003
.20	1.100	.333	.332	1.004
.10	1.100	.182	.181	1.004
.05	1.100	.095	.095	1.004
1.00	1.000	1.000	1.000	1.000
.95	1.000	.974	.974	1.000
.90	1.000	.947	.947	1.000
.80	1.000	.889	.889	1.000
.70	1.000	.824	.824	1.000
.60	1.000	.750	.750	1.000
.50	1.000	.667	.667	1.000
.40	1.000	.571	.571	1.000
.30	1.000	.462	.462	1.000
.20	1.000	.333	.333	1.000
.10	1.000	.182	.182	1.000
.05	1.000	.095	.095	1.000

Further, Figures 1 to 3 indicate that the Spearman-Brown formula's over-estimation tendencies become worse as the standard deviations for the two halves of a test become dissimilar and the correlation between the two halves moves into the moderate (i.e., $> .30 < .70$) and low ranges (i.e., $\leq .30$). Thus, in these circumstances, the use of the Flanagan-Rulon formula would provide the user with a more accurate estimation of the population ρ .

For example, in Figure 2, when the standard deviation ratio is a moderate 1.5 and the correlation between the two halves of a test is also a moderate .600, use of the Spearman-Brown formula yields an $r = .750$. The Flanagan-Rulon formula produces an $r = .713$,

or a discrepancy of nearly 4% showing a ratio of the Spearman-Brown being 1.052 times higher than the Flanagan-Rulon estimate. Thus, the Flanagan-Rulon formula in this case is the more accurate of the two in terms approximating the population ρ . Looking at Figure 3, when the standard deviation ratio is a large 1.9 and the correlation between the two halves of a test is a low .300, use of the Spearman-Brown formula yields an $r = .462$. The Flanagan-Rulon formula produces an $r = .397$, or an even more prominent discrepancy of 6.5% and a ratio of the Spearman-Brown formula yielding results 1.164 times higher than the Flanagan-Rulon estimation.

Figure 2. Example when $r_{xy} = .600$ and $SB = .750$

Figure 3. Example when $r_{xy} = .300$ and $SB = .462$

Application

The SPSS syntax found in the Appendix serves as a check on the variance parameter ratio range to determine which split half reliability formula to employ. The user types into the Begin Data section of the syntax the two standard deviation values for two halves of a test followed by the correlation between the two halves. The syntax is run and produces the following values: r_{xy} , SD_1 , SD_2 , the ratio of SD_1 to SD_2 , Spearman-Brown, Flanagan-Rulon, and the ratio of Spearman-Brown to Flanagan-Rulon. From these results, the user can determine if the ratio between the standard deviations of the two halves of a test are .9 to 1.1, which would also produce a Spearman-Brown to Flanagan-Rulon ratio $< 1\%$, signifying use of the Spearman-Brown (i.e., no over-estimation of the population ρ). If the ratio range were beyond this threshold, the Spearman-Brown to Flanagan-Rulon ratio would be $\geq 1\%$, which would indicate that the Flanagan-Rulon formula would be the more accurate estimator to use. Future research examining the ratio range of the Spearman-Brown and Flanagan-Rulon formulas should include their performance with empirical data under biased distributional situations, with various sample sizes, and under an assortment of missing data conditions.

References

- Charter, R. A. (1996). Note on the underrepresentation of the split-half reliability formula for unequal standard deviations. *Perceptual and Motor Skills*, 82, 401-402.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R.H. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: Macmillan.
- Gordon, E. (1970). Taking into account musical aptitude differences among beginning instrumental students. *American Educational Research Journal*, 7, 41-53.
- Kelley, T. L. (1942). The reliability coefficient. *Psychometrika*, 7, 75-83.
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9, 99-103.
- Stanley, J. C. (1971). Reliability. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, DC: American Council on Education.

Appendix. Syntax for Calculating Tables 1 and 2 or any SB or FR Value

Copyright David A. Walker, 2005

Contact dawalker@niu.edu

Northern Illinois University, 101J Gabel, DeKalb, IL 60115

APA 5th Edition Citation

Walker, D. A. (2005). Comparison between the Spearman-Brown and Flanagan-Rulon split half reliability formulas [Computer program]. DeKalb, IL: Author.

DATA LIST LIST /sd1 sd2 (2f9.2) r (f9.3).

NOTE: Below, insert the SD1 and SD2 values and the r value.

BEGIN DATA

15 10 .95

14 10 .90

13 10 .80

12 10 .70

11 10 .60

10 10 .50

END DATA.

COMPUTE FR1 = (4*r)*(sd1)*(sd2).

COMPUTE FR2 = (sd1**2)+(sd2**2)+(2*r)*(sd1)*(sd2).

COMPUTE FR = FR1/FR2.

COMPUTE SBPF2 = (2*r)/(1+r).

COMPUTE SDRATIO = sd1/sd2.

COMPUTE RATIO = SBPF2/FR.

EXECUTE.

FORMAT FR TO RATIO (f9.3).

VARIABLE LABELS sdratio 'SD Ratio'/ratio 'Ratio of SB to FR'/sbpf2 'Spearman-Brown'/r

'Correlation Between the Two Halves of a Test (rxy)'/sd1 'Standard Deviation for Test 1'/sd2 'Standard Deviation for Test 2'/fr 'Flanagan-Rulon'.

REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)

MARGINS (*,120)

/VARIABLES= r sd1 sd2 sdratio sbpf2 fr ratio

/TITLE "Comparison Between Spearman-Brown and Flanagan-Rulon".

AB/BA Crossover Trials - Binary Outcome

James F. Reed III

Interim Chief of Health Studies and Director of Research
Lehigh Valley Hospital and Health Network

On occasion, the response to treatment in an AB/BA crossover trial is measured on a binary variable - success or failure. It is assumed that response to treatment is measured on an outcome variable with (+) representing a treatment success and a (-) representing a treatment failure. Traditionally, three tests for comparing treatment effect have been used (McNemar's, Mainland-Gart, and Prescott's). An issue arises concerning treatment comparisons when there may be a residual effect (carryover effect) of a previous treatment affecting the current treatment. A general consensus as to which procedure is preferable is debatable. However, if both group and carry-over effects are absent, Prescott's test is the best one to use. Under a model with residual effects, Prescott's test is biased. Therefore, a conservative approach includes testing for residual effects. When there is no period effect, McNemar's test is optimal, while McNemar's test is biased.

Key words: Crossover trial, AB/B McNemar's Test

Introduction

In a classic binary crossover trial, patients are randomly assigned to receive either treatment A (τ_A) in the first period followed by treatment B (τ_B) in a second period, or treatment B in the first period followed by treatment A in the second period. On occasion, the response to treatment is measured on a binary variable - success or failure. In a crossover design each subject acts as their own control, which is valuable in the presence of substantial between subject variability. The purpose of the crossover trial is to estimate the treatment contrast $\tau_A - \tau_B$. Period effects, π_1 and π_2 , as well as carryover effects, λ_A and λ_B , are regarded as nuisance

parameters. Traditionally, three tests for comparing treatment effect have been used (McNemar's, Mainland-Gart, and Prescott's). An issue arises concerning treatment comparisons when there may be a residual effect (carryover effect) of a previous treatment affecting the current treatment. Any residual effect, positive or negative, can bias the estimate of treatment effect. The best way of dealing with a potential residual effect is to design the study so that a suitable washout period is included so that the effect of the first period treatment does not affect the treatment applied to the second treatment period. If the researchers cannot safely insure this, then the crossover design is not an appropriate design.

Numerical Methods

For a binary crossover trial, assume that the response data may be arranged as shown in Table 1. Here the response to treatment is measured on an outcome variable with (+) representing a treatment success and a (-) representing a treatment failure.

James Reed III is the Interim Chief of Health Studies and Director of Research at Lehigh Valley Hospital and Health Network. He has published over 100 journal articles and book chapters. His interests include applied statistical analyses, medical education, and statistical methods in simulation studies. Email: James_F.Reed@lvh.com

Mainland - Gart $H_0: \{(\tau_B - \tau_A) = 0 \text{ and } (\lambda_A - \lambda_B) = 0\}$

Gart (1963) gave a rigorous description of two procedures which had been proposed earlier by Mainland (1963) that simultaneously considers the hypothesis $H_0: \{(\tau_B - \tau_A) = 0 \text{ and } (\lambda_A - \lambda_B) = 0\}$. The first, an exact test, is the product of two symmetric binomial variables, while the second is a goodness-of-fit χ^2 with 2 degrees of freedom for large n 's (Table 2a).

$$X^2_{H_0} = [(n_{12} - m_{12})^2 / (n_{12} + m_{12})] + [(m_{21} - n_{21})^2 / (m_{21} + n_{21})]$$

Rejection of $X^2_{H_0}$ indicates that there is some deviation from the null hypothesis in either or both of the components of the two-sided hypothesis.

Gart, Fleiss $H_{0\lambda}: \lambda_A - \lambda_B = 0$

Gart (1963) described an exact test and a normal deviate test statistic for order or carry-over effect (δ) (e.g. $H_{0\lambda}: \lambda_A - \lambda_B = 0$). The exact test corresponds to Fisher's Exact test for a 2×2 table with fixed marginals given in Table 2b.

The normal deviate test statistic ($Z_{G\lambda}$) is defined as:

$$Z_{G\lambda} = \{|n_{12} - ((n_{12}+m_{12})(n_{12}+m_{21})/M)| - 1/2\} / \{[(n_{12}+m_{12})(n_{21}+m_{21})(n_{12}+n_{21})(M-n_{12}-m_{21})] / [M^2(M-1)]^{1/2}\}.$$

Fleiss (1986) also described a statistic (X^2_λ) for testing the hypothesis of equal residual effects ($H_{0\lambda}: \lambda_A - \lambda_B = 0$). Let $P_1 = (2n_{11} + n_{12} + n_{21})/2n_{..}$, $P_2 = (2m_{11} + m_{12} + m_{21})/2m_{..}$, $V(P_1) = 1/4n_{..}^3 (n_{..}[n_{12} + n_{22}] - [n_{11} - n_{22}]^2)$, and $V(P_2) = 1/4m_{..}^3 (m_{..}[m_{12} + m_{22}] - [m_{11} - m_{22}]^2)$, then

$$X^2_\lambda \text{ is: } X^2_\lambda = (|P_1 - P_2| - 1/4 (1/n_{..} + 1/m_{..}))^2 / (V(P_1) + V(P_2)). X^2_\lambda$$

is compared to a χ^2 distribution with 1 degree of freedom.

Gart, Prescott, Fleiss $H_{0\tau}: \{\tau_B - \tau_A = 0\}$

The hypothesis of the equality of treatment effects (τ) between the two groups may be tested by one of several procedures. Gart proposed a one-tail exact test (Fisher's exact test applied to a 2×2 - Table 2a) and a normal deviate with a continuity correction when the sample size is large ($Z_{G\tau}$) (Gart, 1963).

$$Z_{G\tau} = \{|n_{12} - ((n_{12}+m_{12})(n_{12}+n_{21})/M)| - 1/2\} / \{[(n_{12}+m_{12})(n_{21}+m_{21})(n_{12}+n_{21})(M-n_{12}-n_{21})] / [M^2(M-1)]^{1/2}\}.$$

Prescott (1981) derived a normal deviate test statistic (Z_P) for testing $H_{0\tau}: \{\tau_B - \tau_A = 0\}$ (Table 3). $Z_P = (|T - E(T)| - 1/2)/V(T)$, where: $T = n_{12} - m_{12}$, $E(T) = n_{.1} (n_{1.} - n_{3.})/n_{..}$, and $V(T) = n_{.1} n_{.2} \{ (n_{1.} - n_{3.}) - (n_{1.} - n_{3.})^2/n_{..} \} / \{ n_{..} (n_{..} - 1) \}$.

Table 1. Data layout for a two-period crossover with a binary response variable.

AB				BA			
π_2 Response				π_2 Response			
π_1 Response	+	-	Total	π_1 Response	+	-	Total
+	n_{11}	n_{12}	$n_{1.}$	+	m_{11}	m_{12}	$m_{1.}$
-	n_{21}	n_{22}	$n_{2.}$	-	m_{21}	m_{22}	$m_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$	Total	$m_{.1}$	$m_{.2}$	$m_{..}$

Table 2. Classification of binary outcomes by period preference (Mainland-Gart Test).

a) Test $H_{0\tau}: \{\tau_B - \tau_A = 0\}$

Preference			
Sequence	Period 1 (π_1)	Period 2 (π_2)	Total
AB	n_{12}	n_{21}	$n_{12} + n_{21}$
BA	m_{12}	m_{21}	$m_{12} + m_{21}$
Total	$n_{12} + m_{12}$	$n_{21} + m_{21}$	M

b) Test $H_{0\delta}: \{\lambda_A - \lambda_B = 0\}$

Preference			
Sequence	Period 1 (π_1)	Period 2 (π_2)	Total
AB	n_{12}	m_{21}	$n_{12} + m_{21}$
BA	m_{12}	n_{21}	$m_{12} + n_{21}$
Total	$n_{12} + m_{12}$	$n_{21} + m_{21}$	M

Fleiss (1986) describes a statistic ($Z_{F\tau}$) that tests the hypothesis of equal treatment effects ($H_{0\lambda}: \lambda_A - \lambda_B = 0$). Let $p'_A = n_{12}/n'$, $n' = n_{12} + n_{21}$, $p'_B = m_{12}/m'$, $m' = m_{12} + m_{21}$, $p' = (n_{12} + m_{12})/(n' + m')$, then $Z_{F\tau} = \{[|p'_A - p'_B| - \frac{1}{2}(1/n' + 1/m')]/\sqrt{p'(1-p')}\}/\sqrt{n'm'/(n'+m')}$. The hypothesis of equal treatment effects is rejected if $Z_{F\tau} > Z_{\alpha/2}$.

Fleiss $H_{0\tau|\lambda}$

If there were a significant residual effect, the general advice is that the appropriate test for testing $H_{0\tau|\lambda}$ is to use data from the first treatment period. In this case $H_{0\tau|\lambda}$ may be tested using the following test statistic ($Z_{\tau|\lambda}$) described by Fleiss (1986). $Z_{\tau|\lambda} = \{|p_A - p_B| - [1/n_{..} + 1/m_{..}]/2\} / (p'(1-p')^{1/2}) \{n_{..}m_{..}/(n_{..} + m_{..})\}^{1/2}$, with $p_A = n_{1.}/n_{..}$, $p_B = m_{1.}/m_{..}$, and $p' = (n_{1.} + m_{1.})/(n_{..} + m_{..})$. $H_{0\tau|\lambda}$ is rejected if $Z_{\tau|\lambda} > Z_{\alpha/2}$.

Example

Consider the data given by Fleiss (1986, pg 277).

AB				BA			
π_2 Response				π_2 Response			
π_1 Response	+	-	Total	π_1 Response	+	-	Total
+	25	15	40	+	30	0	30
-	5	5	10	-	5	15	20
Total	30	20	50	Total	35	15	50

The test statistics are:

$H_0: \{\tau_A = \tau_B, \lambda_A = \lambda_B\}$

Mainland-Gart chi-square = 6.00, p-value = 0.04841

$H_0: \{\lambda_A = \lambda_B\}$

Mainland-Gart E (1-tail), p-value = 0.29181

Mainland-Gart Z = 0.61, p-value = 0.54029

Fleiss $X^2 = 0.26$, p-value = 0.61914

$H_0: \{\tau_A = \tau_B\}$

Mainland-Gart E (1-tail), p-value = 0.00474

Mainland-Gart Z = 2.50, p-value = 0.00621

Prescott Z = 2.92, p-value = 0.00349

Fleiss Z = 2.55, p-value = 0.01072

$H_0: \{\tau_{A|\lambda} = \tau_{B|\lambda}\}$

Fleiss-T|Residual = 1.96, p-value = 0.04953

The simultaneous test of $H_0: \{\tau_A = \tau_B, \lambda_A = \lambda_B\}$ indicates that there is a difference in treatment effects or that there is a significant residual effect (or both). All of the test statistics and associated p-values associated with $H_0: \{\lambda_A = \lambda_B\}$ are not statistically significant at any reasonable level. Further, the hypothesis $H_0: \{\tau_A = \tau_B\}$ is also significant (0.01072).

Discussion

Traditionally, three tests for comparing treatment effect have been used to analyze data from a binary two period crossover trial (McNemar, Mainland-Gart, and Prescott). McNemar's test in effect ignores information on periods and tests for difference between expected values of $n_{21} + m_{12}$ and $n_{12} + m_{21}$. The Mainland-Gart test is Fisher's exact test applied to a 2 x 2 table (Table 2). Prescott's test expands the Mainland-Gart test by using tied responses and is a test for linear trend in a 2 X 3 table (Table 3).

Other models have been proposed to analyze binary data from a two period crossover trial. For instance, Becker and Balagtas (1993) proposed a model in terms of linear models for marginal logits and linear models for log-odds ratios for the analysis of two-period binary crossover trials. Their simulation results demonstrated that their likelihood ratio approach compared favorably with standard procedures such as the Mainland-Gart test for a treatment difference, Prescott's test for a treatment

difference and the Hills-Armitage test for treatment by period interaction.

Which, if any, of the procedures is best has been debated for nearly 40 years and the debate will probably continue. A general consensus is that if both group and carry-over effects are absent, Prescott's test is the best one to use. Under the model with residual effects, Prescott's test is biased, as it can give p-values that are too high or too low (Fidler, 1984). Estimates of treatment effects are biased when residual effects are present. With positive residual effects, the treatment effect is underestimated and the crossover test is conservative; with negative residual effects, the treatment effect is overestimated and the crossover test is liberal (Lehmacher, 1991). Unlike the group effect, the residual effect cannot be eliminated by randomization alone. Therefore, a conservative approach includes testing for residual effects. A non-significant result from the residual effect and external evidence for the absence of a residual effect is essential when using Prescott's test. When there is no period effect, McNemar's test is optimal, while McNemar's test is biased (Fidler, 1984). When a crossover trial has been properly designed, it remains a powerful instrument in the analysts' toolbox and is a reasonable and sensitive means of determining the efficacy of two treatment regimens (Cleophas & Val Lier, 1996).

Table 3. Classification of binary outcomes by period preference, including tied responses (Prescott's Test).

Test $H_{0\tau}: \{\tau_B - \tau_A = 0\}$

Sequence	Preference			Total
	Period 1 (π_1)	Tied	Period 2 (π_2)	
AB	n_{12}	$n_{11} + n_{22}$	n_{21}	$n_{1.}$
BA	m_{12}	$m_{11} + m_{22}$	m_{21}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

Program

An executable program that produces the statistics outlined in this article is available from the author on request. The input file is in free format (sequence AB=1, BA=2; observation 1, 2, n, 1, 2, m; period 1 outcome (1 = success, 0 = failure); period 2 outcome (1 = success, 0 = failure)). Data must be in an integer format.

References

Becker, M. P. & Balagtas, C. C. (1993). Marginal modeling of binary cross-over data. *Biometrics*, 49, 997-1009.

Cleophas, T. J. & van Lier, H. J. (1996). Crossover trials with a binary response: A powerful method despite the carryover effect. *Journal of Clinical Pharmacology*, 36, 198-202.

Fidler, V. (1984). Change-over clinical trial with binary data: Mixed-model-based comparison of tests. *Biometrics*, 40, 1063-1070.

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York, N.Y.: John Wiley & Sons.

Gart, J. J. (1996). An exact test for comparing matched proportions in crossover designs. *Biometrika*, 56, 75-80.

Lehmacher, W. (1991). Analysis of the crossover design in the presence of residual effects. *Statistics in Medicine*, 10, 891-899.

Mainland, D. (1963). *Elementary medical statistics* (2nd ed.). Philadelphia, P.A.: Saunders.

Prescott, R. J. (1981). The comparison of success rates in cross-over trials in the presence of an order effect. *Applied Statistics*, 30, 9-15.

Correlation Between the Number of Epileptic and Healthy Children in Family Size that Follows a Size-Biased Modified Power Series Distribution

Ramalingam Shanmugam
Texas State University

Anwar Hassan Peer Bilal Ahmad
University of Kashmir

An expression for the correlation between the random number of epileptic and healthy children in family whose size follows a size-biased Modified Power Series Distribution (SBMPSD) is obtained and illustrated. As special cases, results are extracted for size biased Modified Negative Binomial Distribution (SBGNBD), size biased Modified Poisson Distribution (SBGPD) and size biased Modified Logarithmic Series Distribution (SBGLSD).

Key words: Size-biased MPSD, GNBD, GLSD, GPD correlation

Introduction

Weighted distributions arise when observations are generated from a stochastic process without an equal chance of being selected from a population. When the sampling mechanism selects units with probability proportional to some measure of the unit size, the resulting distribution is called size-biased. Such distributions arise in life several studies (see Blumenthal (1967), Scheaffer (1972), Gupta (1975, 1984) for details).

Kojima and Kellehar (1962) showed that Negative Binomial Distribution (NBD) is appropriate distribution for the discrete type random observation. The random family size could follow a NBD. Gupta (1976) obtained a

general expression for the correlation coefficients 'ρ' between the random number of boys and girls in family whose size follows a Modified Power Series Distribution (mpsd). Gupta's (1974) introduced mpsd and explored its properties. The size biased modified power series distribution (SBMPSD) is considered as a distribution for family size. This class includes among others, size biased GNBD, GPD and GLSD. A general expression for the correlation coefficient for a random number of epileptic and healthy children in a family when the family size follows a size biased Modified Power Series Distribution (SBMPSD) is obtained and illustrated. As special cases, results are extracted for size biased Modified Negative Binomial Distribution (SBGNBD), size biased Modified Poisson Distribution (SBGPD) and size biased Modified Logarithmic Series Distribution (SBGLSD).

Ramalingam Shanmugam is Professor in the College of Health Professions. His research interests include multivariate data modeling and interpretations, informatics, optimal decision making with health/medical data support, health and medical applications of statistics. Anwar Hassan is a member of Faculty in the Department of Statistics. His research interests include discrete statistical distributions and genetic applications of statistics. Peer Bilal Ahmad works in the Department of Statistics. Ahmad's research interests include statistical inference, distribution theory among others.

Main Result: General Expression for ρ

Let N be a discrete random variable denoting the family size. Assume that this random variable is governed by a size-biased Modified Power Series Distribution (SBMPSD) whose probability mass function is

$$P[N = n] = \frac{b(n)(g(\alpha))^n}{\mu f(\alpha)}, \quad n \in T$$

where T is a subset of the set of positive integers, $b(n) > 0$, $b(n) = n a(n)$, $g(\alpha)$ and $f(\alpha)$ are positive, finite and differentiable functions. It is easy to establish that

$$\mu = \frac{f'(\alpha) g(\alpha)}{g'(\alpha) f(\alpha)}$$

is the mean of MPSD (see Gupta 1974). The mean and variance of size biased version of the above modified power series distribution are respectively

$$E_{\alpha}(N) = 1 + \frac{g(\alpha) f''(\alpha)}{g'(\alpha) f'(\alpha)} - \frac{g(\alpha) g''(\alpha)}{g'^2(\alpha)} \quad (1)$$

and

$$V_{\alpha}(N) = \frac{g(\alpha)}{g'(\alpha)} E'_{\alpha}(N) \quad (2)$$

Where

$$E'_{\alpha}(N) = \frac{\partial}{\partial \alpha} E(N).$$

Let X be the random number of epileptic children in a family of size N . Assume that X follows a binomial random variable with parameters N and p . The correlation coefficient, ρ , between X and $N-X$ (the number of healthy children in a family of size N) is known (Rao et al [1973]) to be

$$\rho = \frac{(p[1-p])^{1/2} [V(N) - E(N)]}{\left[\frac{pV(N) + (1-p)V(N)}{[1-p]E(N)} \right]^{1/2} \left[\frac{pE(N)}{pE(N)} \right]^{1/2}}. \quad (3)$$

When $p = q = 1/2$, the correlation coefficient reduces to

$$\rho = \frac{V(N) - E(N)}{V(N) + E(N)} \quad (4)$$

Similar results for the correlation coefficient in the case of size biased MPSD can be obtained. After algebraic simplifications, it is

$$\rho = \frac{(pq)^{1/2} [g(\alpha)g'(\alpha)f'(\alpha)E'_{\alpha}(N) - \{g''(\alpha)f'(\alpha) + f''(\alpha)g'(\alpha)g(\alpha) - g''(\alpha)g(\alpha)f'(\alpha)\}]}{[pg(\alpha)g'(\alpha)f'(\alpha)E'_{\alpha}(N) + q\{g''(\alpha)f'(\alpha) + f''(\alpha)g'(\alpha)g(\alpha) - g''(\alpha)g(\alpha)f'(\alpha)\}]^{1/2}} \times$$

$$\frac{1}{\left[(1-p)g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N) \right]^{1/2} + p \left\{ \begin{array}{l} g'^2(\alpha)f'(\alpha) \\ +f''(\alpha)g'(\alpha)g(\alpha) \\ -g''(\alpha)g(\alpha)f'(\alpha) \end{array} \right\}} \quad (5)$$

where the prime denotes derivate with respect to α . For $p = q = 1/2$, expression (5) reduces to

$$\rho = \frac{\left[g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N) - \left\{ \begin{array}{l} g'^2(\alpha)f'(\alpha) + \\ f''(\alpha)g'(\alpha)g(\alpha) - \\ g''(\alpha)g(\alpha)f'(\alpha) \end{array} \right\} \right]}{\left[g(\alpha)g'(\alpha)f'(\alpha)E'_\alpha(N) + \left\{ \begin{array}{l} g'^2(\alpha)f'(\alpha) + \\ f''(\alpha)g'(\alpha)g(\alpha) - \\ g''(\alpha)g(\alpha)f'(\alpha) \end{array} \right\} \right]} \quad (6)$$

Particular Cases

In this section, expressions are obtained for particular cases such as modified negative binomial, Poisson, and log series distributions.

Modified Negative Binomial Distribution:

In this section, a particular case of the modified negative binomial distribution (see Jain and Consul 1971) is obtained. That is,

$$P[N = n] = \frac{\delta \Gamma(\delta + \beta n)}{x! \Gamma(\delta + (\beta - 1)n + 1)} \frac{[\alpha(1 - \alpha)^{\beta-1}]^n}{(1 - \alpha)^{-\delta}}$$

Here $f(\alpha) = (1 - \alpha)^{-\delta}$, $g(\alpha) = \alpha(1 - \alpha)^{\beta-1}$.

Note that

$$E_\alpha(N) = \frac{\delta \alpha}{(1 - \beta \alpha)} + \frac{1 - \alpha}{(1 - \beta \alpha)^2}$$

and

$$E'_\alpha(N) = \frac{\delta - \delta \alpha \beta + 2\beta - 1 - \alpha \beta}{(1 - \alpha \beta)^3}$$

Using these values in (5) and (6), the following is obtained

$$\rho = \frac{(pq)^{1/2} \left[\begin{array}{l} \alpha(1 - \alpha)(\delta - \delta \alpha \beta + 2\beta - 1 - \alpha \beta) \\ - (1 - \alpha \beta)^2(1 + \delta \alpha - \alpha - \delta \alpha^2 \beta) \end{array} \right]}{\left[\begin{array}{l} p\alpha(1 - \alpha)(\delta - \delta \alpha \beta + 2\beta - 1 - \alpha \beta) \\ + q(1 - \alpha \beta)^2(1 + \delta \alpha - \alpha - \delta \alpha^2 \beta) \end{array} \right]^{1/2}} \times$$

$$\frac{1}{\left[\begin{array}{l} q\alpha(1 - \alpha)(\delta - \delta \alpha \beta + 2\beta - 1 - \alpha \beta) \\ + p(1 - \alpha \beta)^2(1 + \delta \alpha - \alpha - \delta \alpha^2 \beta) \end{array} \right]^{1/2}}$$

For $p = q = 1/2$, the following is obtained

$$\rho = \frac{\left[\begin{array}{l} \alpha(1 - \alpha)(\delta - \delta \alpha \beta + 2\beta - 1 - \alpha \beta) \\ - (1 - \alpha \beta)^2(1 + \delta \alpha - \alpha - \delta \alpha^2 \beta) \end{array} \right]}{\left[\begin{array}{l} \alpha(1 - \alpha)(\delta - \delta \alpha \beta + 2\beta - 1 - \alpha \beta) \\ + (1 - \alpha \beta)^2(1 + \delta \alpha - \alpha - \delta \alpha^2 \beta) \end{array} \right]}$$

Modified Poisson Distribution

The pdf. of GPD given by Consul and Jain (1971) and Shoukri and Consul (1989) is slightly altered to represent in terms of α and β . In this section, a particular case of modified Poisson distribution (see Shoukri & Consul, 1989) is obtained. That is,

$$P[N = n] = \frac{(1 + \beta n)^{n-1} \alpha^n e^{-\alpha(1 + \beta n)}}{n!};$$

$$\alpha > 0, |\alpha \beta| < 1, n = 0, 1, 2, \dots$$

Here $f(\alpha) = e^\alpha$, $g(\alpha) = \alpha e^{-\alpha \beta}$

$$E_\alpha(N) = \frac{\alpha}{(1 - \alpha \beta)} + \frac{1}{(1 - \alpha \beta)^2}$$

$$V_{\alpha}(N) = \frac{2\alpha\beta}{(1-\alpha\beta)^4} + \frac{\alpha}{(1-\alpha\beta)^3}$$

Using these values in (5) and (6), the following is obtained

$$\rho = \frac{(pq)^{1/2} [\alpha(2\beta+1-\alpha\beta) - (1-\alpha\beta)^2 \{1+\alpha-\alpha^2\beta\}]}{[p\alpha(2\beta+1-\alpha\beta) + q(1-\alpha\beta)^2 \{1+\alpha-\alpha^2\beta\}]^{1/2}} \times \frac{1}{[q\alpha(2\beta+1-\alpha\beta) + p(1-\alpha\beta)^2 \{1+\alpha-\alpha^2\beta\}]^{1/2}}$$

For $p = q = 1/2$, the following is obtained

$$\rho = \frac{[\alpha(2\beta+1-\alpha\beta) - (1-\alpha\beta)^2 \{1+\alpha-\alpha^2\beta\}]}{[\alpha(2\beta+1-\alpha\beta) + (1-\alpha\beta)^2 \{1+\alpha-\alpha^2\beta\}]}$$

Modified Logarithmic Series Distribution.

In this section, a particular case of modified logarithmic series distribution (see Jain & Gupta, 1973) is obtained. That is,

$$P[N = n] = \frac{\Gamma(n\beta)}{\Gamma(n+1)\Gamma([\beta-1]n+1)} \frac{\alpha^n (1-\alpha)^{(\beta-1)n}}{[-\log(1-\alpha)]}$$

Here

$$f(\alpha) = -\log(1-\alpha) = \theta^{-1}(\text{say}), \quad g(\alpha) = \alpha(1-\alpha)^{\beta-1}$$

$$E_{\alpha}(N) = \frac{1-\alpha}{(1-\alpha\beta)^2}$$

$$V_{\alpha}(N) = \frac{\alpha(1-\alpha)(2\beta-\alpha\beta-1)}{(1-\alpha\beta)^4}$$

Using these values in (5) and (6), the following is obtained

$$\rho = \frac{(pq)^{1/2} \left[\frac{\alpha(1-\alpha)(2\beta-\alpha\beta-1)}{-(1-\alpha\beta)^2(1-\alpha)} \right]}{\left[\frac{p\alpha(1-\alpha)(2\beta-\alpha\beta-1)}{+q(1-\alpha\beta)^2(1-\alpha)} \right]^{1/2} \left[\frac{q\alpha(1-\alpha)(2\beta-\alpha\beta-1)}{+p(1-\alpha\beta)^2(1-\alpha)} \right]^{1/2}}$$

$$\rho = \frac{(pq)^{1/2} \left[\frac{\alpha(2\beta-\alpha\beta-1)}{-(1-\alpha\beta)^2} \right]}{\left[\frac{p\alpha(2\beta-\alpha\beta-1)}{+q(1-\alpha\beta)^2} \right]^{1/2} \left[\frac{q\alpha(2\beta-\alpha\beta-1)}{+p(1-\alpha\beta)^2} \right]^{1/2}}$$

For $p = q = 1/2$, the following is obtained

$$\rho = \frac{[\alpha(2\beta-\alpha\beta-1) - (1-\alpha\beta)^2]}{[\alpha(2\beta-\alpha\beta-1) + (1-\alpha\beta)^2]}$$

Illustration.

In this section, the main results of Section 3.2 are illustrated using data in Table 1. Results of Section 3.1 and 3.3 can be similarly made but are omitted.

Note that $\hat{p} = 73/136 = 0.54$. Using Moment Method of estimation for estimate of α and β , the following is obtained

$$\mu'_1 = \frac{\alpha}{(1-\alpha\beta)} + \frac{1}{(1-\alpha\beta)^2}$$

$$\mu_2 = \frac{2\alpha\beta}{(1-\alpha\beta)^4} + \frac{\alpha}{(1-\alpha\beta)^3}$$

For convenience Let $1-\alpha\beta=\theta$ or $\alpha\beta = 1-\theta$ the following is obtained

$$\mu'_1 = \frac{\alpha\theta + 1}{\theta^2}$$

$$\mu_2 = \frac{2(1-\theta) + \alpha\theta}{\theta^4}$$

Table 1. Number of Epileptic (x) and Healthy (N-X) children in random sample of family size (N) among 50 families

N=n	X=x	f=# families
1	0	9
2	1	13
3	2	15
4	2	9
5	3	4
#children= $\sum nf = 136$	# Epileptic Children = $\sum xf = 73$	# families= $\sum f = 50$

These give an equation in θ as

$$\mu_2 \theta^4 - \mu_1' \theta^2 + 2\theta - 1 = 0$$

Replacing μ_1' and μ_2 by corresponding sample values $\bar{x}$ and S^2 respectively, the following is obtained

$$S^2 \theta^4 - \bar{X} \theta^2 + 2\theta - 1 = 0$$

It is a polynomial of degree four and can be solved by using the Newton –Raphson method and so an estimate of θ can be obtained. An estimate of α is then obtained as

$$\hat{\alpha} = \frac{\hat{\theta}^2 \bar{x} - 1}{\hat{\theta}}$$

After estimating $\hat{\alpha}$ and $\hat{\theta}$, $\hat{\beta}$ is obtained as

$$\hat{\beta} = \frac{1 - \hat{\theta}}{\hat{\alpha}}$$

After using Revised Table -1, the estimated values of α and β are $\hat{\alpha} = 2.0807$, $\hat{\beta} = -0.0478$, and $\rho = -0.310$.

In comparison to the critical chi-squared score $\chi_{2df}^2 = 5.991$, the computed chi-squared score for testing the goodness of fit is $\chi^2(\text{calculated}) = 0.825$ highly insignificant implying the best fit.

References

- Blumenthal, S. (1967). Proportional sampling in life length studies. *Technometrics*, 9, 205-219.
- Gupta, R. C. (1975). Some characterizations of discrete distributions. *Communications in Statistics*, A5(1), 45-48.
- Gupta, R. C. (1984). Some characterizations of renewal densities with emphasis in reliability. *Mathematische Operations Far Schung Und Statistics*, 15, 571-579.
- Gupta, R. C. (1974). Modified power series distributions and some of its applications. *Sankhya, series B*, 36, 288-298.
- Gupta, R. C. (1976). Some applications of modified power series distributions in genetics. *Sankhya, series B*, 38, 187-191.

Kojima, K. I. & Kelleher, T. M. (1962). Survival of mutant genes. *The American Naturalist*, *XCVI*, 329-45.

Jain, G. C. & Consul, P. C. (1971). A modified negative binomial distribution. *SIAM, Journal of Applied Mathematics*, *21*, 501-513

Jain, G. C. & Gupta, R. P. (1973). A logarithmic type distribution. *Trabjos Estadist.*, *24*, 99-105.

Shoukri, M. M. & Consul, P. C. (1989). Bayesian analysis of modified Poisson distribution. *Communication and Statistical Theoretical Mathematics*, *18*(4), 1465-1480.

Scheaffer, R. L. (1972). Size biased sampling. *Technometrics*, *14*, 635-644.

Rao, B. R., Mazumdar, S., Waller, J. H., & Li, C. C. (1973). Correlation between the numbers of two types of children in a family, *Biometrics*, *29*, 271-279.

A Robust Exponentially Weighted Moving Average Control Chart for the Process Mean

Michael B.C. Khoo S.Y. Sim
School of Mathematical Sciences,
Universiti Sains Malaysia

To date, numerous extensions of the exponentially weighted moving average, EWMA charts have been made. A new robust EWMA chart for the process mean is proposed. It enables easier detection of outliers and increase sensitivity to other forms of out-of-control situation when outliers are present.

Key words: Exponentially weighted moving average (EWMA), cumulative sum (CUSUM), Shewhart, process mean, sample mean, sample range, average run length (ARL)

Introduction

The EWMA chart is a good alternative to the Shewhart chart in the detection of small shifts. The EWMA chart constructed from the sample mean is first developed by Roberts (1959). Since then various extensions of the EWMA charts have been proposed. Sweet (1986) proposed two models to construct simultaneous control charts to monitor the mean and the variance of a process using the EWMA. Crowder (1987 & 1989) provided average run length (ARL) tables and graphs for the selection of the optimum values of the EWMA control chart parameters in the design of an EWMA chart. Ng & Case (1989) presented several EWMA control chart schemes based on individual measurement,

sample mean, sample range and moving range statistics. Lucas & Saccucci (1990) showed that a fast initial response (FIR) feature is useful for the EWMA chart, especially for small values of smoothing constants. Rhoads, Montgomery & Mastrangelo (1996) proposed a scheme which is superior to that of Lucas & Saccucci (1990). MacGregor & Harris (1993) suggested an approach of using the EWMA based statistics in the monitoring of the process standard deviation. Gan (1990) proposed three modified EWMA charts for the Poisson data. A better procedure for using the EWMA chart for Poisson count is given by Borror, Champ & Rigdon (1998). Somerville, Montgomery & Runger (2002) developed a smoothing and filtering method using the EWMA and Poisson probabilities which separates the two distributions in a particle count data stream into a base process and an outlier process followed by applying statistical monitoring schemes to each of them. A Bernoulli EWMA is suggested to monitor the outlier process.

Michael B. C. Khoo is a member of the American Society for Quality (ASQ) and the American Statistical Association (ASA). He also serves as a member of the editorial boards of Quality Engineering, Quality Management Journal and Journal of Modern Applied Statistical Methods. His research interest is in statistical process control. S. Y. Sim graduated from the School of Mathematical Sciences, Universiti Sains Malaysia with an M.Sc. degree.

The EWMA control chart scheme for the sample mean proposed by Ng & Case (1989) is constructed by assuming that the data used in the computation of the limits are outlier free. This assumption may not be true in real situations since outliers often occur in the data used to compute the control limits. Outliers may consist of single unusual values which happen due to a sporadic special cause. Such outliers act only on occasional observations in a subgroup

and not on subgroups as a whole. These outliers have to be detected, investigated and the special cause removed if possible. The presence of outliers will reduce the sensitivity of a control chart because the control limits are stretched so that the detection of the outliers themselves become more difficult. Furthermore, these stretched limits also make it more difficult for other types of out-of-control signals to be detected (Rocke, 1989, 1992). The purpose of this article is to propose a robust EWMA (EWMA_{SMQ}) chart for the process mean as an alternative which is superior to the standard EWMA (EWMA_{SM}) chart for the process mean. The EWMA_{SMQ} chart is constructed based on the limits that are set using an estimate of the process standard deviation using the average of the subgroup interquartile ranges (IQRs) rather than the average of the subgroup ranges in the case of the EWMA_{SM}. Thus, the EWMA_{SMQ} chart is less affected by outliers compared to the EWMA_{SM} chart. The next section gives an overview of the EWMA_{SM} chart.

The EWMA_{SM} Chart

The EWMA_{SM} chart is based on the following statistic (Ng & Case, 1989):

$$\hat{W}_t = \alpha \bar{X}_t + (1 - \alpha) \hat{W}_{t-1} \quad (1)$$

where $\bar{X}_t$ is the mean of sample t and α is a weighting constant. In the estimation of the limits based on the process data, m subgroups of size n each are taken and then the average of the m sample means, $\bar{\bar{X}}$ is computed. $\bar{\bar{X}}$ is used as the starting point, i.e., $\hat{W}_0 = \bar{\bar{X}}$. The average of the m sample ranges is computed to be $\bar{R}$.

The upper and lower control limits of the EWMA_{SM} chart are

$$UCL = \bar{\bar{X}} + F_1 \bar{R} \quad (2a)$$

and

$$LCL = \bar{\bar{X}} - F_1 \bar{R} \quad (2b)$$

respectively, where

$$F_1 = \frac{3}{d_2 \sqrt{n}} \sqrt{\frac{\alpha}{2 - \alpha}} \quad (3)$$

In equation (3), d_2 is a standard constant whose value depends on the sample size n . The values of d_2 and F_1 (Ng & Case, 1989) for the various sample sizes n are given in Tables 1 and 2 respectively.

A Proposed EWMA_{SMQ} Chart For The Process Mean

Similar to the EWMA_{SM} chart, every observed sample mean is transformed into a corresponding EWMA before it is plotted on the EWMA_{SMQ} chart. Let $\bar{X}_t$ represents the mean of sample t . Every $\bar{X}_t$ will be transformed into a corresponding EWMA, $\hat{W}_t$, using the transformation

$$\hat{W}_t = \alpha \bar{X}_t + (1 - \alpha) \hat{W}_{t-1}, \quad t = 1, 2, \dots \quad (4)$$

When developing the EWMA_{SMQ} chart based on process data, m subgroups of size n each must be taken to compute the estimate of the process mean. The average of the m sample means will be used as the starting point, i.e., $\hat{W}_0 = \bar{\bar{X}}$. For the EWMA_{SMQ} chart, the interquartile range, IQR is defined as $X_{(b)} - X_{(a)}$, where $()$ denotes the order statistics $a = [n/4] + 1$ and $b = n - a + 1$. Here, $[y]$ represents the greatest integer that is less than or equal to y . It is shown by Rocke (1992) that the mathematical expectation of IQR can be defined as

$$E(\text{IQR}) = d_2^0 \sigma_x \quad (5)$$

Table 1. Factors for the EWMA_{SM} and EWMA_{SMQ} Charts

n	a	b	d_2	d_2^Q
2	1	2	1.128	1.1284
3	1	3	1.693	1.6926
4	2	3	2.059	0.5940
5	2	4	2.326	0.9900
6	2	5	2.534	1.2835
7	2	6	2.704	1.5147
8	3	6	2.847	0.9456
9	3	7	2.970	1.1439
10	3	8	3.078	1.3121
11	3	9	3.173	1.4577
12	4	9	3.258	1.0737
13	4	10	3.336	1.2057
14	4	11	3.407	1.3235
15	4	12	3.472	1.4298
16	5	12	3.532	1.1400
17	5	13	3.588	1.2389
18	5	14	3.640	1.3269
19	5	15	3.689	1.4132
20	6	15	3.735	1.1806

Table 2. Factors of Control Limits for the EWMA_{SM}, F_1

n	α										
	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
2	0.000	0.431	0.627	0.790	0.940	1.085	1.231	1.380	1.535	1.701	1.880
3	0.000	0.235	0.341	0.430	0.512	0.591	0.670	0.751	0.835	0.925	1.023
4	0.000	0.167	0.243	0.306	0.365	0.421	0.477	0.535	0.595	0.659	0.729
5	0.000	0.132	0.192	0.242	0.289	0.333	0.378	0.423	0.471	0.522	0.577
6	0.000	0.111	0.161	0.203	0.242	0.279	0.316	0.354	0.394	0.437	0.483
7	0.000	0.096	0.140	0.176	0.210	0.242	0.274	0.307	0.342	0.379	0.419
8	0.000	0.086	0.124	0.157	0.187	0.215	0.244	0.274	0.305	0.337	0.373
9	0.000	0.077	0.112	0.142	0.169	0.195	0.221	0.247	0.275	0.305	0.337
10	0.000	0.071	0.103	0.129	0.154	0.178	0.202	0.226	0.251	0.279	0.308
11	0.000	0.065	0.095	0.120	0.143	0.165	0.187	0.209	0.233	0.258	0.285
12	0.000	0.061	0.089	0.112	0.133	0.154	0.174	0.195	0.217	0.241	0.266
13	0.000	0.057	0.083	0.105	0.125	0.144	0.163	0.183	0.203	0.225	0.249
14	0.000	0.054	0.078	0.099	0.118	0.136	0.154	0.172	0.192	0.213	0.235
15	0.000	0.051	0.074	0.094	0.112	0.129	0.146	0.164	0.182	0.202	0.223
16	0.000	0.049	0.071	0.089	0.106	0.122	0.139	0.156	0.173	0.192	0.212
17	0.000	0.047	0.068	0.085	0.102	0.117	0.133	0.149	0.166	0.184	0.203
18	0.000	0.045	0.065	0.081	0.097	0.112	0.127	0.142	0.158	0.175	0.194
19	0.000	0.043	0.062	0.079	0.094	0.108	0.122	0.137	0.153	0.169	0.187
20	0.000	0.041	0.060	0.076	0.090	0.104	0.118	0.132	0.147	0.163	0.180

where d_2^0 is a constant whose value depends on the sample size n (see Table 1). The standard deviation, σ_x is estimated by

$$\hat{\sigma}_x = \frac{\overline{\text{IQR}}}{d_2^0} \quad (6)$$

where $\overline{\text{IQR}}$ is the average of the subgroup interquartile ranges.

Assuming that all the observed data are independent from one sample to another, then as t increases,

$$\text{Var}(\hat{W}_t) = \text{Var}(\bar{X}) \frac{\alpha}{2-\alpha} \quad (7)$$

and

$$\sigma_{\hat{W}_t} = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{\alpha}{2-\alpha}}. \quad (8)$$

From equation (6), by using the average interquartile range to estimate the standard deviation of $\hat{W}_t$, $\hat{\sigma}_{\hat{W}_t}$, the following is obtained

$$\hat{\sigma}_{\hat{W}_t} = \frac{\overline{\text{IQR}}}{d_2^0 \sqrt{n}} \sqrt{\frac{\alpha}{2-\alpha}}. \quad (9)$$

Thus, the limits of the EWMA_{SMQ} chart are

$$\text{Center Line} = \hat{W}_0 = \bar{X} \quad (10)$$

$$\begin{aligned} \text{Control Limits} &= \hat{W}_0 \pm 3\hat{\sigma}_{\hat{W}_t} \\ &= \bar{X} \pm 3 \frac{\overline{\text{IQR}}}{d_2^0 \sqrt{n}} \sqrt{\frac{\alpha}{2-\alpha}} \end{aligned} \quad (11)$$

If

$$G_1 = \frac{3}{d_2^0 \sqrt{n}} \sqrt{\frac{\alpha}{2-\alpha}} \quad (12)$$

then the control limits calculated based on equation (11) are

$$\text{UCL}_{\text{SMQ}} = \bar{X} + G_1 \overline{\text{IQR}} \quad (13a)$$

and

$$\text{LCL}_{\text{SMQ}} = \bar{X} - G_1 \overline{\text{IQR}}. \quad (13b)$$

Values of G_1 for different sample sizes, n , and smoothing constants, α , are listed in Table 3.

Table 3. Factors of Control Limits for the EWMA_{SMQ}, G_1

Sample Size, n	Value of α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
2	0.0000	0.4313	0.6266	0.7897	0.9400	1.0854	1.2307	1.3795	1.5350	1.7005	1.8799
3	0.0000	0.2348	0.3411	0.4299	0.5117	0.5908	0.6699	0.7509	0.8355	0.9256	1.0233
4	0.0000	0.5793	0.8418	1.0608	1.2626	1.4580	1.6532	1.8530	2.0619	2.2842	2.5253
5	0.0000	0.3109	0.4517	0.5693	0.6776	0.7824	0.8872	0.9944	1.1065	1.2258	1.3552
6	0.0000	0.2189	0.3181	0.4009	0.4771	0.5509	0.6247	0.7002	0.7791	0.8631	0.9542
7	0.0000	0.1717	0.2495	0.3145	0.3743	0.4322	0.4901	0.5493	0.6112	0.6771	0.7486
8	0.0000	0.2573	0.3739	0.4712	0.5608	0.6476	0.7343	0.8231	0.9158	1.0146	1.1217
9	0.0000	0.2006	0.2914	0.3672	0.4371	0.5047	0.5723	0.6415	0.7138	0.7907	0.8742
10	0.0000	0.1659	0.2410	0.3037	0.3615	0.4174	0.4733	0.5306	0.5903	0.6540	0.7230
11	0.0000	0.1424	0.2068	0.2607	0.3103	0.3583	0.4062	0.4553	0.5067	0.5613	0.6205
12	0.0000	0.1850	0.2689	0.3388	0.4033	0.4657	0.5280	0.5919	0.6586	0.7296	0.8066
13	0.0000	0.1583	0.2300	0.2899	0.3450	0.3984	0.4518	0.5064	0.5635	0.6242	0.6901
14	0.0000	0.1390	0.2019	0.2545	0.3029	0.3498	0.3966	0.4445	0.4946	0.5480	0.6058
15	0.0000	0.1243	0.1806	0.2276	0.2709	0.3128	0.3547	0.3975	0.4423	0.4900	0.5418
16	0.0000	0.1509	0.2193	0.2764	0.3289	0.3798	0.4307	0.4828	0.5372	0.5951	0.6579
17	0.0000	0.1347	0.1958	0.2467	0.2937	0.3391	0.3845	0.4310	0.4795	0.5312	0.5873
18	0.0000	0.1223	0.1776	0.2239	0.2665	0.3077	0.3489	0.3910	0.4351	0.4820	0.5329
19	0.0000	0.1117	0.1623	0.2046	0.2435	0.2812	0.3188	0.3574	0.3976	0.4405	0.4870
20	0.0000	0.1304	0.1894	0.2387	0.2841	0.3281	0.3720	0.4169	0.4639	0.5140	0.5682

Comparison Of Performances

The performance of the EWMA_{SMQ} chart is compared to that of the EWMA_{SM} by performing a Monte-Carlo simulation. The following four different conditions are considered for the two control charts for the process mean:

- (i) The In-control situation where the data are all standard normal random variables.
- (ii) The Outliers situation where the data are a mixture of 95% standard normal and 5% data with five times the standard deviation. These outliers might represent an episodic phenomenon resulting from a sporadic special cause that control charts should detect.
- (iii) The Special Cause situation where there is an additional $N(\delta, 1)$ source of variability added to the subgroup means. Here, $\delta = \frac{\mu - \mu_0}{\sigma_0}$ where $\mu_0 = 0$ and $\sigma_0 = 1$ represent the nominal mean and standard deviation respectively while μ is the off-target mean. The values of $\delta \in \{0, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4\}$ are considered. The control charts should detect this special cause.
- (iv) The Outliers and Special Cause situation which consists of the data that are a mixture of 95% standard normal and 5% data with five times the standard deviation together with an additional component of variation, $N(\delta, 1)$ added to the subgroup means.

Simulation studies based on the above four conditions are conducted using SAS version 8. Repeatedly, $m = 10$ and 20 subgroups of size $n = 5$ observations each are generated, control limits computed and the number of subgroups that fall outside the control limits are calculated. This procedure is repeated 10 000 times for a total of $10\,000 \times m$ subgroups. The proportions of out-of-control subgroups (based on $10\,000 \times m$ subgroups) are computed for the four different conditions and two different charting methods.

The results for the EWMA_{SM} chart are displayed in Tables 4 and 5 for $m = 10$ and 20 respectively. Similarly, Tables 6 and 7 give the results of the EWMA_{SMQ} chart for $m = 10$ and 20 respectively. An ideal procedure is the chart which gives a higher proportion for detecting out-of-control signals for the three conditions of Outliers, Special Cause and Outliers and Special Cause and a lower signal proportion for the In-control situation.

A comparison of the results in Tables 4 and 6 for $m = 10$ show that for fixed values of α and δ , the out-of-control proportions of the EWMA_{SMQ} chart are higher than the corresponding values of the EWMA_{SM} chart for the two out-of-control conditions of Outliers and Outliers and Special Cause. For example, when $\alpha = 0.5$, the out-of-control proportion in Table 6 is 0.0242 while that in Table 4 is much lower at only 0.0047 for the Outliers condition. For the Outliers and Special Cause condition, the proportions of out-of-control in Table 4 when $\alpha = 0.5$ are $\{0.0194, 0.0388, \dots, 0.9976\}$ while the corresponding proportions in Table 6 are $\{0.0554, 0.0878, \dots, 0.9992\}$ where the values of the former are all lower than that of the latter for $\delta \in \{0, 0.25, \dots, 4\}$.

This shows that the EWMA_{SMQ} chart is more sensitive to out-of-control conditions when outliers are present in the data. The limits computed from the estimate of the interquartile ranges for the EWMA_{SMQ} chart are less influenced by outliers, compared to that of the EWMA_{SM} chart whose limits are computed based on the sample ranges. Thus, the EWMA_{SMQ} chart is more robust than the

EWMA_{SM} and the former is a better alternative in the detection of a special cause when outliers are present. The EWMA_{SMQ} chart is also superior to the EWMA_{SM} when only outliers are present. The results in both Tables 4 and 6 indicate that for the Special Cause condition, the EWMA_{SMQ} chart is superior to the EWMA_{SM} for smaller values of δ and that both charts have comparable performances for larger values of δ .

Table 4. Proportions of Out-of-Control for the EWMA_{SM} Chart Under Four Different Conditions based on $m = 10$ and $n = 5$

		EWMA _{SM}								
		α								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
In-Control		0.0000	0.0001	0.0003	0.0005	0.0009	0.0013	0.0017	0.0022	0.0025
Outliers		0.0000	0.0002	0.0009	0.0023	0.0047	0.0066	0.0093	0.0107	0.0126
	δ									
Special Cause	0	0.0030	0.0106	0.0168	0.0214	0.0256	0.0286	0.0311	0.0327	0.0338
	0.25	0.0377	0.0601	0.0638	0.0618	0.0600	0.0568	0.0550	0.0528	0.0504
	0.5	0.2395	0.2786	0.2502	0.2114	0.1816	0.1546	0.1353	0.1187	0.1056
	0.75	0.5162	0.5703	0.5337	0.4633	0.3946	0.3317	0.2829	0.2399	0.2045
	1	0.6813	0.7482	0.7484	0.7035	0.6341	0.5529	0.4772	0.4090	0.3483
	1.5	0.8246	0.8803	0.9051	0.9150	0.9097	0.8809	0.8296	0.7606	0.6802
	2	0.8858	0.9323	0.9559	0.9684	0.9750	0.9760	0.9684	0.9478	0.9077
	2.5	0.9181	0.9661	0.9839	0.9899	0.9929	0.9942	0.9948	0.9929	0.9847
	3	0.9469	0.9880	0.9953	0.9977	0.9985	0.9990	0.9992	0.9993	0.9986
	4	0.9904	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Outliers and Special Cause	0	0.0013	0.0056	0.0108	0.0154	0.0194	0.0234	0.0262	0.0287	0.0303
	0.25	0.0159	0.0296	0.0366	0.0381	0.0388	0.0399	0.0397	0.0401	0.0394
	0.5	0.1241	0.1568	0.1457	0.1255	0.1097	0.0961	0.0853	0.0773	0.0700
	0.75	0.3492	0.3908	0.3535	0.2971	0.2480	0.2071	0.1749	0.1490	0.1268
	1	0.5464	0.6025	0.5726	0.5070	0.4363	0.3643	0.3064	0.2574	0.2176
	1.5	0.7440	0.8079	0.8214	0.8027	0.7630	0.7000	0.6265	0.5484	0.4728
	2	0.8309	0.8847	0.9084	0.9169	0.9114	0.8881	0.8511	0.7953	0.7276
	2.5	0.8792	0.9270	0.9499	0.9609	0.9644	0.9605	0.9487	0.9243	0.8859
	3	0.9091	0.9553	0.9739	0.9819	0.9857	0.9857	0.9824	0.9728	0.9569
	4	0.9542	0.9872	0.9944	0.9967	0.9976	0.9981	0.9979	0.9964	0.9933

Table 5. Proportions of Out-of-Control for the EWMA_{SM} Chart Under Four Different Conditions based on $m = 20$ and $n = 5$

		EWMA _{SM}								
		α								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
In-Control		0.0000	0.0003	0.0007	0.0012	0.0017	0.0020	0.0024	0.0026	0.0029
Outliers		0.0001	0.0017	0.0048	0.0080	0.0113	0.0140	0.0160	0.0180	0.0191
	δ									
Special Cause	0	0.0079	0.0170	0.0228	0.0258	0.0287	0.0306	0.0320	0.0330	0.0338
	0.25	0.1162	0.1045	0.0879	0.0753	0.0678	0.0612	0.0574	0.0538	0.0507
	0.5	0.5068	0.4269	0.3252	0.2510	0.2035	0.1677	0.1425	0.1222	0.1064
	0.75	0.7492	0.7439	0.6492	0.5316	0.4343	0.3545	0.2948	0.2463	0.2078
	1	0.8381	0.8695	0.8503	0.7784	0.6844	0.5848	0.4961	0.4182	0.3514
	1.5	0.9114	0.9392	0.9516	0.9547	0.9427	0.9071	0.8477	0.7712	0.6855
	2	0.9424	0.9658	0.9776	0.9836	0.9870	0.9861	0.9772	0.9533	0.9116
	2.5	0.9588	0.9831	0.9918	0.9951	0.9965	0.9973	0.9974	0.9951	0.9867
	3	0.9728	0.9943	0.9979	0.9991	0.9995	0.9996	0.9997	0.9996	0.9990
	4	0.9958	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Outliers and Special Cause	0	0.0039	0.0117	0.0183	0.0227	0.0268	0.0293	0.0318	0.0330	0.0338
	0.25	0.0545	0.0582	0.0524	0.0488	0.0465	0.0446	0.0442	0.0430	0.0420
	0.5	0.3232	0.2577	0.1893	0.1464	0.1188	0.0997	0.0867	0.0767	0.0684
	0.75	0.6173	0.5575	0.4401	0.3358	0.2653	0.2095	0.1731	0.1443	0.1225
	1	0.7609	0.7622	0.6824	0.5685	0.4679	0.3764	0.3086	0.2497	0.2067
	1.5	0.8699	0.9021	0.9029	0.8721	0.8161	0.7331	0.6411	0.5487	0.4626
	2	0.9154	0.9414	0.9537	0.9555	0.9462	0.9196	0.8770	0.8109	0.7322
	2.5	0.9392	0.9629	0.9749	0.9808	0.9815	0.9772	0.9635	0.9400	0.8995
	3	0.9534	0.9772	0.9878	0.9921	0.9936	0.9930	0.9893	0.9810	0.9664
	4	0.9763	0.9947	0.9979	0.9988	0.9991	0.9993	0.9990	0.9980	0.9954

Table 6. Proportions of Out-of-Control for the EWMA_{SMQ} Chart Under Four Different Conditions based on $m = 10$ and $n = 5$

		EWMA _{SMQ}								
		α								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
In-Control		0.0000	0.0002	0.0006	0.0013	0.0023	0.0030	0.0038	0.0045	0.0052
Outliers		0.0010	0.0050	0.0119	0.0189	0.0242	0.0299	0.0336	0.0366	0.0387
	δ									
Special Cause	0	0.0059	0.0152	0.0228	0.0286	0.0329	0.0362	0.0388	0.0406	0.0417
	0.25	0.0465	0.0703	0.0745	0.0727	0.0697	0.0666	0.0639	0.0615	0.0591
	0.5	0.2462	0.2858	0.2588	0.2220	0.1912	0.1661	0.1459	0.1296	0.1162
	0.75	0.5140	0.5661	0.5314	0.4666	0.4001	0.3410	0.2915	0.2500	0.2161
	1	0.6794	0.7448	0.7439	0.6994	0.6302	0.5523	0.4802	0.4134	0.3562
	1.5	0.8228	0.8802	0.9040	0.9122	0.9038	0.8730	0.8208	0.7527	0.6738
	2	0.8859	0.9322	0.9546	0.9660	0.9726	0.9728	0.9639	0.9404	0.8989
	2.5	0.9205	0.9639	0.9816	0.9884	0.9918	0.9936	0.9940	0.9913	0.9817
	3	0.9468	0.9855	0.9945	0.9972	0.9983	0.9987	0.9990	0.9990	0.9980
	4	0.9869	0.9990	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Outliers and Special Cause	0	0.0087	0.0241	0.0377	0.0487	0.0554	0.0618	0.0652	0.0694	0.0711
	0.25	0.0454	0.0731	0.0852	0.0888	0.0878	0.0878	0.0868	0.0859	0.0852
	0.5	0.2216	0.2616	0.2424	0.2168	0.1924	0.1717	0.1563	0.1426	0.1317
	0.75	0.4726	0.5166	0.4805	0.4215	0.3636	0.3173	0.2744	0.2425	0.2128
	1	0.6425	0.7021	0.6873	0.6360	0.5669	0.4970	0.4334	0.3775	0.3288
	1.5	0.8011	0.8574	0.8773	0.8759	0.8537	0.8085	0.7480	0.6814	0.6076
	2	0.8699	0.9178	0.9399	0.9499	0.9512	0.9429	0.9213	0.8841	0.8322
	2.5	0.9084	0.9521	0.9711	0.9791	0.9826	0.9824	0.9777	0.9653	0.9437
	3	0.9357	0.9757	0.9876	0.9923	0.9942	0.9946	0.9933	0.9899	0.9822
	4	0.9769	0.9961	0.9984	0.9991	0.9992	0.9994	0.9994	0.9991	0.9982

It is shown by the results in Tables 4 and 6 that the EWMA_{SM} chart has lower false alarm (Type-I error) rates than the EWMA_{SMQ} chart. However, it should be noted that the superiority of the EWMA_{SMQ} chart in comparison to the EWMA_{SM} chart for the two out-of-control conditions of Outliers and Outliers and Special Cause far outweighs its deficiency in terms of a higher rate of false alarm. For example, if $\alpha = 0.3$, the proportion of false alarm of the EWMA_{SMQ} chart is twice that of the EWMA_{SM} but the former is thirteen times more efficient in detecting outliers compared to the

latter. Because a comparison between Tables 5 and 7 shows similar trend, it is suffice to discuss and draw conclusions based on the above comparisons.

Note that the evaluation of the performances of the control charts discussed in this section are made based on the proportions of subgroup points falling outside the control limits and are not based on the average run length (ARL) values because the charts' parameters are estimated from the subgroup data, i.e., the nominal values of these parameters are assumed to be unknown.

Table 7. Proportions of Out-of-Control for the EWMA_{SMQ} Chart Under Four Different Conditions based on $m = 20$ and $n = 5$

		EWMA _{SMQ}								
		α								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
In-Control		0.0001	0.0006	0.0014	0.0020	0.0026	0.0030	0.0035	0.0038	0.0040
Outliers		0.0024	0.0115	0.0204	0.0268	0.0321	0.0350	0.0378	0.0402	0.0417
Special Cause	δ									
	0	0.0099	0.0197	0.0257	0.0295	0.0323	0.0343	0.0360	0.0369	0.0379
	0.25	0.1205	0.1093	0.0925	0.0809	0.0723	0.0661	0.0617	0.0581	0.0550
	0.5	0.5026	0.4270	0.3282	0.2570	0.2085	0.1733	0.1468	0.1273	0.1114
	0.75	0.7462	0.7395	0.6438	0.5312	0.4346	0.3581	0.2987	0.2511	0.2122
	1	0.8370	0.8680	0.8461	0.7756	0.6797	0.5825	0.4955	0.4195	0.3549
	1.5	0.9106	0.9389	0.9511	0.9541	0.9398	0.9025	0.8423	0.7658	0.6817
	2	0.9421	0.9657	0.9771	0.9832	0.9862	0.9851	0.9751	0.9500	0.9064
	2.5	0.9590	0.9822	0.9914	0.9948	0.9963	0.9970	0.9970	0.9943	0.9850
	3	0.9730	0.9935	0.9977	0.9989	0.9994	0.9996	0.9997	0.9995	0.9988
	4	0.9948	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Outliers and Special Cause	0	0.0181	0.0379	0.0503	0.0576	0.0626	0.0657	0.0673	0.0692	0.0698
	0.25	0.1183	0.1210	0.1101	0.1022	0.0973	0.0919	0.0895	0.0860	0.0845
	0.5	0.4536	0.3845	0.3052	0.2486	0.2090	0.1788	0.1589	0.1415	0.1295
	0.75	0.7082	0.6774	0.5755	0.4747	0.3945	0.3303	0.2797	0.2412	0.2100
	1	0.8139	0.8330	0.7857	0.7015	0.6075	0.5216	0.4444	0.3798	0.3252
	1.5	0.8981	0.9274	0.9352	0.9240	0.8918	0.8387	0.7714	0.6939	0.6137
	2	0.9345	0.9583	0.9695	0.9746	0.9717	0.9597	0.9365	0.8976	0.8420
	2.5	0.9533	0.9761	0.9860	0.9900	0.9914	0.9896	0.9833	0.9716	0.9492
	3	0.9672	0.9884	0.9917	0.9965	0.9973	0.9970	0.9956	0.9917	0.9844
	4	0.9894	0.9984	0.9993	0.9995	0.9997	0.9997	0.9996	0.9993	0.9983

An Example of Application

An example will be given to illustrate how the EWMASMQ chart is put to work in a real situation. The EWMASM chart is also constructed so that a comparison between the two approaches can be made. This example is based on the data from Wadsworth, Stephens and Godfrey (1986) and concerns the melt index of an extrusion grade polyethylene compound. As part of a study of the process, 20 subgroups of four each are taken. Table 8 gives the data of this process.

The limits of the EWMASM chart are computed using equations (2a) and (2b) while that of the EWMASMQ chart are calculated from equations (13a) and (13b). Figure 1 shows the EWMASM and EWMASMQ charts together with their respective control limits. The limits of the EWMASM chart are represented by UCL/LCL and those of the EWMASMQ chart by UCL_{SMQ}/LCL_{SMQ} .

Figure 1. The EWMASM and EWMASMQ Charts for Melt Index Data

The EWMA_{SM} chart with $\alpha = 0.2$ detects out-of-control points at subgroups 8 and 9 while the corresponding EWMA_{SMQ} chart signals at subgroups 8, 9, 14 and 15. The EWMA_{SMQ} chart detects two additional out-of-control points (i.e., subgroups 14 and 15) besides the two points at subgroups 8 and 9 which are also detected by the EWMA_{SM} chart. From Figure 1 for $\alpha = 0.2$, a shift is observed that is gradually increasing between subgroups 1 and 9 followed by a shift which is gradually decreasing until subgroup 15. Here, both the EWMA_{SM} and EWMA_{SMQ} charts have successfully detected the upward shift but only the EWMA_{SMQ} chart managed to detect the downward shift. The presence of sample ranges with large values such as those in subgroups 3, 4, 6 and 8 cause the average sample range, $\bar{R}$ to be overestimated, hence widening the limits of the EWMA_{SM} chart so that the chart is less sensitive in detecting shifts in the mean. On the contrary, the EWMA_{SMQ} chart does not face

this problem since its limits are computed based on the average sample interquartile range (IQR).

The EWMA_{SM} and EWMA_{SMQ} charts with $\alpha = 0.8$ give more weight to the current sample average, $\bar{X}_t$ compared to the charts with $\alpha = 0.2$. Thus, a weighting constant of $\alpha = 0.8$ makes the two charts more sensitive to single subgroup averages with big or small values. Out-of-control points are detected at subgroups 1, 6, 8, 9, 11, 13, 14 and 17 by the EWMA_{SMQ} chart and at only subgroup 8 by the EWMA_{SM} chart. From the sample averages, $\bar{X}_t$ in Table 8, we notice that the values for subgroups 6, 8, 9 and 17 are somewhat bigger while those for subgroups 1, 11, 13 and 14 are somewhat smaller than the other sample averages. Investigations need to be carried out to search for assignable causes before these subgroups are classified as out-of-control points. It should be noted that the EWMA_{SM} chart with $\alpha = 0.8$ fails to detect any subgroup average that plots below the LCL.

Table 8. Subgroups of Melt Index Measurements and the Computed Sample Means, Sample Ranges, Sample Interquartile Ranges and $\hat{W}_t$ Statistics

Sub. No., t	Observations				$\bar{X}_t$	R_t	IQR _{t}	$\hat{W}_t$	
	X_1	X_2	X_3	X_4				$\alpha=0.2$	$\alpha=0.8$
1	218	224	220	231	223.25	13	4	232.66	225.60
2	228	236	247	234	236.25	19	2	233.38	234.12
3	280	228	228	221	239.25	59	0	234.55	238.22
4	210	249	241	246	236	39	5	234.84	236.45
5	243	240	230	230	235.75	13	10	235.02	235.89
6	225	250	258	244	244.25	33	6	236.87	242.58
7	240	238	240	243	240.25	5	0	237.55	240.72
8	244	248	265	234	247.75	31	4	239.59	246.34
9	238	233	252	243	241.5	19	5	239.97	242.47
10	228	238	220	230	229	18	2	237.78	231.69
11	218	232	230	226	226.5	14	4	235.52	227.54
12	226	231	236	242	233.75	16	5	235.17	232.51
13	224	221	230	222	224.25	9	2	232.98	225.90
14	230	220	227	226	225.75	10	1	231.54	225.78
15	224	228	226	240	229.5	16	2	231.13	228.76
16	232	240	241	232	236.25	9	8	232.15	234.75
17	243	250	248	250	247.75	7	2	235.27	245.15
18	247	238	244	230	239.75	17	6	236.17	240.83
19	224	228	228	246	231.5	22	0	235.23	233.37
20	236	230	230	232	232	6	2	234.59	232.27
					$\bar{\bar{X}} =$ 235.0125	$\bar{\bar{R}} =$ 18.75	$\bar{\bar{IQR}} =$ 3.5		

Conclusion

This article discusses a new robust EWMA control chart for the sample mean which is referred to as the EWMA^{SMQ} chart. It is shown by simulation that the EWMA^{SMQ} chart is a superior alternative to the EWMASM chart when one is concerned with the presence of outliers. Generally, the new EWMA^{SMQ} chart allows easier detection of outliers in the subgroups and is also more sensitive to other forms of out-of-control behavior when outliers are present. An example is given to show how the EWMA^{SMQ} chart works in a real situation. This example also illustrates the superiority of the EWMA^{SMQ} chart to the EWMASM chart, hence making the EWMA^{SMQ} as an attractive alternative to quality control practitioners.

References

- Borror, C. M., Champ, C. W., & Rigdon, S. E. (1998). Poisson EWMA control charts. *Journal of Quality Technology*, 30, 352–361.
- Crowder, S. V. (1987). A simple method for studying run-length distributions of exponentially weighted moving average charts. *Technometrics*, 29, 401–407.
- Crowder, S. V. (1989). Design of exponentially weighted moving average schemes. *Journal of Quality Technology*, 21, 155–162.
- Gan, F. F. (1990). Monitoring Poisson observations using modified exponentially weighted moving average control charts. *Communications in Statistics - Simulation and Computation*, 103–124.
- Lucas, J. M. & Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32, 1–12.
- MacGregor, J. F. & Harris, T. J. (1993). The exponentially weighted moving variance. *Journal of Quality Technology*, 25, 106–118.
- Ng, C. H. & Case, K. E. (1989). Development and evaluation of control charts using exponentially weighted moving averages. *Journal of Quality Technology*, 21, 242–250.
- Rhoads, T. R., Montgomery, D. C., & Mastrangelo, C. M. (1996). A fast initial response scheme for the exponentially weighted moving average control chart. *Quality Engineering*, 9, 317–327.
- Roberts, S. W. (1959). Control chart based on geometric moving averages. *Technometrics*, 1, 239–250.
- Rocke, D. M. (1989). Robust control charts. *Technometrics*, 31, 173–184.
- Rocke, D. M. (1992). $\bar{X}_Q$ and R_Q charts: Robust control charts. *The Statistician*, 41, 97–104.
- Somerville, S. E., Montgomery, D. C., & Runger, G. C. (2002). Filtering and smoothing methods for mixed particle count distributions. *International Journal of Production Research*, 40, 2991–3013.
- Sweet, A. L. (1986). Control chart using coupled EWMA. *IIE Transactions*, 18, 26–33.
- Wadsworth, H. M., Stephens, K. S., & Godfrey, A. B. (1986). *Modern methods for quality control and improvement*. New York, N.Y.: John Wiley & Sons.

Interaction Graphs for $4^r 2^{n-p}$ Fractional Factorial Designs

M.L. Aggarwal
Department of Mathematical Sciences,
University of Memphis

S. Roy Chowdhury
Department of Statistics,
University of Delhi

Anita Bansal
Department of Statistics,
University of Delhi

Neena Mital
Department of Statistics,
University of Delhi

Interaction graphs have been developed for two-level and three-level fractional factorial designs under different design criteria. A catalogue is presented of all possible non-isomorphic interaction graphs for $4^r 2^{n-p}$ ($r=1$; $n=2, \dots, 10$; $p=1, \dots, 8$ and $r=2$; $n=1, \dots, 7$; $p=1, \dots, 7$) fractional factorial designs, and non-isomorphic interaction graphs for asymmetric fractional factorial designs under the concept of combined array.

Key words: Minimum aberration designs, asymmetric orthogonal arrays, combined array, control and noise factors, pattern and extended pattern.

Introduction

Taguchi (1959, 1987) introduced the concept of linear graphs associated with various orthogonal arrays. Linear graphs are the graphical representation of allocation of main effects and two-factor interactions among various columns of orthogonal array. Ankenman and Dean (2003) have given an excellent review on Taguchi's methodology. Joglekar and Kacker (1989) and Kacker and Tsui (1990) discussed the concept of linear graphs for planning industrial

experiments. Barton (1999) and Wu and Hamada (2000) discussed the concept of linear graphs. Li *et al.* (1991), Wu and Chen (1992), Chen, Sun and Wu (1993) and Sun and Wu (1994) developed interaction graphs (linear graphs) for two-level and three-level fractional factorial designs under different design criteria. These designs enable one to estimate all main effects and required two-factor interactions Aggarwal, Gupta, and Chowdhury (2001) developed interaction graphs which enable one to estimate three factor interactions along with all main effects and required two-factor interactions.

M.L. Aggarwal is Professor of Statistics at The University of Memphis. He is Associate Editor of *Journal of Statistical Theory & Practice* and is the co-organizer of Design & Analysis of Experiments (DAE2007) Conference. S. Roy Chowdhury is Reader in Statistics. Anita Bansal is Reader in Statistics. Neena Mital is Lecturer in Statistics. Acknowledgements: The authors are grateful to the University of Memphis for providing necessary research facilities and to University Grants Commission, New Delhi, INDIA for supporting the Research under Teacher's Fellowship Scheme.

In the literature, interaction graphs are available for either two level or three level fractional factorial designs. Dey and Mukerjee (1999) discussed the concept of asymmetric orthogonal arrays which has been extensively used in industrial experiments for quality improvement. Further, Dey, Suen and Das (2005) have developed asymmetric fractional factorials plans which are universally optimal using the concept of finite projection geometry. Wu and Hamada (2000) extended the minimum aberration criterion of 2^{n-p} designs to $4^r 2^{n-p}$ asymmetric designs for $r = 1$ and 2. Xu and Wu (2001) generalized minimum aberration criterion for asymmetrical fractional factorial

designs. In this article below, an algorithm has been developed to generate all possible non-isomorphic interaction graphs for $4^r 2^{n-p}$ ($r=1$; $n=2, \dots, 10$; $p=1, \dots, 8$ and $r=2$; $n=1, \dots, 7$; $p=1, \dots, 7$) fractional factorial designs. Some of these designs satisfy minimum aberration criterion. Interaction graphs were developed for only minimum aberration designs for $r = 2$. These designs can estimate all main effects and required two-factor interactions.

Many times, in industry, there are uncontrollable (noise) factors which induce variations in the system. Taguchi (1959) introduced robust design methodology to develop experiment that allows us to identify the settings of the control factors that make the product or process insensitive to the effects of the noise factors. Taguchi suggested the use of crossing of two OA's, inner array involving control factors and outer array involving noise factors and named it Crossed array. Welch *et al.* (1990), Shoemaker *et al.* (1991) and Montgomery (1991) independently proposed the concept of combined array where they combine control factors and noise factors in a single design matrix. This approach postulated a single response model of the type:

$$Y = f(X, Z) \quad (1)$$

where X and Z represent the settings in the control and noise variables respectively. Aggarwal *et al.* (2002) have developed a catalogue of all possible non-isomorphic interaction graphs for two level combined arrays. In this article, non-isomorphic interaction graphs were developed for asymmetric designs where one of the four levels factor may be treated as noise or control factor along with other two level control and noise factors. These designs enable one to estimate:

- (a) Control and Noise main effects
- (b) Control-by-Noise interactions
(CxN)
- (c) Control-by-Control interactions
(CxC)

Also below, an algorithm is developed to generate non-isomorphic interaction graphs for $4^1 2^{n+m-p}$ combined array fractional factorial

designs, where 'n' and 'm' are the number of control factors and noise factors respectively. A catalogue has been developed giving the number of non-isomorphic interaction graphs for $n=2, \dots, 6$; $m=1, \dots, 3$; $p=1, \dots, 3$ when a four level factor can be treated as a control or a noise factor.

Algorithm for developing interaction graphs for $4^r 2^{n-p}$ fractional factorial designs.

In case of interaction graphs for two level symmetric fractional factorial designs, there exists only one edge representing two-factor interaction corresponding to linear x linear component. Whereas, in case of asymmetric designs with one factor at four-level its interactions with other two-level factors will generate at the most triple edges representing linear x linear, quadratic x linear and cubic x linear components. Similarly, when two factors are at four levels then the interaction between two four level factors will generate at the most nine edges. Therefore, the interaction graphs differ from each other on the basis of number of edges depending on type of interaction effects between two factors.

The method for developing non-isomorphic interaction graphs for $4^r 2^{n-p}$ fractional factorial designs is an extension of the algorithm for symmetric designs based on the technique given by Li *et al.* (1991) and Wu and Chen (1992).

Non-isomorphic Interaction graphs for $4^1 2^{n-p}$ fractional factorial designs

Consider defining relation with one four levels and 'n' two level factors. First allocate the linear, quadratic and cubic effects viz. A_1 , A_2 and A_3 in column 1, 3 and 2 respectively of an orthogonal array. The advantage of the (A_1 , A_3 , A_2) system is that the three vectors are of the form $(\alpha, \beta, \alpha\beta)$. This relationship makes it easier to relate and trace each A_i to a factorial effect in the original two-level design from which $4^1 2^{n-p}$ design has been generated. Then two level factors are allocated to remaining columns of orthogonal array depending on the defining relation. Develop alias structure neglecting three and higher factor interactions. Select one two-factor interaction from each of the aliased two-factor interactions along with all clear two-factor

interactions. Construct interaction matrix for each design with column and row headed alphabetically (representing both four level and two level factors) and $(ij)^{th}$ entry value 1 if one of the component is present between i^{th} row and j^{th} column, value 2 if two components are present, value 3 if all the three components are present and 0 otherwise. Corresponding to the interaction matrix, calculate pattern P_4 and P_2 which are the column total of the interaction matrix corresponding to four level factor and two level factors respectively. In other words P_4 is just the number of edges from four level factors to two level factors and vice-versa. Similarly, P_2 is the number of edges between two level factors only.

Next calculate the extended pattern in two parts viz. EP_4 defined as $D_i = \sum d_{ij}$, where d_{ij} are the P_4 patterns of j^{th} factor adjacent to i^{th} factor and EP_2 defined as $D_k = \sum d_{kl}$, where d_{kl} are the P_2 patterns of l^{th} factor adjacent to k^{th} factor. To obtain non-isomorphic interaction graphs sort the patterns (P_2 and P_4) and extended pattern (EP_2 and EP_4) in ascending order separately. Repeat the procedure for all combinations. The combinations are non-isomorphic if the patterns are distinct or if patterns are same but the corresponding extended patterns are distinct. Corresponding to each distinct combination develop an interaction graph. The following

example gives the procedure to develop non-isomorphic interaction graphs for $4^1 2^{n-p}$ fractional factorial designs.

Example1

Consider a $4^1 2^{3-1}$ minimum aberration design with 16 runs where A is a four-level factor and B, C and D are two-level factors with the defining relation: $I = A_1 BCD$. The alias structure is shown in Table 1.

Here, three sets of aliased two-factor interactions have been provided. This gives 8 combinations of nine two-factor interactions taking six clear two-factor interactions together. Corresponding to each unique combination there is a distinct interaction graph. Consider one of the combinations:

CD A_3B A_2B BD A_3C A_2C BC A_2D A_3D

There are 3 double edges representing interactions between four level factor and three two-level factors and 3 single edges representing interactions among two-level factors. The interaction matrix with pattern and extended pattern for this combination is shown in table 2.

Repeat the procedure for remaining combinations. This gives four non-isomorphic interaction graphs corresponding to four unique combinations. Table 3 gives pattern and extended pattern of all unique combinations for the design $I = A_1 BCD$.

Table 1. Alias Structure

A_1B	A_3B	A_2B	A_1C	A_3C	A_2C	A_1D	A_2D	A_3D
CD			BD			BC		

Table 2. Interaction Matrix

	A	B	C	D
A	0	2	2	2
B	2	0	1	1
C	2	1	0	1
D	2	1	1	0
P_4	6	2	2	2
P_2	0	2	2	2
EP_4	12	16	16	16
EP_2	12	4	4	4

Table 3: Patterns and Extended Patterns for the design $I = A_1BCD$

S.No.	Design Combinations	P_4	P_2	EP_4	EP_2
1	$A_1B \ A_3B \ A_2B \ A_1C \ A_3C \ A_2C \ A_1D \ A_2D \ A_3D$	9 3 3 3	0 0 0 0	27 27 27 27	0 0 0 0
2	$A_1B \ A_3B \ A_2B \ A_1C \ A_3C \ A_2C \ BC \ A_2D \ A_3D$	8 2 3 3	0 0 1 1	22 16 27 27	6 0 1 1
3	$A_1B \ A_3B \ A_2B \ BD \ A_3C \ A_2C \ BC \ A_2D \ A_3D$	7 2 2 3	0 1 1 2	17 17 17 25	10 2 2 2
4	$CD \ A_3B \ A_2B \ BD \ A_3C \ A_2C \ BC \ A_2D \ A_3D$	6 2 2 2	0 2 2 2	12 16 16 16	12 4 4 4

Figure 1: Interaction Graphs for the design $I = A_1BCD$.

Note: Interaction graphs correspond to the designs outlined in Table 3

Non-isomorphic Interaction graphs for 4^{2n-p} fractional factorial designs

Suppose that there are two factors at four levels and n factors at two levels. Optimal minimum aberration designs developed by Wu and Hamada (2000) have been considered, to allocate the two four level factors viz. A in column 1, 2 and 3, B in columns 4, 8 and 12. Depending on the defining relation, n two-level factors are allocated in the remaining columns of orthogonal array. Develop alias structure neglecting three and higher factor interactions. Construct interaction matrix and calculate pattern and extended pattern in the similar manner as discussed above. Due to two factors at four level the entries in the interaction matrix corresponding to interaction between the two four level factors varies from 0 to 9 depending upon the presence of the component

combination. The rest of the procedure remains the same as when only one factor is at four levels. The following example gives the procedure to develop non-isomorphic interaction graphs for 4^{2n-p} fractional factorial designs.

Example 2

Consider a 4^{22-1} design with 32 runs where A and B are four level factors and C and D are two level factors with the defining relation: $I = A_3B_1CD$. The alias structure as shown in Table 4.

Here, there are 16 clear two-factor interactions and three sets of aliased two-factor interactions. This gives all together 8 combinations. For this design, there are 4 unique combinations, which give 4 non-isomorphic interaction graphs. Table 5 gives pattern and extended pattern of all unique combinations for the design $I = A_3B_1CD$.

Table 4. Alias Structure

A_1B_1	A_3B_1	A_2B_1	A_1B_3	A_3B_3	A_2B_3	A_1B_2	A_3B_2	A_2B_2	A_1C	A_3C	A_2C
	CD									B_1D	
A_2D	A_3D	A_1D	B_3C	B_2D	B_2C	B_3D					
	B_1C										

Table 5. Patterns and Extended Patterns for the design $I = A_3B_1CD$

S.No.	Design Combinations	P_4	P_2	EP_4	EP_2
1	A_1B_1 A_3B_1 A_2B_1 A_1B_3 A_3B_3 A_2B_3 A_1B_2 A_3B_2 A_2B_2 A_1C A_3C A_2C A_3D A_2D A_1D B_3C B_2D B_2C B_3D	13 15 5 5	0 0 0 0	147 155 71 71	0 0 0 0
2	A_1B_1 CD A_2B_1 A_1B_3 A_3B_3 A_2B_3 A_1B_2 A_3B_2 A_2B_2 A_1C A_3C A_2C A_3D A_2D A_1D B_3C B_2D B_2C B_3D	12 14 5 5	0 0 1 1	126 132 71 71	4 6 1 1
3	A_1B_1 CD A_2B_1 A_1B_3 A_3B_3 A_2B_3 A_1B_2 A_3B_2 A_2B_2 A_1C B_1D A_2C A_3D A_2D A_1D B_3C B_2D B_2C B_3D	13 13 4 6	0 0 1 1	130 130 58 82	5 5 1 1
4	A_1B_1 A_3B_1 A_2B_1 A_1B_3 A_3B_3 A_2B_3 A_1B_2 A_3B_2 A_2B_2 A_1C B_1D A_2C A_3D A_2D A_1D B_3C B_2D B_2C B_3D	14 14 4 6	0 0 0 0	152 152 56 84	0 0 0 0

Figure 2. Interaction Graphs for the above $I = A_3B_1CD$

Note. All the non-isomorphic interaction graphs correspond to the designs outlined in Table 5.

A catalogue has been developed for $4^1 2^{n-p}$ and $4^2 2^{n-p}$ designs giving the number of non-isomorphic interaction graphs. A part of the catalogue is shown in Appendix – I. A complete catalogue is available with the authors.

Algorithm for developing non-isomorphic interaction graphs for $4^1 2^{n+m-p}$ combined array fractional factorial designs.

Consider $4^1 2^{n+m-p}$ combined array fractional factorial designs where n and m are the number of control and noise factors respectively. If one of the control factors is at four-level then the model will be of the form:

$$\begin{aligned}
 y = & b_0 + b_1 x_1 + \sum_{i=2}^n b_i x_i + \sum_{j=1}^m b_j z_j + \\
 & \sum_{j=1}^m b_{1j} x_1 z_j + \sum_{j=2}^n \gamma_{1j} x_1 x_j + \sum_{i=2}^n \sum_{j=1}^m b_{ij} x_i z_j \\
 & + \sum_{i=2}^n \sum_{j>i}^n \gamma_{ij} x_i x_j + \varepsilon
 \end{aligned} \tag{2}$$

where x_1 denotes a four-level control factor, x_i denotes two-level control factors and z_j denotes two-level noise factors.

Whereas if one of the noise factors is at four levels then the model will be of the form:

$$\begin{aligned}
 y = & b_0 + b_1 z_1 + \sum_{j=2}^m b_j z_j + \sum_{i=1}^n b_i x_i \\
 & + \sum_{i=1}^n b_{i1} x_i z_1 + \sum_{i=1}^n \sum_{j=2}^m b_{ij} x_i z_j \quad (3) \\
 & + \sum_{i=1}^n \sum_{j>i}^n \gamma_{ij} x_i x_j + \epsilon
 \end{aligned}$$

where z_1 denotes a four level noise factor.

The method of developing non-isomorphic interaction graphs for $4^1 2^{n+m-p}$ combined array designs is an extension of the algorithm based on the technique given by Aggarwal *et al.* (2002) for symmetric designs. In this case, some of the two-level factors are treated as control and noise factor along with a four-level factor which may be treated as control or noise factor. Capital letters indicate control factors and small letters indicate noise factors. For a given defining relation, an alias structure is first constructed with pre-defined number of control and noise factors neglecting two-factor interactions of type noise $\times$ noise and all three and higher order interactions. In order to define non-isomorphic alias structure for a given defining relation, the following is first counted:

- (a) Number of clear $C_2 \times C_2$, $C_2 \times C_4$, $C_2 \times N_2$, $C_2 \times N_4$ and $C_4 \times N_2$ interactions.
- (b) Number of alias $C_2 \times C_2$ with $C_2 \times C_2$, $C_2 \times C_2$ with $C_2 \times C_4$, $C_2 \times C_2$ with $C_2 \times N_2$, $C_2 \times C_2$ with $C_2 \times N_4$, $C_2 \times C_2$ with $C_4 \times N_2$, $C_2 \times N_2$ with $C_2 \times N_2$, $C_2 \times N_2$ with $C_2 \times N_4$, $C_2 \times N_2$ with $C_4 \times N_2$ interaction (any two-factor interaction aliased with $N_2 \times N_2$ and $N_2 \times N_4$ interaction is assumed to be clear two-factor interaction).

where C_2 and N_2 are control and noise factors at two level respectively and C_4 and N_4 are control and noise factors at four level respectively.

The above counting technique gives all possible non-isomorphic alias structures for different number of control and noise factors at

four-level and two-level for a given defining relation.

Calculate P_4 , P_2 , EP_4 and EP_2 as mentioned earlier. While sorting the patterns and extended patterns in ascending order, divide each P_4 , P_2 , EP_4 and EP_2 further into two groups, corresponding to the control factors and other corresponding to noise factors. The combinations are non-isomorphic if the patterns are distinct or if patterns are same but the corresponding extended patterns are distinct. Corresponding to each distinct combination develop an interaction graph. For various designs with one factor at four-level and other factors at two-levels a catalogue has been developed highlighting the number of non-isomorphic interaction graphs for each design. A part of the catalogue for $4^1 2^{n+m-p}$ combined array fractional factorial designs corresponding to different number of control and noise factors are shown in Appendix II. A complete catalogue is available with the authors. The concept for developing non-isomorphic interaction graphs for $4^1 2^{n+m-p}$ combined array is explained with the help of following example.

Example 3

Consider a $4^1 2^{5-2}$ design with 32 runs with defining relation $I = A_1 BCDE = A_2 CDF = A_3 BEF$. Suppose there are 3 control factors and 3 noise factors, in which one control or noise factor is at 4 levels and rest are at two-level. This gives resolution IV design. There are 6 non-isomorphic defining relations of same word length pattern (0, 0, 0, 2, 1, 0) but with different alias structure according to the criteria given above. These defining relations are defined in Table 6.

Now consider the first defining relation i.e., $I = a_1 bcDE = a_2 cDF = a_3 bEF$. The alias structure for the given defining relation is as shown in Table 7.

There are 16 possible combinations of eligible but not clear two-factor interactions along with clear two-factor interactions. For this design there are 9 non-isomorphic interaction graphs corresponding to unique combinations as shown in Figure 3. The pattern and extended pattern of all unique combinations for above design are shown in Table 8.

Figure 3. Interaction Graphs for the Design $I = a_1bcDE = a_2cDF$

Conclusion

The non-isomorphic interaction graphs for $4^{r2^{n-p}}$ ($r=1; n=2, \dots, 10; p=1, \dots, 8$ and $r=2; n=1, \dots, 7; p=1, \dots, 7$) fractional factorial designs developed

in this article will enable the engineers to work on experiments when mixed level factors are present say one or two factors are at four level and rest are at two level. These designs will allow the estimation of all the main effects and

required two-factor interactions for certain type of asymmetric designs.

All possible non-isomorphic interaction graphs for $4^{12^{n+m-p}}$ ($n=2, \dots, 6$; $m=1, \dots, 3$; $p=1, \dots, 3$) combined array fractional factorial designs are also presented which will allow the estimation of all the main effects and required two-factor interactions when one of the control or noise factor is at four level and rest of the control and noise factors are at two level.

References

- Aggarwal, M. L. Gupta, B. C., Roy Chowdhury, S., & Walker H. F. (2002). Interaction graphs for two level combined array experiment design. *Journal of Industrial Technology*, 18(4).
- Aggarwal, M. L., Gupta, B. C., & Roy Chowdhury, S. (2001). Non-Isomorphic Interaction graphs for 2^{n-k} when three factor interactions are present. *Journal of Combinatorics, Information and System Sciences*, 26, 137-147.
- Ankenman, B. C. & Dean, A. M. (2003). Quality improvement and robustness via design of experiments. *Handbook of Statistics*, 22, Elsevier Science.
- Barton, R. R. (1999). Graphical methods for the design of experiments. New York, N.Y.: Springer-Verlag.
- Chen, Y., Sun, D. X., & Wu, C. F. J. (1993). Catalogue of two-level and three-level fractional factorial designs with small runs. *International Statistics Review*, 61, 131-146.
- Dey, A. & Mukerjee, R. (1999). *Fractional factorial plans*. New York, N.Y.: John Wiley.
- Dey, A., Suen, C. Y., & Das, A. (2005). Asymmetric fractional factorial plans optimal for main effects and specified two-factor interactions. *Statistica Sinica*, 15(3), 751-766.
- Joglekar, A. M. & Kacker, R. N (1989). Graphical and computer aided approach to plan experiments. *Quality and Reliability Engineering International*, 8, 113-123.
- Kacker, R. N. & Tsui, K. L. (1990). Interaction graphs: Graphical aids for planning experiments. *Journal of Quality Technology*, 22(1), 1-14.
- Li, C. C., Washio, Y., Iida, T., & Tanimoto, S. (1991). *Linear graphs of resolution IV for orthogonal array $L_{16}(2^{15})$* . Technical Report, Department of Administration Engineering, Faculty of Engineering, Keio University, Japan.
- Miller, A., Sitter, R. R., Wu, C. F. J., & Long, D. (1993). Are large Taguchi-style experiments necessary: A reanalysis of gear and pinion data. *Quality Engineering*, 6, 21-38.
- Montgomery, D. C. (1991). Using fractional factorial design for robust process development. *Quality Engineering*, 3, 193-205.
- Shoemaker, A. C., Tsui, K. L., & Wu, C. F. J. (1991). Economical experimentation methods for robust parameter design. *Technometrics*, 33, 415-427.
- Sun, D. X. & Wu, C. F. J. (1994). Interaction graphs for three-level fractional factorial designs. *Journal of Quality Technology*, 26, 297-307.
- Taguchi, G. (1959). Introduction to Experimental Design. White Plains, N.Y.: UNIPUB.
- Taguchi, G. (1987). System of experimental design: Engineering methods to optimize quality and minimize cost. White Plains, N.Y.: UNIPUB.
- Xu, H. & Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *The Annals of Statistics*, 29(2), 549-560.
- Welch, W. J., Yu, T. K., Kang, S. M., & Sacks, J. (1990). Computer experiments for quality control by parameter design. *Journal of Quality Technology*, 22, 15-22.
- Wu, C. F. J. & Chen, Y. (1992). A graph aided method for planning two-level experiment when certain interactions are important. *Technometrics*, 34, 162-175.
- Wu, C. F. J. & Hamada, M. (2000). *Experiments planning, analysis, and parameter design optimization*. John Wiley and Sons Inc.

Appendix 1
Non-Isomorphic Interaction Graphs for $4^{12^{n-p}}$ Fractional Factorial Designs

S.No	Design	Design generator	No. of 2 level factors (n)	No. of non-isomorphic interaction graphs
1	$4^{12^{2-1}}$	$I=A_1BC$	2	2
2	$4^{12^{3-2}}$	$I=A_1BC=A_2BD$	3	1
3	$4^{12^{3-1}}$	$*I=A_1BCD$	3	4
4	$4^{12^{4-2}}$	$*I=A_1BD=A_2BCE$	4	25
5	$4^{12^{4-2}}$	$I=A_2BD=A_3BE$	4	1
6	$4^{12^{5-3}}$	$*I=A_3BD=A_2CE=A_1BCF$	5	75
7	$4^{12^{5-3}}$	$I=A_1BD=A_2BE=A_3BF$	5	1
8	$4^{12^{5-3}}$	$I=BCD=A_1CE=A_2BF$	5	69
9	$4^{12^{5-3}}$	$I=A_2BD=A_1BE=BCF$	5	14
10	$4^{12^{5-3}}$	$I=A_1BCD=A_2BCE=A_3BCF$	5	35
11	$4^{12^{6-4}}$	$*I=A_3BD=A_2CE=A_1BCF=A_2BCG$	6	92
12	$4^{12^{6-4}}$	$I=A_1BCD=A_2BCE=A_3BCF=BCG$	6	13
13	$4^{12^{6-4}}$	$I=A_1BCD=A_2CE=A_3BF=BCG$	6	49
14	$4^{12^{7-5}}$	$*I=A_1BD=A_1CE=A_3BF=A_2CG=A_2CH$	7	12
15	$4^{12^{7-5}}$	$I=A_1BCD=A_2BCE=A_3BCF=BCG=A_2CH$	7	42
16	$4^{12^{7-5}}$	$I=A_1BD=A_1CE=A_3BF=A_2CG=A_2BH$	7	12
17	$4^{12^{8-6}}$	$*I=A_1BD=A_1CE=A_3BF=A_2CG=A_2BH=A_2CJ$	8	6
18	$4^{12^{8-6}}$	$I=A_1BCD=A_2BCE=A_3BCF=A_1BG=A_2BH=A_2BJ$	8	26
19	$4^{12^{8-6}}$	$I=A_1BCD=A_2BCE=A_3BCF=BCG=A_2CH=A_2BJ$	8	26
20	$4^{12^{9-7}}$	$*I=A_1BD=A_2BE=BCF=A_2BG=A_1BCH=A_3BCJ=A_2BCK$	9	8
21	$4^{12^{9-7}}$	$I=A_1BCD=A_2BCE=A_1BF=A_2BG=A_1CH=A_2CJ=BCK$	9	10
22	$4^{12^{10-8}}$	$*I=A_1BD=A_1CE=A_3BF=BCG=A_2BH=A_1BCJ=A_3BCK=A_2BCL$	10	8
23	$4^{12^{10-8}}$	$I=A_1BCD=A_2BCE=A_3BCF=A_1BG=A_2BH=A_3BJ=A_1CK=A_2CL$	10	5
24	$4^{12^{4-1}}$	$*I=A_1BCDE$	4	1
25	$4^{12^{4-1}}$	$I=BCDE$	4	2
26	$4^{12^{5-2}}$	$*I=A_2CDE=A_1BCDF$	5	19
27	$4^{12^{5-2}}$	$I=A_1BCDE=A_3BCDF$	5	2
28	$4^{12^{6-3}}$	$*I=A_2BDE=A_2CDF=A_1BCDG$	6	392
29	$4^{12^{6-3}}$	$I=A_1BCDE=A_2BCDF=A_3BCDG$	6	1
30	$4^{12^{6-3}}$	$I=A_1BCDE=A_2BCDF=BCDG$	6	7
31	$4^{12^{7-4}}$	$I=A_1BCDE=A_2BCDF=A_3BCDG=BCDH$	7	4
32	$4^{12^{7-4}}$	$I=A_1BCDE=A_2BCDF=A_3BCDG=A_3CDH$	7	124
33	$4^{12^{5-1}}$	$*I=A_2BCDEF$	5	1
34	$4^{12^{5-1}}$	$I=CDEF$	5	2
35	$4^{12^{5-1}}$	$I=A_3BCF$	5	4
36	$4^{12^{6-2}}$	$*I=A_1BCDF=A_2CDEG$	6	1
37	$4^{12^{6-2}}$	$I=A_1BCDEF=A_2BCDEG$	6	2
38	$4^{12^{6-2}}$	$I=A_1BCDEF=A_2BCDG$	6	4
39	$4^{12^{6-2}}$	$I=A_1BCDF=A_2BCG$	6	19
40	$4^{12^{6-2}}$	$I=BCDEF=CDEG$	6	4
41	$4^{12^{6-2}}$	$I=BCDF=CDEG$	6	7
42	$4^{12^{7-3}}$	$*I=A_1BCDF=A_2BCEG=A_2BDEH$	7	1
43	$4^{12^{7-3}}$	$I=A_1BCDEF=A_2BCDEG=A_2BCDEH$	7	2

44	$4^{12^{7-3}}$	$I = A_1BCDF = A_2BCEG = BCDH$	7	11
45	$4^{12^{7-3}}$	$I = BCDEF = BCDG = CDEH$	7	150
46	$4^{12^{8-4}}$	$I = BDEF = A_1BCDG = A_3CDEH = A_5BCEJ$	8	3
47	$4^{12^{8-4}}$	$I = A_1BCDEF = A_2BCDEG = A_3BCDEH = BCDEJ$	8	1
48	$4^{12^{8-4}}$	$I = A_1BCDEF = A_2BCDEG = A_1BCDH = A_5BCEJ$	8	15
49	$4^{12^{8-4}}$	$I = A_1BCDEF = A_1BCDG = A_1CDEH = A_1BDEJ$	8	6
50	$4^{12^{8-4}}$	$I = A_2BCEF = A_1BCDG = A_3CDEH = A_3BDEJ$	8	7
51	$4^{12^{8-4}}$	$I = A_1BCDF = A_2BCDG = A_3BCDH = BCDJ$	8	4
52	$4^{12^{8-4}}$	$I = A_1BCDF = A_2BCDG = A_3BCDH = A_3DEJ$	8	4
53	$4^{12^{8-4}}$	$I = A_1BCDF = BCEG = BDEH = CDEJ$	8	17
54	$4^{12^{8-4}}$	$I = BCDF = BCEG = BDEH = CDEJ$	8	26
55	$4^{12^{9-5}}$	$I = BDEF = A_2BDG = A_1BCDH = A_3CDEJ = A_5BCEK$	9	660

Note. (* Designs are Minimum Aberration Designs (MAD))

Non-Isomorphic Interaction Graphs for $4^2 2^{n-p}$ Minimum Aberration Designs

S.No.	Design	Design generator	No. of 2 level factors (n)	No. of non-isomorphic interaction graphs
1	$4^2 2^{1-1}$	$I = A_1B_3C$	1	6
2	$4^2 2^{2-2}$	$I = A_1B_3C = A_3B_1D$	2	42
3	$4^2 2^{3-3}$	$I = A_1B_3C = A_3B_1D = A_5B_2E$	3	110
4	$4^2 2^{4-4}$	$I = A_1B_3C = A_3B_1D = A_5B_2E = A_7B_4F$	4	117
5	$4^2 2^{5-5}$	$I = A_1B_3C = A_3B_1D = A_5B_2E = A_7B_4F = A_9B_6G$	5	37
6	$4^2 2^{6-6}$	$I = A_1B_3C = A_3B_1D = A_5B_2E = A_7B_4F = A_9B_6G = A_{11}B_8H$	6	17
7	$4^2 2^{7-7}$	$I = A_1B_3C = A_3B_1D = A_5B_2E = A_7B_4F = A_9B_6G = A_{11}B_8H = A_{13}B_{10}J$	7	10
8	$4^2 2^{2-1}$	$I = A_3B_1CD$	2	4
9	$4^2 2^{3-2}$	$I = A_3B_1CD = A_5B_2CE$	3	58
10	$4^2 2^{4-3}$	$I = A_3B_1CD = A_5B_2CE = A_7B_4CF$	4	1730
11	$4^2 2^{3-1}$	$I = A_5B_2CDE$	3	1
12	$4^2 2^{4-2}$	$I = A_1B_1CDE = A_3B_3CDF$	4	4
13	$4^2 2^{5-3}$	$I = A_1B_1CDE = A_3B_3CDF = A_5B_5DG$	5	223

Appendix 2

Non-Isomorphic Interaction Graphs for $4^1 2^{n+m-p}$ Combined Array Fractional Factorial Designs

S. N	Design	Defining Relation	No. of Control Factors (n)	No. of Noise Factors (m)	No. of Non- isomorphic Interaction graphs
1	$4^1 2^{2-1}$	I= A ₃ bC	2	1	3
2	$4^1 2^{3-2}$	I= A ₁ bC = A ₂ bD	3	1	2
3	$4^1 2^{3-2}$	I= a ₁ BC = a ₂ BD	3	1	1
4	$4^1 2^{3-2}$	I= A ₁ bc = A ₂ bD	2	2	2
5	$4^1 2^{3-2}$	I= a ₁ bC = a ₂ bD	2	2	1
6	$4^1 2^{4-2}$	I= A ₂ bD = A ₃ bE	4	1	2
7	$4^1 2^{4-2}$	I= a ₂ BD = a ₃ BE	4	1	1
8	$4^1 2^{4-2}$	I= A ₁ cE = BcD	4	1	9
9	$4^1 2^{4-2}$	I= A ₁ Ce = BCD	4	1	9
10	$4^1 2^{4-2}$	I= a ₁ CE = BCD	4	1	8
11	$4^1 2^{4-2}$	I= a ₂ BD = a ₂ CE	4	1	6
12	$4^1 2^{4-2}$	I= a ₂ bD = a ₃ bE	3	2	1
13	$4^1 2^{4-2}$	I= a ₂ BD = a ₃ BE	3	2	1
14	$4^1 2^{4-2}$	I= A ₂ bD = A ₃ bE	3	2	2
15	$4^1 2^{4-2}$	I= A ₂ bd = A ₃ bE	3	2	2
16	$4^1 2^{4-2}$	I= a ₁ BCE = a ₁ CD	4	1	8
17	$4^1 2^{4-2}$	I= A ₁ BCE = A ₁ Cd	4	1	9
18	$4^1 2^{4-2}$	I= a ₁ bCE = a ₁ CD	3	2	6
19	$4^1 2^{4-2}$	I= a ₁ BcE = a ₁ Cd	3	2	3
20	$4^1 2^{4-2}$	I= a ₁ BCE = a ₁ Cd	3	2	3
21	$4^1 2^{5-3}$	I= a ₁ BD = a ₂ BE = a ₃ BF	5	1	1
22	$4^1 2^{5-3}$	I= A ₁ bD = A ₂ bE = A ₃ bF	5	1	1
23	$4^1 2^{5-3}$	I= a ₂ BD= a ₁ BE = BCF	5	1	14
24	$4^1 2^{5-3}$	I= A ₂ bD= A ₁ bE = bCF	5	1	17
25	$4^1 2^{5-3}$	I= a ₁ bD = a ₂ bE = a ₃ bF	4	2	1
26	$4^1 2^{5-3}$	I= A ₁ bd = A ₂ bE = A ₃ bF	4	2	1
27	$4^1 2^{5-3}$	I= A ₁ Bd = A ₂ Be = A ₃ BF	4	2	1
28	$4^1 2^{5-3}$	I= a ₁ bCD= a ₂ bCE = a ₃ bCF	4	2	10
29	$4^1 2^{5-3}$	I= a ₂ bD= a ₁ bE = bcF	3	3	4
30	$4^1 2^{5-3}$	I= a ₂ bd= a ₁ bE = bCF	3	3	10
31	$4^1 2^{5-3}$	I= a ₂ Bd= a ₁ Be = BCF	3	3	7
32	$4^1 2^{5-3}$	I= A ₂ Bd= A ₁ BE = Bcf	3	3	8
33	$4^1 2^{5-3}$	I= a ₁ bcD= a ₂ bcE = a ₃ bcF	3	3	1
34	$4^1 2^{4-1}$	I= a ₁ BCDE	4	1	1
35	$4^1 2^{4-1}$	I= A ₁ bCDE	4	1	1
36	$4^1 2^{4-1}$	I= A ₁ BcDE	4	1	1
37	$4^1 2^{4-1}$	I= a ₁ bCDE	3	2	1
38	$4^1 2^{4-1}$	I= A ₁ bcDE	3	2	1
39	$4^1 2^{4-1}$	I= a ₁ BCE	4	1	4

40	$4^{12^{4-1}}$	$I = A_1bCE$	4	1	6
41	$4^{12^{4-1}}$	$I = a_1bCE$	3	2	3
42	$4^{12^{4-1}}$	$I = A_1bcE$	3	2	3
43	$4^{12^{4-1}}$	$I = BCDE$	4	1	2
44	$4^{12^{4-1}}$	$I = bCDE$	4	1	4
45	$4^{12^{4-1}}$	$I = bCDE$	3	2	4
46	$4^{12^{4-1}}$	$I = bcDE$	3	2	2
47	$4^{12^{5-2}}$	$I = a_1BCDE = a_2CDF$	5	1	19
48	$4^{12^{5-2}}$	$I = A_1bCDE = A_2CDF$	5	1	44
49	$4^{12^{5-2}}$	$I = A_1BCDE = A_2CDf$	5	1	21
50	$4^{12^{5-2}}$	$I = a_1BCDE = a_2CDf$	4	2	6
51	$4^{12^{5-2}}$	$I = A_1bCDe = A_2CDF$	4	2	16
52	$4^{12^{5-2}}$	$I = A_1bCDe = A_2CDf$	4	2	23
53	$4^{12^{5-2}}$	$I = a_1bCDe = a_2CDF$	3	3	4
54	$4^{12^{5-2}}$	$I = a_1bCDe = a_2CDf$	3	3	3
55	$4^{12^{5-2}}$	$I = A_1bcDE = A_2CDf$	3	3	9
56	$4^{12^{5-2}}$	$I = A_1bCDe = A_2CDf$	3	3	6
57	$4^{12^{5-2}}$	$I = a_1BCDE = a_3BCDF$	5	1	2
58	$4^{12^{5-2}}$	$I = A_1bCDE = A_3bCDF$	5	1	2
59	$4^{12^{5-2}}$	$I = A_1BCDe = A_3BCDF$	5	1	3
60	$4^{12^{5-2}}$	$I = a_1BCDe = a_3BCDF$	4	2	1
61	$4^{12^{5-2}}$	$I = A_1bCDe = A_3bCDF$	4	2	3
62	$4^{12^{5-2}}$	$I = A_1BCDe = A_3BCDF$	4	2	2
63	$4^{12^{5-2}}$	$I = a_1bcDE = a_3bcDF$	3	3	2
64	$4^{12^{5-2}}$	$I = a_1bCDe = a_3bCDF$	3	3	1
65	$4^{12^{5-2}}$	$I = A_1bcdE = A_3bcdF$	3	3	2
66	$4^{12^{5-2}}$	$I = A_1bCDe = A_3bCDF$	3	3	2
67	$4^{12^{5-2}}$	$I = a_2BCE = BCDF$	5	1	30
68	$4^{12^{5-2}}$	$I = A_2BCe = BCDF$	5	1	36
69	$4^{12^{5-2}}$	$I = a_2bCE = bCDF$	4	2	49
70	$4^{12^{5-2}}$	$I = a_2BCe = BCDF$	4	2	24
71	$4^{12^{5-2}}$	$I = A_2bCe = bCDF$	4	2	53
72	$4^{12^{5-2}}$	$I = a_2bCE = bCdF$	3	3	10
73	$4^{12^{5-2}}$	$I = A_2bcE = bcdF$	3	3	22
74	$4^{12^{5-2}}$	$I = A_2bCe = bCdF$	3	3	10
75	$4^{12^{6-3}}$	$I = a_1BCDE = a_2BCDF = a_3BCDG$	6	1	1
76	$4^{12^{6-3}}$	$I = A_1bCDE = A_2bCDF = A_3bCDG$	6	1	1
77	$4^{12^{6-3}}$	$I = A_1BCDe = A_2BCDF = A_3BCDG$	6	1	2
78	$4^{12^{6-3}}$	$I = a_1bCDE = a_2bCDF = a_3bCDG$	5	2	1
79	$4^{12^{6-3}}$	$I = A_1bCDe = A_2bCDF = A_3bCDG$	5	2	2

80	$4^1 2^{6-3}$	$I = a_1 bcDE = a_2 bcDF = a_3 bcDG$	4	3	1
81	$4^1 2^{6-3}$	$I = a_1 bCDe = a_2 bCDF = a_3 bCDG$	4	3	1
82	$4^1 2^{6-3}$	$I = A_1 bcdE = A_2 bcdF = A_3 bcdG$	4	3	1
83	$4^1 2^{6-3}$	$I = A_1 bcDe = A_2 bcDF = A_3 bcDG$	4	3	2
84	$4^1 2^{6-3}$	$I = a_1 BCDE = a_2 BCDF = BCDG$	6	1	7
85	$4^1 2^{6-3}$	$I = A_1 BCDE = A_2 BCDF = BCDg$	6	1	8
86	$4^1 2^{6-3}$	$I = a_1 BCDe = a_2 BCDF = BCDG$	5	2	7
87	$4^1 2^{6-3}$	$I = A_1 BcdE = A_2 BcdF = BcdG$	5	2	5
88	$4^1 2^{6-3}$	$I = a_1 bcDE = a_2 bcDF = bcDG$	4	3	5
89	$4^1 2^{6-3}$	$I = A_1 BCDe = A_2 BCDf = BCDg$	4	3	7
90	$4^1 2^{6-3}$	$I = a_2 BCE = a_3 CDF = a_1 BDG$	6	1	92
91	$4^1 2^{6-3}$	$I = A_2 BCe = A_3 CDF = A_1 BDG$	6	1	251
92	$4^1 2^{6-3}$	$I = a_2 bCE = a_3 CDF = a_1 bDG$	5	2	69
93	$4^1 2^{6-3}$	$I = A_2 bcE = A_3 cDF = A_1 bDG$	5	2	136
94	$4^1 2^{6-3}$	$I = a_2 bcE = a_3 cDF = a_1 bDG$	4	3	10
95	$4^1 2^{6-3}$	$I = A_2 bcE = A_3 cdF = A_1 bdG$	4	3	16
96	$4^1 2^{6-3}$	$I = a_2 BCE = a_3 CDF = BCDG$	6	1	392
97	$4^1 2^{6-3}$	$I = a_2 bCE = a_3 CDF = bCDG$	5	2	438
98	$4^1 2^{6-3}$	$I = A_2 bCE = A_3 CdF = bCdG$	5	2	446
99	$4^1 2^{6-3}$	$I = a_2 bcE = a_3 cDF = bcDG$	4	3	86
100	$4^1 2^{6-3}$	$I = A_2 bCe = A_3 CdF = bCdG$	4	3	264

Corrections for Type I Error in Social Science Research: A Disconnect between Theory and Practice

Kenneth Lachlan
Department of Communication
Boston College

Patric R. Spence
Department of Communication
Western Kentucky University

Type I errors are a common problem in factorial ANOVA and ANOVA based analyses. Despite decades of literature offering solutions to the Type I error problems associated with multiple significance tests, simple solutions such as Bonferroni corrections have been largely ignored by social scientists. To examine this discontinuity between theory and practice, a content analysis was performed on 5 flagship social science journals. Results indicate that corrections for Type I error are seldom utilized, even in designs so complicated as to almost guarantee erroneous rejection of null hypotheses.

Key words: Type I error, false positive, Bonferroni, Type II error

Introduction

Despite the breadth of literature in statistical and methodological research detailing the problems associated with Type I error and multiple F tests in factorial ANOVA (Fletcher, et al. 1989, Keppel, 1991, Cohen, 1994, Agresti & Finlay, 1997, Mulaik, Raju, & Harshman, 1997; Smith et al., 2002, Padilla & Algina, 2004), a cursory examination of social science literature suggests that these warnings have been largely ignored. This article briefly reviews some of the literature concerning Type I error rates, then offers an ad hoc content

analysis of several leading journals in different social science disciplines. The results of this content analysis suggest that there is a serious split between statistical literature warning researchers about the Type I error problems associated with multiple F tests in factorial ANOVA, and the actual practice of statistical inference in social scientific research.

Type I errors refer to instances in which a null hypothesis is erroneously rejected. Type I error may be the result of several factors (such as a high alpha level or the violation of statistical assumptions), but the most common source appears to be the number of significance tests that are calculated (Steinfatt, 1979). Although it is well documented that multiple tests along different levels of a single factor will produce Type I errors, less documented is the fact that multiple F tests alone will increase the probability of Type I error (Fletcher, et al., 1989). When testing at the commonly accepted criterion of $p < .05$, one out of every twenty tests will produce an error of Type I (assuming the null hypothesis is always true). Calculations can be performed to compute the expected

Ken Lachlan is Assistant Professor in the Communication Department at Boston College. He earned his doctoral degree from Michigan State University in 2003 email: lachlan@bc.edu. Patric Spence is Assistant Professor in the Department of Communication at Western Kentucky University. He earned his doctoral degree from Wayne State University in 2005.

probability of Type I error through the equation $1 - (1 - \alpha)^c$ where c represents the number of independent comparisons. (Keppel, 1991, Steinfatt, 1979, Smith et al., 2002).

The most commonly used correction for Type I error is a simple reduction of alpha, usually through Bonferroni corrections. These corrections divide the alpha level by the number of tests being performed, then set each test accordingly (Agresti & Finlay, 1997, Cohen & Cohen, 1983, Keppel, 1991).

Fletcher et al. (1989) performed a series of Monte Carlo simulations demonstrating a substantial increase in the number of Type I errors corresponding with the number of factors in a given model. With regard to Bonferroni corrections, Fletcher et al. (1989) reported that Type I error rates dropped from 32 percent to 11 percent through the use of these corrections, using a three-factor ANOVA model in which the null was assumed to be true.

Smith et al. (2002) attempted to extend the work of Fletcher and colleagues by conducting a series of similar Monte Carlo simulations using three and four factor models in which the null is sometimes assumed true and sometimes assumed false. They reported that the addition of main effects into multi-factor models, the use of larger samples, and Bonferroni corrections substantially reduce Type I error rates, to levels as low as 2% across 500 trial models. They caution, however, that Bonferroni corrections may in fact be too conservative and in turn inhibit the detection of true effects, increasing errors of Type II.

Concern over the hypersensitivity of Bonferroni corrections is nothing new. Simes (1986), Hochberg (1988), and Hommel (1988) offer more mathematically sophisticated means of adjusting alpha levels based on sequential adjustments relative to the number of tests that have been

performed, rather than the total number of tests performed on a given model. Keppel (1991) offered a modified Bonferroni adjustment that is based on the number of groups used in the model, as opposed to the total number of tests. Monte Carlo simulations of these techniques demonstrate their effectiveness, and they have been lauded for their ability to effectively reduce Type I error without excessive Type II risk (e.g. McDonald, Seifert, Lorenzet, Givens, & Jaccard, 2002).

As outlined, a substantial body of research has been devoted to identifying and correcting for Type I errors in social science research. Although scholars in applied statistics have debated whether to use the original Bonferroni formula or some type of adjusted formula, the fact remains that correction for Type I errors across multiple tests in multi-factor ANOVA has been identified as a necessary and important component of factorial inference. Without consideration of Type I error, statistical conclusion validity (see Cook & Campbell, 1979) is called into question, with grave implications for the usefulness and validity of findings that are based solely on estimations of the likelihood that they are false (Nickerson, 2000).

However, it is likely that the reader can think of dozens of articles he or she has read recently which have used multi-factor ANOVA procedures and performed numerous F tests, with no regard for Type I errors or the necessary adjustments. Indeed, Smith et al. (2002) in a review of Communication research report that about a quarter of the articles examined featured ANOVA designs of 3 or more factors, with almost none adjusted for the error rates produced by multiple F tests.

The goal of the current analysis to examine a few major journals in the social sciences in order to obtain an estimate of the frequency with which Type I corrections-

Bonferroni or otherwise- are considered and implemented in contemporary research. To do so, a content analysis was performed by the authors examining quantitative research articles in each of the following journals during the 2004 calendar year: Journal of Personality and Social Psychology, Personality and Individual Differences, Human Communication Research, Educational and Psychological Measurement, and the American Journal of Public Health.

Methodology

An initial examination of all articles appearing in these journals during the 2004 calendar year was conducted in order to identify articles using some kind of ANOVA or related analysis. A total of 6 articles were identified among 423 articles appearing in AJPH (1.42%); 36 out of 58 were found for JPSP (62.1%), 4 of 61 (6.6%) for EPM, 10 of 22 (45.5%) for HCR, and 96 of 296 (32.4%) for PID.

Two coders were then given the task of coding several content features of each article. Specifically, they were asked to identify whether or not the article reported ANOVA, ANCOVA, MANOVA, or MANCOVA procedures, the total number of analyses, the number of F tests reported, an estimate of the largest single cell size across all analyses, and whether or not Bonferroni or forms of Type I error correction were employed. Intercoder reliability was calculated using Scott's Pi for categorical variables and Kronbach's alpha for continuous variables; reliability checks on 10% of the sample produced coefficients of at least .87 for all variables.

It should be noted that for definitional purposes coding was completed solely for the number of F tests reported, not an estimated number of total possible F tests. This decision was made for two

reasons: first, to produce a conservative estimate of the number of F tests that were run in each study; and second, because there were numerous instances in which the statistical reporting was so ambiguous that it was impossible to estimate the total number of tests that could have been run.

Results

Results indicate that Bonferroni and other corrections for Type I error are generally absent in these journals, as only 15.8% of the identified articles reported such a correction. More specifically, only 2.8% of the identified JPSP articles reported these corrections, along with 10% of HCR, 16.7% of AJPH, 20.8% of PID, and 25% of the EPM articles. It is perhaps not surprising that Educational and Psychological Measurement, a journal in which psychometric pieces are quite common, would have the highest incidence of reports in which Type I error corrections were performed. But the general tenor of these findings is that Bonferroni and similar procedures are under reported in social science literature.

These results would not be particularly alarming if the studies featured in these journals performed a small number of F-tests with p set at .05. However, there were numerous instances in which this was not the case. Across the entire sample, the average study contained 5.86 ANOVA or ANOVA related analyses, and the average number of reported F-tests was found to be 14.51. Given that a p value criterion of .05 should produce one false positive out of every 20 tests by chance alone, simple frequency distributions were used to determine the number of articles reporting 20 or more tests; in total, 34 of 152 (22.4%) or the articles reported enough F-tests without corrections that at least one Type I error could be expected. A few were

particularly notable, including one study that featured an incredible 111 F-tests.

An argument could be made that ANOVA procedures in the social sciences experience a higher incidence of Type II errors due to small sample sizes and resultant lack of statistical power, thus justifying numerous F-tests without oversensitive corrective measures. Although the authors would argue that this logic leads to errors of both Type I and Type II and actually leads to even less statistical conclusion validity, the concerns associated with underpowered analyses can be seen. However, this analysis suggests that at least in these journals, Type II errors are likely less of a concern.

Across the entire sample, the median score for maximum cell size was found to be 28.5. While there are obviously varying scenarios in which this may present an adequately or inadequately powered analysis, it at least is above the minimum criteria set for adequate cell size in ANOVA analysis. It also needs to be noted that the mean score for maximum cell size (104.6) was intensely skewed by a handful of epidemiological studies with samples of over one thousand. For this reason, the median score is reported, which the authors believe to be a better indicator of central tendency.

Another logical question concerns differences between the studies utilizing error corrections and those that do not in terms of the number of reported F-tests. It could be argued that if the studies running dozens of F-tests are the ones controlling for Type I error, then there is no cause for consternation. This is, however, not the case. *T* tests were used to examine differences in these scores. For reported F-tests, significant differences were not detected between those studies using Type I corrections ($M = 11.5$, $SD = 8.81$) and those that did not ($M = 15.08$, $SD = 18.38$), $t(150) = 3.91$, n.s.

Conclusion

The results reported above, while confined to only a few journals in one calendar year, suggest a disconnect between the statistical analysis and reporting procedures commonly advocated in the statistics literature and the actual practices of social scientists. Across this sample of flagship journals in Public Health, Communication, Psychology, and Education, techniques for reducing Type I error in factorial ANOVA that have been advocated, debated, and refined in the statistics literature for decades are going largely unused. Although this may be less of a concern in the Education and Public Health literatures, empirical Psychology and Communication is largely dependent on ANOVA analyses, especially when utilizing experimental designs.

Further, it should be noted that the statistical reporting under scrutiny was often composed of several sets of multi-factor analyses, leading to situations in which dozens of F-tests were reported. In several studies it could be extrapolated that the null hypothesis may have been erroneously rejected upwards of four times simply due to chance. The most plausible solution is to follow years of advice from the statistics literature and correct alpha with Bonferroni or other adjustments for Type I error, allowing the researcher to differentiate between statistically valid findings and falsely rejected null hypotheses. Although some have suggested that Bonferroni corrections and other adjustments may be overly sensitive and in fact lead to an increase in Type II error (Smith et al., 2002), this increase in Type II error can be avoided through the use of larger samples.

Although obtaining large samples for experimental studies can often be difficult, costly, and time-consuming, it is the opinion of the researchers that the best possible solution is to use enough subjects to provide

for adequate statistical power (see Keppel, 1991) and perform corrections for Type I errors. In instances in which this is not possible it is recommended that social scientists report their effect sizes and establish an a priori criterion for findings that will be considered relevant based on effect size (see Cohen, 1977 for suggested effect size criteria). While not a formal part of this study's coding scheme, it should be noted that the coders observed many articles in which multiple F-tests were reported with no regard for effect size. Reporting only those findings that are statistically significant at .05 and that meet an established criterion for the amount of variance accounted for may be one more solution to the prevalence of Type I error.

References

- Agresti, A., & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd Edition). Prentice Hall.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J.: Erlbaum.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation*. Boston: Houghton Mifflin.
- Hochberg, Y. (1988). A sharper procedure for multiple tests of significance. *Biometrika*, 75, 800-802.
- Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75, 383-386.
- Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Upper Saddle River, N.J.: Prentice Hall.
- McDonald, R. A., Seifert, C. F., Lorenzet, S. J., Givens, S., & Jaccard, J. (2004). The effectiveness of methods for analyzing multivariate factorial data. *Organizational Research Methods*, 5, 255-274.
- Mulaik, S. A., Raju, N. S., & Harshman, R. A. (1997). There is a time and place for significance testing. In L.L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.) *What if there were no significance tests?* Mahwah, N.J.: Earlbaum.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301.
- Padilla, M. A., & Algina, J. (2004). Type I error rates for a one-factor within-subjects design with missing values. *Journal of Modern Applied Statistical Methods*, 3, 406-416.
- Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73, 751-754.
- Smith, R. A., Levine, T. R., Lachlan, K. A., & Fediuk, T. A. (2002). The high cost of complexity in experimental design and data analysis: Type I and Type II error rates in multiway ANOVA. *Human Communication Research*, 28, 515-530.
- Fletcher, H. J., Daw, H., & Young, J. (1989). Controlling multiple F test errors with an overall F test. *Journal of Applied Behavioral Science*, 25, 101-108.
- Steinfatt, T. M. (1979). The alpha percentage and experimentwise error rates in communication research. *Human Communication Research*, 5, 366-374.

Statistical Methods and Artificial Neural Networks

Mammadagha Mammadov Berna Yazıcı Şenay Yolaçan
Atilla Aslanargun Ali Fuat Yüzer Embiya Ağaoğlu
Anadolu University

Artificial Neural Networks and statistical methods are applied on real data sets for forecasting, classification, and clustering problems. Hybrid models for two components are examined on different data sets; tourist arrival forecasting to Turkey, macro-economic problem on rescheduling of the countries' international debts, and grouping twenty-five European Union member and four candidate countries according to macro-economic indicators.

Key words: Time series, ARIMA, neural networks, hybrid models, logistic and probit regression, rescheduling and non-rescheduling of the international debts, Kohonen nets, cluster analysis, Maastricht criteria

Introduction

Artificial Neural Networks (ANN) is efficiently being used as an alternative of statistical methods for different problems like estimation, classification, clustering analysis, sample recognition and etc. Since ANN models are usually nonlinear, those models give better estimates in application.

In this study for real data sets forecasting, classification and cluster problems and hybrid models for two components are examined. The statistical models (ARIMA) used for time series analysis are usually linear. Therefore, using ANN models that can impress

the nonlinear structure has a great importance. Zhang et al. (2001) presented a recent review in this area. There are numerous studies to compare the performances of ANN and traditional time series techniques. For example, the empirical results in Ansuj et al. (1996), Caire et al. (1992), Chin and Arthur (1996), Hill and O'Connor (1996), Kohzadi et al. (1996), Maier and Dandy (1996) showed that the ANN gave improved results in terms of forecasting accuracy.

Recently, an approach was recommended that uses ARIMA and ANN models together for time series forecasts (Tseng et. al., 2002; Zhang, 2003). Ginzburg and Horn (1994) and Pelikan et. al. (1992) proposed to combine several feedforward neural networks to improve time series forecasting accuracy. Wedding and Cios (1996) described a combining methodology using radial basis function networks and Box and Jenkins models. Voort et al. (1996) used a Kohonen self-organizing map as an initial classifier; with each class having an individually tuned ARIMA model associated with it. Luxhoj et al. (1996) presented a hybrid econometric and ANN approach for sales forecasting. Wang and Leu (1996) developed neural network trained by features extracted from ARIMA analysis. Their results showed that the neural network trained by different data produced better predictions than otherwise trained by raw data. Results from Su et al. (1997) showed that the hybrid model

Mammadagha Mammadov is Professor of Statistics. Email: mmammadov@anadolu.edu.tr
Berna Yazici is Assistant Professor of Statistics. Email: bbaloglu@anadolu.edu.tr. Senay Yolaçan is Research Assistant. Email: senayyolacan@anadolu.edu.tr.
Atilla Aslanargun is Assistant Professor of Statistics. Email: aaslanar@anadolu.edu.tr. Ali Fuat Yüzer is Professor of Statistics. Email: afyuzer@anadolu.edu.tr. Embiya Ağaoğlu is Professor Statistics. Email: eagaoglu@anadolu.edu.tr

produced better forecasts than either the ARIMA model or the neural network by itself. Thus, the combined model has produced promising results in these studies.

According to these results on hybrid models, constructed by combining ANN and statistical models or different ANN models makes more accurate time series forecasts. Choosing the hybrid model depends on the statistical and mathematical structure of the data. Of course, usually this relationship can not be explained. Therefore, the components of the hybrid models are chosen by experiments.

Economics, finance, and business have great importance among the problems examined using ANN techniques (Wilson & Sharda, 1992, Kuan & White, 1994, Sharda, 1994, Wong et al., 1994). There are an increasing number of articles on economics that applied ANN models in recent years. Leung et al. (2000) compared the performance of different forecasting techniques and general regression in forecasting exchange rates. Kim et al. (2004) mentioned the usefulness of neural networks for early warning system of economic crisis. They studied the Korean economy as an example of world economic crisis.

One of the other macroeconomic problems that are examined using ANN is country's international debt problems. There are also numerous researches using statistical methods on this subject. Frank and Cline (1971) used multiple discriminant analysis to predict debt-servicing difficulties. Dhonte (1975) used principal components analysis to obtain a description of a country's debt position. Feder and Just (1977) used logit analysis in determining debt servicing capacity. Yoon-Dae Euh (1979) made multiple regression analysis in determining of a country's creditworthiness. Kharas (1984) made probit analysis to assess the probability of a country's becoming uncreditworthy. Cooper (1985) used canonical correlation analysis to examine the relationship between country creditworthiness and a country's recent economic performance.

Rescheduling and non-rescheduling of the international debts of countries is examined. As a result of classification, more successful architectures and algorithms are defined.

Another macroeconomic problem taken into account is whether the Economic and Monetary Union (EMU) appears to be a wholly homogenous group of countries due to Optimal Currency Area (OCA) or not. This is an important question because the sustainability of EMU depends on the existence of a reasonable degree of homogeneity (Artis & Zhang, 2001). A long debate on this problem has taken place after the first publication on optimum currency area among different countries by Mundell (1961) and followed by McKinnon (1963), with important elaborations by, among others, Kenen (1969) and Krugman (1990). Artis and Zhang (2001) examined the status of the EMU member countries by the time the euro was launched with criteria being measured according to optimal currency area criteria using a cluster analysis approach.

In this article, homogeneities in the actual and candidate members of the EMU are examined using Neural Network techniques and cluster analysis, traditional statistical technique.

Artificial Neural Network Algorithms

In this part of the study, a brief discussion is given for ANN as an alternative of statistical methods for estimation, classification and cluster analysis.

Standard Backpropagation (BPM)

First, multi-layer feedforward perceptron (MLP) ANN models that are widely used for estimation and classification problems are taken into account. Backpropagation is the widespread approximation approach for training of the multi-layer feedforward neural networks. The main idea here is minimizing the sum of square error at each epoch (Bishop, 1995; Haykin, 1999). In fact, this algorithm is an application of gradient descent methods at numerical optimization to error function of which the variables are the weights (Nocedal & Wright, 1999). At each epoch the weight vector is changed in the direction of the error function gradient in defined ratio at the currency point and the value of error is lessened:

$$w_{k+1} = w_k - \alpha \cdot g_k$$

where w_k is the current vector of weights and biases, g_k is the current gradient of error function at the point w_k , and α is the learning (training) rate.

In the simple gradient method there are some difficulties about local minimum and speed of convergence. Thus, the momentum term is added to weight change formula:

$$\Delta w_{k+1} = -\alpha \cdot g_k + \mu \Delta w_k, \quad (1)$$

where $\Delta w_k = w_k - w_{k-1}$ is the weight change in the previous iteration, μ is the momentum coefficient. The spread algorithm with momentum is called the standard backpropagation algorithm. In this algorithm, the better choice of α and μ constants speed of convergence and stability of the algorithm (Yu & Chen, 1997; Qian, 1999; Bhaya et al., 2004).

Resilient Backpropagation (RP)

MLP's typically use S-shaped sigmoid activation functions in the hidden layers. For this reason, their slope must approach zero as the input gets large. The purpose of the resilient backpropagation (RP) training algorithm is to eliminate these injury effects of the magnitudes of the partial derivatives. In RP algorithm the magnitude of the derivative has no effect on the weight update. Only the sign of the derivative is used to determine the direction of the weight update. If the derivative is zero, then the update value remains the same. Whenever the weights are oscillating the weight change will be reduced. If the weight continues to change in the same direction for several iterations, then the magnitude of the weight change will be increased.

There are some different versions of backpropagation ANN algorithm like Conjugate Gradients and Newton method. Those algorithms which are used for the experimental studies in the article are briefly given in this section.

Conjugate Gradients (CG) Algorithm

This is a special version of conjugate direction method. For a starting point $w_0 \in \mathbb{R}^n$,

the method defined by the formulas below called conjugate direction method (Nocedal & Wright, 1999):

$$w_{k+1} = w_k + \alpha_k p_k, \quad \alpha_k = -\frac{p_k^T g_k}{p_k^T H p_k} \quad (2)$$

Here, α_k is defined by the minimum problem of one variable function of $\varphi(\alpha) = F(w_k + \alpha p_k)$ and $\{p_i\}_{i=1}^n$ conjugate directions. For any starting point $w_0 \in \mathbb{R}^n$, the sequence $\{w_k\}$ generated by the conjugate direction algorithm converges to the minimum point w^* of the problem (2.3) in at most n steps (Nocedal & Wright, 1999). The conjugate gradient method is a conjugate direction method; start out by searching in the gradient direction on the first iteration

$$p_0 = -g_0 \quad (3)$$

The conjugate current p_k vector is calculated by using the previous p_{k-1} vector and current gradient

$$p_k = -g_k + \beta_k p_{k-1} \quad (4)$$

Coefficient β_k is selected by the condition of p_{k-1} and p_k vectors' being conjugate with respect to the symmetric positive definite $n \times n$ Hessian matrix H ($p_{k-1}^T H p_k = 0$):

$$\beta_k = \frac{g_k^T H p_{k-1}}{p_{k-1}^T H p_{k-1}}. \quad (5)$$

There are various CG algorithms depending on different calculation formulas of β_k coefficients, for example Fletcher-Reeves, Polak and Ribere conjugate gradient algorithms. Some important examples are due to Bishop (1995), Haykin (1999), and Nocedal and Wright (1999).

Another algorithm of CG algorithms is Scaled Conjugate Gradients (SCG) Algorithm. CG algorithms reviewed above makes a line search at each iteration and makes calculations

very expensive. Moller (1993) put forward scaled conjugate gradient algorithm that does not use line search procedure in traditional CG algorithms. The idea of SCG is to combine the model trust region approach with the conjugate gradient approach (Bishop, 1995; Nocedal & Wright, 1999).

Quasi-Newton (QN) Algorithms
Newton's method is

$$w_{k+1} = w_k - H_k^{-1} g_k, \quad (6)$$

where H_k is the Hessian matrix in current point w_k . Newton's method often converges faster than conjugate gradient methods. Unfortunately, it is complex and expensive to compute the Hessian matrix. Quasi-Newton (or secant) methods are based on Newton's method, but don't require the calculation of second derivatives. They update an approximate Hessian matrix at each iteration of the algorithm (Nocedal & Wright, 1999). (The update is computed as a function of the gradient.)

The Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm is one of the successful QN algorithms. The one step secant (OSS) algorithm is an attempt to bridge the gap between the CG algorithms and Quasi-Newton algorithms. This algorithm accepts that the previous Hessian is the identity matrix and so inverse matrixes are not calculated to choose the new search direction. This algorithm requires very light storing and requires less storing at each step than the CG algorithms.

The Levenberg-Marquart (LM) algorithm is one of the required QN algorithms (Bishop, 1995). Hessian matrixes are not calculated and not taken into account of its being second order, the Hessian matrix can be approximated as $H = J^T J$, where J is the Jacobean matrix that contains first derivatives of the network errors with respect to the weights and biases. The gradient can be computed as $g = J^T \mathcal{E}$, where $\mathcal{E}$ is a vector of network error. The Levenberg-Marquart algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$w_{k+1} = w_k + [J^T J + \mu I]^{-1} J^T \mathcal{E}. \quad (7)$$

This algorithm runs fast for moderate-dimensioned feedforward neural networks for regression problems.

Radial Basis Function Networks (RBFN)

RBFN (Haykin, 1999; Bishop, 1995) are also used, aside from MLP networks, in regression and classification problems. In RBFN, one hidden layer with required number of units is enough in order to model a function. The activations of hidden (radial) units are defined depending on the distance of the input vector and the center vector. Typically, the radial layer has exponential activation functions and the output layer a linear activation function.

Education is made in three stages in RBFN. In the first stage by unsupervised education, radial basis function centers are optimized using all $\{x^{(i)}\}$, $i = 1, 2, \dots, N$ education data. Centers can be assigned by a number of algorithms: Sub-Sampling, K-means, Kohonen training, or Learned Vector Quantization. In the second stage σ_j , $j = 1, 2, \dots, p$ deviation (or width) parameters can be assigned by algorithms such as Explicit, Isotropic, K-nearest neighbor. In the third stage of education, the basis functions that are obtained for adjusting the appropriate weights for output units are taken fixed and deviation parameters are added to linear sum. Optimum weights are obtained by minimization of the sum of square errors. The output layer is usually optimized using the Pseudo-Inverse technique.

MLP with a defined architecture is given by the appropriate weights and the biases of the units, but in RBFN, it is given by the center and the deviation of the radial units and by the weights and biases of the output units. Since point is given by n coordinates in n -dimensional space, the number of the coordinates are equal to the linear input units n . So in ST Neural Network software, the coordinates of the center radial unit are taken as weights, and the deviation of the radial unit is taken as bias. As a result, radial weights denotes the center point, radial bias denotes the deviation.

Having only one hidden layer and making faster education than MLP can be taken as advantages of RBFN. The linear modeling methods are more useful in output layers of RBFN that removes the difficulties which occur about the local minimums in MLP.

However, RBFN has some disadvantages compared with MLP. In order to correctly modeling a typical function in RBFN, many more hidden (radial) units can be required than appropriate MLP model. That may cause running the model slower and more memory can be required. RBFN is very sensitive to increment of the network dimension and some difficulties can be occurred as a result of increment in the number of input units.

RBFN is unsuccessful in extrapolation in its nature. MLP networks are more successful in extrapolation problems than RBFN. Because, when the input data are far from the radial centers, than the output signal is 0, and this may not show the required result.

Kohonen Networks (Kohonen Self-Organizing Maps (SOM))

The algorithms given above belong to the supervised training rule. That training is evaluated due to a given target. Nevertheless Kohonen networks are fulfilled a topologic structure among the cluster units by unsupervised training. That property is observed as one characteristic of the brain and does not occur at other ANNs. These Networks are based on competitive learning. The neurons of competitive Networks learn to recognize groups of similar input vectors. An output neuron that wins the competition is called a winning neuron. The weights of winning neuron with index j and its specified i^{th} unit neighborhood in training algorithm are updated as follows:

$$w_{ij}(k+1) = w_{ij}(k) + \alpha[x_i - w_{ij}(k)] \quad (8)$$

where α is the learning ratio. At each epoch of algorithm, the radius R and learning ratio α are changed by decreasing. There are some alternative structures in order to decrease R and α . Kohonen (1989) showed that α is satisfactory for linearly decreasing functions,

and geometrically decreasing can produce similar results.

An Application Study on Time Series Using ARIMA, ANN and Hybrid Models

The statistical models used for time series analysis are usually linear. Thus, using ANN models that can impress the nonlinear structure is of great importance. Zhang et al. (2001) presented a recent review in this area. Experimental results with real data sets indicate that a hybrid methodology that combines both ARIMA and ANN models can be an effective way to improve forecasting accuracy achieved by either of the models used separately. G. P. Zhang (2003) explained the reasons for using hybrid models in detail.

According to hybrid model results, combining ANN and statistical models or different ANN models produces more accurate forecasting for time series. Choosing the hybrid model depends on the statistical and mathematical structure of the data. Of course, this relationship can not usually be explained. As a result, the components of the hybrid models are chosen by experiment.

Some articles on tourism forecasting problems mention ANN as having better performance than statistical techniques. Law and Au (1999) found that using the feedforward neural network model to forecast annual Japanese arrivals for travel to Hong Kong outperformed multiple regression models, naive, moving average, and exponential smoothing in terms of forecasting accuracy. The empirical results of Law (2000) showed that BP neural network outperformed regression models, time series models and feedforward neural networks in terms of forecasting accuracy. In his article, Vincent Cho (2003) investigated the applications of three time series forecasting techniques, namely exponential smoothing, univariate ARIMA, and Elman's model of ANN. Neural networks seem to be the best method for tourist arrival forecasting. Kim et al. (2003) provided a short introduction to the use of Kohonen's SOM algorithm in tourism and presents a descriptive analysis of the ANN methodology. It provides a research technique that assesses the weighting of different attributes, and uses an unsupervised

ANN model to describe a consumer product relationship.

The ANN approach and the hybrid methodology to time series modeling

For one hidden layer network architecture $n:p:1$ (n : number of inputs, p : number of hidden units and 1 : number of outputs), inputs are the observed values of n^{th} previous time points and outputs (targets) are $(n+1)^{\text{th}}$ the observed value. ANN are nonlinear functions of previous observations $(y_{t-1}, y_{t-2}, \dots, y_{t-n})$ to y_t future observations (Zhang, 2003):

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}, w) + \varepsilon_t, \quad (9)$$

where $(y_{t-1}, y_{t-2}, \dots, y_{t-n})$ are input values, y_t is target value, w are weights of the network, ε_t are the vector of biases at time point t . The $\hat{y}_t$ prediction value is calculated as follows:

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}, w). \quad (10)$$

In training procedure, with the help of different backpropagation algorithms, the parameters (weights and biases) of the network are obtained by getting closer to the minimum value of the

sum of the square error $SSE = \sum_{t=n+1}^N (y_t - \hat{y}_t)^2$.

In recent articles (Tseng et. al., 2002; Zhang, 2003), more importance is given to hybrid models that are composite of ARIMA and NN models. However, the proposed hybrid model does not always show a better performance. The idea of model combination in

forecasting is to use each model's unique feature to capture different patterns in the data (Zhang, 2003).

On time series sampling in this study, ARIMA&NN and different NN&NN hybrid models are evaluated with experimental calculations, with interesting results obtained. As a result of the experiment, it is seen that hybrid models like MLP&MLP and MLP&RBFN with two nonlinear components show better performance for forecasting problems. So, the hybrid model structure of Zhang (2003) can be extended. This means that both of the components of the hybrid model may be nonlinear at the same time (Aslanargun et al., 2007):

$$y_t = y_t^1 + y_t^2 \quad (11)$$

In this model, y_t is the observation value at time point t , y_t^1 and y_t^2 are linear or nonlinear model components, and superscripts denote the row number of the model. Firstly, the model with 1-indiced is applied to the observation data and $e_t^1 = y_t - \hat{y}_t^1$, then the others are calculated.

Here, $\hat{y}_t^1$ is the forecast value of the first model at time point t . If the first model contains m_1 input units, the number of e_t^1 units will be $N - m_1$. If the second model contains m_2 input units, the number of $\hat{y}_t^2$ forecast values will be $N - m_1 - m_2$. In this case, the forecast values appropriate for the second model are calculated as follows:

Figure 1. Number of monthly tourist arrivals to Turkey (1984-2001).

$$\hat{y}_t^2 = f_2(e_{t-1}, e_{t-2}, \dots, e_{t-m_2}) \quad (12)$$

where f_2 is the function obtained from the second model. The forecast for the combined model is defined as follows:

$$\hat{y}_t = \hat{y}_t^1 + \hat{y}_t^2 \quad (13)$$

The adjusted forecasts are calculated as the sums of the first model and the second model. The hybrid model with good performance is obtained by the evaluation measure for the forecasting.

Experimental evaluation

In this part of the study, the time series of number of monthly tourist arrivals to Turkey was examined. Appropriate ARIMA, ANN and hybrid models were chosen by doing experiments in order to make forecasts, and these models are also compared. For the analyses STATISTICA Neural Networks and SPSS statistical packages are used.

The data set, the number of monthly tourist arrivals to Turkey between January 1984 and December 2003, is taken from the Republic of Turkey, Prime Ministry State Institute of Statistics (www.die.gov.tr).

The graph of 216 month period for January 1984 to December 2001 is given in Figure 1. Examining Figure 1 it is seen that the series is not stable in variance, after the logarithmic transformation variance stability criterion is provided, and it can also be seen that the series includes seasonal and trend effects. Thus, it was decided that the most appropriate model for this series is $ARIMA(1,1,1)(1,1,0)_{12}$.

The 216 monthly tourist arrivals to Turkey data for January 1984 to December 2001 period was used in training of the network. An evaluation of the model was made dependent on the forecasts for the 24 month period between January 2002 and December 2003. The choice of the best model depends on a comparison of statistics such as the MSE (RMSE) and MAE. Since the initial weight and bias values of the network were random, 150 replications were

Table 1. The performance values for the selected models

Models	MSE	RMSE	MAE
ARIMA(1,1,1)(1,1,0) ₁₂	3.06E+10	174926.3	137589.4
(12:1)LinearNN	2.93E+10	171046.8	144662.4
(12:1:1)MLP	2.19E+10	147909.3	127838.8
(12:48:1)RBFN	5.40E+10	232280.8	176592.2
ARIMA(1,1,1)(1,1,0) ₁₂ &(6:8:1)MLP	2.79E+10	167068.6	128148.6
ARIMA(1,1,1)(1,1,0) ₁₂ &(6:7:1)RBFN	2.86E+10	169058.0	129721.5
(12:1)LinearNN&(6:8:1)MLP	2.73E+10	165220.0	141377.4
(12:1)LinearNN&(9:15:1)RBFN	2.57E+10	160479.0	139907.8
(12:1:1)MLP&(5:1:1:1)MLP	2.09E+10	144410.7	123104.7
(12:1:1)MLP&(6:9:1)RBFN	2.07E+10	143958.1	123184.7
(12:48:1)RBFN&(12:8:8:1)MLP	4.78E+10	218540.3	160724.3
(12:48:1)RBFN&(12:12:1)RBFN	4.99E+10	223470.6	167694.1

made for the same network structure and the models giving the best forecasts were determined.

Because the tourist arrival data in question included seasonality, after trying many neural networks with different numbers of input units, as expected, the number of input units was determined as 12. Various neural network algorithms with single layer, with one or two hidden layers, multilayer feedforward algorithms and RBFN models were applied on the sample data set. Twelve were lost due to seasonality, so 204 of 216 were used to adjust the weights. In the training stage of the network, data were divided into two parts; 132 of the 204 data were used for training and 72 data were used for validation. This division was used to restrict memorization of the network and provided for better forecasts (Bishop, 1995, Haykin, 1999).

The single layer neural network with (12:1) architecture (12 input units, 1 output unit and 1 bias) showed better performance for the forecast data as a result of adjustment of the suitable weights in training. The neural network that showed the best performance among the MLP for the time series in question was found to be the (12:1:1) MLP. The QN method showed the best performance in 33 epochs. Among the RBF Networks the (12:48:1) RBFN showed the best performance. The hyperbolic-tangent function is applied in the hidden layer and the linear activation function is applied in the output unit. The biases and weights resulting from training are obtained.

Hybrid models such as the ARIMA&MLP, LinearNN&MLP, ARIMA&RBFN, LinearNN&RBFN and MLP&MLP, MLP&RBFN, RBFN&MLP, RBFN&RBFN were taken into account. In order to determine the ARIMA&MLP and the ARIMA&RBFN hybrid models, firstly, the ARIMA(1,1,1)(1,1,0)₁₂ model was applied to the data. A series of residuals, the differences between the observed values and the estimates of these models were obtained. In the series of residuals, since a 1st order difference and a 1st order seasonal differences were taken in the ARIMA model, 13 data were lost and 203 data remained. At the next step, the MLP or RBFN models were applied to the series of residuals. The same procedure is applied for other hybrid models.

The hybrid models with the best performance are given in Table 1. The weights and biases are also obtained for all the second components of hybrid models.

Results Part 1

The observed and forecasted values of number of monthly tourist arrivals to Turkey for January, 2002 and December, 2003 period is obtained. Also MSE, RMSE and MAE values for the forecasts are obtained in order to compare the performances of ARIMA, ANN and hybrid models mentioned in previous sections and results are given in Table 1.

As seen in Table 1, due to the MSE and RMSE measures, (12:1:1)MLP&(6:9:1)RBFN,

(12:1:1)MLP&(5:1:1:1)MLP hybrid models and (12:1:1)MLP ANN models respectively; due to the MAE measure (12:1:1)MLP&(5:1:1:1)MLP, (12:1:1)MLP&(6:9:1)RBFN hybrid and (12:1:1)MLP ANN models respectively make the best forecasts. The graph of forecasted values is given in Figure 2.

Among the models with one component (the models that are not hybrid models), due to the MSE and MAE measures, the (12:1:1)MLP showed the best performance, and the (12:48:1)RBFN showed the worst performance. It is known that the RBFN is usually unsuccessful in extrapolation problems (Bishop, 1995). In this manner, the result from Table 1 for the RBFN is an expected result. Although the RBFN makes bad forecasts by itself, as can be seen from Table 1, when it is used as a second component, the (12:1:1)MLP&(6:9:1)RBFN hybrid model showed the best performance when it is used as

a second component. When only the RBFN is used, the number of radial units is defined as 48. In other words, when the start data set is used in education and forecast, 48 radial centers are selected. In the (12:1:1)MLP&(6:9:1)RBFN hybrid model, when the second component RBFN is applied to the remaining data, the number of radial centers are decreased to 9. This situation can be explained thus; the differences between the residuals are getting smaller.

An Application Study on a Classification Problem

Interest on macro economic problems using ANN has increased since 1990s. Roy and Cosset (1990) developed a neural network to estimate sovereign credit ratings. Burrell and Folarin (1997) used neural networks in order to maintain competitiveness in a global economy and collected the studies on global economy using neural networks until then. Cooper (1999)

Figure 2. The observed and forecasted values for the January 2002 to December 2003 period.

compared the performance of artificial neural networks with logistic, probit regression and discriminant analysis to classify the countries to two classes; reschedule or nonreschedule the international debts.

Rescheduling and non-rescheduling of the international debts of countries is examined using data for 1982. To analyze the data taken into account, logistic and probit regression among the statistical methods and different backpropagation algorithms of multilayer networks as ANN are used. As a result of classification, more successful architectures and algorithms are defined.

Definition of Macro Economic Problem

Because international debt crisis effects all economical and political balances of countries, in this study, countries' rescheduling and non-rescheduling of their international debts is examined. In 1982, international debt problem became such a big problem that detailed studies required for lending phase of countries (Cooper, 1999). It became an investigation topic whether the countries will pay their debt in time or not, whether they reschedule their debt or not. Hence, in this study countries' rescheduling and non-rescheduling of the international debts is considered as a classification problem and the data for 70 countries of 1982 are used.

The factors that affect the rescheduling and non-rescheduling of the international debts defined as follows and those factors are taken as independent variables for logistic and probit regression, and inputs for ANN:

X_1 : Average increase in GDP growth (annual %) over 1960-1982

X_2 : The ratio of short term debt (%GDP) to export of goods and services (%GDP)

X_3 : Debt service ratio, the interest on total external debt plus amortization on long-term debt as a percentage of the exports of goods and services in 1982.

X_4 : The import of cover and equals international reserves divided by total imports in 1982.

The data for this study which were collected from Morgan Guaranty Trust Company and World Bank is taken from Cooper (1999).

The number of input variables is four for MLP model since all those variables are taken as input variables. The output variable can either be 1 for rescheduled debts if probability is equal or greater than 0.5 or 0 for non-rescheduled debts if probability is less than 0.5. So in output layer of the network, there is only one neuron.

Analysis

Cooper (1999) applied logistic, probit, discriminant analysis and gradient descent with momentum with one hidden layer 4:8:1 algorithm for countries divided to 3 groups of 20 units. But, 4:8:1 architected resilient backpropagation of ANN algorithms that is applied in this study give 100% correct classification.

In this study, logistic and probit regression analysis are made for modeling of the data related with 70 countries. The obtained models are given as follows:

$$L = -3.4610 - 0.1872X_1 + 0.0228X_2 + 0.0869X_3 - 0.1782X_4$$

$$P = -1.9650 - 0.0879X_1 + 0.0117X_2 + 0.0471X_3 - 0.0669X_4$$

To the same data, gradient descent with momentum algorithm (GDM), scaled conjugate gradient algorithm (SCG) and resilient backpropagation algorithm (RP) of backpropagation algorithms of ANN are applied. In order to construct models for extrapolation with ANN, it is useful to divide the data to three different groups for training, validation and test. During the process of modeling, training is made by the training data and total square error related to validation data is calculated in each epochs. In a certain epoch, training is stopped when the error related with the validation data start to increase and the weights of the net is determined. If the errors corresponding to training, validation and test data have small values in the same time, the architecture of determined neural net and the algorithm is respected to be good (Demuth & Beale, 1996). The model constructed by this training method, satisfy the opportunity to make more successful extrapolation.

Table 2. Extrapolation for the data consist of 1988-2001 related to Turkey

	Data			
	X_1	X_2	X_3	X_4
1988	2.1	0.8	2.4	2.1
1989	0.03	0.09	2.4	3.5
1990	9.3	1.4	2.3	4.2
1991	0.09	1.3	2.3	3.9
1992	0.06	1.6	2.4	4.4
1993	0.8	2	2.5	4
1994	-5.5	0.8	2.1	4.2
1995	7.2	1.1	0.2	5.7
1996	7	1	1.8	6.4
1997	7.5	0.9	1.8	6.4
1998	3.1	0.9	1.7	7.4
1999	-4.7	1	2.3	9
2000	7.4	1	2	7.4
2001	-7.5	0.4	2.9	6.3

Extrapolation						
	SCG		GDM		RP	
	(4:8:1)	(4:11:1)	(4:8:1)	(4:11:1)	(4:8:1)	(4:11:1)
1988	0.9931 (1)	0.9990 (1)	0.9945 (1)	0.8473 (1)	0.9899 (1)	0.9937 (1)
1989	0.7857 (1)	0.9017 (1)	0.8438 (1)	0.5536 (1)	0.9694 (1)	0.9806 (1)
1990	0.9981 (1)	0.9994 (1)	0.9982 (1)	0.8648 (1)	0.9940 (1)	0.9995 (1)
1991	0.8176 (1)	0.9764 (1)	0.8912 (1)	0.5742 (1)	0.9535 (1)	0.9053 (1)
1992	0.8255 (1)	0.9828 (1)	0.8981 (1)	0.5862 (1)	0.9516 (1)	0.9110 (1)
1993	0.8991 (1)	0.9969 (1)	0.9704 (1)	0.7703 (1)	0.9819 (1)	0.9706 (1)
1994	0.0003 (0)	0.0009 (0)	0.0003 (0)	0.0045 (0)	0.0002 (0)	0.0033 (0)
1995	0.9979 (1)	0.9994 (1)	0.9981 (1)	0.8503 (1)	0.9941 (1)	0.9993 (1)
1996	0.9978 (1)	0.9994 (1)	0.9980 (1)	0.8507 (1)	0.9941 (1)	0.9995 (1)
1997	0.9979 (1)	0.9994 (1)	0.9981 (1)	0.8531 (1)	0.9941 (1)	0.9995 (1)
1998	0.9957 (1)	0.9994 (1)	0.9965 (1)	0.8483 (1)	0.9939 (1)	0.9990 (1)
1999	0.1303 (0)	0.1080 (0)	0.1261 (0)	0.1936 (0)	0.3335 (0)	0.2378 (0)
2000	0.9979 (1)	0.9994 (1)	0.9980 (1)	0.8529 (1)	0.9941 (1)	0.9995 (1)
2001	0.0003 (0)	0.0001 (0)	0.0001 (0)	0.0045 (0)	0.0003 (0)	0.0019 (0)

The data set of 70 units is divided to 4 parts, the half of the data set is used for training, the quarter of the data set is used for validation and the remained quarter is used for testing. GDM, SCG and RP algorithms are applied as ANN algorithms. Consequently, although ANN models constructed by using the whole data give successful result, they give unsuccessful result for extrapolation. However, ANN models constructed by division of the data set can make successful extrapolation. Extrapolation for the data consist of 1988-2001 related to Turkey is given in Table 2.

Results Part 2

The 4:8:1 and 4:11:1 architected ANN are selected from variety of experiments. The obtained correct classification ratios are summarized in Table 3. 4:8:1 and 4:11:1 with one hidden layer model which give better performance are selected as feed forward ANN architecture. The correct classification ratio for each algorithm is calculated and results are given in Table 4.

Comparisons are made according to correct classification ratios calculated by using probabilistic values related with the results of logistic and probit regression analysis of statistical techniques and ANN algorithms. It is obtained as a conclusion that ANN models give better results in classification. While SCG of ANN algorithms is giving more successful correct classification ratio, logistic regression analysis gives more successful results than probit regression.

ANN models constructed by using whole data give unsuccessful results for extrapolation. This situation can be explained as follows; ANN models constructed for whole data memorize the data, hence they don't give successful result for extrapolation. However, ANN models constructed by the divided data (training-validation-test) give successful result for extrapolation. If the results of the application are examined, it can be seen that only the target results for 1994, 1999 and 2001 can not be obtained. The reason can be explained by the economic crisis in these years.

Table 3. Results of ANN models (whole data) and Logistic, Probit Regression

Model Criteria for Goodness of fit	SCG		GDM		RP		Logistic Regression	Probit Regression
	4:8:1	4:11:1	4:8:1	4:11:1	4:8:1	4:11:1		
Correct Classification Ratio (%)	95.7	97	90	95.7	92.9	94.3	90	88.6

Table 4. Results of ANN models (training-validation-test)

Model Criteria for Goodness of fit	SCG		GDM		RP	
	4:8:1	4:11:1	4:8:1	4:11:1	4:8:1	4:11:1
Correct Classification Ratio (%)	87	85.7	82.9	84.3	82.9	85.7

Clustering actual and candidate members of EU due to economical indicators

Countries are different from each other due to economical indicators such as inflation, budget deficit, gross domestic product etc. In order to construct monetary union, countries try to minimize their economical differences. For example, the Economic and Monetary Union (EMU) is one of the constructed unions. Studies on optimality (based on the similarity of economic indicators) of currency areas have started by the article published by Mundell in 1961 and followed by various studies. Since then, it is general investigation topic whether euro area countries or non-euro area countries satisfy the Maastricht Criteria or not.

Definition of the problem of clustering countries

Maastricht criteria describe the necessary conditions that actual member of EU have to satisfy for EMU entry, according to the Maastricht Treaty of February 7th, 1992. The Maastricht criteria are the five conditions set that countries had to meet if they want to take part in full economic and monetary union. These are:

1. Inflation - no more than 1.5% above the average inflation rate of the lowest 3 inflation countries in the EU
2. National debt - no more than 60% of GDP
3. Budget deficit - no more than 3% of GDP

4. Interest rates - the long-term rate should be no more than 2% above the average of the three countries with the lowest inflation rates

5. Exchange rates - currency within the normal bands of the ERM with no re-alignments for at least 2 years

Only 12 actual members of EU within 25 members are in the EMU and are in the euro area. These members are as follows: EU countries in euro area are Austria, Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Portugal, Spain and Holland. EU countries in non-euro area are Cyprus, Czech Republic, England, Estonia, Finland, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia and Sweden.

For evaluation of 25 actual members and 4 candidate members of EU based on Maastricht criteria for EMU, 4 economical indicators (inflation rate, budget deficit, interest rate, public debt) are taken into account.

Additionally 3 different variables (total employment rate, share of employment of corresponding country versus EU, share of external trade of corresponding country versus EU) are considered and the variables for data set are defined as follows:

X_1 : Total Employment Rate (%)

X_2 : Inflation Rate

X_3 : Interest Rate

Table 5. Clusters by Ward's cluster technique

Cluster No	Countries in the cluster
1	Greece, Italy, Hungary, Slovenia, Malta, Portugal
2	Belgium, Sweden, Lithuania, Czech Republic, Germany, Estonia, France, Luxembourg, Latvia, Ireland, Spain
3	Austria, Finland, Cyprus, Slovakia, Denmark, Poland, Holland, England

Figure 3 - Cluster dendrogram obtained by Ward's cluster technique

- X₄: Budget Deficit/GDP
- X₅: Public Debt/GDP
- X₆: Share of employment
- X₇: Share of external trade

(The data set of economical variables is collected from the EUROSTAT of 2003.)

Kohonen Nets of Neural Network Architectures (Kohonen, 1990) is recently a new cluster technique. This technique groups the data such as holding similar properties. Based on this property of Kohonen nets, in first stage this net is applied to cluster 25 actual members of EU and besides traditional statistical techniques of cluster analysis are applied. Then, in order to determine (classify) which set the 4 candidate members of EU belong, multilayer ANN techniques are applied. In this part of study, depending on the obtained results, which of those 29 countries can really join to EMU under the certain circumstances is identified and recommended. Applications are realized by using STATISTICA 6.0.

Analysis

In order to separate 25 member countries of EU in homogenous sets based on main macro economical indicators, use Ward's cluster technique of Hierarchical cluster analysis and then apply the k-means technique. Sets obtained by Ward's cluster technique are given in Table 5. A dendrogram obtained by Ward's cluster technique is given in Figure 3.

In this part of the application, K-means technique is applied to check the correctness of the clusters. The outcome of this technique is the same as the outcome of Ward's technique.

Kohonen Nets of ANN are applied to the same data in order to cluster the countries. To realize clustering using Kohonen nets the number of sets is given before as k-means technique. Because three sets are obtained by Ward's Clustering Technique, the same number of set is used to apply Kohonen Nets. After the training Kohonen Net, SOFM7:7-3:1 ANN model is obtained.

According to the win frequencies of this model, there are 6 countries in the first cluster, 7 countries in the second cluster and 12 countries in the third cluster. Observing the topological

map of appropriate Kohonen Net, the clusters of the countries are given in Table 6.

After clustering 25 actual member of EU based on Ward's technique and K-means technique, at first, in order to determine (classify) which cluster the 4 candidate members of EU belong, multilayer ANN techniques are applied. MLP 6:6-6-3:1 model is selected as appropriate model within 1000 simulation. According to this model, clustering of the candidate countries of EU is given in Table 7.

Later, depending on cluster analysis by Kohonen Net, in order to determine (classify) which cluster the 4 candidate members of EU belong, multilayer ANN techniques are applied. RBF 6:6-3-3:1 is selected as appropriate model within 1000 simulation. According to this model, clustering of the candidate countries of EU is given in Table 8.

Results

Traditional statistical techniques of cluster analysis and Kohonen Nets of ANN are applied for clustering 25 actual members of EU. By each of the techniques, three separate clusters are performed, and especially it is seen that the third cluster of the both techniques consists of approximately the same countries. It is investigated that in which group of 4 candidate countries will take place according to basic macro economical indicators of 25 EU countries, and as another result of both of the techniques Croatia, Romania and Turkey take place in the same cluster and Bulgaria takes place in another cluster.

Table 6. Clustering countries using Kohonen Net

Cluster No	Countries in the Cluster
1	Greece, Hungary, Malta, Portugal, Slovenia, Spain
2	Cyprus, Czech Republic, Estonia, Italy, Latvia, Lithuania, Sweden
3	Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Luxembourg, Poland, Slovakia, Holland, England

Table 7. Clustering of the candidate countries of EU based on Ward's technique due to ANN

Cluster No	Countries in the cluster
2	Bulgaria
3	Croatia, Romania, Turkey

Table 8. Clustering of the candidate countries of EU based on Kohonen Net due to ANN

Cluster No	Countries in the cluster
1	Croatia, Romania, Turkey
3	Bulgaria

The unique euro-area country that satisfies all of the Maastricht Criteria in 2003 is Portugal. By 2003, actual members of EU that don't entry EMU but satisfy all of the Maastricht criteria are Czech Republic, Finland and Lithuania. Other countries' currency is Euro although they don't satisfy the criteria for that data set but they used to satisfy the criteria since 1993. On the other hand, with regard to some countries' not using Euro, although they satisfy the criteria, they can be explained by the political reasons.

According to traditional statistical techniques of cluster analysis, by 2003, countries in the first cluster satisfy the 3rd and 4th criteria, countries in second and third clusters satisfy at least one criterion. Bulgaria is a single candidate country of EU that mostly satisfies the Maastricht criteria, except for 1st criteria.

Conclusion

It is seen that for forecasting, classification and clustering problems, when suitable ANN algorithms are chosen better estimates are obtained compared to the statistical methods. Also, ANN has a great advantage of not requiring the assumptions which are necessary for statistical methods. Moreover, ANN has another advantage for extraneous data and prediction. However, the choice of suitable architecture and algorithm has a great importance. Depending on the idea of getting better results using RP and SCG algorithms for classification problems, those algorithms are suggested for that kind of problems. It is also necessary to be cautious about the QN algorithms that can give good results at the stage of choosing the suitable network algorithms. On the other hand, for the problem related with time series, it is seen that the hybrid models with both components are statistical or ANN models, give better performances. Also, the hybrid model's good performance with both nonlinear components changes the notion that the first component is linear.

Although in extrapolation problems when RFBN is applied by itself it is found unsuccessful, when it is used as second component of hybrid model, it provides better performance. In our opinion the reason of that is,

after applying the first component to the data set, the difference between the units of neighborhood of the residuals lessens. That provides the opportunity of accumulation of residuals at a defined central (radial) points' neighborhood.

The result of being the hybrid model's more advantageous for forecasting problems; it gives the idea of constructing and using suitable hybrid models for other statistical problems.

The best model should be chosen depending on the statistical and the mathematical structure of the data set. Unfortunately, there aren't certain criteria or test about the selection of appropriate model subject to nature of the data. The numerous experimental studies which will be done on this subject can provide a definite and useful aggregation on a criterion.

References

- Ansuj, A. P., Camargo, M. F., Radharamanan, R. & Petry, D. G. (1996). Sales forecasting using time series and neural networks. *Comput. Ind. Eng.* 31 (1), 421-424.
- Aslanargun, A., Mammadov, M., Yazici, B. and Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series tourist arrival forecasting. *Journal of Statistical Computation and Simulation*, Vol.77, Issue 1, 29-53.
- Bhaya, A. & Kaszkurewicz, E. (2004). Steepest descent with momentum for quadratic functions is a version of the conjugate gradient method. *Neural Networks*, 17, 65-71.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.
- Box, G. P. & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco, CA: Holden-Day.
- Caire, D., Hatabian, G., & Muller, C. (1992). Progress in forecasting by neural networks. *International Conference on Neural Networks II*, 540-545.
- Chin, K. & Arthur, R. (1996). Neural network vs. conventional methods of forecasting. *Journal of Business Forecasting*, 14 (4), 17-22.

- Cho, V. (2003). A comparison of different approaches to tourism arrival forecasting. *Tourism Management*, 24, 323-330.
- Cooper, J. C. B. (1985). Country creditworthiness: A canonical correlation analysis. *Glasgow Caledonian University Economics Discussion Paper*, No. 3.
- Cooper, J. C. B. (1999). Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*, 26, 909-921.
- Dhonte, P. (1975). Describing external debt situations: A roll over approach. *International Monetary Fund Staff Papers*, 24, 159-186.
- Feder, G. & Just, R. (1977). A Study of debt servicing capacity applying logit analysis. *Journal of Development Economics*, 4, 25-38.
- Frank, C. R. & Cline, W. R. (1971). Measurement of debt service capacity: An application of discriminant analysis. *Journal of International Economics*, 1, 327-344.
- Ginzburg, I. & Horn, D. (1994). Combined neural networks for time series analysis. *Advancement in Neural Information Processing Systems*, 6, 224-231.
- Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*. Boston, MA: PWS Publishing.
- Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice Hall Publishing Co.
- Hill, T., O'Connor, M., & Remus, W. (1996). Neural network models for time series forecasts. *Management Science*, 42(7), 1082-1092.
- Kenen, P. B. (1969). The theory of optimal currency areas: An eclectic view, in Mundell, R. A. & Swoboda, A. K. (eds). *Monetary Problems of the International Economy*, Chicago: University of Chicago Press.
- Kharas, H. (1984). The long run creditworthiness of developing countries: Theory and practice. *Quarterly Journal of Economics*, 99, 415-439.
- Kim, T. Y., Oh, K. J., Sohn, I., & Hwang, C. (2004). Usefulness of artificial neural networks for early warning system of economic crisis. *Expert Systems with Applications*, 26(4), 583-590.
- Kim, J., Wei, S., & Ruys, H. (2003). Segmenting the market of West Australian senior tourists using an artificial neural network. *Tourism Management*, 24, 25-34.
- Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series model for forecasting commodity prices. *Neurocomputing*, 10 (2), 169-181.
- Krugman, P. R. (1990). Policy problems of a monetary Union, in de Grauwe, P. & Papademos, L. (eds). *The European monetary system in the 1990s*. Harlow: Longman, (reprinted in P. Krugman (ed) *Currencies and crises*. Cambridge, M.S.: MIT Press, 1992).
- Kuan, C. M. & White, H. (1994). Artificial neural networks: An economic Perspective. *Economic Reviews*, 13 (1), 1-91.
- Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management*, 21, 331-340.
- Law, R. & Au, N. (1999). A neural networks model to forecast Japanese demand for travel to Hong Kong. *Tourism Management*, 20 (1), 89-97.
- Leung, M. T., Chen, A., & Daouko, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers and Operations Research*, 27(11-12), 1093-1110.
- Luxhoj, J. T., Riis, J. O., & Stensballe, B. (1996). A hybrid econometric-neural-network modelling approach for sales forecasting. *Int. J. Prod. Econ.* 43, 175-192.
- Maier, H. R. & Dandy, G. C. (1996). Neural network models for forecasting univariate time series. *Neural Networks World*, 6 (5), 747-772.
- McKinnon, R. I. (1963). Optimum currency areas. *American Economic Review*, 53, 717-725.
- Moller, M. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4), 525-533.
- Mundell, R. A. (1961). A theory of optimum currency areas. *American Economic Review*, 51, 657-64.
- Nocedal, J. & Wright, S. J. (1999). *Numerical optimization*. New York, N.Y.: Springer-Verlag Publishing.

- Pelikan, E., Groot, C. D., & Wurtz, D. (1992). Power consumption in West-Bohemia: Improved forecasts with decorrelating connectionist networks. *Neural Networks World*, 2, 701-712.
- Qian, N. (1999). Efficient backpropagation learning using optimal learning rate and on the momentum term in gradient descent learning algorithms. *Neural Networks*, 102(1), 145-151.
- Roy, J. & Cosset, J. (1990). Forecasting country risk ratings using a neural network. *Proceedings of the 23rd Hawaii International Conference on Systems Sciences*, 4, 327-334 (Alamitos, CA, IEEE Pres).
- Sharda, R. (1994). Neural networks for the MS / OR analyst: An application bibliography. *Interfaces*, 24 (2), 116-130.
- Su, C. T., Tong, L. I., & Leou, C. M. (1997). Combination of time series and neural network for reliability forecasting modeling. *J.Chin. Inst. Ind. Eng.* 14 (4), 419-429.
- Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2002). Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting & Social Change*, 69, 71-87.
- Voort, V. D., Dougherty, M., & Watson, M. (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transp. Res. Circ. (energe Technol.)*, 4C (5), 307-318.
- Wang, J. H. & Leu, J. Y. (1996). Stock market trend prediction using ARIMA-based neural networks. *IEEE Int. Conf. Neural Networks*, 4, 2160-2165.
- Wedding, D. K. & Cios, K. J. (1996). Time-series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model. *Neurocomputing*, 10, 149-168.
- Wilson, R. & Sharda, R. (1992). Neural networks, *OR / MS today*. 36-42.
- Wong, B., Bodnovich, T., & Selvi, Y. (1995). A bibliography of neural networks business application research. *Expert Systems*, 12 (3), 253-262.
- www.worldbank.org/data/countrydata
- www.die.gov.tr
- Yoon-Dae, E. (1979). *Commercial banks and the creditworthiness of less developed countries*. Ann Arbor, MI: UMI Research Press.
- Yu X., H. & Chen, G. A. (1997). Efficient backpropagation learning using optimal learning rate and momentum. *Neural Networks*, 10 (3), 517-527.
- Zhang, G. P., Patuwo, E. B., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35-62.
- Zhang, G. P., Patuwo, E. B., & Hu, M. Y. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computer & Operations Research*, 28, 381-396.
- Zhang, G. P. (2003). Time-series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.
- Zhang, G. P. & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160, 501-514.
- Zhang, G. P. (2001). An investigation of neural networks for linear time-series forecasting. *Computer & Operations Research*, 28, 1183-1202.

A Comparison of Risk Classification Methods for Claim Severity Data

Noriszura Ismail Abdul Aziz Jemain
Universiti Kebangsaan Malaysia

The objective of this article is to compare several risk classification methods for claim severity data by using weighted equation which is written as a weighted difference between the observed and fitted values. The weighted equation will be applied to estimate claim severities which is equivalent to the total claim costs divided by the number of claims.

Key words: Risk classification, claim severity, claim cost.

Introduction

The process of establishing premium rates for insuring uncertain events requires estimates which were made of two important elements; the probabilities or frequencies associated with the occurrence of such event, and the magnitude or severities of such event. The process of grouping risks of similar risk characteristics for the frequencies or severities is also known as risk classification. The risks may be categorized according to risk or rating factors. In motor insurance for instance, the driver's gender and claim experience, or the vehicle's make and capacity, may be considered as rating factors.

In the last forty years, researchers suggested various statistical procedures for risk classification. For example, Bailey and Simon

(1960) suggested the minimum chi-squares, Bailey (1963) proposed the zero bias, Jung (1968) produced a heuristic method for minimum modified chi-squares, Ajne (1975) applied the method of moments also for minimum modified chi-squares, Chamberlain (1980) used the weighted least squares, Coutts (1984) produced the method of orthogonal weighted least squares with logit transformation, Harrington (1986) suggested the maximum likelihood procedure for models with functional form, and Brown (1988) proposed the bias and likelihood functions.

In the recent actuarial literature, research on risk classification methods is still continuing and developing. For example, Mildenhall (1999) studied the relationship between minimum bias and Generalized Linear Models (GLMs), Feldblum and Brosius (2003) provided minimum bias procedures for practicing actuary, Anderson *et al.* (2004) provided practical insights for GLMs analysis also for practicing actuary, Fu and Wu (2005) developed and generalized the minimum bias models, Ismail and Jemain (2005a) bridged the minimum bias and maximum likelihood methods for claim frequency data, and Ismail and Jemain (2005b) proposed the Negative Binomial and Generalized Poisson regressions as alternatives for the Poisson to handle over-dispersion.

In addition to statistical procedures, research on multiplicative and additive models has also been carried out. Bailey and Simon (1960) compared systematic bias and found that

Noriszura Ismail is a Lecturer teaching in Actuarial Science Department. Currently, she is pursuing her PhD in Statistics. Abdul Aziz Jemain is Associate Professor in Statistics Department. The authors gratefully acknowledge financial support received (IRPA RMK8: 09-02-02-0112-EA274) from the Ministry of Science, Technology and Innovation (MOSTI), Malaysia. The author also thanks the General Insurance Association of Malaysia (PIAM), in particular Mr. Carl Rajendram and Mrs. Addiwyah.

the multiplicative model overestimates the high risk classes, Jung (1968) and Ajne (1975) also found that the estimates for multiplicative model are positively biased, Bailey (1963) compared the models by producing two statistical criteria, i.e., minimum chi-squares and average absolute difference, Freifelder (1986) predicted the pattern of over and under estimation of the models if they were misspecified, Brown (1988) discussed the additive and multiplicative models which were derived from the maximum likelihood and minimum bias approaches, Jee (1989) compared the predictive accuracy of the models, Holler *et al.* (1999) compared their initial values sensitivity, and Mildenhall (1999) identified the GLMs with the additive and multiplicative models.

Based on the actuarial literature, studies for risk classification were centered on two main areas; risk classification methods, and multiplicative vs. additive models. The objective of this study is to compare several risk classification methods for multiplicative and additive models by using weighted equation which is written as a weighted difference between the observed and fitted values. In addition, the parameter solution for multiplicative and additive models will also be compared by using weighted solution. The weighted solution for multiplicative model is in the form of a weighted proportion of observed over fitted values, whereas for additive model, it is in the form of a weighted difference between observed and fitted values.

Although the weighted equation was previously suggested by Ismail and Abdul Aziz (2005a), the application was implemented on claim frequency data. Therefore, this study differs such that the weighted equation will be applied to estimate claim severity or average claim cost which is also equivalent to the total claim costs divided by the number of claims. Because the nature of claim frequency and severity data is different, the approach taken is also slightly modified.

It is well established that the claim cost distributions generally have positive support and are positively skewed. Because of these desired properties, the Gamma and Lognormal distributions have been widely used by the practitioners for modeling claim severities. As a

comparison, several actuarial studies also reported severity results from Normal distribution. For example, Baxter *et al.* (1980) fit the U.K. own damage costs for privately owned and comprehensively insured vehicles to the weighted linear (additive) regression model by assuming that the variance is constant across the classes, McCullagh and Nelder (1989) reanalyzed the same data by fitting the costs to the Gamma by assuming that the coefficient of variation is constant across the classes and the mean is linear on reciprocal scale, Brockman and Wright (1992) fit the U.K. own damage costs for comprehensive policies also to the Gamma by using a log-linear (multiplicative) regression model, and Renshaw (1994) fit the U.K. motor insurance claim severity also to the Gamma log-linear regression model.

The fitting procedure for this study will be carried out by using two different approaches; classical and regression. The advantage of using the regression fitting procedure is that it can also be extended to other regression models, as long as the function of the fitted value is written in a specified linear form. In addition, the computation of the regression fitting procedure provides a faster convergence compared to the classical.

In this study, the risk classification methods will be compared on three types of severity data; Malaysian data, U.K. data (McCullagh & Nelder, 1989), and Canadian data (Bailey & Simon, 1960).

Methodology

The related data sets for claim severity are (c_i, y_i) , $i = 1, 2, \dots, n$, where c_i and y_i denotes the average claim cost adjusted for inflation and the claim count for the i th rating class. Therefore, the total claim cost is equal to the product of the claim count and the average claim cost, $y_i c_i$.

Let $\mathbf{x}_i$ be the vector of explanatory variables for the i th rating class, $\boldsymbol{\beta}$ the $p \times 1$ vector of regression parameters, and $\mathbf{f}$ the vector of fitted values.

If the model is assumed to be multiplicative, the fitted value is $f_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, which can also be written as,

$$f_i = f_{i(-j)} \exp(\beta_j x_{ij}), \quad (1)$$

where $f_{i(-j)}$ is the multiplicative fitted value without the j th effect.

For an additive model, the fitted value is $f_i = \mathbf{x}_i^T \boldsymbol{\beta}$, so that it can be written as,

$$f_i = f_{i(-j)} + \beta_j x_{ij}, \quad (2)$$

where $f_{i(-j)}$ is the additive fitted value without the j th effect.

Minimum Bias Models

The parameters for zero bias model are solved by equating (Bailey 1963),

$$\sum_i y_i c_i x_{ij} = \sum_i y_i f_i x_{ij}, \quad j = 1, 2, \dots, p \quad (3)$$

Therefore, Eq.(3) can also be written as a weighted difference between observed and fitted values,

$$\sum_i w_i (c_i - f_i) = 0, \quad j = 1, 2, \dots, p, \quad (4)$$

where the weight, w_i , is equal to $y_i x_{ij}$.

Substituting Eq.(1) for multiplicative model, Eq.(4) can be rewritten as,

$$\sum_i y_i c_i x_{ij} = \sum_i y_i f_{i(-j)} \exp(\beta_j x_{ij}) x_{ij}, \quad j = 1, 2, \dots, p \quad (5)$$

The solution, $\exp(\beta_j)$, may be calculated from Eq.(5) because the value for x_{ij} is either one or zero. The solution may be written as a weighted proportion,

$$\exp(\beta_j) = \sum_i v_i \frac{c_i}{f_{i(-j)}}, \quad j = 1, 2, \dots, p, \quad (6)$$

where the weight, v_i , is equal to,

$$v_i = \frac{z_i}{\sum_i z_i}, \quad (7)$$

and z_i is $y_i f_{i(-j)} x_{ij}$.

If Eq.(2) is substituted for additive model, Eq.(4) can be rewritten as,

$$\sum_i y_i (c_i - f_{i(-j)}) x_{ij} = \sum_i y_i \beta_j x_{ij} x_{ij}, \quad j = 1, 2, \dots, p$$

Again, because the value for x_{ij} is either one or zero, the solution, β_j , is obtainable and it is in the form of a weighted difference,

$$\beta_j = \sum_i v_i (c_i - f_{i(-j)}), \quad j = 1, 2, \dots, p, \quad (8)$$

where the weight, v_i , is also equal to Eq.(7). However, the equation for z_i is $y_i x_{ij}$.

The parameters for minimum chi-squares model are solved by minimizing the chi-squares (Bailey & Simon 1960),

$$\chi^2 = \sum_i \frac{y_i (c_i - f_i)^2}{f_i},$$

or by equating,

$$\frac{\partial \chi^2}{\partial \beta_j} = \sum_i w_i (c_i - f_i) = 0, \quad j = 1, 2, \dots, p, \quad (9)$$

where the weight, w_i , is

$$\frac{y_i(c_i + f_i)}{f_i^2} \frac{\partial f_i}{\partial \beta_j}.$$

The first derivative of the fitted value is equal to,

$$\frac{\partial f_i}{\partial \beta_j} = f_i x_{ij}, \quad (10)$$

for multiplicative model and,

$$\frac{\partial f_i}{\partial \beta_j} = x_{ij}, \quad (11)$$

for additive model.

Substituting Eq.(1) and Eq.(10) into Eq.(9) for multiplicative model, $\exp(\beta_j)$ is equal to Eq.(6), where the weight, v_i , is equal to Eq.(7). However, the value for z_i is $y_i(c_i + f_i)x_{ij}$. If Eq.(2) and Eq.(11) are substituted into Eq.(9) for additive model, β_j is equal to Eq.(8), where the weight, v_i , is also equal to Eq.(7). The value for z_i is

$$\frac{y_i(c_i + f_i)}{f_i^2} x_{ij}.$$

Maximum Likelihood Models

Let $T_i = y_i C_i$ be the random variable for total claim costs. If T_i is assumed to follow Normal distribution with mean $E(T_i) = y_i f_i$ and variance $Var(T_i) = \sigma^2$ (Brown 1988), the parameters are solved by using the likelihood equations,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_i w_i (c_i - f_i) = 0, \quad j = 1, 2, \dots, p, \quad (12)$$

where w_i is $y_i^2 \frac{\partial f_i}{\partial \beta_j}$. Substituting Eq.(1) and

Eq.(10) into Eq.(12) for multiplicative model, $\exp(\beta_j)$ is equal to Eq.(6), where v_i is equal to

Eq.(7). However, the value for z_i is $y_i^2 f_{i(-j)} x_{ij}$.

If Eq.(2) and Eq.(11) are substituted into Eq.(12) for additive model, β_j is equal to Eq.(8), where

v_i is equal to Eq.(7). The value for z_i is $y_i^2 x_{ij}$.

Let T_i be Poisson distributed with mean $y_i f_i$. The likelihood equations can also be written as Eq.(12), but the value for w_i is

$\frac{y_i}{f_i} \frac{\partial f_i}{\partial \beta_j}$. Following the same procedure as the

Normal distribution, the parameters for multiplicative model, $\exp(\beta_j)$, are equal to

Eq.(6), where v_i is equal to Eq.(7). However,

the value for z_i is $y_i f_{i(-j)} x_{ij}$. Therefore, the

parameters for Poisson multiplicative are shown to be equivalent to the zero bias multiplicative.

For additive model, β_j is equal to Eq.(8), where

v_i is equal to Eq.(7). The value for z_i is $\frac{y_i}{f_i} x_{ij}$.

If T_i is exponentially distributed with mean $y_i f_i$, the likelihood equations can also be written as Eq.(12). However, the value for w_i is

$\frac{1}{f_i^2} \frac{\partial f_i}{\partial \beta_j}$. The parameters for multiplicative

model, $\exp(\beta_j)$, are equal to Eq.(6), where v_i

is equal to Eq.(7). However, the value for z_i is

x_{ij} . For additive model, β_j is equal to Eq.(8),

where v_i is equal to Eq.(7). The value for z_i is

$\frac{1}{f_i^2} x_{ij}$.

Let T_i be Gamma distributed with mean $y_i f_i$ and variance $\sigma^2 y_i f_i^2$. The likelihood

equations can also be written as Eq.(12), but the

value for w_i is $\frac{y_i}{f_i^2} \frac{\partial f_i}{\partial \beta_j}$. The parameters for

multiplicative model, $\exp(\beta_j)$, are equal to

Eq.(6), where v_i is equal to Eq.(7). However, the value for z_i is $y_i x_{ij}$. For additive model, β_j is equal to Eq.(8), where v_i is equal to Eq.(7).

The value for z_i is $\frac{y_i}{f_i^2} x_{ij}$.

Other Models

The weighted equations shown by Eq.(4), Eq.(9) and Eq.(12) may also be extended to other error functions. For example, if the sum squares error is defined as (Brown 1988), $S = \sum_i y_i (c_i - f_i)^2$, the parameters are solved by using the least squares equations,

$$\frac{\partial S}{\partial \beta_j} = \sum_i w_i (c_i - f_i) = 0, \quad j = 1, 2, \dots, p, \quad (13)$$

where w_i is $y_i \frac{\partial f_i}{\partial \beta_j}$. The parameters for multiplicative model, $\exp(\beta_j)$, are equal to Eq.(6), where v_i is equal to Eq.(7). However, the value for z_i is $y_i f_{i(-j)}^2 x_{ij}$. For additive model, β_j is equal to Eq.(8), where v_i is equal to Eq.(7). The value for z_i is $y_i x_{ij}$. Therefore, the parameters for least squares additive are shown to be equivalent to the zero bias additive.

If the function of errors is a modified chi-squares which is defined as,

$$\chi_{\text{mod}}^2 = \sum_i \frac{y_i}{c_i} (c_i - f_i)^2,$$

the weighted equation is equal to,

$$\frac{\partial \chi_{\text{mod}}^2}{\partial \beta_j} = \sum_i w_i (c_i - f_i) = 0, \quad j = 1, 2, \dots, p \quad (14)$$

where w_i is

$$\frac{y_i}{c_i} \frac{\partial f_i}{\partial \beta_j}.$$

The parameters for multiplicative model, $\exp(\beta_j)$, are equal to Eq.(6) where v_i is equal to Eq.(7). However, the value for z_i is $\frac{y_i f_{i(-j)}^2}{c_i} x_{ij}$. For additive model, β_j is equal to Eq.(8), where v_i is equal to Eq.(7). The value for z_i is $\frac{y_i}{c_i} x_{ij}$.

Table 1 summarizes the weighted equations and parameter solutions for all of the models discussed above. Based on the weighted equations and parameter solutions, the following conclusions can be made regarding the comparison of several risk classification methods which were discussed above:

- The parameter estimates for zero bias and Poisson multiplicative are equal. The parameter estimates for zero bias and least squares additive are also equal.
- The weighted equations and parameter solutions indicate that all models are similar. Each model is distinguished only by its weight.

Classical Fitting Procedure

In this study, the multiplicative and additive models will be fitted by using two different procedures; classical and regression. The classical fitting procedure was introduced by Bailey and Simon (1960). The procedure involves sequential iterations where each parameter, β_j , $j = 1, 2, \dots, p$, is calculated individually in each sequence. In the first sequence, the value for $\beta_1^{(1)}$ is calculated by using the initial values, $\beta^{(0)}$. The sequence is then repeated until the p th sequence, where $\beta_p^{(1)}$ is calculated and vector $\beta^{(1)}$ is produced. The sequential iteration is then repeated until the values for β converged.

Table 1. Weighted equations and parameter solutions

Models	w_i for weighted equation, $\sum_i w_i (c_i - f_i) = 0$	z_i for multiplicative parameter solution, $\exp(\beta_j) = \sum_i v_i \frac{c_i}{f_{i(-j)}},$ where $v_i = \frac{z_i}{\sum_i z_i}$	z_i for additive parameter solution, $\beta_j = \sum_i v_i (c_i - f_{i(-j)}),$ where $v_i = \frac{z_i}{\sum_i z_i}$
Zero bias	$w_i = y_i x_{ij}$	$z_i = y_i f_{i(-j)} x_{ij}$	$z_i = y_i x_{ij}$
Poisson	$w_i = \frac{y_i}{f_i} \frac{\partial f_i}{\partial \beta_j}$	$z_i = y_i f_{i(-j)} x_{ij}$	$z_i = \frac{y_i}{f_i} x_{ij}$
Least squares	$w_i = y_i \frac{\partial f_i}{\partial \beta_j}$	$z_i = y_i f_{i(-j)}^2 x_{ij}$	$z_i = y_i x_{ij}$
Minimum χ^2	$w_i = \frac{y_i (c_i + f_i)}{f_i^2} \frac{\partial f_i}{\partial \beta_j}$	$z_i = y_i (c_i + f_i) x_{ij}$	$z_i = \frac{y_i (c_i + f_i)}{f_i^2} x_{ij}$
Normal	$w_i = y_i^2 \frac{\partial f_i}{\partial \beta_j}$	$z_i = y_i^2 f_{i(-j)}^2 x_{ij}$	$z_i = y_i^2 x_{ij}$
Exponential	$w_i = \frac{1}{f_i^2} \frac{\partial f_i}{\partial \beta_j}$	$z_i = x_{ij}$	$z_i = \frac{1}{f_i^2} x_{ij}$
Gamma	$w_i = \frac{y_i}{f_i^2} \frac{\partial f_i}{\partial \beta_j}$	$z_i = y_i x_{ij}$	$z_i = \frac{y_i}{f_i^2} x_{ij}$
Minimum modified χ^2	$w_i = \frac{y_i}{c_i} \frac{\partial f_i}{\partial \beta_j}$	$z_i = \frac{y_i f_{i(-j)}^2}{c_i} x_{ij}$	$z_i = \frac{y_i}{c_i} x_{ij}$

As an example, the programming for zero bias multiplicative will be discussed here. The parameter solution is

$$\exp(\beta_j) = \sum_i v_i \frac{c_i}{f_{i(-j)}},$$

where

$$v_i = \frac{z_i}{\sum_i z_i}$$

and

$$z_i = y_i f_{i(-j)} x_{ij}.$$

Let $f_{i(-j)}$ be the i th row of vector $\mathbf{f}_{(-j)}$. For multiplicative model, $\mathbf{f}_{(-j)} = \exp(\mathbf{X}_{(-j)} \boldsymbol{\beta}_{(-j)})$, where $\mathbf{X}_{(-j)}$ is the matrix of explanatory variables without the j th column and $\boldsymbol{\beta}_{(-j)}$ the vector of regression parameters without the j th row. Let $\mathbf{x}_j$ be the vector which is equivalent to the j th column of matrix $\mathbf{X}$. Therefore, x_{ij} is equal to the i th row of vector $\mathbf{x}_j$. For each j , let $z_i = y_i f_{i(-j)} x_{ij}$. Therefore, the weight, v_i , is equal to z_i divided by sum of z_i for all i . Finally, the parameter solution, $\exp(\beta_j)$, is equal to the sum of $v_i \frac{c_i}{f_{i(-j)}}$ for all i .

An example of S-PLUS programming for zero bias multiplicative is given in Appendix A. Similar programming can also be used for all of the multiplicative and additive models which were discussed in this study. Each model should be differentiated only by three elements:

- The fitted values for multiplicative model are $\mathbf{f} = \exp(\mathbf{X}\boldsymbol{\beta})$ and $\mathbf{f}_{(-j)} = \exp(\mathbf{X}_{(-j)} \boldsymbol{\beta}_{(-j)})$. For additive model, the fitted values are $\mathbf{f} = \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{f}_{(-j)} = \mathbf{X}_{(-j)} \boldsymbol{\beta}_{(-j)}$.

- The parameter solution is $\exp(\beta_j) = \sum_i v_i \frac{c_i}{f_{i(-j)}}$ for multiplicative model, and $\beta_j = \sum_i v_i (c_i - f_{i(-j)})$ for additive model.
- Each model has its own equation for z_i .

Regression Fitting Procedure

The regression fitting procedure involves standard iterations where all of the parameters, β_j , $j = 1, 2, \dots, p$, are calculated simultaneously in each iteration. Because the parameters are solved simultaneously, the regression procedure provides a faster convergence compared to the classical procedure. In the first iteration, $\boldsymbol{\beta}^{(1)}$ are calculated by using initial values of $\boldsymbol{\beta}^{(0)}$. The iteration is then repeated until the values for $\boldsymbol{\beta}$ converged.

The parameters, β_j , $j = 1, 2, \dots, p$, for regression fitting procedure are solved by minimizing $\sum_i w_i (c_i - f_i)^2$ or by equating,

$$\sum_i w_i (c_i - f_i) \frac{\partial f_i}{\partial \beta_j} = 0, \quad j = 1, 2, \dots, p. \quad (15)$$

Therefore, the weighted equations for risk classification models shown by Eq.(4), Eq.(9), Eq.(12), Eq.(13) and Eq.(14) are also equivalent to Eq.(15). By using Taylor series approximation, the values for $\boldsymbol{\beta}$ in the r th iteration is equal to,

$$\boldsymbol{\beta}^{(r)} = (\mathbf{Z}^{(r-1)\text{T}} \mathbf{W}^{(r-1)} \mathbf{Z}^{(r-1)})^{-1} \mathbf{Z}^{(r-1)\text{T}} \mathbf{W}^{(r-1)} (\mathbf{c} - \mathbf{s}^{(r-1)}) \quad (16)$$

where $\boldsymbol{\beta}^{(r)}$ and $\boldsymbol{\beta}^{(r-1)}$ are the values for $\boldsymbol{\beta}$ in the r th and $r-1$ th iterations, $\mathbf{Z}^{(r-1)}$ the $n \times p$ matrix whose ij th element is $\left. \frac{\partial f_i(\boldsymbol{\beta})}{\partial \beta_j} \right|_{\boldsymbol{\beta} = \boldsymbol{\beta}^{(r-1)}}$,

$\mathbf{W}^{(r-1)}$ the diagonal weight matrix evaluated at

$\beta^{(r-1)}$, and $s^{(r-1)}$ the vector whose i th row is equal to

$$f_i(\beta^{(r-1)}) - \sum_{j=1}^p \beta_j^{(r-1)} z_{ij}^{(r-1)}.$$

As an example, the programming for additive least squares will be discussed here. The weighted equation is

$$\sum_i y_i (c_i - f_i) \frac{\partial f_i}{\partial \beta_j} = 0, j = 1, 2, \dots, p.$$

Therefore, the i th diagonal element of the weight matrix is equal to y_i , which is free of $\beta^{(r-1)}$. For an additive model, the ij th element of matrix $Z^{(r-1)}$ is equal to $\left. \frac{\partial f_i(\beta)}{\partial \beta_j} \right|_{\beta=\beta^{(r-1)}} = x_{ij}$,

which is also free of $\beta^{(r-1)}$. Because x_{ij} is the ij th element of matrix X and the dimensions for $Z^{(r-1)}$ and X are equal, $Z^{(r-1)} = X$ and $s^{(r-1)} = f(\beta^{(r-1)}) - X\beta^{(r-1)} = 0$. Therefore, Eq.(16) for additive least squares is simplified into,

$$\beta^{(r)} = \beta = (X^T W X)^{-1} X^T W c, \quad (17)$$

which is also equal to the Normal equation in the standard linear regression model. Eq.(17) also indicates that the parameters for additive least squares can be solved without any iteration.

For multiplicative model, the ij th element of matrix $Z^{(r-1)}$ is equal to, $\left. \frac{\partial f_i(\beta)}{\partial \beta_j} \right|_{\beta=\beta^{(r-1)}} = f_i(\beta^{(r-1)}) x_{ij}$. Therefore, the

equation for $Z^{(r-1)}$ may be written as,

$$Z^{(r-1)} = F^{(r-1)} X, \quad (18)$$

where $F^{(r-1)}$ is the diagonal matrix whose i th diagonal elements is $f_i(\beta^{(r-1)})$. The vector for $s^{(r-1)}$ may be written as

$$s^{(r-1)} = f(\beta^{(r-1)}) - F^{(r-1)} X \beta^{(r-1)}.$$

The advantage of using the regression fitting procedure is that besides multiplicative and additive models, the fitting can also be extended to other regression models as well. Therefore, the regression fitting procedure allows a variety of regression model to be created and applied, as long as the function of the fitted value is written as

$$f_i = \left(\sum_{j=1}^p \beta_j x_{ij} \right)^b, -1 \leq b < 0, 0 < b \leq 1.$$

For example, if the fitted value is assumed to follow an inverse function, i.e., $b = -1$, the ij th element of matrix $Z^{(r-1)}$ is equal to $\left. \frac{\partial f_i(\beta)}{\partial \beta_j} \right|_{\beta=\beta^{(r-1)}} = -\{f_i(\beta^{(r-1)})\}^2 x_{ij}$. Therefore, the

equation for $Z^{(r-1)}$ may also be written as Eq.(18). However, the i th diagonal element of matrix $F^{(r-1)}$ is equal to $-\{f_i(\beta^{(r-1)})\}^2$.

An example of S-PLUS programming for least squares multiplicative is given in Appendix B. Similar programming can also be used for either of the multiplicative, additive or inverse models. Each programming should be differentiated only by three elements:

- The vector for the fitted values is equal to $f = \exp(X\beta)$ for multiplicative model, $f = X\beta$ for additive model, and $f = (X\beta)^{-1}$ for inverse model.

- $Z^{(r-1)} = X$ for additive model, and $Z^{(r-1)} = F^{(r-1)} X$ for multiplicative and inverse models. However, the i th diagonal element of matrix $F^{(r-1)}$ is equal to $f_i(\beta^{(r-1)})$ for multiplicative model, and $-\{f_i(\beta^{(r-1)})\}^{-2}$ for inverse model.

- Each model has its own weight matrix.

Results

Malaysian Data

The risk classification methods will be compared on the Malaysian private car Third Party Property Damage (TPPD) average claim costs data. Specifically, the TPPD claim covers the legal liability for third party property loss or damage caused by or arising out of the use of an insured motor vehicle. The data, which was obtained from an insurance company in Malaysia and was supplied by the General Insurance Association of Malaysia (PIAM), was based on 170,000 private car policies in a three-year period of 1998-2000. The data consists of claim counts and average claim costs which were already paid as well as outstanding. The average claim costs, which were already adjusted for inflation, were given in Ringgit Malaysia (RM). The risks for the claims were associated with five rating factors; coverage

type, vehicle make, vehicle use and driver's gender, vehicle year, and location. Altogether, there were $2 \times 2 \times 3 \times 4 \times 5 = 240$ cross-classified rating classes of claim severities to be estimated. The complete data is available by contacting the author.

The claim severities were fitted to all of the multiplicative and additive models which were discussed in this study. However, the fitting involves only 108 data points because 132 of the rating classes have zero claim count. In addition, the models will be evaluated by using two different tests; chi-squares and average absolute difference. The average absolute difference is equal to (Bailey and Simon 1960)

$$\frac{\sum_i y_i |c_i - f_i|}{\sum_i y_i c_i}.$$

Table 2 and Table 3 give the parameter estimates, chi-squares and average absolute difference for multiplicative and additive models of the Malaysian data.

Table 2: Multiplicative models for Malaysian data

Parameters	Zero bias /Poisson	Least squares	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
exp(β_1) Intercept	7,467.43	7,459.97	7,460.65	7,493.35	7,229.48	7,480.11	7,486.69
exp(β_2) Non-comp	1.15	1.15	1.17	1.13	1.16	1.15	1.10
exp(β_3) Foreign	1.08	1.07	1.08	1.07	1.20	1.08	1.08
exp(β_4) Female	0.90	0.90	0.90	0.93	0.80	0.89	0.88
exp(β_5) Business	0.20	0.20	0.20	0.20	0.21	0.20	0.20
exp(β_6) 2-3 years	0.78	0.78	0.78	0.79	0.74	0.78	0.78
exp(β_7) 4-5 years	0.69	0.70	0.70	0.69	0.66	0.69	0.69
exp(β_8) 6+ years	0.73	0.73	0.73	0.72	0.72	0.73	0.73
exp(β_9) North	0.94	0.94	0.94	0.93	0.92	0.94	0.93
exp(β_{10}) East	0.86	0.85	0.87	0.84	0.88	0.87	0.83
exp(β_{11}) South	0.93	0.93	0.94	0.94	1.04	0.93	0.93
exp(β_{12}) East M'sia	0.94	0.95	0.97	0.94	1.06	0.94	0.89
χ^2 (10^{-5})	3.76	3.77	3.73	3.89	6.69	3.77	4.08
Absolute difference	0.07	0.06	0.07	0.07	0.11	0.07	0.07

Table 3: Additive models for Malaysian data

Parameters (10^{-2})	Zero bias /Least squares	Poisson	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
$\exp(\beta_1)$ Intercept	74.08	74.07	74.07	74.74	72.82	74.09	74.10
$\exp(\beta_2)$ Non-comp	8.06	8.03	9.31	7.23	5.37	8.01	5.26
$\exp(\beta_3)$ Foreign	4.36	4.55	4.67	4.10	10.16	4.72	4.34
$\exp(\beta_4)$ Female	-6.18	-6.51	-6.26	-4.77	-12.32	-6.77	-7.16
$\exp(\beta_5)$ Business	-40.79	-40.83	-40.75	-40.13	-39.69	-40.88	-41.09
$\exp(\beta_6)$ 2-3 years	-15.51	-15.60	-15.56	-15.42	-18.34	-15.70	-15.78
$\exp(\beta_7)$ 4-5 years	-21.90	-22.03	-21.90	-22.64	-23.23	-22.19	-22.34
$\exp(\beta_8)$ 6+ years	-19.53	-19.54	-19.72	-20.38	-19.15	-19.56	-19.15
$\exp(\beta_9)$ North	-3.76	-3.70	-3.60	-4.23	-3.99	-3.64	-3.86
$\exp(\beta_{10})$ East	-8.53	-7.77	-6.88	-8.78	-6.44	-7.18	-9.69
$\exp(\beta_{11})$ South	-3.92	-3.83	-3.72	-3.87	2.42	-3.75	-4.03
$\exp(\beta_{12})$ East M'sia	-3.36	-3.53	-1.97	-3.45	2.27	-3.72	-6.46
χ^2 (10^{-5})	3.71	3.70	3.66	3.81	6.21	3.70	4.01
Absolute difference	0.06	0.06	0.06	0.07	0.11	0.06	0.07

The classical and regression fitting procedures give equal values for parameter estimates. However, the regression procedure provides a faster convergence.

The multiplicative and additive models give similar parameter estimates. The smallest chi-squares is given by the minimum chi-squares model. Except for the exponential model, all models provide similar values for absolute difference.

U.K. Data

The U.K. data provides information on the Own Damage claim counts and average claim costs for privately owned and comprehensively insured vehicles (McCullagh & Nelder 1989). The average claim costs, which were already adjusted for inflation, were given in Pound Sterling. The risks for the claims were

associated with three rating factors; policyholder's age, car group and vehicle age. Altogether, there were $8 \times 4 \times 4 = 128$ cross-classified rating classes of claim severities to be estimated.

The claim severities were fitted to all of the multiplicative and additive models which were discussed in this study. In addition, the severities were also fitted to the inverse models because McCullagh and Nelder (1989) also fit the severities to the Gamma regression model by assuming that the regression effects were linear on the reciprocal scale. The fitting involves only 123 data points because five of the rating classes have zero claim count.

Table 4, Table 5 and Table 6 give the parameter estimates, chi-squares and average absolute difference for multiplicative, additive and inverse models of the U.K. data.

Table 4: Multiplicative models for U.K. data

Parameter	Zero bias /Poisson	Least squares	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
exp(β_1) Intercept	297.57	309.81	313.59	279.34	302.38	286.75	257.91
exp(β_2) 21-24 yrs	0.98	0.94	0.95	1.05	0.90	1.00	1.08
exp(β_3) 25-29 yrs	0.91	0.88	0.87	0.97	1.01	0.94	1.04
exp(β_4) 30-34 yrs	0.88	0.86	0.84	0.96	0.75	0.89	1.01
exp(β_5) 35-39 yrs	0.70	0.67	0.67	0.75	0.72	0.73	0.79
exp(β_6) 40-49 yrs	0.77	0.75	0.73	0.81	0.76	0.79	0.89
exp(β_7) 50-59 yrs	0.78	0.76	0.75	0.83	0.79	0.80	0.89
exp(β_8) 60+ yrs	0.78	0.77	0.74	0.82	0.75	0.80	0.90
exp(β_9) B	0.99	0.98	0.99	0.96	1.06	1.00	0.99
exp(β_{10}) C	1.16	1.15	1.16	1.14	1.17	1.17	1.16
exp(β_{11}) D	1.48	1.48	1.50	1.53	1.60	1.49	1.45
exp(β_{12}) 4-7 yrs	0.91	0.90	0.91	0.95	0.89	0.92	0.91
exp(β_{13}) 8-9 yrs	0.70	0.69	0.70	0.74	0.66	0.71	0.69
exp(β_{14}) 10+ yrs	0.49	0.48	0.51	0.50	0.48	0.50	0.46
χ^2 (10^{-4})	3.10	3.13	3.07	3.27	4.50	3.12	3.40
Absolute difference	0.08	0.08	0.08	0.08	0.11	0.08	0.08

Table 5: Additive models for U.K. data

Parameter	Zero bias /Least squares	Poisson	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
exp(β_1) Intercept	298.67	288.34	303.94	273.49	291.89	278.98	241.88
exp(β_2) 21-24 yrs	-5.60	0.31	-7.53	17.58	-10.84	4.96	34.01
exp(β_3) 25-29 yrs	-24.64	-16.95	-30.52	-2.01	15.31	-9.91	26.37
exp(β_4) 30-34 yrs	-33.22	-29.34	-43.39	-7.76	-47.35	-26.59	14.17
exp(β_5) 35-39 yrs	-87.89	-75.74	-89.26	-64.78	-44.23	-64.82	-33.45
exp(β_6) 40-49 yrs	-66.99	-60.27	-75.55	-50.51	-45.84	-54.15	-13.68
exp(β_7) 50-59 yrs	-63.35	-55.64	-70.12	-45.49	-36.19	-48.60	-10.87
exp(β_8) 60+ yrs	-63.15	-56.91	-72.15	-47.39	-44.32	-51.10	-10.32
exp(β_9) B	-2.46	-0.21	-0.50	-7.03	8.19	2.04	-0.30
exp(β_{10}) C	34.18	35.45	35.05	33.89	25.86	36.41	35.84
exp(β_{11}) D	108.66	108.76	113.74	123.07	97.83	108.90	96.09
exp(β_{12}) 4-7 yrs	-24.21	-21.54	-21.98	-10.57	-30.60	-19.62	-20.39
exp(β_{13}) 8-9 yrs	-76.75	-72.26	-71.63	-59.08	-96.51	-69.12	-74.38
exp(β_{14}) 10+ yrs	-126.63	-121.21	-118.78	-111.15	-147.85	-117.94	-128.54
χ^2 (10^{-4})	3.41	3.35	3.32	3.55	4.88	3.40	3.77
Absolute difference	0.09	0.09	0.09	0.09	0.11	0.09	0.09

Table 6: Inverse models for U.K. data

Parameter (10^4)	Poisson	Least squares	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
exp(β_1) Intercept	32.79	31.30	31.23	35.10	37.44	34.11	37.44
exp(β_2) 21-24 yrs	2.41	4.16	3.12	-0.28	-0.74	1.01	-0.74
exp(β_3) 25-29 yrs	4.75	6.26	6.11	2.25	0.46	3.50	0.46
exp(β_4) 30-34 yrs	5.30	6.39	6.81	2.50	0.83	4.62	0.83
exp(β_5) 35-39 yrs	14.97	16.61	16.11	12.41	11.36	13.70	11.36
exp(β_6) 40-49 yrs	10.28	11.12	11.73	8.41	5.97	9.69	5.97
exp(β_7) 50-59 yrs	9.96	10.98	11.30	7.78	5.89	9.16	5.89
exp(β_8) 60+ yrs	9.75	10.58	11.26	7.88	5.32	9.20	5.32
exp(β_9) B	0.70	0.93	0.68	2.06	0.65	0.38	0.65
exp(β_{10}) C	-5.68	-5.29	-5.60	-5.11	-5.95	-6.14	-5.95
exp(β_{11}) D	-13.77	-13.55	-13.90	-14.27	-13.60	-14.21	-13.60
exp(β_{12}) 4-7 yrs	3.95	4.21	3.99	2.65	3.88	3.66	3.88
exp(β_{13}) 8-9 yrs	16.83	17.14	16.33	15.45	17.95	16.51	17.95
exp(β_{14}) 10+ yrs	41.74	41.97	38.52	43.50	47.09	41.54	47.09
χ^2 (10^{-4})	3.10	3.13	3.07	3.27	3.37	3.12	3.37
Absolute difference	0.08	0.08	0.08	0.08	0.08	0.08	0.08

As expected, the parameter estimates for classical and regression fitting procedures are equal and the regression fitting procedure provides a faster convergence.

The parameter estimates for multiplicative, additive and inverse models are similar. In particular, the parameter estimates for Gamma inverse model are equal to the parameter estimates produced by McCullagh and Nelder (1989). The smallest chi-squares is also given by the minimum chi-squares model. Except for the exponential model, all models provide equal values for absolute difference.

Canadian Data

The Canadian data, which was obtained from Bailey and Simon (1960), provides information on liability claim counts and

average claim costs for private passenger automobile insurance. The data involves two rating factors; merit and class. Altogether, there were $4 \times 5 = 20$ cross-classified rating classes of claim severities to be estimated.

The claim severities were fitted to all of the multiplicative and additive models which were discussed in this study. Table 7 and Table 8 give the parameter estimates, chi-squares and average absolute difference for multiplicative and additive models of the Canadian data.

As expected, the multiplicative and additive models give similar parameter estimates. The smallest chi-square is also given by the minimum chi-squares model. Except for the exponential model, all models provide equal values for absolute difference.

Table 7: Multiplicative models for Canadian data

Parameter	Zero bias /Poisson	Least squares	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
$\exp(\beta_1)$ Intercept	292.00	292.10	291.97	291.08	294.57	291.92	292.07
$\exp(\beta_2)$ Merit X	0.99	0.99	0.99	1.00	0.97	0.99	0.98
$\exp(\beta_3)$ Merit Y	0.99	0.99	0.99	0.99	1.00	0.99	0.99
$\exp(\beta_4)$ Merit B	1.06	1.05	1.06	1.07	1.05	1.06	1.06
$\exp(\beta_5)$ Class 2							
$\exp(\beta_6)$ Class 3	1.09	1.08	1.09	1.09	1.12	1.09	1.08
$\exp(\beta_7)$ Class 4	1.02	1.02	1.02	1.03	0.98	1.02	1.02
$\exp(\beta_8)$ Class 5	1.17	1.17	1.17	1.18	1.16	1.17	1.17
	0.92	0.92	0.92	0.92	0.92	0.92	0.92
$\chi^2 (10^{-4})$	4.95	4.96	4.95	5.45	8.03	4.95	4.99
Absolute difference	0.01	0.01	0.01	0.01	0.02	0.01	0.01

Table 8: Additive models for Canadian data

Parameter	Zero bias /Least squares	Poisson	Minimum χ^2	Normal	Exponential	Gamma	Minimum modified χ^2
$\exp(\beta_1)$ Intercept	291.95	291.87	291.83	291.06	294.77	291.80	291.94
$\exp(\beta_2)$ Merit X	-4.24	-4.05	-3.38	0.59	-10.11	-3.92	-5.37
$\exp(\beta_3)$ Merit Y	-3.45	-3.58	-3.51	-3.95	1.00	-3.68	-3.71
$\exp(\beta_4)$ Merit B	17.11	17.53	17.58	20.28	15.49	17.92	17.44
$\exp(\beta_5)$ Class 2							
$\exp(\beta_6)$ Class 3	25.16	25.35	25.75	25.13	35.64	25.54	24.63
$\exp(\beta_7)$ Class 4	4.71	4.68	4.80	8.26	-6.92	4.65	4.43
$\exp(\beta_8)$ Class 5	51.08	51.18	51.28	53.30	47.12	51.30	51.01
	-22.92	-22.99	-22.79	-23.62	-25.33	-23.05	-23.38
$\chi^2 (10^{-4})$	4.68	4.67	4.67	5.10	8.20	4.67	4.71
Absolute difference	0.01	0.01	0.01	0.01	0.02	0.01	0.01

Conclusion

This study compared several risk classification methods for multiplicative and additive models by using weighted equation which is written as a weighted difference between the observed and fitted values. In addition, the parameter solutions for multiplicative and additive models were also compared by using weighted solution. The weighted solution for multiplicative model is in the form of a weighted proportion of observed over fitted values, whereas the weighted solution for additive model is in the form of a weighted difference between observed and fitted values.

In this study, the weighted equation was applied to estimate claim severity or average claim cost which is also equivalent to the total claim costs divided by the number of claims. The risk classification methods were compared on three types of severity data; Malaysian private motor third party property damage data, U.K. private vehicles own damage data from McCullagh and Nelder (1989), and data from Bailey and Simon (1960) on Canadian private automobile liability.

The fitting procedure were carried out by using two different approaches; classical and regression. The advantage of using the regression fitting procedure is that besides multiplicative and additive models, the fitting can also be extended to other regression models, as long as the function of the fitted value is written in a specified linear form. The inverse models were also fitted to the U.K. data because McCullagh and Nelder (1989) also fit the same data to the Gamma regression model by assuming that the regression effects were linear on the reciprocal scale.

As expected, the multiplicative and additive models give similar parameter estimates. The smallest chi-squares for multiplicative, additive and inverse models is given by the minimum chi-squares model. Except for the exponential model, all models provide similar values for absolute difference.

References

Ajne, B. (1975). A Note on the Multiplicative Ratemaking Model. *ASTIN Bulletin*, 8(2), 144-153.

Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher, E. & Thandi, N. (2004). A practitioner's guide to generalized linear models. *Casualty Actuarial Society Discussion Paper Program*, 1-115.

Bailey, R. A. & Simon, L. J. (1960). Two studies in automobile insurance ratemaking, *ASTIN Bulletin*, 4(1), 192-217.

Bailey, R. A. (1963). Insurance rates with minimum bias. *Proceedings of the Casualty Actuarial Society*, 50(93), 4-14.

Baxter, L. A., Coutts, S. M., & Ross, G. A. F. (1980). Applications of linear models in motor insurance. *Proceedings of the 21st International Congress of Actuaries*, Zurich. 11-29.

Brockmann, M. J. & Wright, T. S. (1992). Statistical motor rating: Making effective use of your data. *Journal of the Institute of Actuaries*, 119(3), 457-543.

Brown, R. L. (1988). Minimum bias with generalized linear models. *Proceedings of the Casualty Actuarial Society*, 75(143), 187-217.

Chamberlain, C. (1980). Relativity pricing through analysis of variance. *Casualty Actuarial Society Discussion Paper Program*, 4-24.

Coutts, S. M. (1984). Motor insurance rating: An actuarial approach. *Journal of the Institute of Actuaries*, 111, 87-148.

Feldblum, S. & Brosius, J. E. (2003). The minimum bias procedure: A practitioner's guide. *Proceedings of the Casualty Actuarial Society*, 90(172), 196-273.

Freifelder, L. (1986). Estimation of classification factor relativities: A modeling approach. *Journal of Risk and Insurance*, 53, 135-143.

Fu, L. & Wu, C. P. (2005). Generalized minimum bias models. *Casualty Actuarial Society Forum, Winter*, 72-121.

Harrington, S. E. (1986). Estimation and testing for functional form in pure premium regression models. *ASTIN Bulletin*, 16, 31-43.

Holler, K. D., Sommer, D. & Trahair, G. (1999). Something old, something new in classification ratemaking with a novel use of GLMs for credit insurance. *Casualty Actuarial Society Forum, Winter*, 31-84.

Jee, B. (1989). A comparative analysis of alternative pure premium models in the automobile risk classification system. *Journal of Risk and Insurance*, 56, 434-459.

Jung, J. (1968). On automobile insurance ratemaking. *ASTIN Bulletin*, 5(1), 41-48.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models* (2nd Ed.). London: Chapman and Hall.

Mildenhall, S. J. (1999). A systematic relationship between minimum bias and generalized linear models. *Proceedings of the Casualty Actuarial Society*, 86(164), 93-487.

Ismail, N. & Jemain, A. A. (2005a). Bridging minimum bias and maximum likelihood methods through weighted equation. *Casualty Actuarial Society Forum*, Spring, 367-394.

Ismail, N. & Jemain, A. A. (2005b). Handling overdispersion with negative binomial and generalized poisson regression models. *Proposal for the Casualty Actuarial Society Forum 2007*. Sub-committee members: Lewandowski, J.J., Holler, K.D. & White, J.

Renshaw, A. E. (1994). Modelling the claims process in the presence of covariates. *ASTIN Bulletin*, 24(2), 265-285.

APPENDIX A

S-PLUS programming for classical fitting procedure (Zero Bias Multiplicative)

```
ZeroBias.multi <- function(data)
{
# To identify matrix X, vector cost, and vector count from the data
  X <- as.matrix(data[,-(1:2)])
  cost <- as.vector(data[,1])
  count <- as.vector(data[,2])
# To set initial values for vector beta
  new.expbeta <- rep(c(1), dim(X)[2])
# To start the iteration
  for (i in 1:50)
  {
# To start the sequence
    for (j in 1:dim(X)[2])
    {
      expbeta <- new.expbeta
      fitted <- as.vector(exp(X%*%log(expbeta)))
      fitted.noj <- as.vector(exp(X[, -j]%*%log(expbeta[-j])))
      z <- as.vector(count*fitted.noj*X[,j])
      v <- as.vector(z/sum(z))
      new.expbeta[j] <- as.vector(sum(v*(cost/fitted.noj)))
    }
  }
# To calculate fitted values, chi-squares, and absolute difference
  fitted <- as.vector(exp(X%*%log(expbeta)))
  chi.square <- sum((count*(cost-fitted)^2)/fitted)
  abs.difference <- sum(count*abs(cost-fitted))/sum(count*cost)
# To list programming output
  list (expbeta=expbeta, chi.square=chi.square,
        abs.difference=abs.difference)
}
```

APPENDIX B

S-PLUS programming for regression fitting procedure (Least Squares Multiplicative)

```

LeastSquares.Reg <- function(data)
{
# To identify matrix X, vector cost, and vector count
  X <- as.matrix(data[, -(1:2)])
  cost <- as.vector(data[, 1])
  count <- as.vector(data[, 2])
# To set initial values for vector beta
  new.beta <- c(5, rep(c(1), dim(X)[2]-1))
# To start the iteration
  for (i in 1:20)
  {
    beta <- new.beta
    fitted <- as.vector(exp(X%*%beta))
    Z <- diag(fitted)%*%X
    W <- diag(count)
    r.s <- cost-fitted+as.vector(Z%*%beta)
    new.beta <- as.vector(solve(t(Z)%*%W%*%Z)%*%t(Z)%*%W%*%r.s)
  }
# To calculate fitted values, chi-squares, and absolute difference
  fitted <- as.vector(exp(X%*%new.beta))
  chi.square <- sum((count*(cost-fitted)^2)/fitted)
  abs.difference <- sum(count*abs(cost-fitted))/sum(count*cost)
# To list programming output
  list (expbeta=exp(new.beta), chi.square=chi.square,
        abs.difference=abs.difference)
}

```

Model Selection of Meat Demand System Using the Rotterdam Model and the Almost Ideal Demand System (AIDS)

Maria Divina S. Paraguas Anton Abdulbasah Kamil
Universiti Sains Malaysia

Aggregated time series data for differentiated meat products namely, beef, pork, poultry, and mutton were used to estimate and analyze Malaysian market demand for meats. The study aimed to select the most appropriate demand model between the equally popular Rotterdam model and the first difference Linear Approximate Almost Ideal Demand System (LA/AIDS) model by using a non-nested test. Both models were accepted, but further diagnostic tests revealed that the first difference LA/AIDS represents more appropriately the Malaysian market demand for meat than the Rotterdam model. Also, the elasticities from the first difference LA/AIDS were found to be more reliable than the Rotterdam model.

Keywords: Non-nested test, Rotterdam, AIDS, meat demand, Malaysia

Introduction

The consumer demand literature abounds with studies in which different models and estimation techniques of demand functions are applied. The two most widely adopted especially in food demand studies are the Rotterdam model introduced by Theil (1965) and Barten (1969) and Deaton and Muelbauer's (1980) almost ideal demand system (AIDS). Both models are derived from consumer theory, and are used to impose or test behavioral restrictions that are deduced from that theory (Kastens & Brester, 1996). However, neither economic theory nor statistical analysis provides clear *a priori* criteria for choosing between these two models (Lee, Brown, & Seale, 1994). Thus, the choice between which models fits better for a particular data set is an empirical question.

Jung and Koo (2000), in their study of the structure of Korean meat and fish product demand, compared the Linear Approximate Almost Ideal Demand System (LA/AIDS) and Rotterdam model to determine which of the two models is more appropriate for the data. Their study indicated that the LA/AIDS fits better than the Rotterdam model. In the study made by Tridimas (2000) in analyzing the pattern of consumer demand in Greece, the General Dynamic model of the AIDS fits better than the Static AIDS and the Rotterdam model.

In Malaysia, some studies have been conducted to analyze consumer demand for meat. Abdullah (1994) estimated both static and the dynamic AIDS in analyzing demand for fish and meat products in the country using time series data from 1960 to 1990. His results showed that the dynamic AIDS performed better than the static version. In an earlier study, Baharumshah (1993) used LA/AIDS and tested the model for serial correlation. A recent study by Milad (2003) adopted the Rotterdam model using data from 1970-2000. An *ex post* analysis was done to validate the model. In these studies, either only one functional form is used, so the choice of the model is made arbitrarily or the demand model is selected based on diagnostic tests. No study has been done to select the correct model by using a non-nested

Maria Divina S. Paraguas, School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.
Anton Abdulbasah Kamil, Mathematics Program, School of Distance Education, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia. Email: anton@usm.my

hypothesis test. Limited or no study has been done to compare different model specifications that best fit the demand for meat in Malaysia.

The purpose of this article is to analyze meat demand in Malaysia during the period of 1961-2002. The two systems of demand equations, the well-known AIDS and the Rotterdam model, are compared using a non-nested hypothesis test adapted from the compound model approach of Alston and Chalfant (1993). The dynamic structure and the empirical validity of the constraints of demand theory are systematically explored.

Rotterdam versus Almost Ideal Demand System (AIDS)

The estimable absolute price version of the Rotterdam model for n goods is written in the form:

$$\begin{aligned} \bar{w}_{i,t} \Delta \log q_{i,t} \\ = a_i + \sum_{j=1}^n \gamma_{ij} \Delta \log p_{j,t} \\ + \beta_i (\Delta \log X_t \\ - \sum_j \bar{w}_{j,t-1} \Delta \log p_{j,t}) + \varepsilon_{i,t} \end{aligned} \quad (1)$$

where $\bar{w}_{i,t}$ is the average budget share weight between consecutive time periods t and $t-1$ for good i ($i=1, \dots, n$), Δ is the across-periods first difference operator $q_{i,t}$ denotes the quantity demanded on good i at time t , $p_{j,t}$ is the nominal price of good j at time t , X_t is the total expenditure on the n goods at time t , a_i , γ_{ij} and β_i are the parameters to be estimated, and $\varepsilon_{i,t}$ is a zero-mean, normally distributed constant error variance.

The constraints of demand theory can be directly applied to the Rotterdam parameters. These are adding up, $\sum_i \beta_i = 1$, $\sum_i \gamma_{ij} = 0$; homogeneity, $\sum_j \gamma_{ij} = 0$; and symmetry, $\gamma_{ij} = \gamma_{ji}$.

The AIDS model on the other hand derives demand function for each consumption item in budget share form. However, in the time series context, the AIDS model is often

estimated in the first difference form to reduce the autocorrelation effect. And so, to make it consistent with the Rotterdam form, first difference LA/AIDS is then specified as:

$$\begin{aligned} \Delta w_{i,t} \\ = a_i + \sum_{j=1}^n \gamma_{ij} \Delta \log p_{j,t} \\ + \beta_i [\Delta \log X_t \\ - \sum_{j=1}^n w_{j,t-1} \Delta \log p_{j,t}] + \varepsilon_{i,t} \end{aligned} \quad (2)$$

Where the only difference in notation from equation (1) involves w , which is actual expenditure share weight at time t rather than a two-period average in equation (1). The theory of demand implies the following restriction on the parameters: adding up, $\sum_j \gamma_{ij} = 0$; homogeneity, $\sum_j \gamma_{ij} = 0$; and symmetry, $\gamma_{ij} = \gamma_{ji}$.

It is obvious that the Rotterdam model and the first difference LA/AIDS model are non-nested models. They are not directly comparable, because they have different dependent variables. However, comparisons of the right-hand sides of equations (1) and (2) indicate their similarity. Ex post analysis via statistical tests from estimating both models may suggest one is preferable but these kinds of comparisons are necessarily incomplete. Thus, when comparing these models, one needs an alternative procedure for the competing alternatives (Lee et. al, 1994).

Non-nested Test

Non-nested hypothesis tests select between two regression models where one model cannot be written as a special case of the other. In such a case, the models themselves are said to be non-nested (<http://go.okstate.edu/~brorsen/aidsvsrotterdam>). Alston and Chalfant (1993) developed a compound model approach in testing the two alternative models in which the right hand sides are identical but the dependent variables are not.

Let the two models be defined as:

$$\begin{aligned}\text{Model 1: } & y = f(x) \\ \text{Model 2: } & z = f(x)\end{aligned}$$

Using the Box-Cox transformation to nest both alternatives, and to test each against the more general alternative, the compound model is estimated as:

$$\lambda y + (1 - \lambda)z = f(x) \quad (3)$$

Thus, following Alston and Chalfant (1993) in testing for the Rotterdam model, the two alternative models can be combined as:

$$\begin{aligned}\text{Test 1: } & (1 - \phi)\Delta\bar{w}_i \log(q_i) + \phi\Delta w_i \\ & = \sum_{j=1}^n \gamma_{ij} \Delta \log(p_j) \\ & + \beta_i [\Delta \log X - \sum_{j=1}^n \bar{w}_j \Delta \log p_j]\end{aligned} \quad (4)$$

Equation (4) is a linear combination of the first difference LA/AIDS and the Rotterdam model. If $\phi = 0$, Equation (4) reduces to the Rotterdam model. A test of the hypothesis that $\phi = 0$ can be interpreted as a test of the hypothesis that the Rotterdam is the correct model.

The LA/AIDS can be tested directly as well. In the alternative compound model,

$$\begin{aligned}\text{Test 2: } & (1 - \lambda)\Delta w_i + \lambda\Delta\bar{w}_i \log(q_i) \\ & = \sum_{j=1}^n \gamma_{ij} \Delta \log(p_j) \\ & + \beta_i [\Delta \log X - \sum_{j=1}^n w_j \Delta \log p_j]\end{aligned} \quad (5)$$

a $\lambda = 0$ test implies that the first difference LA/AIDS is the correct model. And as with any pair of non-nested models, there are four possible outcomes from such a test: reject both, neither or either one of the two hypotheses. It is only when neither models are rejected that discrimination criteria via diagnostic tests could be used to select the best model (Doran, 1993).

Data

Time series data from 1961-2002 is used to estimate the meat demand model. Beef, pork, mutton, and poultry per capita annual consumption data are obtained from the FAOSTAT database (<http://faostat.fao.org>). The prices are average annual retail prices obtained from various reports of Division of Veterinary Services (DVS) (www.jphpk.gov.my) and Federal Agricultural Marketing Authority (FAMA) of Malaysia.

Model Estimation

The demand model consists of four equations, including beef, pork, poultry, and mutton. The iterated seemingly unrelated regression procedure available in SAS is used as an estimation method. Symmetry, adding up, and homogeneity conditions were all imposed to make the models consistent with underlying economic theory. The mutton equation was not included in the system during the estimation process to avoid singularity in the covariance matrix. The parameters of the deleted equation were recovered using the adding up restriction.

Results

The test for the Rotterdam model as the correct specification is not rejected at any reasonable significance level. The estimated value of ϕ is 0.4853 with a p-value of 0.1658. Therefore, imposing the Rotterdam model as a restriction on the compound model is supported by this data. However, the test on the first difference LA/AIDS as an alternative model, the $\lambda = 0$ test is also not significant. The estimated value of λ is 0.1560 with a p-value of 0.1389. In other words, imposing the LA/AIDS as a restriction on the compound model is also supported by this data. Therefore, the outcome of the tests reveals that both models are accepted. This implies that this specific data is not rich enough to discriminate between the Rotterdam and the first-difference LA/AIDS models.

In order to discriminate between the two systems, the empirical performance was

Table 1. Parameter Estimates with Homogeneity and Symmetry Imposed

	Rotterdam Model				First Difference LA/AIDS			
	Beef	Pork	Poultry	Mutton	Beef	Pork	Poultry	Mutton
γ_{1i}	-0.046 (0.032)				0.086* (0.033)			
γ_{2i}	0.013 (0.016)	0.010 (0.020)			-0.048* (0.017)	0.230* (0.021)		
γ_{3i}	0.013 (0.018)	-0.008 (0.015)	0.005 (0.019)		-0.050* (0.018)	-0.160* (0.014)	0.228* (0.018)	
γ_{4i}	0.018 (0.032)	-0.036 (0.041)	0.025 (0.034)	0.020	0.017 (0.033)	-0.045 (0.042)	0.009 0.032	0.020
β_i	0.103* (0.047)	0.644* (0.060)	0.219* (0.051)	0.017	-0.020 (0.050)	0.182* (0.065)	-0.179* (0.050)	0.017
Constant	0.002 (0.003)	-0.014* (0.004)	0.013* (0.003)	0.999	0.001 (0.003)	-0.014* (0.004)	0.015* (0.003)	0.999
Adj. R ²	0.0307	0.7449	0.2831		0.234	0.806	0.870	
RMSE	0.0165	0.0216	0.0182		0.017	0.023	0.017	

*Denotes significance at the 5 per cent, based on asymptotic t-ratios. $i = 1, 2, 3$ and 4, where 1 = beef, 2 = pork, 3 = poultry, 4 = mutton

examined with regard to goodness-of-fit, forecasting accuracy, and the elasticity behaviors of the demand systems.

The parameter estimates for both models are reported in Table 1. Five of the 18 coefficients are significantly different from zero in the Rotterdam model, although ten coefficients are statistically significant for the first difference LA/AIDS model. No price coefficient is statistically significant in the Rotterdam model.

The first difference LA/AIDS model performs better than the Rotterdam model as indicated by the adjusted R² in each meat equation. The first difference LA/AIDS model has the highest adjusted R².

Based on the predictive accuracy of the model, the RMSE measures the *ex post* forecasting performance. From table 1, the RMSEs are the lowest from the first difference LA/AIDS model, suggesting a better fit than the Rotterdam model.

Based on the predictive accuracy of the model, the RMSE measures the *ex post* forecasting performance. From table 1, the RMSEs are the lowest from the first difference LA/AIDS model, suggesting a better fit than the Rotterdam model.

Demand systems are consistent with the assumptions of utility maximization if they satisfy homogeneity and symmetry restrictions. Testing and imposing of demand restrictions are central to the empirical analysis of demand. Table 2 reports the results of the joint symmetry and homogeneity restriction test. Both models

accept the null hypothesis of joint symmetry and homogeneity at 5 % significance levels. Thus, the data confirms with the theoretical restrictions of demand in both models.

Choosing between alternative specifications of the model by a purely statistical test is the interest of model selection. The influence of model choice on elasticity estimates is also worth considering. The parameter estimates obtained from both models are used to calculate the demand elasticity estimates in each model.

The following elasticity formulas were calculated using the formula from Barten[6] :

Table 2. Joint Symmetry and Homogeneity Restriction Test

Model	Alternative Hypothesis	Number of restrictions	F-Value	Pr>F	Conclusion
Rotterdam Model	No Restriction	6	0.83	0.5517	Accept Ho
LA AIDS Model	No Restriction	6	0.99	0.4360	Accept Ho

	Rotterdam Model	LA/AIDS model	
Expenditure Elasticity	$\eta_i = \beta_i / \bar{w}_i$	$\eta_i = 1 + \beta_i / \bar{w}_i$	(6)
Compensated Elasticity	$e_{ij}^c = \gamma_{ij} / \bar{w}_i$	$e_{ij}^c = -\delta_{ij} + (\gamma_{ij} / \bar{w}_i) + \bar{w}_j$	(7)
Uncompensated Elasticity	$e_{ij}^u = (\gamma_{ij} - \beta_i \bar{w}_j) / \bar{w}_i$	$e_{ij}^u = (\gamma_{ij} - \beta_i \bar{w}_j) / \bar{w}_i$	(8)

Where $\delta = 1$ for $i = j$ and $\delta = 0$ otherwise. $\bar{w}$ is the average budget share in each meat equation a. β_i and γ_{ij} are the estimated parameters.

The estimated elasticities exhibit some similarities and minor differences between the two models. Looking at Table 3, the calculated expenditure elasticity estimates are similar for both models and suggest that beef and poultry are necessities in Malaysia, while pork and mutton are luxury meat products.

Table 3. Estimated Expenditure Elasticities: Rotterdam and Almost Ideal Specification

	Rotterdam LA/AIDS	
	Model	Model
Beef	0.62	0.88
Pork	1.54	1.44
Poultry	0.57	0.54
Mutton	1.13	1.56

Table 4 summarizes the uncompensated and compensated price elasticity estimates of both models. The own-price elasticities of the first difference LA/AIDS model have all the correct negative signs while the Rotterdam model compensated own-price elasticity for pork (0.02) and poultry (0.01) are positive, which are unexpected. All the own-price elasticities are less than 1 implying that these meat commodities are price inelastic. In all cases, the absolute value of own-price elasticity is greater

in the LA/AIDS model (both uncompensated and compensated). Pork has the greatest uncompensated own-price elasticity. Beef, mutton, and poultry follow it.

With respect to the cross price elasticity estimates, the results from the first difference LA/AIDS model are similar to the results obtained from the cross price elasticity estimates of the Rotterdam model. However, they do differ in the value of the estimates generated. The Marshallian cross price elasticity estimates are mostly negative which indicate gross complements among the meat products.

The results are in accordance to the results obtained by Wohlgenant and Hahn (1992) and Alston and Chalfant (1993), in their studies in the US. Their studies have found that the elasticity estimates from the Rotterdam model and first difference LA/AIDS model have minor differences despite the variation in their implications and their consistency with the data. Their results produce very similar elasticities although one model is rejected in favor of the other.

However, the results reported in the preceding paragraphs revealed that though the estimates from both models are quite similar. The estimates from the Rotterdam model are found to be in question based on their signs. This result is comparable to the study made by Lee et al. (1994) on general consumption patterns in Taiwan; they concluded that elasticity estimates from the Rotterdam model are questionable. Thus, choice of functional form and demand elasticity estimates for the Rotterdam and Almost Ideal Demand models may vary with the data set (Xu & Veeman, 1996).

Table 4. Estimated Elasticities: Rotterdam and LA/Almost Ideal Demand Specification

Quantity	Price	Uncompensated		Compensated	
		Rotterdam	LA AIDS	Rotterdam	LA AIDS
Beef	Beef	-0.38	-0.46	-0.28	-0.31
	Pork	-0.22	-0.19	0.03	0.05
	Poultry	-0.06	-0.05	0.03	0.04
	Mutton	0.40	0.46	0.59	0.72
Pork	Beef	-0.18	-0.24	0.08	0.13
	Pork	-0.62	-0.63	0.02	-0.03
	Poultry	-0.26	-0.22	-0.02	0.005
	Mutton	-1.65	-1.73	-1.18	-1.07
Poultry	Beef	-0.16	-0.26	0.08	0.08
	Pork	-0.61	-0.55	-0.02	0.004
	Poultry	-0.21	-0.23	0.01	-0.02
	Mutton	0.39	0.07	0.83	0.67
Mutton	Beef	0.08	0.11	0.11	0.13
	Pork	-0.13	-0.12	-0.08	-0.08
	Poultry	0.05	0.04	0.07	0.05
	Mutton	-0.28	-0.37	-0.24	-0.32

Conclusion

The purpose of the article is to analyze the market demand for differentiated meat products in Malaysia during the period 1961-2002. The functional forms selected have been the popular Rotterdam and Almost Ideal Demand models. Comparison of the two models required the use of a non-nested test. Moreover, economic criteria and the behavior of the elasticity estimates were used to evaluate the demand systems.

For this particular data, the compound model approach suggested by Alston and Chalfant (1993) was used to nest and select the appropriate model in this study. The results suggested that the first difference LA/AIDS or the Rotterdam models are both appropriate to represent Malaysian demand for differentiated meat products. Also, turning to the empirical validity or testing for the joint symmetry and homogeneity restrictions showed that both models satisfy the theoretical restriction of demand.

However, the first difference LA/AIDS gained superiority over the Rotterdam model based on its goodness of fit and reliability of estimates. The first differenced LA AIDS fits well with the data as reflected by its higher Adjusted R^2 and lower RMSE relative to the Rotterdam model. Compensated own- price elasticity estimates of pork and poultry from the Rotterdam model do not carry the expected signs, which render the estimates from the Rotterdam model questionable. Thus, the first difference LA/AIDS is chosen in favor of the Rotterdam model.

References

- <http://faostat.fao.org>. Accessed on August 4, 2004.
- <http://go.okstate.edu/~brorsen/aidsvsroterdam>. Accessed on April 2, 2005.
- Abdullah, N. M. R. (1994). Incorporating habit in the demand for fish and meat products in Malaysia. *Malaysian Journal of Economic Studies*, 10(2), 25-34.
- Alston, J. M. & Chalfant, J. A. (1993). The silence of the lambdas: A test of the almost ideal and Rotterdam models. *American Journal of Agricultural Economics*, 75(2), 304-313.
- Baharumshah, A. Z. (1993). Applying the almost ideal demand systems to meat expenditure data: Estimation and specification issues. *The Malaysian Journal of Agricultural Economics*, 10, 23-37.
- Barten, A. P. (1993). Choice of functional form: Consumer allocation models. *Empirical Economics*, 273-83.
- Barten, A. P. (1969). Maximum likelihood estimation of complete systems of demand equations. *European Economic Review*, 10, 7-73.
- Deaton, A. S. & Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3), 312-326.
- Division of Veterinary Services (DVS). Market price of livestock products. Available: <http://www.jphpk.gov.my>. Accessed on January 19, 2004.
- Doran, H. (1993). Testing nonnested models. *American Journal of Agricultural Economics*, 75(2), 95-103.
- Jung, J. & Koo, W. W. (2000). *An econometric analysis of demand for meat and fish products in Korea*. (Agricultural Economics Report No. 439. North Dakota State University, USA).
- Kastens, T. L. & Brester, G. W. (1996). Model selection and forecasting ability of theory-constrained food demand systems. *American Journal of Agricultural Economics*, 78, 301-312.
- Lee, J., Brown, M. G., & Seale, J. L. (1994). Model choice in consumer analysis: Taiwan, 1970-89. *American Journal of Agricultural Economics*, 76(3), 504-512.
- Milad, M. S. (2003). Application of the Rotterdam model to the meat demand system in Malaysia. (Masters Thesis, Universiti Sains Malaysia).
- Theil, H. (1965). The information approach to demand analysis. *Econometrica*, 33, 67-87.

Tridimas, G. (2000). The analysis of consumer demand in Greece: Model selection and dynamic specification. *Economic Modeling*, 17, 455–471.

Wohlgenant, M. & Hahn, W. (1992). Dynamic adjustment in monthly consumer demand for meats. *American Journal of Agricultural Economics*, 64(3), 553-7.

Xu, X. & Veeman, M. (1996). Model choice and structural specification for Canadian meat consumption. (Staff Paper No. 96-04, University of Alberta, Canada, 1996).

Supporting and Preparing Future Decision-makers with the Needed Tools

Michael Wolf-Branigin
George Mason University

Educational and social service researchers and evaluators continue to develop advanced statistical methods. To ensure that our students have the essential skills as they enter direct service, the focus must be on assuring that they learn readily understandable methods that are appropriate for small samples and use repeated measures.

Keywords: Non-parametric, small samples, repeated measures, decision-making, complexity

Introduction

After conducting behavioral health accreditation surveys over the past nine years and having been academia for five years, it is surprising how infrequently evaluators and administrators of these organizations use inferential statistics. Being able to cross over between the theoretical foundations of academia and the applied realm of social service organizations affords a unique perspective as to how organizations actually apply statistical methods. For example, in only two instances out of 46 opportunities over the past nine years did the social service organization's evaluator or quality assurance administrator apply any inferential statistic when analyzing the outcomes generated by their organization. When a test of significance was applied it was a Student's t-test in order to measure the improvement of client outcomes by comparing a cohort of intake scores to scores at time of program completion. In none of these cases did the organization's evaluator use distribution free statistics or a repeated measures approach to measure their client's outcomes.

Michael Wolf-Branigin is Associate professor of Social Work. His research interests include developmental disabilities and addictions, and applications of complex systems and information technology infusion. He has had prior experience in human service administration, governmental and non-governmental consulting and academia.

Many readers of higher-level statistical and methodological journals are doctoral students with a teaching load of introductory research and statistical classes and must thereby remain cognizant of the analytical abilities and the statistical skills of the students that they need in order to encourage effective decisions. These students frequently enter their respective organization as administrators, consultants, or evaluators. While efforts have been initiated, for example the American Statistical Association (ASA) called for university statistical and other related departments to prepare statisticians on database management; IT through an accreditation process, and another ASA initiative is investigating the need to assure that statistical practitioners develop soft skills among statisticians to improve written, verbal and interpersonal skills (Batcher, 2005) their scope is limited. Additional supports need to be provided directly in the classrooms.

To emphasize the point, as the use of evidence-based practice and approach such as action research gain acceptance by practitioners and educators, gaps in the appropriate use of simple data analysis remain. As the methodological and statistical educators of these individuals, serious attention is needed to improve the link between research and practice. Given the complexities of many modern statistical techniques, now is an opportune time to reinforce some of the basic methods to assure that these future decision-makers are using correct methods. As an educator of students who eventually plan to become social work administrators, I suggest that there is a need to

reflect and ask what useful information do these entry-level professional students gain that can be applied to their organizations or schools? When asked this, three problems frequently occur:

1. Failing to measure the same outcome variables at the beginning and end of an intervention or educational experience
2. Failing to apply tests of significance using repeated measures when appropriate
3. Committing Type II Errors.

Although the topics discussed below will appear obvious to most readers, these elementary concepts need strengthening. Most readers of this journal have experience instructing introductory research methods and statistics classes. Often these courses are for either higher-level undergraduate or master level social and educational services professionals. As instructors, it must be realized that a large proportion of the decision-makers within these social and education related services have a master's degree that required at best two research classes and possible one introductory statistics class. Their future employers assume that this rudimentary level of quantitative education provides a sufficient background for their employment.

Correct and robust application of statistics provides an essential tool for these students to use in assisting their organizations demonstrate statistically significant and measurable outcomes. These measurable outcomes additionally assist the organization in locating, securing and justifying further funding. This becomes increasingly important as these individuals work with smaller data sets in either classroom settings or smaller programs within educational and social service settings. Again, as an individual conducting accreditation surveys, quality improvement, and academic pursuits, it remains apparent how little information practitioners have on the use of robust methods.

Efforts of educators and practitioners in advanced methods can focus on several simple areas that will increase the likelihood that useful statistical techniques are applied in smaller settings, including: using repeated measures, increasing use of nonparametric tests, and

providing an introduction to robust estimators. Although each of us knows these introductory techniques, below are the common and overlooked topics from introductory statistics classes that provide the greatest likelihood of assuring that future decision-makers apply appropriate methods.

Use of Identical Measures

First, the student needs to understand what their organization's monitoring and evaluation system is attempting to measure. Once understood, the student needs to assure that an organization uses the identical instrument at each measurement phase in order to guarantee that data on the same variable are collected. For example, in substance abuse treatment, organizations typically submit admission and discharge forms to their state funding source that in turn are reported to the Federal government. A variable such as "the number of days in the last month that the client has abused substances" is asked by most programs at admission and discharge.

Staff at the organizations can readily compare these measures at the different phases to identify actual changes in substance use by the consumer. Assuming that the organization tracks this information by each individual, use of repeated measures is a simple process. Qualitatively, social service, and educational institutions often have well conceptualized interventions and curricula that lead to positive outcomes; however, future decision-makers have not been provided with the simple tools to measure significant improvements.

Apply the Correct Statistical Methodology

As students prepare to enter the workforce, they need to identify and focus on the outcomes that their program is attempting to achieve. Most often organizations are attempting to compare groups of scores. These groups most often will be a group of scores before the persons served are active in treatment, during treatment, at the end, and finally at follow-up. Evaluators at these organizations also need to differentiate when to use parametric and nonparametric tests. Again, reinforce that the outcomes data or dependent variables when scaled at an ordinal or nominal measurement

level requires using a nonparametric test. Similarly, the actual distributions are rarely normally distributed in these data sets (Micceri, 1989) and therefore are best analyzed using a comparable nonparametric test.

Improve likelihood of identifying significant differences

The third problem will be a bit more abstract for your students, the likelihood of committing type II errors. The two most common reasons include using too small of a sample and applying the incorrect test of significance. To encourage that students gain an introductory understanding of inferential and just as importantly non-parametric statistics, you may encourage them to use the MS Excel Stat Tool Pak on their computers. While obviously not as flexible as more advanced packages such as SPSS or SAS, the package is available to virtually anyone at anytime. As these individuals are instructed on more advanced methods, one must not rush into teaching only empiricism at the expense of constructs (Resich, 2000) and the student's ability to think critically.

Conclusion

Two vital issues must be addressed as the next generation of quantitative researchers and evaluators are prepared to enter into education and social service. Because organizations are asked frequently to conduct evaluations using small samples, and the data from these distributions usually are non-normal, these future leaders need to receive sufficient instruction on the methods most appropriate for dealing with these concerns. Therefore, methods must be provided to link robust methods for organizational (social service and educational) decision-making. As a frequent user of modern statistical methods, the intention of this article is not to disparage our tremendous strides, but rather to suggest that in preparing the next generation of educational and social service decision-makers there must be balance. This may best be achieved through reinforcing the students' knowledge of why repeated measures are more powerful especially given limited samples.

One noticeable and encouraging trend within these organizations is the continually improving and increasing use of information technology that assists in doing a better job of measuring positive changes in their clients' outcomes. This increasing use of evidence to support the design and implementation of interventions may be related to recent developments of evidence-based practice whereby planners and administrators of services identify relevant information for choosing the interventions they provide. Evidence may be in the form of scholarly journal articles, programs having been identified as best practice, and comparing outcomes of program interventions to state and national norms, to name but a few.

These future decision-makers must be assisted with the data they collect and process for organizations to ensure they use appropriate statistical methods so as to minimize the likelihood of producing Type II errors. Researchers need to encourage the use of more robust methods in the statistics and methodology classes in order for their students, who eventually become leaders within these organizations, too become familiar and comfortable in their use. Encouraging students to use these elementary and powerful statistical approaches will assist them in justifying future funding.

As governments increase their efforts to identify educational improvements in student achievement vis-à-vis programs such as *No Child Left Behind* (US Department of Education, 2002), or as child and family service organizations attempt to demonstrate macro level improvements in the outcomes of children, through the US Department of Health and Human Services Child and Family Service Reviews - real change often occurs at the smaller organizational unit. One possible avenue to explore with these students is the use of complexity theory. Although applications remain rare in educational and social services (Hudson, 2000, Proehl, 2001), complexity theory has in recent decades evolved within the natural sciences. As the assumptions pose significant challenges (Patton, 2002), providing students with a framework for applying a complexity approach at a minimum should include reviewing the components of a complex

system. These components include identifying the agent (student or consumer of services), the different choices that the agent can make, the dynamics within the environment, the availability of a feedback mechanism to assist in decision-making, an underlying organizing structure within which all these phenomena occur, and the resulting emergent behavior or outcome.

References

Batcher, M. (2005). Results of the focus groups on accreditation. *Amstat News*. No. 337, 18-19.

CARF: The Rehabilitation Accreditation Commission (2005). *Behavioral Health Accreditation Standards Manual 2005-2006*, 51-60. Tucson, AZ: CARF.

Hudson, C. G. (2000). At the edge of chaos: A new paradigm for social work? *Journal of Social Work Education*, 36(2), 215-230.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156-166.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

Proehl, R. A. (2001). *Organizational change in the human services*. Thousand Oaks, CA: Sage.

Reisch, M. (2000). The future of doctoral social work education in the United States: Questions, issues, and persistent dilemmas. *Journal of Social Work Education*, 36(2), 57-71.

United States Department of Education (2001) *No Child Left Behind Act*. Website <http://www.ed.gov/nclb/overview/intro/index.html> Washington, DC.

The Individuals Control Chart in Case of Non-Normality

Betül Kan Berna Yazıcı
Department of Statistics
Anadolu University

This article examines the effects of non-normality as measured by skewness and provides an alternative method of designing individuals control chart with non-normal distributions. A skewness correction method for constructing the individuals control chart is provided. An example of thickness of biscuit process is presented to illustrate the individuals control chart limits.

Key words: Quality control, shewhart control charts (S), skewness correction (SC).

Introduction

Statistical methods provide many useful applications in industrial process control. One of them is the control charts method, used to detect the occurrence of assignable causes in industry, presented by Shewhart in 1924. The Shewhart control charts are based on the assumption of the distribution of the quality characteristic is normal or approximately normal.

The setting of the control limits to utilize on a control chart assumes the assumption of normality. However, in many situations, this condition does not hold. There are numerous studies on the control charts when the underlying distribution is non-normal.

Burr (1967) studied the non-normality on the constants for the $\bar{x}$ and R charts. For skewed populations, Type I Risk probabilities grow larger as the skewness increases. For highly skewed populations, one possible solution is to increase sample size.

Control limits for median, range, scale, and location charts for the Weibull distribution have been developed, and they match Shewhart's assumption for samples of four or more; normal theory is satisfactory. There has been a suggestion to use a weighted variance method based on the semivariance approximation, and asymmetric control chart limits for the $\bar{x}$ and R charts are available. The geometric midrange and geometric range charts has also been studied for a lognormal population. An arcsin transformation to transform a binomial variable into a normal one has been suggested.

Also, Bai and Choi (1995) proposed a heuristic weighted variance method with no assumption on the form of the distribution. Another approach for non-normality is considered using some transformed Q statistics which are the standard normal or approximately normal. Others chose to transform the data in order to make them quasi-normal, or used systems of the distributions as a general tool for transforming the data to normality.

In this study, the application of generalized Burr and Weibull distributions are demonstrated. First, non-normality and the assumption of the normal are mentioned. Second, the skewness method formulas for an individual control chart, when the distribution of the data is non-normal, is proposed. Third, a data set is illustrated as an example at the application study. Finally, the results of the studies are presented.

Betül Kan is Research Assistant in Department of Statistics. Her research areas are quality control, mathematical statistics and design of experiments. E-Mail: kanbetul@yahoo.com. Berna Yazıcı is Assistant Professor in the Department of Statistics. Her research interests include design of experiments, regression analysis and quality control. E-Mail: bernayazici@yahoo.com.

Methodology

Non-Normality

The assumption of normality means that the probability density function of the quality characteristic X has to be normal or approximately normal. However, this assumption may not be tenable. In many situations, there may be reason to doubt the validity of the normality assumption. For instance, the distribution of measurements from chemical processes or observations of lifetimes are often skewed. If the measurements are really normally distributed, the statistic $\bar{x}$ is also normally distributed. If the measurements are asymmetrically distributed, the statistic $\bar{x}$ will be approximately normally distributed only when the sample size n is sufficiently large (the central limit theorem).

However, the sample size n is usually chosen according to the sampling cost. Therefore, if the sample size n is not sufficient and the measurements are not normally distributed, then using the Shewhart control charts for constructing the control limits may reduce the ability of detecting the assignable causes of the control chart. In order to compute the control chart constants, Weibull and Burr distributions are chosen, because they represent a wide variety of shapes as parameters change.

The Burr Distribution

Burr proposed new $\bar{x}$ charts based on non-normality by using the Burr distribution to modify the usual symmetrical control limits. He developed the tables for constructing the modified control charts. He tabulated the expected value, standard deviation, skewness coefficient, and kurtosis coefficient of the Burr distribution for various combinations of c and q . The user can make a standardized transformation between a Burr variate (Y) and another random variate (X) by these tables. The Burr system covers a wide and important range of the standardized third and fourth central moments and can be used to fit a wide variety of practical data distributions.

The cumulative distribution function (CDF) of the Burr distribution is:

$$F(y) = 1 - (1 + y^c)^{-k} \quad y > 0 \quad (1)$$

where c and k are parameters ($c, k \geq 1$). Also, for large values of k , Burr distribution approaches the Weibull distribution.

The Weibull Distribution

The Weibull distribution is very flexible, and by appropriate selection of the parameters, the distribution can assume a wide variety of shapes. It is used extensively in reliability engineering as a model of time failure for electronic devices. The CDF of the Weibull's distribution is

$$F(y) = 1 - \exp(-[\lambda y]^\beta) \quad y > 0 \quad (2)$$

where λ and β are the scale and shape parameters, respectively.

The Individuals Control Chart

The individuals control chart, which may also be used to observe the magnitude of skewness, is used to find and remove the assignable causes quickly from the process if the sampling is possible, although $\bar{x}$ -R or $\bar{x}$ -s control charts can be constructed for the same situations.

The Process Distribution and the Parameters are Known

Let $x_1, x_2, \dots, x_n$ be a sample from a process distribution whose mean is μ and standard deviation σ is known. The proposed the individuals control chart based on Skewness Correction (I_{sc} Chart) and the individuals control chart based on Shewhart (I_s Chart) methods are:

$$\begin{aligned} UCL_{SC} &= \mu + (3 + c_4^*)\sigma / \sqrt{n} \\ I_{sc} \text{ Chart: } CL_{SC} &= \mu \\ LCL_{SC} &= \mu + (-3 + c_4^*)\sigma / \sqrt{n} \end{aligned}$$

(3)

The constant c_4^* is the skewness correction.

$$c_4^* = \frac{\frac{4}{3}k_3(\bar{x})}{1 + 0.2k_3^2(\bar{x})} \quad (4)$$

$k_3(\bar{x})$ is the skewness (i.e., coefficient of skewness) of sample mean $\bar{x}$. When the process distribution is symmetric, $c_4^*=0$ and I_{SC} chart has the form of Shewhart chart.

The Process Distribution and the Parameters are Unknown

Let $x_{i1}, x_{i2}, \dots, x_{in}$, $i=1, 2, \dots, m$, be m samples of size n from a process distribution with mean μ , standard deviation σ , and skewness k_3 . When the process distribution is normal, Shewhart individuals control chart is

$$\begin{aligned} UCL_S &= \bar{x} + 3\bar{R}/d_2 \\ CL_S &= \bar{x} \\ LCL_S &= \bar{x} - 3\bar{R}/d_2 \end{aligned} \quad (5)$$

Here,

$$\bar{x} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n x_{ij},$$

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i,$$

R_i is the range of the i th sample. d_2 depends on the sample size n and calculated when the distribution is normal.

Let d_2^* , $d_2^* = \mu_R / \sigma_x$, be the constant for the given skewed process distribution corresponding to the role of the constant d_2 for the normal distribution. When $k_3=0$, the constant d_2^* is close to d_2 for the normal process distribution. Therefore, the skewness correction

method for individuals control chart is practically the Shewhart individuals control chart. The sample size n , is determined for the samples in the control process, the estimated value for the skewness of the sample averages

$$\hat{k}_{3\bar{x}} = \hat{k}_3 / \sqrt{n} \quad (6)$$

In many cases, the skewness k_3 needs to be estimated. It can be estimated by the sample skewness

$$\hat{k}_3 = \frac{1}{mn-3} \sum_{i=1}^m \sum_{j=1}^n \left(\frac{x_{ij} - \bar{x}}{\sqrt{\frac{1}{mn-1} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2}} \right)^3 \quad (7)$$

It is known that the estimator $\hat{k}_3$ of the k_3 is consistent. As well, it is known that the sample skewness $\hat{k}_3$ is essentially the third moment estimator. Based on the skewness correction method for the individuals control chart, the following is proposed

I_{SC} Chart:

$$\begin{aligned} UCL_{SC} &= \bar{x} + \left(3 + \frac{4k_3/(3\sqrt{n})}{1 + 0.2k_3^2/n} \right) \frac{\bar{R}}{d_2^*} = \bar{x} + A_{SC}^U \bar{R} \\ CL_{SC} &= \bar{x} \\ LCL_{SC} &= \bar{x} + \left(-3 + \frac{4k_3/(3\sqrt{n})}{1 + 0.2k_3^2/n} \right) \frac{\bar{R}}{d_2^*} = \bar{x} + A_{SC}^L \bar{R} \end{aligned} \quad (8)$$

Table 1 gives the values of the constants A_{SC}^U and A_{SC}^L . If the skewness $k_3 < 0$, the A_{SC}^U is the same as the A_{SC}^L for $-k_3$ in Table 1, and vice versa. For instance, when $n=2$ and $k_3=0.8$, $A_{SC}^U=3.35$ and $A_{SC}^L=2.08$, with $k_3=-0.8$, $A_{SC}^U=2.08$ and $A_{SC}^L=3.35$.

Table 1. The constants A_{SC}^U and A_{SC}^L for n=2,3,4,5,7,10

n	2		3		4		5		7		10	
k_3	A_{SC}^U	A_{SC}^L	A_{SC}^U	A_{SC}^L	A_{SC}^U	A_{SC}^L	A_{SC}^U	A_{SC}^L	A_{SC}^U	A_{SC}^L	A_{SC}^U	A_{SC}^L
0,0	2,66	2,66	1,78	1,78	1,46	1,46	1,30	1,30	1,11	1,11	0,98	0,98
0,4	3,03	2,36	1,96	1,59	1,64	1,38	1,41	1,19	1,19	1,03	1,04	0,92
0,8	3,35	2,08	2,17	1,45	1,74	1,22	1,52	1,12	1,27	0,98	1,11	0,89
1,2	3,69	1,87	2,37	1,33	1,90	1,14	1,65	1,03	1,38	0,93	1,17	0,82
1,6	4,00	1,73	2,58	1,25	2,06	1,08	1,77	0,98	1,48	0,87	1,23	0,79
2,0	4,27	1,63	2,77	1,18	2,20	1,02	1,90	0,94	1,56	0,85	1,33	0,79
2,4	4,51	1,58	2,93	1,13	2,36	0,98	2,03	0,89	1,67	0,79	1,39	0,73
2,8	4,70	1,60	3,08	1,11	2,48	0,94	2,12	0,87	1,75	0,77	1,45	0,70
3,2	4,88	1,64	3,22	1,11	2,58	0,94	2,24	0,85	1,83	0,77	1,52	0,66
3,6	4,98	1,70	3,33	1,13	2,68	0,94	2,33	0,83	1,90	0,74	1,58	0,66
4,0	5,08	3,36	3,41	1,14	2,78	0,94	2,39	0,83	1,98	0,71	1,61	0,38

If the k_3 value is not in this table, the user can take the nearest k_3 value or use interpolation. For instance, if a control chart for individuals n= 2 and $k_3 = 1,35$ is created, then one can take $A_{SC}^U = 3,69$ and $A_{SC}^L = 1,87$, or linearly interpolated values,

$$A_{SC}^U = \left(\frac{1,35-1,6}{1,2-1,6} \right) 3,69 + \left(\frac{1,35-1,2}{1,6-1,2} \right) 4,00$$

$$= 0,625 \cdot 3,69 + 0,375 \cdot 4,00 = 3,806$$

$$A_{SC}^L = \left(\frac{1,35-1,6}{1,2-1,6} \right) 1,87 + \left(\frac{1,35-1,2}{1,6-1,2} \right) 1,73$$

$$= 0,625 \cdot 1,87 + 0,375 \cdot 1,73 = 1,817$$

Here, the user can also take the nearest k_3 value from Table 1, as $k_3=1,2$. Then, the values $A_{SC}^U=3,69$ and $A_{SC}^L=1,87$ are obtained.

Results

The measurement of the thickness of biscuit was obtained at a random time from a process line in a factory. ISO 9002 Quality

Assurance System is applied in this factory. The production of the dough is automatically materialized. The manufacturer uses an $\bar{x}$ chart to monitor the process. The data used in the study are independently taken in by equal time intervals. Based on an analysis of the skewness correction method, it is estimated that for $n=5$ with $\bar{x}=15,355$, $\bar{R}=0,90$, $k_3=0,051$ from the Weibull distribution, by interpolation, the constants are $A_{SC}^U=1,30$ and $A_{SC}^L=1,29$ from Table 1. It is noted that the constant is $d_2=2,326$ for $n=5$. Figure 1 shows the process for the Weibull distribution, $n=5$. Also the two different control limits are pointed out. The control limits of the individuals control chart based on the Shewhart method and skewness correction method are:

$$UCL_{SC} = \bar{x} + A_{SC}^U \bar{R} = 16,525$$

$$\text{I}_{SC} \text{ Chart: } CL_{SC} = \bar{x} = 15,355$$

$$LCL_{SC} = \bar{x} - A_{SC}^L \bar{R} = 14,194$$

$$UCL_S = \bar{x} + 3\bar{R}/d_2 = 16,51$$

$$\text{I}_S \text{ Chart: } CL_S = \bar{x} = 15,355$$

$$LCL_S = \bar{x} - 3\bar{R}/d_2 = 14,20$$

At the second stage, it is estimated that for $n=7$ with $\bar{x}=15,434$, $\bar{R}=1,065$, $k_3=0,056$ for the Weibull distribution, by interpolation, the constants are $A_{SC}^U=1,11$ and $A_{SC}^L=1,11$ from Table 1. It is noted that the constant is $d_2=2,704$ for $n=7$. Figure 2 shows the process for the Weibull distribution, $n=7$. Also, the two different control limits are pointed out. The control limits of the individuals control chart are based on the Shewhart method and skewness correction method which are:

$$UCL_{SC} = \bar{x} + A_{SC}^U \bar{R} = 16,61$$

$$\text{I}_{SC} \text{ Chart: } CL_{SC} = \bar{x} = 15,434$$

$$LCL_{SC} = \bar{x} - A_{SC}^L \bar{R} = 14,25$$

$$UCL_S = \bar{x} + 3\bar{R}/d_2 = 16,61$$

$$\text{I}_S \text{ Chart: } CL_S = \bar{x} = 15,434$$

$$LCL_S = \bar{x} - 3\bar{R}/d_2 = 14,25$$

Figure 1. Individuals chart for Weibull Distribution, n=5

Figure 2. Individuals chart for the Weibull Distribution, n=7

If a comparison is made between the control limits constructed by two methods, it may be said that for $n=5$, the range of control limits have expanded but for $n=7$ the range between the control limit have not changed.

On the other hand, based on an analysis of skewness correction method, it is estimated that for $n=5$ with $\bar{x}=14,599$, $\bar{R}=0,74$ from the Burr distribution, by interpolation, the constants are $A_{SC}^U=1,48$ and $A_{SC}^L=1,14$ from Table 1. For the set of data, the sample skewness $=0,67$ and kurtosis $=0,64$ coefficients are estimated, the tables given in Burr (1942) may be used to obtain the mean and the standard deviation of the corresponding Burr distribution. Then the data set may be described by the Burr distribution with the parameters $c=3$ and $k=4$. Figure 3 shows the process for the Weibull

distribution, $n=5$. Also, the two different control limits are pointed out. The control limits of the individuals control chart based on the Shewhart method and skewness correction method are

$$\begin{aligned} I_{SC} \text{ Chart: } UCL_{SC} &= \bar{x} + A_{SC}^U \bar{R} = 15,69 \\ CL_{SC} &= \bar{x} = 14,59 \\ LCL_{SC} &= \bar{x} - A_{SC}^L \bar{R} = 13,75 \end{aligned}$$

$$\begin{aligned} I_S \text{ Chart: } UCL_S &= \bar{x} + 3\bar{R}/d_2 = 15,51 \\ CL_S &= \bar{x} = 14,59 \\ LCL_S &= \bar{x} - 3\bar{R}/d_2 = 13,64 \end{aligned}$$

Figure 3. Individuals chart for the Burr Distribution, $n=5$

At the second stage, it is estimated that for $n=7$ with $\bar{x}=14,646$, $\bar{R}=0,855$ from the Burr distribution, by interpolation, the constants are $A_{SC}^U = 1,21$ and $A_{SC}^L = 1,02$ from Table 1. For the data set, the sample skewness = 0,484 and kurtosis = 0,384 coefficients are estimated, the tables given in Burr (1942) may be used to obtain the mean and the standard deviation of the corresponding Burr distribution. Then, the data set may be described by the Burr distribution with the parameters $c=3$ and $k=6$. Figure 4 shows the process for the Weibull distribution, $n=7$. Also, the two different control

limits are pointed out. The control limits of the individuals control chart based on the Shewhart method and the skewness correction method are:

$$UCL_{SC} = \bar{x} + A_{SC}^U \bar{R} = 15,68$$

$$I_{SC} \text{ Chart: } CL_{SC} = \bar{x} = 14,64$$

$$LCL_{SC} = \bar{x} - A_{SC}^L \bar{R} = 13,77$$

$$UCL_S = \bar{x} + 3\bar{R}/d_2 = 15,59$$

$$I_S \text{ Chart: } CL_S = \bar{x} = 14,64$$

$$LCL_S = \bar{x} - 3\bar{R}/d_2 = 13,70$$

Figure 4. Individuals chart for the Burr Distribution, $n=7$

If a comparison is made between the control limits constructed by two methods, it may be stated that for $n=5$ and $n=7$, the range of control limits have expanded. Differing from the Weibull distribution's results for $n=7$, the Burr distribution's results are not satisfactory.

Conclusion

The development of the skewness correction method for the individuals control charts when the distribution of the data is non-normal has been illustrated. The function used here depends on Chan and Cui's (2003) model. The Burr and the Weibull distributions are applied to derive the control limits. An example of the biscuit thickness process is presented to illustrate the individuals control chart. For $n = 5$, the individuals control chart limits obtained by the skewness correction method are larger than the limits of Shewhart method. However, for $n = 7$, when the data set is Weibull distributed, there is no difference between the two limit's values. There is a difference when the distribution of the data set is Burr distributed. In other words, for the Burr distributed data for $n = 7$, the limits in question are expanded for both methods.

In this study, according to the control charts which are compared, when the distribution of the data set is Weibull, the individuals control chart limits calculated by the skewness correction method are close to individuals control chart limits calculated by the Shewart method for the larger values of sample

size n . Hence, if the data set is Weibull distributed, researchers may use individuals control chart with skewness correction instead of Shewhart individuals control chart.

References

- Bai, D. S. & Choi, I. S. (1995). $\bar{x}$ and R control charts for skewed populations. *Journal Of Quality Technology*, 27(2).
- Burr, W. I. (1967). The effect of non-normality on constants for $\bar{x}$ and R charts. *Industrial Quality Control*, 563.
- Burr, W. I. (1973). Parameters for a general system of distributions to match a grid of α_3 and α_4 . *Communications In Statistics*, 2(1), 1-21.
- Burr, W. I. (1942). Cumulative frequency distribution. *Annals Of Mathematical Statistics*, 13, 215-232.
- Chan, L. K., & Cui, H. J. (2003). *Skewness correction $\bar{x}$ and R charts for skewed distributions*. *Naval Research Logistics*. Wiley Periodicals. Inc, 50.
- Chou, C., Chen, C., & Liu, H. (2000). Economic-statistical design of $\bar{x}$ charts for non-normal data by considering quality loss. *Journal of Applied Statistics*, 27, 8.

Statistical Software Applications and Review Statistical Tests, Tests of Significance, and Tests of a Hypothesis Using Excel

David A. Heiser
Environmental Management
United States Air Force, Retired

Microsoft's spreadsheet program Excel has many statistical functions and routines. Over the years there have been criticisms about the inaccuracies of these functions and routines (see McCullough 1998, 1999). This article reviews some of these statistical methods used to test for differences between two samples. In practice, the analysis is done by a software program and often with the actual method used unknown. The user has to select the method and variations to be used, without full knowledge of just what calculations are used. Usually there is no convenient trace back to textbook explanations. This article describes the Excel algorithm and gives textbook related explanations to bolster Microsoft's Help explanations.

Key words: Excel, spreadsheets, statistical functions, hypothesis testing, t test

Introduction

Testing any commercial/academic statistically oriented computer program for correctness and accuracy runs directly into the questions, what is correctness and what is accuracy. Unfortunately, the answers are user dependent in the sense that each user has a different answer. The fact is that all commercial/academic software at sometime gives incorrect values, but that doesn't stop users from using it.

"There's a credibility gap: We don't know how much of the computer's answers to believe. Novice computer users solve this problem by implicitly trusting in the computer as an infallible authority; they tend to believe that all digits of a printed answer are significant. Disillusioned computer users have just the opposite approach; they are constantly afraid that their answers are almost meaningless" (Knuth 1998, p229).

David A. Heiser, B.S, University of Wisconsin, M. S., California Institute of Technology, both in Chemical Engineering. He maintains the web page on using Excel in statistics at <http://www.daheiser.info/excel/frontpage.html>. Email: dah_box1@innercite.com

The question is here, how much of Excel's computed output is believed to be correct and just what is correct?

The EXCEL Spreadsheet Program

Microsoft's Excel spreadsheet program is an inexpensive program for doing many kinds of calculations in business, engineering, and science. Excel has functions and data analysis routines for doing statistical calculations. There are many introductory statistics books that include instructions for solving problems using Excel. Excel also has basic chart and graph capabilities for displaying data and results.

Excel remains very popular, because it allows easy integration with Microsoft's Word and with Microsoft's Access (large business data bases). Results in the form of tables and charts can be easily integrated with Microsoft's PowerPoint presentation software. The pivot table feature as a means of analyzing data is a very popular feature.

Excel's capabilities are limited by the fact that it only does simple statistics. It does not include a lot of additional functions and routines that reflect current commonly used statistical procedures. It was programmed prior to 1992 and version 4.0 in 1994 was the first fully documented version (Excel 1992). It has had essentially no major improvements in statistical capabilities since then. Significant changes

(corrections and improvements) were made for the Excel 1997 and Excel 2003 versions, but the basic module remained the same.

The Computer Environment

It is important for people who deal with numerical computations to understand that the computer works only with a subset of real numbers {IR}. It is a special kind of mathematical object, a field. The computer software uses a different object {IF} to simulate {IR} objects. These objects are called floating point numbers. The object defined by {IF} is a finite subset of {IR}, it is not however, a field (nor any other object that mathematicians commonly define and study) (Gentle, 2004).

In computer software, addition and multiplication of {IF} objects are not associative. The summation in {IF} is not well defined, and usually is taken as a number when its value no longer changes. This no-further-change limit is referred to as being {IF}-convergent, which is different from {IR}-convergent. The harmonic series (sum of $1/i$) in {IR} is divergent, but in {IF}, it is {IF}-convergent. The {IF}-convergent value can be different, depending on how the internal algorithm does associations. The sum of integers is {IF}-convergent, because there is a limit on the size of integers that can be represented as {IF} objects (Gentle, 2004).

The Excel functions and routines handle numbers as the IEEE-754 64 bit standard floating point double precision number. The following are descriptions from KBA 78113:

“A floating-point number is stored in binary in three parts within a 65-bit range: the sign, the exponent, and the mantissa.

1 Sign Bit	11 Bit Exponent	1 Hidden Bit	52 Bit Mantissa
---------------	--------------------	-----------------	--------------------

The sign stores the sign of the number (positive or negative), the exponent stores the power of 2 to which the number is raised or lowered (the maximum/minimum power of 2 is +1,023 and -1,022), and the mantissa stores the actual number. The finite storage area for

the mantissa limits how close two adjacent floating point numbers can be (that is, the precision). (KBA 78113)

The mantissa and the exponent have fixed sizes. As a result, the amount of precision possible may vary depending on the size of the number (the mantissa) being manipulated. Whenever a computation is made (or a value input), the mantissa bits are moved left one at a time and the exponent bits are re-set until the left most bit is a one. Then one more shift is made, transforming this one-bit of information to the hidden bit. Zero bits are added on the right to fill out the 52-bit mantissa.” (KBA 78113)

An augmented mantissa of 53 bits corresponds to 15.7 decimal digits. Excel only displays the rounded 15 decimal digits.

“Every decimal integer can be exactly represented by a binary integer; however, this is not true for fractional numbers. In fact, every number that is irrational in base 10 will also be irrational in any system with a base smaller than 10.

For binary, in particular, only fractional numbers that can be represented in the form p/q , where q is an integer power of 2, can be expressed exactly, with a finite number of bits.

Even common decimal fractions, such as decimal 0.0001, cannot be represented exactly in binary. (0.0001 is a repeating binary fraction with a period of 104 bits).” (KBA 78113)

Errors occur during computer arithmetic {IF} operations.

Round off error.

Results when addition and subtraction are performed. Also occurs in multiplication and division when the sequences involve interchanges between internal 80 bit registers and external 64 bit memory storage. The Excel display also involves another round off.

Overflow and underflow.

Results when the sequence of instructions results in one of the intermediate values either exceeding 1.797693134862315E + 308 (fpmax) or being less than 4.940656458412465E-324 (fpmin). An error return does not always occur. Changing the associations will result in different results.

Quantizing error.

Results when the decimal number cannot be exactly represented by the IEEE-754 binary representation.

The IEEE-754 standard also has an 80-bit floating-point standard. This standard retains the same bit pattern as the 64-bit standard, but extends the mantissa (to the right) an additional 16 bits to a total of 68-bits. Microsoft uses the 80-bit standard for the machine registers that contain the floating-point numbers. At the machine level, computations are done using the 80-bit standard. If however in the sequence of instructions, one of these registers has to be stored in memory, the 80-bit number is rounded to the 64-bit standard and transferred to memory. A multiply-divide sequence that transfers intermediate values to memory will have a different result than one in which the intermediate values are held in the 80 bit floating-point registers. The issue on round-off errors comes from the conversion of the 80 bit number to a 64 bit number.

KBAs 42980, 78113, 145889, 125056 and 214118 are some good sources of information on the {IF} problem. McCullough (1998) also discussed this problem. Knuth (1998) presented the basic theoretical problems of accurately adding, subtraction, multiplying and dividing using floating point numbers as the {IF} object. Higham (1993) also found that there is no universal way to correct for addition (and subtraction) errors in long lists in floating point form.

Algorithms and Computer Programs

This is the area where the mathematics is converted into computer instructions. The general process is to take the mathematics (the equations) and to break the sequences into a series of computing blocks (i.e. subroutines).

Then for each of the subroutines, develop (or find in the literature) algorithms made up of fundamental arithmetic type operations (addition, subtraction, multiplication, division, etc) that will perform the desired computations. Subroutines will be written using a computer language such as Fortran, C++, or Visual Basic. The final step is then a conversion (compiling) to a sequence of binary machine instructions (i.e. Intel chip level).

Building a robust algorithm that always gives correct values is not an easy task. For example, take the simple computation of the standard deviation of a list of numbers.

$$\sigma = \sqrt{(\sum (x_i - x_{ave}) / (n-1))} \quad (1)$$

This computation would be done using the calculator formula

$$\sigma = \sqrt{\{ [(n \sum x_i^2) - (\sum x_i)^2] / [n(n-1)] \}} \quad (2)$$

with internal summation loops (Knuth, 1998, p 232). This will occasionally require a square root of a negative number, and the overall accuracy is poor. Excel 2000 and earlier versions used this calculator formula to calculate standard deviation values. Excel 2003 uses a two pass method, first calculating an average, then in the second pass calculating deviations from the average, a sum of squares of the deviations and then the standard deviation (KBAs 828888 and 826248). An improved algorithm is Welford's (1962), which is recommended by Knuth (1998). Knuth's form of the algorithm is provided below. Both the mean and the standard deviation are outputted values.

```

DIM Data X(1 to N) As Double
DIM M1, M2 ,S1, S2 as Double
DIM N, K As Integer
M1 = X(1)
S1 = 0
FOR K = 2 to N
M2 = M1 + (X(K)-M1) / CDBL(K)
S2 = S1 + (X(K)-M1) * (X(K) - M2)
M1=M2
S1=S2
NEXT K
AVERAGE = M2
STDEV = SQR(S2/ CDBL(N - 1) )

```

Note: CDBL converts integers to a floating point numbers

Use of the third algorithm substantially improves the accuracy of the result in Excel 2000, but only slightly in Excel 2003. Other statistical computer programs use other algorithms. Maechler (2005) chose West's modification of this algorithm. As he stated, "I'd conclude from Communications ACM, Vol 22, No. 9, page 531, that Welford's algorithm is a bit less accurate than the (very similar) 'West' version, and we (the R developers) should rather implement the latter."

Algorithms sometimes show strange results for an unusual set of input values. For example, enter three identical values, 1E+30, 1E+30 and 1E+30 into Excel cells and do a STDEV function on this range. The result is 1.72368E+14, not zero as expected. Also, do a VAR on this range and 2.97106E+28 will appear.

This raises an important issue. When input of parameter values from one narrow, unusual region of input parameter space results in a wrong output, does one conclude that the computer program should never be used because it returns wrong values?

The Display Of The Result

Within the computer program there are internal subroutines that convert the binary floating point word (64 bits) to a string of ASCII characters (text) which are displayed/printed. The user can (in Excel) chose how the text is formatted as to text type, size, bold, italic, floating point or fixed point and the number of decimals to the right of the decimal point. In Excel there is a default set (Arial, 10, regular), a default cell width of 8.43 points, and the default General format. For numbers from 1 to 0.0001, the General display will show 6 decimal digits. Below 0.0001, a floating point display of 3 digits (plus exponent) will be displayed.

There have been articles published criticizing the accuracy of computer software based solely on the default display (e.g., Altman 2002, Hilbe, 2002, McCullough, 1998, 1999, McCullough & Wilson, 1999, 2000, 2004; Knüsel, 1998, 2003).

Methodology

McCullough (1998, 1999) pioneered some of the basic methods of conducting tests on software. He used the NIST suite of data-bases with known statistics to test several software programs. His two articles are good background and methodology sources.

Testing methods

Any testing of statistical software programs involves the exercise of selection to get down to the area or routines to be tested. With respect to Excel these are functions and data analysis routines. For other programs, there may be all kinds of decision trees and selections to arrive at the test objective or method to be tested.

What is the function/routine actually doing? In most cases, the developer says very little regarding the specifics of what the program does, but a great deal is said on marketing (selling) how good and comprehensive is the program. For proprietary reasons, of course, very little should be said. For that reason, some testing has to be done to find out just exactly what is being calculated, how to get as many digits as possible, and to find some boundaries on the ranges of input parameters. This is exploratory testing.

The next level is accuracy testing. For accuracy testing the software will require a test database and a parameter and selection vector. In some cases only a test database is needed and in some others such as the distribution functions, only a parameter vector is needed. In all cases there has to be an output vector that can be compared to a reference standard vector, such that a difference can be obtained as a measure of the accuracy of the method. In the case of Excel functions, this output vector has only one value (the exception is the array functions that output a range, matrix or a table of values). The Excel Data Analysis routines also may output a table, which is the output vector formatted to be readable.

Standard values of summary statistics from a data set may come from several sources.

1. Theoretical values manually calculated or selected (by theory) that are valid accurate reference values. For example one can construct a list of data values that has a theoretical precise mean and a precise standard deviation. (Method: A).
2. Values calculated by an external software program, chosen to be the reference (Method: B).
3. Data and values published as part of a standard. (Method C).
4. Comparing the results from many different software programs on the same data set and deciding on “correctness” (Method D). Altman and McDonald (2000).

The NIST Tests

The National Institute of Technology (NIST) established datasets for software tests, the StRD series (NIST nd).

“For all datasets multiple precision calculations (accurate to 500 digits) were made using the post-processor and FORTRAN subroutine package of Bailey (1995, available from NETLIB). Data were read in exactly as multiple precision numbers and all calculations were made with this very high precision. The results were output in multiple precision, and only then rounded (without error) to fifteen significant digits. These multiple precision results are an idealization. They represent what would be achieved if calculations were made without round-off or other errors. Any typical numerical algorithm (i.e. not implemented in multiple precision) will introduce round-off error, and will produce results that differ slightly from these certified values.” (NIST, nd)

The NIST data sets covered univariate analysis, linear regression, non-linear equation fitting, ANOVA and correlations. This has been the essential test method (method C) to test Excel. McCullough (1998, 1999) pioneered the basic method of conducting tests on software using the

NIST test sets. McCullough and Wilson (1999, 2000, 2004) also presented a series of papers on tests made on Excel using the NIST and other test data .

Other Previous Excel Tests

Some of the early testing (Excel 1995) was done by the Center for Information Systems Engineering, (Britain) in 1999 (CISE 27/99). They used the IMSL Fortran 90 Math/Library (version 3.0) provided by the Digital Equipment Corporation to do testing (Method B).

A number of email messages, web site reports (papers), and discussions on the newsgroups and on the statistical lists (since 1998) described tests on some of the Excel functions and routines. These included cases where a particular (real) data set, when analyzed using Excel, gave results different from some other software package. Most of these were casual tests, based on a particular data set.

Significance Test Methods

The NIST data sets and their computed statistics were not useable on the family of significance tests in Excel. NIST did not provide paired or dual data sets for testing significance test functions/routines. The literature does not report on specific testing of Excel significance test functions and routines. Therefore, test data sets for testing the Excel family of significance tests had to be built, and ways to arrive at accurate statistical values found

Because the outputs from some of the significance tests are p values, a set of Visual Basic statistical distribution functions provided by Smith (2002) were used to calculate accurate reference p values. The Excel distribution functions are not accurate enough to be used to obtain accurate p values.

Two approaches were taken, one of exploratory testing to identify just what the function was returning (e.g., the proper tail area). The other was to do accuracy testing. This required the development of more extensive data sets to stress the functions/routines.

The NIST approach was to use several types of test data sets. One of these types was to build patterned data tables of data. A patterned number can be considered as having a whole number part and a fractional part where the

numbers to the right of the decimal point is the fractional part. A patterned data table has patterned numbers all with the same whole number, but with different fractional values. For the NIST SmLs01 to SmLs09 data sets, the fractional part had specific alternating values (0.3 and 0.5 or 0.2 and 0.4), and then with one odd value for each set, gave a data set with theoretical, precise means, variances and standard deviation values. By increasing the magnitude of the whole number from 1 to 1E+09, and by changing the size of the set, the overflow effect on floating point number computations and algorithms could be determined.

The NIST approach to the SmLs sets suggested ways to build test data sets with accurate statistics to test the Excel family of significance tests. The theory behind it comes from the basic way numbers are represented in Excel.

In terms of floating point numbers, a larger whole number part of the patterned number pushes the mantissa bits (these are on a number base of 2, not on a number base of 10) off the right end, characteristic of overflow. This overflow of floating point numbers is one of the causes of errors. However, there are other causes of errors that are not brought out by the use of patterned numbers, and other methods have to be used. Good algorithms are those that minimize the overflow effect. The charts in Heiser (2005) show the loss of accuracy of many Excel functions due to this type of overflow.

Measures Of Accuracy - Log Relative Error (LRE)

The measure of the accuracy of the information from a computed value is by a calculation called Log Relative Error or LRE. This was introduced by McCullough in his 1998 paper. The LRE value represents a measure of how many significant (accurate) digits (decimal) there are in the output parameter values.

$$\text{LRE} = -\text{LOG}_{10} (\text{abs} (\text{CV} - \text{RV}) / \text{RV}))$$

CV is the computed value and RV is the reference or true value. LRE values vary from 0 to 15 on the McCullough scale. 15 can be considered as an exact match.

LRE values from the statistical distributions present problems, because of the 9's problem. Here, a leading sequence of 9's really are leading zeros, and should not be considered as significant digits, but mathematically they are. Excel computes p values above 0.5 as 1 minus the corresponding below 0.5 value, for all symmetric probability distributions. Consequently, p values above 0.5 have uncertain accuracies, depending on the user's view. Smith's (2002) distribution functions calculate p and q values by separate algorithms.

The LRE values approximate the number of accurate digits in the Excel cell value, independent of how it is displayed. For the floating point form, (select Format→ Cells→ Number→ Scientific→ Decimal Places→ 14) it approximately represents the number of accurate digits, including the digits to the left of the decimal point, and the digits to the right of the decimal point.

Results of Tests

This study examined the errors from the Excel VAR algorithm and Welford's algorithm on a patterned data set. In this case, two sets of random fractional numbers, one uniform u(0-1) and the other normal n(0,1) with 1001 values of each set were generated in a column (Please note that for all test data sets with random numbers, Marsaglia's MWC256 RNG, Marsaglia (1995, 2002) was used. For random normal, Smith's (2002) precise inverse normal function was used). The variance value of the base case from either of the two functions was the identical. Whole number sets (from 1 to 1E+15) were added forming 15 additional columns. Variances from each function were then calculated. Figure 1 shows the result.

Given the nature of the input data and the basic structure of a patterned number in terms of the decimal system, the data from a good algorithm should closely follow a straight line from 16 on the y axis to 16 on the x axis.

The Excel 2003 algorithm, although an exact algorithm, shows some unexpected behavior in the region below an exponent of 8. This behavior generally occurs also for other Excel functions when the whole number is less than 1E+08. The inaccuracies at the right end

are expected. Welford's algorithm in general is close to the expected line and shows consistent behavior, typical of a good algorithm.

Figure 1: Comparison of Algorithms

The Excel Significance Test Functions And Routines

Excel 2003 provides 80 direct functions and 19 Data Analysis routines that can be used in statistical data analysis. Only a part of the available functions and routines are directly applicable to tests of significance and hypothesis testing. The functions and routines useful for significance testing are:

CHITEST - This is a Chi-square Goodness-of-Fit test for grouped data. It does not support general Chi Square tests on variances. The test will only work on 2 way contingency tables. The test cannot be applied to single lists of observed and expected values. The first input, actual range is the range of the observed values, as a 2-way contingency table. The second input is expected range, the range of a separate contingency table giving the expected values.

FTEST - Returns the one-tailed probability value of an F test on two separate ranges of data. The ranges may be of different lengths.

TTEST - Returns the probability value of a t test on two separate data sets. Function allows for 1 or 2 tail tests, paired data and equal-unequal variances. The function has two parts internally, one to calculate a t value from the two separate data sets, and the other to calculate internally a p value from the t value.

ZTEST - Returns the two-tailed probability of a normal distribution z test on a range of data with respect to a known population mean and standard deviation. If the standard deviation field is left blank, the routine used the standard deviation of the data. The function has three parts internally:

- 1 To calculate a mean value (and a standard deviation) from the input data set.
- 2 To calculate $z = [(\text{input mean value}) - (\text{data set mean})] / [(\text{data set standard deviation or input standard deviation}) / \text{Square Root} (\text{size of the data set, } n)]$.
- 3 To calculate a p value from **NORMSDIST(z)**.

All of the other Excel functions can be used to build up intermediate values for significance test inputs. They can also be used along with new VBA functions and subroutines to build new significance tests beyond the limited capability of Excel.

Data Analysis Routines

These are routines called by selecting the *Tools* menu and then selecting *Data Analysis* and then selecting one of the listed routines.

F-Test Two-Sample for Variances

t-Test: Paired Two Sample for Means

t-Test: Two-Sample Assuming Equal Variances

t-Test: Two-Sample Assuming Unequal Variances

z-Test: Two-Sample for Means:

After inputting the requested data, they return a table.

Tests on the Accuracies of Functions and Data Analysis Routines

The CHIDIST, FTEST and TTEST functions were tested. There were differences found between the results of these tests for Excel 2000 and Excel 2003. The Excel 2000 tests show relatively low LRE values. As explained by Microsoft in KBA 828888, the problem was the low accuracy of the VAR and STDEV functions that were used inside the routines. Rather than take up a great deal of space to show both 2000 and 2003 outputs, only the Excel 2003 values are shown in the following tables.

There were 4 data sets used for testing as follows:

Set 1 (columns A and B) represented paired data, integers with blank spaces.

Set 2 (columns C and D) represented unequal length data from two different populations. Integers.

Set 3 (columns E and F) represented patterned data of two samples from one population with equal sample sizes. The whole number was 1000, and the fractional numbers were uniformly distributed (0-1) random numbers.

Set 4 (columns G and H) represented a variable length set (up to 2000). The first column represented the control data set, and the second column represented the treatment data set. The base case was where the numbers in both columns were all random normal (0,1) z values from one population. Whole numbers were added as described previously.

Testing The Difference Between Variances

CHITEST

Tests indicated that the Excel algorithm in the CHITEST function is the correct one. Errors occur from errors in the inputted expected values table and in the CHIDIST function. CHITEST returns correct values if the Expected Values table is correct.

FTEST

The function description (Excel, 1992) suggests that the FTEST function just computes the ratio of two variances where the variances come from the VAR function. Neither Excel Help nor the KBAs provide any additional information. The VAR function holds up well against overload as shown in figure 1, but does introduce some error.

Given the ratio, the FINV function then was used to arrive at a p value. The F distribution FINV generally has p value accuracies above an LRE value of 8, over the entire range of input parameters (see Heiser, 2005) for specific details. The output then of FTEST should be an accurate p value with at least 8 accurate decimal digits. The actual output for data set 2 indicates that FTEST returns wrong values.

Table 1: FTEST Function Response

Cell Entry	=FTEST(C,D)
Returned Value	0.9425381810184540
Correct Value	0.481410961628470

FTEST outputs an incorrect p value, corresponding to a two-tail test. The problem is with Microsoft.

In Excel (1992), the function description says, "Returns the results of an F-test. An F-test returns the one-tailed probability that the variances in array 1 and array 2 are not significantly different".

In Excel Help (2006), "Returns the result of an F-test. An F-test returns the two-tailed probability that the variances in array1 and array2 are not significantly different. Use this function to determine whether two samples have different variances."

The standard for the F test on a ratio of variances is the one tailed test. It is a test on all values of the ratio from 0 to the critical value. On this basis, the only valid test is the one-tailed test. The workaround here is to always divide the FTEST p value by 2 to get the correct q value of the right tail. This has been reported before.

Test On The Data Analysis F-Test: Two-Sample For Variances:

Here Excel returns an accurate value.

Table 2: Excel Data Analysis Routine Output, Actual Excel Output

F-Test Two-Sample for Variances		
	C	D
Mean	1000.503767	1000.696727
Variance	0.092055155	0.090461689
Observations	30	30
Df	29	29
F	1.017614821	
P(F<=f) one-tail	0.481410962	
F Critical one-tail	1.860811434	

The true p value is 0.481410961628470. The Data Analysis F test on two variances gives the correct p value (excluding the argument on the correctness of all displayed digits). Differences are only due to the inaccuracies in FINV (the function that uses the df and F ratio values to arrive at a p value)

Testing the Difference Between Mean Values the Basic Problems and Solutions

There are three possible situations or problems here with tests on the differences in means.

(1) Dependent, Paired values,

(2) Independent, Two sample sets, each coming from different (or the same) population with possible differences in means, but both populations having the same variance

(3) Independent, Two sample sets, each coming from different populations with different variances (The Fisher-Behrens problem).

These are the three classical situations, which require different test methods.

Excel provides a function (TTEST) with 3 options and three Data Analysis routines for statistical solutions for the basic problem. The questions here are just what do these functions and routines do, and do they compute the statistics correctly in terms of theory, and are the results numerically accurate. Other concerns include: how robust are they on non-normal data, how stable are the results in terms of type I error rates, and what the power is.

In traditional statistics, the three possible situations are considered as separate, important classical problems for analysis in introductory statistics. In introductory statistics, the assumption of normality is made, and this results in a simplification of the statistical tests. The test is usually put in terms of a test of a hypothesis. The discussion below is based on the traditional tests using the t distribution and the assumption of normality.

The paired values (or dependent data values) solution problem (1) is straightforward, and is given in textbooks. The test is to determine if the sum of the differences between each pair is zero or is some preset difference, depending on the hypothesis made.

For problems (2) and (3), the test is on the differences of the means, using a joint measure of variation from both samples. Problem (2) where the variance does not change and approximately equal sample sizes are involved has a very robust t test solution under sample departures from normality. However, if the variances are not truly equal, and substantially different sample sizes are involved, the normal t test solution loses its robustness and the true alpha may be quite different from the selected alpha. The third problem is the Behrens-Fisher problem, which does not have a direct theoretical solution.

The Behrens-Fisher Problem

For the Behrens-Fisher problem, there is no uniformly most powerful (unbiased) test for

all sample sizes. There are several approximations found in textbooks and in the literature, and this complicates the assessing of Excel's accuracy on problem (3). This impacts the decision to fault Excel or not. Sawilowsky (2002) is an excellent review of the attempts to come up with more exact solutions since 1929.

Fisher's Solution Of The Behren's Problem

"For samples from a single population, the effect of eliminating the unknown variance σ^2 , by Student's method, on the distribution of the error of the mean, is to replace, in the specification of this error,

$$\sigma * x / \sqrt{N}$$

where x is normally distributed with unit variance, but σ is unknown by

$$s * t / \sqrt{N}$$

where t is distributed in Student's distribution, for the appropriate number of degrees of freedom N , and s is the estimate of σ available from N degrees of freedom."

For two samples from populations having a common mean, the deviations will be independent, and the data will supply values s_1 , based on n_1 degrees of freedom, and s_2 based on n_2 . The difference between the observed means is the sum (or difference) of the two deviations from the true mean, so that on the null hypothesis considered, namely that the two populations means are equal, we have

$$x_1 - x_2 = (s_1 * t_1 / \sqrt{n_1}) - (s_2 * t_2 / \sqrt{n_2})$$

where t_1 and t_2 are distributed independently in the two distributions.

If the frequency is small, such as 1%, that the expression on the right, which has a known distribution, for the observed values s_1 and s_2 , shall exceed

the observed difference in the sample means, this difference may be judged significant." (Fisher 1973, p. 98).

This is the same as the confidence interval method described in Schenker and Gentleman (2001), where the s values are population values and the t values are z values. Fisher's method does not lend itself to a direct solution and is not referred to as a solution in the literature.

The Welch-Aspin-Satterthwaite Solution

The Welch-Aspin-Satterthwaite solution is a solution to the Behrens-Fisher problem. It evolved over the years from Satterthwaite's ideas in 1941 to Welch's ideas in 1937 -1949, with Aspin's inputs during 1948-1949. It is commonly referred to as the Aspin-Welch test or the Welch test in research papers. However, some statistics textbooks authors (i.e. Moore & McCabe, 2003) ignore this and use "pooled df for this test, also known as the computer solution". There is no consistency in the literature between the names or terms used and which of six computational methods it applies to.

One of the inherent problems with the Welch-Aspin-Satterthwaite approximate solution is that it is not robust to departures from normality. Sawilowsky (2002), stated, "I would be remiss if I failed to note that numerous Monte Carlo studies have shown that the nonparametric Wilcoxon Rank Sum test can be three to four times more powerful in detecting differences in location parameters when the normality assumption is violated....Therefore the Wilcoxon procedure should be the test of choice".

However this does not resolve the fundamental problem as to whether the difference should be determined based on the medians (Wilcoxon) or on the means (Welch). The predominate applications in psychology and related behavioral research are based on the difference in means, the standard error of the means and on effect size. There is little concern about non-normality and equality of variances. Effect sizes are more important than p values. (Sprinthall, 2000, Kline, 2004)

Excel does not provide the Wilcoxon Rank Sum Test, which can be considered a fault in Excel. For samples that have large differences in sample sizes and have asymmetry, the Balkin and Mallows (2001) approach should be considered.

The Six Solutions

The range of possible solutions to the three situations identified above has to be limited specifically to what Excel has provided. Within the context of what was discussed above, there are 6 possible solutions to the Behrens-Fisher problem.

In general, the p value (compared to the alpha value) comes from the t distribution, and therefore for each problem, a df value and a t value has to be calculated. A decision also has to be made, on whether a single tail or a two-tail test is required.

Methods to obtain a t value from the difference between the two means are listed in table 3. There are others such as the Score statistic that are not considered here, because they are not found in or introduced in introductory statistics textbooks.

Table 3: Combined Variance Measures

t-Value Method	Term 1	Term 2	Common names
1	var_1 / n_1	var_2 / n_2	Un-equal variances
2	$\text{var}_{\text{pooled}} / n_1$	$\text{var}_{\text{pooled}} / n_2$	Equal variances, pooled t test
3	$\text{var}_1 / (2n_1 + 1)$	$\text{var}_2 / (2n_2 + 1)$	Fisher's 1939 form
4	$\text{var}_1 * (n_1 - 1) / (n_1^2 - 3n_1)$	$\text{var}_2 * (n_2 - 1) / (n_2^2 - 3n_2)$	Fenstad's Statistic
5	$\text{var}_1 * (n_1 - 1) / n_1^2$	$\text{var}_2 * (n_2 - 1) / n_2^2$	Wald Statistic

The pooled variance in method 2 is:

var_1 = Variance of Sample 1

var_2 = Variance of Sample 2

$\text{var}_{\text{pooled}} = ((n_1 - 1) * \text{var}_1 + (n_2 - 1) * \text{var}_2) / (n_1 + n_2 - 2)$

The combining of terms 1 and 2 for a t value are as follows:

Methods 1-4:

t value = Difference in Means / Square Root (Term 1 + Term 2)

Method 5:

t value = (Differences in Means)² / (Term 1 + Term 2)

Currently, only t-value methods 1 and 2 are considered. The different degrees of freedom used are given in table 4.

Table 4: Degrees-of-Freedom Values Used In The Tests

df-Method	Used on Problems	df value used to obtain the t distribution p value
1	(1)	= n-1
2	(2)	= $n_1 + n_2 - 2$
3	(3)	= Smaller of either $n_1 - 1$ or $n_2 - 1$
4	(3)	= Welch df

This then gives six ways to calculate a p value, as shown in table 5.

Table 5: The Six Calculation Combinations for Problems (2) and (3)

Calculation	df Method	t Value method
1	2	1
2	3	1
3	4	1
4	2	2
5	3	2
6	4	2

Calculation 3 is generally referred to as the Welch test.

The maximum power here for any of the calculation methods is at sample sizes related to the ratio of the known variances of the samples.

κ = variance population 2 / variance population 1

Where the optimum sample sizes (n_1 and n_2) come from the following equation:

$$n1 / (n1+n2) = 1 / (1 + \sqrt{\kappa})$$

However, the local optimal design is sensitive to the misspecification of the κ value. (Dette & O'Brien, 2004)

The Welch df Value (df method 4)

$$u_1 = (s_1^2 * s_1) / n_1$$

$$u_2 = (s_2^2 * s_2) / n_2$$

$$df = (u_1 + u_2)^2 / [(u_1^2 / (n_1 - 1)) + (u_2^2 / (n_2 - 1))]$$

There are other textbooks and statistics course handouts that give a different formula and also may call it by other names.

Figure 2 shows how the Welch df value varies as the ratio of the variances varies. It is the factor that when multiplied by the df value of one sample (i.e. $n1-1$) gives the Welch df value.

Figure 2: Welch df Factor Changes

Values between 0 and 1 are a mirror image of the values from 1 to infinity, with the x axis values the reciprocals of the x axis values greater than 1. When the variance ratio is 1, the pooled df value is equal to the df method 2 value. As the ratio increases, the pooled df value becomes asymptotic to the df method 3 value.

For example, given equal samples of 30, the F test would probably indicate that variance ratios greater than 2, would indicate a high probability of the variances being unequal. One could conclude then a factor of 1 would be appropriate. However, the Welch-Aspin-Satterthwaite df gives a more conservative estimate that in a sense, compensates for the fact that it is not truly known that the variances are equal.

There have been many articles over the years that point out that if the F test is used to decide on equal/unequal variances at an alpha level, and then do the t test at the same alpha level, there is a subsequent loss of control of the Type I and Type II error rates (e.g., Sawilowsky, 2002).

There are three views regarding the actual Welch df value to be input to the t distribution. The calculated Welch df value is not an integer. The options are to truncate the computed df value to an integer, round to an integer, or interpolate (in tables) to obtain a value for a fractional df. Moore and McCabe (2003) recommended that interpolation be used when only tables are available. Most software routines that calculate the t distribution p value require that the df value be an integer, although the basic computing algorithm will take fractional df values. Excel's t distribution functions will only allow integer df values to be entered.

The Common Textbook Df Value

For unequal variance problems, df method 3 corresponding to calculation 2 is usually given. This results in a conflict here, because Excel follows df method 4.

Best and Rayner (1987) identified four t statistic measures that should be considered:

- (V) The common statistic:
- (W) The Wald statistic:
- (L) The likelihood statistic:
- (S) The score statistic:

The common statistic is (V) which corresponds to calculation 3. Best and Rayner (1987) defined the other three (W, L and S), but concluded that for their $n_1=4$ and $n_2=8$ sample sets (from Monte Carlo sets), the power of the test for differences was about the same.

Best and Rayner (1987) defined calculation method 3 as the V statistic. They found that calculation 3 gives results that closely follow the preset alpha value, whereas calculation 4 results vary considerably from the designated alpha value when the population variance ratio departs from 1. The V statistic was their choice, because it can be used for both tests involving equal and unequal variances.

Some Textbook Directions

Moore and McCabe (2003) suggested the use of calculation 4 (e.g., Table 6) for equal variances and calculation 2 or 3 for unequal variances. Calculation 3 is preferred for unequal variances. Larson and Farber (2003) said to use calculation 4 for equal variances and calculation 2 for unequal variances. Triola (2001) said to use calculation 4 for equal variances and calculation 2 for unequal variances. Lind, Marchal, and Mason (2001) said to use calculation 4 for equal variance (they do not say anything about unequal variances). Pelosi and Sandiffer (2000) said to use calculation 4 for equal variances and calculation 3 for unequal variances. Levine, Berenson, and Stephen (1999) said to use calculation 4 for equal variances. Unequal variances were not covered. In Sprinthall (2000) the equal/not equal variance issue is never brought up. The standard error of the difference in means is from calculation 4.

The general consensus among textbook authors is to use calculation 4 for equal variances, because it is based on accepted practice. Calculation 2 is more frequently recommended than calculation 3 for unequal variances. In some textbooks, the distinction between equal and unequal variance is not made and calculation 1 is given for all tests on two means from independent samples. This suggests there is a wide range of practices, all derived from whatever was said in the textbook used in the course.

In doing the calculations for tables covering the six methods, calculation method 2

gives higher p values than calculation method 3. Consequently using textbook recommendations may not be the best solution method. They also may lead to false claims about the accuracy of Excel's TTEST and Data Analysis routines.

The Best Approach

In applied studies and research, the current view is that the real problem is that both a shift in location and a change in scale occur simultaneously when a treatment is applied. Consequently both a change in the mean and a change in variance occur. The occurrence of a change in variance without a change in means or a change in means without a change in variance is very rare (Sawilowsky 2002). The third problem then is the main view when dealing with real data.

If the assumption of normality is valid, then the best method is the V test or calculation 3 for all tests on the difference in means, regardless if the variances are equal or unequal. If the test is not a zero difference, but a test on a predetermined (theory) difference (d), then the non-central distribution has to be used rather than the central t distribution. (Steiger & Fouladi, 1997) Excel only has the central t distribution, and therefore Excel cannot be used to test for d.

Computed Reference Values

Computed reference values from each of the six methods for each of the four reference data sets were calculated as a means of finding out which of the methods are used in Excel.

Significance Test Functions and Routines.

The Excel TDIST function (which uses the BETADIST function to derive p values) appears to be used in all cases. There are errors in BETADIST that carry over to the problem solution. An analysis of these function errors and inaccuracies are in Heiser (2005).

Fisher's Solution

Fisher's equation obviously represents confidence intervals, but the signs are a problem. If the left hand sign is taken as a plus (adding two confidence intervals), it is possible to obtain a p value, when the sum of the confidence intervals equals the difference in means. The

theoretical problem is, should the non-central t distributions be used (see Steiger & Fouladi, 1997).

Tests on the TTEST Function.

The TTEST function has three options corresponding to the three possible solutions to differences in means, as discussed above.

Option 1: Paired Sample

The literature has commented on this test, primarily on its failure to give correct values when a BLANK occurs in a cell (indicating a missing value). Both the Excel 2000 and 2003 versions have this problem. Therefore, it is important to never have blank cells in the input ranges.

KBA 829252 describes the odd behavior of TTEST when there is missing data.

When there is no missing data, TTEST returns correct values. The algorithm is correct. The main contributor to errors is the inaccuracy in the BETADIST function that is used to obtain t distribution p values.

Option 2, Two-Sample Equal Variance

Returns correct values. The algorithm is correct. The main contributor to errors is the inaccuracy in the BETADIST function that is used to obtain t distribution p values.

Option 3, Two-Sample Unequal Variance.

Uses calculation 3, the Welch-Aspin-Satterthwaite solution to the Fisher-Behrens problem. Returns correct values. The main contributor to errors is the inaccuracy in the BETADIST function that is used to obtain t distribution p values.

Comments on TTEST

The primary source of errors in TTEST is that from BETADIST. The algorithms used for this function are poor, and consequently often give inaccurate results (see Heiser, 2005). It is not unusual to get LRE values down to 4 with actual data from TTEST because of this problem. Tests on different test data sets generally return p values with LRE values in the 7-10 range. Microsoft has no plans to fix BETADIST, so the use of TTEST will always have this uncertainty.

A fix for this problem is to download an accurate beta distribution function as a vba module or as an *.xla addin from another source. Most t distributions in add-ins or modules are blocked against non-integer df values. Non-integer df values are required for the Welch solution. The relationship (from Abramowitz and Stegun, 1963, eq. 26.5.27) is

$$1 - \text{tdist}(t, df) = \text{Beta}(X, A, B)$$

where

$$\begin{aligned} A &= df / 2 \\ B &= 1 / 2 \\ X &= df / (df + t * t) \end{aligned}$$

Beta is the cumulative or incomplete beta function

You will have to modify the left hand side to get the correct one or two tailed probabilities. Smith (2002) has accurate beta distribution vba functions. However, his t distribution functions are restricted to integer df values.

Tests on The Data Analysis Routines

These are routines from the Tools → Data Analysis menu:

t-Test: Paired Two Sample for Means

t-Test: Two-Sample Assuming Equal Variances

t-Test: Two-Sample Assuming Unequal Variances

z-Test: Two-Sample for Means:

They are programmed macros written prior to Excel 4, are not in vba and never have been fixed. Microsoft has issued KBA's on the problems with these macro's but has never fixed the problems.

A consistent error with all four of these routines is the output table where the cells:

P(T<=t) one-tail
t Critical one-tail
P(T<=t) two-tail
t Critical two-tail

appear. The T and t relationships in the first and third cell are usually wrong in terms of the

values given in the cells to the right of this group. KBA 829252 talks about one instance of this problem.

The $P(t \leq T)$ two-tail statement is in error. It should be $P(T = t)$ two-tail. A two tailed test in regard to a hypothesis is a test on a null hypothesis of equality. The alternate hypothesis is $T \neq t$. Microsoft is wrong in their help narrative.

t-Test: Paired Two Sample for Means

The Data Analysis Macro uses the TTEST function with option 1, and as a result, the p values from the two are the same.

However there is a difference when blank cells occur. KBA 829252 describes what happens. "First, this Analysis ToolPak tool counts the number of subjects with Before measurements and the number of subjects with After measurements. If these totals are different, you receive an error message and this Analysis ToolPak tool does not continue." Therefore, this routine should not be used when there are missing values in the data.

t-Test: Two-Sample Assuming Equal Variances

The Data Analysis Macro uses the TTEST function with option 2, and as a result, the p values from the two are the same.

t-Test: Two-Sample Assuming Unequal Variances

The Data Analysis Macro uses the TTEST function with option 3. However, there is a difference here. The macro takes the computed Welch df value and converts it by rounding to an integer. This integer value then goes into BETADIST and comes out with a p value different from that coming out of TTEST, option 3. The p values are different here, so in a sense, the Data Analysis macro is in error. For correct Welch p values, fractional df values must be retained. Therefore, this macro gives inaccurate results.

z-Test: Two-Sample for Means

This routine uses the normal distribution function NORMDIST which has LRU accuracies of 7 or more in the z range of -3 to -5, and 12 or more outside of this range. Some tests indicated no algorithm problems, and

output p value accuracies corresponding to the NORMSDIST accuracies. However again the $P(t \leq T)$ two tail statement in the table is in error. It should be $P(T \neq t)$ two-tail.

Excel 2007 (Formerly Excel 12)

This is the new version that will be available in 2007 to work with Windows Vista. Microsoft has made no changes to Excel 2003 in the statistics area for the 2007 version (Gainer, 2006)

References

- Abramowitz, M. & Stegun, I. A. (1964). Handbook of mathematical functions with formulas, graphs, and mathematical tables. *National Bureau of Standards, Applied Mathematics*, Series 55.
- Altman, M. (2002). A review of JMP 4.03 with special attention to its numerical accuracy. *The American Statistician*, 56(1).
- Altman, M. & McDonald, M. P. (2000). *The robustness of statistical abstractions: A look under the hood of statistical models and software*. http://data.fas.harvard.edu/numerical_stability/g3009.pdf
- Balkin, S. D. & Mallows, C. L. (2001). An adjusted, asymmetric two-sample t test. *The American Statistician*, 55(3), 203.
- Best, D. J. & Rayner, J. C. W. (1987). Welch's approximate solution for the Behrens-Fisher problem. *Technometrics*, 29(2), 205-210.
- Cook, H. R, Cox, M. G., Dainton, M. P., & Harris, P. M. (1999). *Testing of spreadsheets and other packages used In metrology, testing the intrinsic functions of Excel*. Center for Information Systems Engineering, NPL Report CISE 27/99.
- Dette, H. & O'Brien, T. E. (2004). Efficient experimental design for the Behrens-Fisher problem with application to bioassay. *The American Statistician*, 58(2), 138.
- Microsoft Excel function reference for Excel version 4.0 (1992). Manual published by Microsoft Corporation.
- Gainer, D. (2006). Email message from Gainer, D, Group Program Manager for Microsoft Excel. See <http://blogs.msdn.com/excel/archive/2005/09/23/473185.aspx>

- Heiser, D. A. (2005). Microsoft Excel 2000 and 2003 faults, problems, workarounds and fixes. <http://www.daheiser.info/excel/frontpage.html>.
- Fisher, R. A. (1973b). *Statistical methods and scientific inference* (3rd Ed). New York, N.Y.: Hafner Press.
- Gentle, J. E. (2004). Courses in statistical computing and computational statistics. *The American Statistician*, 58(1), 2-5.
- Higham, N. J. (1993). *The accuracy of floating point summation*. Department of Mathematics, University of Manchester Press.
- Hilbe, J. B. (2002). Section editor's notes. *The American Statistician*, 56(2), 148.
- Microsoft knowledge base articles. <http://support.microsoft.com/kb/.....the numberen-us>
- Kline, R. B. (2004). Beyond significance testing, reforming data analysis methods in behavioral research. Washington, D.C.: American Psychological Association.
- Knüsel, L. (1998). On the accuracy of statistical distributions in Microsoft Excel 97. *Computational Statistics and Data Analysis*, 26, 375-377.
- Knüsel, L. (2003). *Numerical accuracy of distributions in statistical packages*. Paper presented at the International Statistical Institute in Helsinki (1999), 147-148.
- Knuth, D. E. (1998). *The art of computer programming (vol 2.). Semi-numerical algorithms* (3rd Ed.). Addison-Wesley
- Larson, R. & Farber, B. (2003). *Elementary statistics: Picturing the world* (2nd Ed.). Upper Saddle River, NJ: Prentice-Hall.
- Levine, D. M., Berenson, M. L., & Stephan, D. (1999). *Statistics for managers using Microsoft Excel* (2nd Ed.). Upper Saddle River, NJ: Prentice-Hall.
- Lind, D. A., Marchal, W. C., & Mason, R. D. (2002). *Statistical techniques in business and economics* (11th Ed.). New York, N.Y.: McGraw-Hill-Irwin.
- Maechler, M. (2005). Welford's algorithm. Email: Seminar fuer Statistik, ETH-Zentrum. Zurich, Switzerland.
- Marsaglia, G. (1995). *The Marsaglia random number CD-ROM with the diehard battery of tests of randomness*. Available at www.stat.fsu.edu/diehard.
- Marsaglia, G. (2003). Random number generators. *Journal of Modern Applied Statistical Methods*, 2(1), 2-13.
- McCullough, B. D. (1998). Assessing the reliability of statistical software: Part 1. *The American Statistician*, 52(4), 358-366.
- McCullough, B. D. (1999). Assessing the reliability of statistical software: Part 2. *The American Statistician*, 53(2), 149.
- McCullough, B. D. & Wilson, B. (1999). On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics and Data Analysis*, 31, 27-37.
- McCullough, B. D. & Wilson, B. (2000). On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics and Data Analysis*, 40(4), 713-721.
- McCullough, B. D. & Wilson, B. (2004). On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics and Data Analysis*.
- Moore, D. S. & McCabe, G. P. (2003). *Introduction to the practice of statistics* (4th Ed.). New York, N.Y.: W.H. Freeman and Company.
- National Institute of Science and Technology. Standard Data. <http://www.itl.nist.gov/div898/strd/general/dataarchive.html>
- Pelosi, M. K., & Sandifer, T. M. (2000). *Doing statistics for business with Excel*. New York, N.Y.: John Wiley and Sons.
- Sawilowsky, S. S. (2002). Fermat, Schubert, Einstein, and Behrens-Fisher: The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$. *Journal of Modern Applied Statistical Methods*, 1(2), 461-472.
- Schenker, N. & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3), 182.
- Smith, I. (2002). A set of statistical distribution functions in Excel VBA, (Version 1.0.24). Copyright © Ian Smith 2002-2003
- Sprinthall, R. C. (2000). *Basic statistical analysis* (6th Ed.). Needham Heights, MA: Allyn and Bacon.
- Steiger, J. H. & Fouladi, R. T. (1997). Noncentral interval estimation and the evaluation of statistical models. In *What if there were no significance tests?* Mulaik, H. & Steiger, L. E. M. New Jersey.
- Triola, M. F. (2001). *Elementary statistics* (8th Ed.). Addison-Wesley Longham.
- Welford, B. P. (1962). Note on a method for calculating corrected sums of squares and products. *Technometrics*, 4(3), 419.

Algorithms and Code JMASM24: Numerical Computing for Third-Order Power Method Polynomials (Excel)

Todd C. Headrick
Southern Illinois University-Carbondale

The power method polynomial transformation is a popular procedure used for simulating univariate and multivariate non-normal distributions. It requires software that solves simultaneous nonlinear equations. Potential users of the power method may not have access to commercial software packages (e.g., Mathematica, Fortran). Therefore, algorithms are presented in the more commonly available Excel 2003 spreadsheets. The algorithms solve for (1) coefficients for polynomials of order three, (2) intermediate correlations and Cholesky factorizations for multivariate data generation, and (3) the values of skew and kurtosis for determining if a transformation will produce a valid power method probability density function (pdf). The Excel files are available at <http://www.siu.edu/~epse1/headrick/PowerMethod3rd/> or can be requested from the author at headrick@siu.edu.

Key words: Correlated data, cumulants, Monte Carlo methods, polynomial transformations, non-normality

Introduction

The power method polynomial transformation (Fleishman, 1978, Headrick, 2002) is a popular moment-matching technique used for simulating continuous non-normal distributions in the context of Monte Carlo or simulation studies. The primary advantage of this transformation is that it provides computationally efficient algorithms for generating univariate or multivariate distributions with arbitrary correlation matrices (Headrick, 2002, Headrick & Sawilowsky, 1999, Vale & Maurelli, 1983).

For the univariate case, the power method transformation can be summarized by the polynomial

$$Y = \sum_{n=1}^r c_n Z^{n-1} \quad (1)$$

where $Z \sim i.i.d. N(0,1)$. Setting $r=4$ in (1) gives the Fleishman (1978) class of distributions associated with polynomials of order three. It is noted that setting $r=6$ gives the larger Headrick (2002) class of distributions.

The power method has been used in studies that have included such topics or techniques as: ANCOVA (Harwell & Serlin, 1988, Headrick & Sawilowsky 2000a, Headrick & Vineyard, 2001, Klockars & Moses, 2002), computer adaptive testing (Zhu, Yu, & Liu, 2002), hierarchical linear models (Shieh, 2000), item response theory (Stone, 2003), logistic regression (Hess, Olejnik, & Huberty, 2001), microarray analysis (Powell, Anderson, Chen, & Alvord, 2002), regression (Headrick & Rotou, 2001), repeated measures (Beasley & Zumbo, 2003, Harwell & Serlin, 1997, Lix, Algina, & Keselman, 2003; Kowalchuk, Keselman, & Algina, 2003), structural equation modeling (Hipp & Bollen, 2003, Reinartz, Echambadi, & Chin, 2002), and other univariate or multivariate (non)parametric tests (Beasley, 2002, Finch, 2005, Habib & Harwell, 1989, Harwell & Serlin, 1989, Rasch & Guiard, 2004, Steyn, 1993).

The shape of the non-normal distribution Y in (1) is contingent on the constant

Todd C. Headrick is Associate Professor of Statistics. Address: Section on Statistics and Measurement, Department of EPSE, 222-J Wham Building, Mail Code 4618, Southern Illinois University-Carbondale, IL, 62901. His areas of research interest are statistical computing, nonparametric statistics, and optimization. Email address: headrick@siu.edu.

coefficients c_n . These coefficients are determined by simultaneously solving a system of nonlinear equations (described below for a polynomial of order three) for specified values of skew (γ_3) and kurtosis (γ_4). Further, the multivariate extension of equation (1) also requires a nonlinear equation solver for determining intermediate correlations and an algorithm to decompose the associated intermediate correlation matrix.

Purpose of the Study

In view of the above, because not all potential users of the power method have access to commercial software (e.g., Mathematica, Fortran) for solving nonlinear equations, algorithms are provided in Excel 2003 spreadsheets and are made available for free. These spreadsheets determine for polynomials of order three (a) the approximate lower boundary of kurtosis for a given value of skew, (b) a set coefficients for a polynomial and if these coefficients will also produce a valid pdf, and (c) intermediate correlations and Cholesky factorizations for multivariate data generation.

The Third-Order Polynomial

In the context of a third-order polynomial (Fleishman, 1978), the constant coefficients c_2 and c_4 in equation (1) can be determined by simultaneously solving the following two nonlinear equations (Headrick & Sawilowsky, 2000b, Equations 11 and 13)

$$\begin{aligned} \gamma_3 = & (8 - 6c_2^4 - 2c_2^6 + 144c_2^3c_4 - \\ & 108c_2^5c_4 + 720c_4^2 - 540c_2^4c_4^2 - \\ & 2178c_2^4c_4^2 + 2160c_2c_4^3 - 20952c_2^3c_4^3 + \\ & 9450c_2^4 - 106110c_2^2c_4^4 - \\ & 283500c_2c_4^5 - 330750c_4^6)^{1/2} \end{aligned} \quad (2)$$

and

$$\begin{aligned} \gamma_4 = & 12(1 - c_2^4 + 24c_2c_4 - \\ & 34c_2^3c_4 + 150c_4^2 - 324c_2^2c_4^2 - \\ & 1170c_2c_4^3 - 1665c_4^4) \end{aligned} \quad (3)$$

where γ_3 and γ_4 are the specified values of skew and kurtosis, respectively. On solving (2) and (3), the values of c_3 and c_1 are obtained by evaluating the expression (Fleishman, 1978, Equation 17)

$$c_3 = \frac{\gamma_3}{2(c_2^2 + 24c_2c_4 + 105c_4^2 + 2)} \quad (4)$$

and then setting $c_1 = -c_3$. It is noted that only positive values of γ_3 are necessary to consider because negative skew can be obtained by simultaneously reversing the signs between c_1 and c_3 . Further, equations (2), (3), and (4) ensure that all power method distributions in (1) will have a mean of zero ($\gamma_1 = 0$), unit variance ($\gamma_2 = 1$), and where kurtosis is scaled such that $\gamma_4 = 0$ for the unit normal distribution.

The probability density function (pdf) and cumulative distribution function (cdf) associated with Y in (1) are given in parametric form ($\mathbb{R}^2$) generally as (Headrick & Kowalchuk, 2007, Equation 11 and 12)

$$\begin{aligned} f_{Y(Z)}(Y(z)) &= f_{Y(Z)}(Y(x, y)) \\ &= f_{Y(Z)}(Y(z), \frac{f_Z(z)}{Y'(z)}) \end{aligned} \quad (5)$$

and

$$\begin{aligned} F_{Y(Z)}(Y(z)) &= F_{Y(Z)}(Y(x, y)) \\ &= F_{Y(Z)}(Y(z), F_Z(z)) \end{aligned} \quad (6)$$

where $-\infty < z < +\infty$, the derivative $Y'(z) > 0$ (i.e. Y is a strictly increasing monotonic function in Z) and where $f_Z(z)$ and $F_Z(z)$ are the standard normal pdf and cdf.

In terms of a third-order polynomial, the transformation in (1) will produce a valid non-normal power method pdf if and only if (Headrick & Kowalchuck, 2006, Property 4.3)

$$c_4 > (5 + 7c_2^2)^{1/2} / (5\sqrt{3}) - 2c_2 / 5 \quad (7)$$

and

$$0 < c_2 < 1.$$

Figure 1 gives a parametric plot of the parameter space of valid power method pdfs in the γ_3 and γ_4 plane as a function of c_2 . The region inside the enclosed boundary represents the asymmetric distributions that are valid pdfs. Symmetric distributions with $0 < \gamma_4 < 216/5 = 43.2$ are also valid pdfs. Power method distributions with negative kurtosis $\gamma_4 < 0$ are not valid pdfs. The origin represents the standard normal distribution where $c_2 = 1$ and $\gamma_3 = \gamma_4 = 0$.

One of the limitations associated with the power method is that equations (2) and (3) will not provide solutions of coefficients for all possible combinations of γ_3 and γ_4 in the plane defined by the boundary condition

$$\gamma_4 \geq \gamma_3^2 - 2. \quad (8)$$

The lower boundary of kurtosis (γ_4) for given values of skew (γ_3) for the third-order power method transformation was derived by Headrick and Sawilowsky (2000b, see Table 1). A good approximation of the lower boundary of kurtosis derived by Headrick and Sawilowsky (2000b) is the following equation for $0 \leq \gamma_3 \leq 3$

$$\begin{aligned} \gamma_4 \geq & -1.151000 + 0.018264 * \gamma_3 + \\ & 1.637951 * \gamma_3^2 - 0.118343 * \gamma_3^3 + \\ & 0.040960 * \gamma_3^4. \end{aligned} \quad (9)$$

Another limitation associated with the third-order power method polynomial is that some combinations of γ_3 and γ_4 in this class of distributions defined by (9) will not produce valid power method pdfs i.e. will not satisfy the conditions for equation (7). For example, consider an exponential distribution which has standardized cumulants of $\gamma_3 = 2$ and $\gamma_4 = 6$. From Table 1 of Headrick and Sawilowsky (2000b), the lower boundary of kurtosis is $\gamma_4 = 5.151620$ and therefore equations (2) and (3) will yield coefficients for this combination of cumulants (the coefficients are $c_1 = -0.3137$, $c_2 = 0.8263$, $c_3 = 0.3137$, $c_4 = 0.02271$). However, these coefficients will not produce a valid power method pdf because $Y(z)$ in (5) is not a strictly increasing monotonic function for all $z \in (-\infty, +\infty)$ i.e. $Y'(z) = 0$ at $z = -1.592$ and $z = -7.620$.

A technique that can often be used to mitigate the aforementioned problem is to increase γ_4 , ceteris paribus. For example, the values of $\gamma_3 = 2$ and $\gamma_4 = 6.5$ will produce a valid power method pdf. A parametric plot of this power method's pdf and cdf is given in Figure 2.

Multivariate Data Generation

The power method can be extended from univariate to multivariate non-normal data generation by specifying k equations of the form in (1) for third-order polynomials as

$$Y_i = \sum_{n=1}^4 c_{ni} Z_i^{n-1} \quad (10)$$

$$Y_j = \sum_{n=1}^4 c_{nj} Z_j^{n-1} \quad (11)$$

where $i \neq j$. A controlled correlation between two non-normal distributions Y_i and Y_j is accomplished by making use of the following equation (Headrick & Sawilowsky, 1999, Equation 7b; Vale & Maurelli, 1983, Equation 11)

Figure 1. Parameter space for valid pdfs in the skew (γ_3) and kurtosis (γ_4) plane for third-order polynomials. For asymmetric distributions, valid pdfs exist in the region *inside* the graphed boundary. For symmetric distributions, valid pdfs exist for γ_4 between $0 < \gamma_4 < 43.2$. Points on the boundary can be determined exactly by evaluating the following equations (Headrick & Kowalchuk, 2007) for $0 \leq c_2 \leq 1$

$$\gamma_3 = 2(3/5)^{1/2} (c_2(15 + 21c_2^2)^{1/2} (508c_2^4 + 332c_2^2 + 27) - 6c_2(388c_2^4 + 392c_2^2 + 87))^{1/2}$$

and

$$\gamma_4 = (8/125)(-13101c_2^4 + c_2(15 + 21c_2^2)^{1/2}(2866c_2^2 + 45) - 5040c_2^2 + 675).$$

Figure 2. Parametric plots of a valid third-order power method's pdf and cdf.

$$\begin{aligned} \rho_{Y_i Y_j} = & \rho_{Z_i Z_j} (c_{2i} c_{2j} + 3c_{2j} c_{4i} + \\ & 3c_{2i} c_{4j} + 9c_{4i} c_{4j} + 2c_{1i} c_{1j} \rho_{Z_i Z_j} + \\ & 6c_{4i} c_{4j} \rho_{Z_i Z_j}^2) \end{aligned} \quad (12)$$

where $\rho_{Z_i Z_j}$ is referred to as the intermediate correlation. More specifically, the left-hand side of (12) is set to a specified correlation $\rho_{Y_i Y_j}$, the coefficients c_{ni} and c_{nj} are substituted into the right-hand side, and then (1) is numerically solved for the intermediate correlation $\rho_{Z_i Z_j}$. This process is repeated for all $k(k-1)/2$ specified correlations of $\rho_{Y_i Y_j}$. The solved intermediate correlations $\rho_{Z_i Z_j}$ are assembled into a $k \times k$ matrix that is subsequently decomposed (e.g., a Cholesky decomposition). The results from the decomposition are used to generate standard normal deviates Z_i and Z_j , correlated at the intermediate levels, that are then transformed by k polynomials of the form in (10) and (11) such that Y_i and Y_j have their specified shapes and correlation.

The Excel 2003 Spreadsheets

The Excel file PowerMethod3.xls will solve for two sets of polynomial coefficients and a specified bivariate correlation between the two polynomials (if needed). More specifically, this program will

- (a) Determine if a given combination of γ_3 and γ_4 will produce a power method non-normal distribution based on equation (9).
- (b) Solve for the coefficients c_{ni} and c_{nj} in (10) and (11) using equations (2), (3), and (4). And, determine if these coefficients will produce a valid power method pdf using equation (7) on the condition that $0 < c_{2i(j)} < 1$.

- (c) Solve for an intermediate correlation $\rho_{Z_i Z_j}$ for a specified correlation $\rho_{Y_i Y_j}$ between (10) and (11) using equation (12).

On solving for all $k(k-1)/2$ pairwise intermediate correlations $\rho_{Z_i Z_j}$ using step (c), the user can then load the values of $\rho_{Z_i Z_j}$ into CholskeyD.xls which will decompose the intermediate correlation matrix.

Numerical Example

Suppose that four nonnormal distributions $Y_1 \sim (\gamma_3 = 1, \gamma_4 = 2)$, $Y_2 \sim (\gamma_3 = 2, \gamma_4 = 7)$, $Y_3 \sim (\gamma_3 = 0, \gamma_4 = 25)$, $Y_4 \sim (\gamma_3 = 3, \gamma_4 = 25)$ are desired with specified intercorrelations of $\rho_{Y_1 Y_2} = 0.30$, $\rho_{Y_1 Y_3} = 0.70$, $\rho_{Y_1 Y_4} = 0.30$, $\rho_{Y_2 Y_3} = 0.30$, $\rho_{Y_2 Y_4} = 0.70$, and $\rho_{Y_3 Y_4} = 0.30$.

Using PowerMethod3.xls, the Excel solver (under the Tools menu), and the steps from the previous section we have:

- (a) All distributions will produce power method distributions because all values of γ_4 are greater than the lower boundry values of kurtosis (γ_4^*) based on equation (9): ($\gamma_3 = 1, \gamma_4^* = 0.428$), ($\gamma_3 = 2, \gamma_4^* = 5.146$), ($\gamma_3 = 0, \gamma_4^* = -1.151$), and ($\gamma_3 = 3, \gamma_4^* = 13.768$).

- (b) The coefficients are: $c_{11} = -0.1472$, $c_{12} = 0.9048$, $c_{14} = 0.02386$; $c_{21} = -0.2600$, $c_{22} = 0.7616$, $c_{24} = 0.05306$; $c_{31} = 0.0$, $c_{32} = 0.2553$, $c_{34} = 0.2038$; and $c_{41} = -0.2283$, $c_{42} = 0.3342$, $c_{44} = 0.1713$. These coefficients will also satisfy the conditions for producing valid power method pdfs.

- (c) The intermediate correlations are: $\rho_{Z_1 Z_2} = .3244$, $\rho_{Z_1 Z_3} = .8091$, $\rho_{Z_1 Z_4} = .3510$, $\rho_{Z_2 Z_3} = .3718$, $\rho_{Z_2 Z_4} = .7731$, $\rho_{Z_3 Z_4} = .3911$.

The intermediate correlations are loaded into the CholeskeyD.xls which will decompose the 4×4 intermediate correlation matrix. The elements (a_{ij}) in the upper triangular matrix are: $a_{11} = 1$, $a_{12} = .3244$, $a_{13} = .8091$, $a_{14} = .3510$, $a_{22} = .9459$, $a_{23} = .1156$, $a_{24} = .6969$, $a_{33} = .5762$, $a_{34} = .0461$, $a_{44} = .6237$.

The results from the decomposition are then used in an algorithm (e.g., Fortran) to produce standard normal deviates $Z_1, \dots, Z_4$ having the intermediate correlations $\rho_{Z_i Z_j}$ by applying the formulae

$$\begin{aligned} Z_1 &= a_{11}V_1 \\ Z_2 &= a_{12}V_1 + a_{22}V_2 \\ Z_3 &= a_{13}V_1 + a_{23}V_2 + a_{33}V_3 \\ Z_4 &= a_{14}V_1 + a_{24}V_2 + a_{34}V_3 + a_{44}V_4 \end{aligned}$$

where $V_1, \dots, V_4$ are independent standard normal random deviates. The extension to k distributions is shown in the structure of the formulae for $Z_{i=1, \dots, 4}$. The $Z_{i=1, \dots, 4}$ are then substituted into equations of the form in (10) and (11) to produce the specified power method distributions.

Instructions

The spreadsheet PowerMethod3.xls is set up for the bivariate case (i.e. distribution Y_i and Y_j) which is all that is necessary. The user should first enter the desired values of skew and kurtosis and determine if it is possible for these combinations to produce coefficients for the polynomials.

Given that a combination of skew and kurtosis will produce a power method distribution, the user then solves for the coefficients. To do this, the user will need use Solver under the Tools menu. For either distribution, the user will want to specify the objective function to zero (i.e. cell B19 equal to zero), specify c_{2i} (cell K19) and c_{4i} (cell N19) as initial guesses (0.50 and 0.10 are, in general, good initial guesses, respectively), and then add

as constraints equation for skew and equation for kurtosis and set both of these constraints equal to zero. On clicking Solve, the equations solver will yield the coefficients and then determine if the distribution is also a valid power method pdf.

To solve for the intermediate bivariate correlation the user must first specify a correlation (cell B39), set the objective function to zero (cell H39), specify an initial guess for the intermediate correlation (cell E39), and then click Solve. On solving separately for all $k(k-1)/2$ pairwise intermediate correlations, these correlations can then be loaded into the CholskeyD.xls spreadsheet.

More specifically, Choleskey D.xls is set to a 4×4 correlation matrix in cells D14:G17 using the numerical example from the previous section. If a smaller matrix is needed, (a) enter the required intermediate correlations, (b) click in cell D25, (c) click in the $\sqrt{x}$ tool bar and the correlation matrix will be highlighted, (d) click on the correlation matrix and reduce it to a smaller size (e.g. 3×3), and then (e) compute the Cholskey Decomposition by pressing Control-Shift-Enter on the keyboard.

If a larger intermediate correlation is needed then (a) type in the new correlation matrix e.g. a 5×5 matrix, cells D14:H18, (b) specify a Cholskey decomposition matrix below i.e., highlight cells D30:H34, and (c) type in the $\sqrt{x}$ tool bar =Cholskey(D14:H18,H8,H7) and compute the decomposition by pressing Control-Shift-Enter on the keyboard.

References

- Beasley, T. M. (2002). Multivariate aligned rank test for interactions in multiple group repeated measures. *Multivariate Behavioral Research*, 37, 197-226.
- Beasley, T. M. & Zumbo, B. D. (2003). Comparison of aligned Friedman rank and parametric methods for testing interactions in split-plot designs. *Computational Statistics and Data Analysis*, 42, 569-593.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43, 521-532.

Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics with assumptions violated. *Methodology*, 1, 27-38.

Habib, A. R., & Harwell, M. R. (1989). An empirical study of the Type I error rate and power of some selected normal theory and nonparametric tests of independence of two sets of variables. *Communications in Statistics: Simulation and Computation*, 18, 793-826.

Harwell, M. R. & Serlin, R. C. (1988). An experimental study of a proposed test of nonparametric analysis of covariance. *Psychological Bulletin*, 104, 268-281.

Harwell, M. R. & Serlin, R. C. (1989). A nonparametric test statistic for the general linear model. *Journal of Educational Statistics*, 14, 351-371.

Harwell, M. R. & Serlin, R. C. (1997). An empirical study of five multivariate tests for the single-factor repeated measures model. *Communications in Statistics: Simulation and Computation*, 26, 605-618.

Headrick, T. C. (2002). Fast fifth-order polynomial transforms for generating univariate and multivariate non-normal distributions. *Computational Statistics and Data Analysis*, 40, 685-711.

Headrick, T. C. & Kowalchik, R. K. (2007). The power method transformation: Its probability density function, distribution function, and its further use for fitting data. *Journal of Statistical Computation and Simulation*, 77, 229-249.

Headrick, T. C. & Rotou (2001). An investigation of the rank transformation in multiple regression. *Computational Statistics and Data Analysis*, 38, 203-215.

Headrick, T. C., & Sawilowsky, S. S. (1999). Simulating correlated non-normal distributions: extending the Fleishman power method. *Psychometrika*, 64, 25-35.

Headrick, T. C. & Sawilowsky, S. S. (2000a). Properties of the rank transformation in factorial analysis of covariance. *Communications in Statistics: Simulation and Computation*, 29, 1059-1088.

Headrick, T. C., & Sawilowsky, S. S. (2000b). Weighted simplex procedures for determining boundary points and constants for the univariate and multivariate power methods. *Journal of Educational and Behavioral Statistics*, 25, 417-436.

Headrick, T. C. & Vineyard, G. (2001). An empirical investigation of four tests for interaction in the context of factorial analysis of covariance. *Multiple Linear Regression Viewpoints*, 27, 3-15.

Hipp, J. R., Bollen, K. A. (2003). Model fit in structural equation models with censored, ordinal, and dichotomous variables: Testing vanishing tetrads. *Sociological Methodology*, 33, 267-305.

Hess, B., Olejnik S., Huberty, C. J. (2001). The efficacy of two improvement-over-chance effect sizes for two group univariate comparisons under variance heterogeneity and non-normality. *Educational and Psychological Measurement*, 61, 909-936.

Klockars, A. J. & Moses, T. P. (2002). Type I error rates for rank-based tests of homogeneity of regression slopes. *Journal of Modern Applied Statistical Methods*, 1, 452-460.

Kowalchuk, R. K. Keselman, H. J., Algina, J. (2003). Repeated measures interaction test with aligned ranks. *Multivariate Behavioral Research*, 38, 433-461.

Lix, L. M., Algina, J., Keselman, H. J. (2003). Analyzing multivariate repeated measures designs: A comparison of two approximate degrees of freedom procedures. *Multivariate Behavioral Research*, 38, 403-431.

Powell, D. A., Anderson, L. M., Chen, R.Y.S., Alvord, W. G. (2002). Robustness of the Chen-Dougherty-Bittner procedure against non-normality and heterogeneity in the coefficient of variation. *Journal of Biomedical Optics*, 7, 655-660.

Rasch, D., & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science*, 46, 175-208.

Reinartz, W. J., Echambadi, R., Chin, W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research*, 37, 227-244.

Shieh, Y. (April, 2000). The effects of distributional characteristics on multi-level modeling parameter estimates and Type I error control of parameter tests under conditions of non-normality. Paper presented at the annual meeting of the American Educational Research Association: New Orleans.

Stone, C. (2003). Empirical power and Type I error rates for an IRT fit statistic that considers the precision and ability estimates. *Educational and Psychological Measurement*, 63, 566-583.

Steyn, H. S. (1993). On the problem of more than one kurtosis parameter in multivariate analysis. *Journal of Multivariate Analysis*, 44, 1-22.

Vale, C. D. & Maurelli, V. A. (1983). Simulating univariate and multivariate non-normal distributions. *Psychometrika*, 48, 465-471.

Zhu, R., Yu, F., & Liu S. (April, 2002). Statistical indexes for monitoring item behavior under computer adaptive testing environment. Paper presented at the annual meeting of the American Educational Research Association: New Orleans

JMASM25: Computing Percentiles of Skew-Normal Distributions

Sikha Bagui Subhash Bagui
The University of West Florida

An algorithm and code is provided for computing percentiles of skew-normal distributions with parameter λ using Monte Carlo methods. A critical values table was created for various parameter values of λ at various probability levels of α . The table will be useful to practitioners as it is not available in the literature.

Key words: Skew normal distribution, critical values, visual basic.

Introduction

The skew normal (SN) class of densities is a family of densities that includes the normal density, with an extra parameter than the normal one to regulate skewness. In other words, SN class of densities extends the normal model by allowing an extra shape parameter to account for skewness. Let Z be a standard normal random variable with probability density function (pdf) ϕ and cumulative distribution function (cdf) Φ . A random variable Z' is said to have the one-parameter skew-normal distribution with parameter λ , (say, $Z' \sim SN(\lambda)$), if it has the following density function

$$f_{Z'}(z; \lambda) = 2\phi(z)\Phi(\lambda z), -\infty < z < \infty \quad (1)$$

where λ is a real valued parameter. When $\lambda = 0$, one gets the standard normal density, for $\lambda > 0$, one gets the positively skewed

distribution, and for $\lambda < 0$, one gets the negatively skewed distribution. Graphs of skew-normal densities for different values of λ are shown in Figure 1.

Figure 1: Skew-Normal Plot with $\lambda = -4, -1, 0, 1, 4$

The skew-normal distributions first indirectly appeared in Roberts (1966), O'Hagan and Leonard (1976), and Aigner and Lovell (1977). The landmark article by Azzalini (1985, 1986) gave a systematic treatment of this distribution and studied its fundamental properties. Recent work on characterization of skew-normal distribution can be found in Arnold and Lin (2004) and Gupta, Nguyen, and Sanqui (2004a). Azzalini and Dalla (1996), Azzalini and Capitanio (1999), and Gupta, Gonzalez-Farias, and Dominguez-Molina (2004b) investigated the multivariate extensions of this distribution. There is always a growing interest in modeling non-normal distributions. This

Sikha Bagui is an Assistant Professor in the Department of Computer Science. Her areas of research are database and database design, data mining, pattern recognition, and statistical computing. email: bagui@uwf.edu. Subhash Bagui is Professor of Statistics at the University of West Florida. His areas of research are statistical classification and pattern recognition, bio-statistics, construction of designs, tolerance regions, and statistical computing. email:sbagui@uwf.edu

model can be very useful to capture non-gaussian behavior of the data. It already proved to be useful in modeling real data sets (Azzalini & Capitanio, 1999). To capture non-gaussian behavior of the data, Aigner and Lovell (1977) used this density in stochastic frontier models, Adcock and Shutes (1999) used it in portfolio selection of financial assets, and Bartolucci, De Luca, and Loperfido (2000) used it in detection of skewness in stock returns. Below some important properties of $SN(\lambda)$ are stated without proof. Let

$$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$$

and

$$\Phi_\lambda(z) = \int_{-\infty}^z f_{Z'}(t; \lambda) dt.$$

1. $SN(0)$ has the density $N(0,1)$.
2. If $Z' \sim SN(\lambda)$, then $-Z' \sim SN(-\lambda)$.
3. $f_{Z'}(z; \lambda)$ is strongly unimodal, i.e. $\log f_{Z'}(z; \lambda)$ is a concave function of z .
4. $\Phi_\lambda(-z) = 1 - \Phi_{-\lambda}(z)$, i.e. $f_{Z'}(z; \lambda)$ is a mirror image of $f_{Z'}(z; -\lambda)$.
5. If $Z' \sim SN(\lambda)$, then $Z'^2 \sim \chi_1^2$ (chi-square with one d.f.).
6. If $Z' \sim SN(\lambda)$, then Z' has the m.g.f. $M_{Z'}(t) = 2\exp(\frac{t^2}{2})\Phi(\delta t)$.
7. If $Z' \sim SN(\lambda)$, then $E(Z') = \sqrt{\frac{2}{\pi}}\delta$ and $\text{Var}(Z') = 1 - \frac{2}{\pi}\delta^2$.
8. Let Z_1 and Z_2 be independent $N(0,1)$ random variables, then $Z' = \delta|Z_1| + \sqrt{1-\delta^2}Z_2 \sim SN(\lambda)$.
9. Let Z_1 and Z_2 be independent $N(0,1)$ random variables. Set
$$Z' = \begin{cases} Z_2 & \text{if } Z_1 < \lambda Z_2 \\ -Z_2 & \text{if } Z_1 > \lambda Z_2. \end{cases}$$
 Then $Z' \sim SN(\lambda)$.
10. Let (Z_1, Z_2) be a bivariate normal with $E(Z_1) = E(Z_2) = 0$, $\text{Var}(Z_1) = \text{Var}(Z_2) = 1$, and $\text{corr}(Z_1, Z_2) = \rho$. Set $Z' = \max(Z_1, Z_2)$. Then $Z' \sim SN(\lambda)$, where $\lambda = \sqrt{\frac{1-\rho}{1+\rho}}$.

Property (4) tells us that $f_{Z'}(z; \lambda)$ is a mirror image of $f_{Z'}(z; -\lambda)$. So it is sufficient to find percentiles only for positive values of λ . For negative values of λ , one may use property (4) to derive the percentiles. In order to simulate skew-normal random variate one may use either properties (8), or (9), or (10). It is also important that sufficient number of random variates must be used to model the distribution of Z' adequately. Below we extend one parameter skew-normal distribution to three parameters skew-normal distributions.

Definition. (Three-parameter skew-normal). A random variable X' is said to have the three-parameter skew-normal distribution (denoted by $X' \sim SN(\mu, \sigma, \lambda)$) if $X' = \mu + \sigma Z'$, where $Z' \sim SN(\lambda)$, i.e. if X' has the following density function

$$f_{X'}(x; \lambda) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\lambda \frac{x-\mu}{\sigma}\right), -\infty < x < \infty \quad (2)$$

Methodology

Property (8) will be used in order to generate skew-normal random variate Z' . Before this property is used, standard normal variates must be simulated. The following theorem explains how to generate standard normal variates.

Theorem.

(Box & Muller, 1958). Let U_1 and U_2 be independent random variates from uniform $(0,1)$ (i.e. $U_1, U_2 \sim U(0,1)$), then the variates $Z_1 = (-2 \ln U_1)^{\frac{1}{2}} \cos 2\pi U_2$; $Z_2 = (-2 \ln U_1)^{\frac{1}{2}} \sin 2\pi U_2$ are independent standard normal variates.

Proof.

The joint density of U_1 and U_2 is given by $f_{U_1, U_2}(u_1, u_2) = 1$. Also, from the above transformation note that $u_1 = e^{-\frac{1}{2}(z_1^2 + z_2^2)}$ and $u_2 = \frac{1}{2\pi} \tan^{-1} \frac{z_2}{z_1}$. Thus, the Jacobian of this transformation is given by $J = -\frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2 + z_2^2)}$. Hence, the joint density of Z_1 and Z_2 is given by

$$\begin{aligned} f_{Z_1, Z_2}(z_1, z_2) &= f_{U_1, U_2}(u_1, u_2) \times |J| \\ &= \frac{1}{2\pi} e^{-\frac{1}{2}(z_1^2 + z_2^2)}. \end{aligned} \quad (3)$$

The equation (3) confirms that Z_1 and Z_2 is independent standard normal variates.

Now using the above theorem and the property (8), the skew-normal variates Z' shall be generated. The algorithm is as follows.

Algorithm

1. Define $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$
2. Generate two independent random variates U_1 and U_2 from $U(0,1)$.
3. Set $Z_1 \leftarrow (-2 \ln U_1)^{\frac{1}{2}} \cos 2\pi U_2$ and $Z_2 \leftarrow (-2 \ln U_1)^{\frac{1}{2}} \sin 2\pi U_2$.
4. Deliver $Z' = \delta |Z_1| + \sqrt{1-\delta^2} Z_2$. $\square$

This process is repeated until the density of Z' is adequately modeled. The code is written in Visual Basic. After a sufficient number of runs the program will return percentiles corresponding to the probability level $\alpha = 0.01, 0.02, 0.025, 0.05, 0.1, 0.99, 0.995, 0.975, 0.98, 0.99$ for a given λ value. For $\lambda = 0.01$ to 50 , in Table 1A, the percentiles of Z' are provided for $\alpha = 0.01, 0.02, 0.025, 0.05, 0.1$; and in Table 1B, the percentiles are provided for $\alpha = 0.99, 0.995, 0.975, 0.98, 0.99$. Let the notation $Z'_\alpha(\lambda)$ denote the $\alpha(100)\%$ percentile of Z' for a given value of λ .

Percentiles for negative λ :

Let $\lambda > 0$, then $Z'_\alpha(-\lambda)$ can be derived by means of $Z'_\alpha(-\lambda) = -Z'_{1-\alpha}(\lambda)$, as $f_{Z'}(z; \lambda)$ is a mirror image of $f_{Z'}(z; -\lambda)$.

Percentiles of X' :

For the three-parameter skew normal r.v. $X' \sim SN(\mu, \sigma, \lambda)$, let $X'_\alpha(\lambda)$ denote the $\alpha(100)\%$ percentile of X' for a given value of λ . Since $X' = \mu + \sigma Z'$ is a monotonically increasing transformation, the percentile $X'_\alpha(\lambda)$ of X' can be obtained from the equation

$$X'_\alpha(\lambda) = \mu + \sigma Z'_\alpha(\lambda). \quad (4)$$

Table 1A. Percentiles of Skew-normal distributions

$\lambda \downarrow \mid \alpha \rightarrow$	0.01	0.02	0.025	0.04	0.05	0.10
0.01	-2.316601	-2.043600	-1.949637	-1.740196	-1.634294	-1.270585
0.02	-2.308165	-2.035310	-1.941678	-1.732143	-1.626163	-1.262473
0.03	-2.299909	-2.027109	-1.933209	-1.723888	-1.617982	-1.254325
0.04	-2.291742	-2.018600	-1.924809	-1.715525	-1.609591	-1.246031
0.05	-2.282902	-2.010202	-1.916329	-1.707010	-1.601178	-1.237716
0.06	-2.273634	-2.001495	-1.907718	-1.698454	-1.592648	-1.229322
0.07	-2.264764	-1.992613	-1.899061	-1.689889	-1.584055	-1.220915
0.08	-2.255799	-1.983832	-1.890096	-1.681037	-1.575384	-1.212358
0.09	-2.246697	-1.974694	-1.881151	-1.672253	-1.566573	-1.203799
0.10	-2.237531	-1.965540	-1.872131	-1.663274	-1.557718	-1.195168
0.15	-2.189265	-1.918519	-1.825301	-1.617421	-1.512277	-1.151156
0.20	-2.137967	-1.869018	-1.776368	-1.569685	-1.465172	-1.106155
0.25	-2.084749	-1.817640	-1.725684	-1.520615	-1.416825	-1.060535
0.30	-2.029456	-1.764838	-1.673795	-1.470527	-1.367758	-1.014637
0.35	-1.972884	-1.711116	-1.620986	-1.420003	-1.318257	-0.968773
0.40	-1.915085	-1.656844	-1.567945	-1.369303	-1.268771	-0.923367
$\lambda \downarrow \mid \alpha \rightarrow$	0.01	0.02	0.025	0.04	0.05	0.10
0.45	-1.857159	-1.602308	-1.514616	-1.318676	-1.219467	-0.878647
0.50	-1.79914	-1.548298	-1.461832	-1.268712	-1.170971	-0.834859
0.55	-1.741834	-1.494818	-1.409644	-1.219545	-1.123254	-0.792200
0.60	-1.684862	-1.442066	-1.358373	-1.171422	-1.076717	-0.750845
0.65	-1.629129	-1.390497	-1.308256	-1.124578	-1.031460	-0.710951
0.70	-1.574326	-1.340210	-1.259576	-1.079053	-0.987594	-0.672611
0.75	-1.521160	-1.291533	-1.212292	-1.035095	-0.945313	-0.635839
0.80	-1.469659	-1.244274	-1.166618	-0.992756	-0.904630	-0.600690
0.85	-1.419618	-1.198740	-1.122574	-0.951997	-0.865531	-0.567114
0.90	-1.371167	-1.154904	-1.080178	-0.912938	-0.828105	-0.535187
0.95	-1.324828	-1.112776	-1.039482	-0.875538	-0.792291	-0.504804
1.0	-1.280150	-1.072379	-1.000580	-0.839787	-0.758158	-0.475989
1.1	-1.195920	-0.996480	-0.927569	-0.773014	-0.694478	-0.422673
1.2	-1.118556	-0.927121	-0.860912	-0.712349	-0.636756	-0.374875
1.3	-1.047759	-0.863910	-0.800243	-0.657286	-0.584519	-0.331973
1.4	-0.982884	-0.806225	-0.744969	-0.607346	-0.537221	-0.293562
1.5	-0.923732	-0.753687	-0.694679	-0.562104	-0.494470	-0.259078
1.6	-0.869510	-0.705780	-0.648922	-0.521012	-0.455715	-0.228142
1.7	-0.819759	-0.661996	-0.607135	-0.483662	-0.420568	-0.200318
1.8	-0.774230	-0.621985	-0.569028	-0.449680	-0.388656	-0.175231
1.9	-0.732466	-0.585328	-0.534144	-0.418682	-0.359602	-0.152622
2.0	-0.694041	-0.551756	-0.502219	-0.390393	-0.333112	-0.132153
2.1	-0.658558	-0.520897	-0.472897	-0.364488	-0.308919	-0.113589
2.2	-0.625889	-0.492486	-0.445939	-0.340752	-0.286783	-0.096753
2.3	-0.595671	-0.466286	-0.421117	-0.318922	-0.266441	-0.081409
2.4	-0.567697	-0.442084	-0.398212	-0.298842	-0.247777	-0.067438
2.5	-0.541678	-0.419685	-0.376990	-0.280305	-0.203571	-0.054627
2.6	-0.517496	-0.398873	-0.357355	-0.263187	-0.214697	-0.042918
2.7	-0.495071	-0.379556	-0.339106	-0.247320	-0.200020	-0.032172
2.8	-0.474034	-0.361518	-0.322117	-0.232604	-0.186429	-0.022296

Table 1A. Continued

2.9	-0.454425	-0.344773	-0.306319	-0.218931	-0.173802	-0.013191
3.0	-0.436086	-0.329099	-0.291554	-0.206190	-0.162091	-0.004783
3.1	-0.418815	-0.314420	-0.277760	-0.194313	-0.151160	0.002981
3.2	-0.402685	-0.300694	-0.264865	-0.183231	-0.140979	0.010144
3.3	-0.387500	-0.287812	-0.252750	-0.172862	-0.131481	0.016818
3.4	-0.373191	-0.275705	-0.241405	-0.163161	-0.122593	0.022994
3.5	-0.359675	-0.264295	-0.230702	-0.154042	-0.114251	0.028741
3.6	-0.346930	-0.253540	-0.220636	-0.145471	-0.106434	0.034091
3.7	-0.334911	-0.243416	-0.211150	-0.137420	-0.099090	0.039075
3.8	-0.323507	-0.233839	-0.202214	-0.129843	-0.092200	0.043727
3.9	-0.312720	-0.224801	-0.193740	-0.122698	-0.085694	0.048080
4.0	-0.302460	-0.216232	-0.185747	-0.115946	-0.079572	0.052150
4.1	-0.292758	-0.208106	-0.178181	-0.109578	-0.073800	0.055953
4.2	-0.283507	-0.200385	-0.170970	-0.103532	-0.068333	0.059509
4.3	-0.274730	-0.193095	-0.164176	-0.097830	-0.063178	0.062874
4.4	-0.266349	-0.186124	-0.157692	-0.092408	-0.058288	0.066000
4.5	-0.258354	-0.179500	-0.151528	-0.087270	-0.053652	0.068956
4.6	-0.250720	-0.173180	-0.145664	-0.082397	-0.049265	0.071733
4.7	-0.243464	-0.167145	-0.140082	-0.077752	-0.045097	0.074366
4.8	-0.236517	-0.161429	-0.134749	-0.073343	-0.041141	0.076857
4.9	-0.229873	-0.155942	-0.129676	-0.069135	-0.037373	0.079159
5.0	-0.223494	-0.150707	-0.124817	-0.065139	-0.033789	0.081360
$\lambda \downarrow \mid \alpha \rightarrow$	0.01	0.02	0.025	0.04	0.05	0.10
5.1	-0.217406	-0.145728	-0.120181	-0.061327	-0.030392	0.083439
5.2	-0.211567	-0.140914	-0.115759	-0.057684	-0.027137	0.085408
5.3	-0.205935	-0.136321	-0.111519	-0.054198	-0.024034	0.087269
5.4	-0.200537	-0.131905	-0.107427	-0.050879	-0.021072	0.089026
5.5	-0.195384	-0.127710	-0.103541	-0.047696	-0.018256	0.090667
5.6	-0.190408	-0.123636	-0.099800	-0.044650	-0.015548	0.092261
5.7	-0.185623	-0.119771	-0.096229	-0.041742	-0.012973	0.093766
5.8	-0.181004	-0.116017	-0.092782	-0.038952	-0.010507	0.095178
5.9	-0.176603	-0.112419	-0.089485	-0.036272	-0.008136	0.096523
6.0	-0.172317	-0.108956	-0.086301	-0.033701	-0.005888	0.097816
6.1	-0.168194	-0.105632	-0.083230	-0.031240	-0.003710	0.099030
6.2	-0.164207	-0.102420	-0.080282	-0.028865	-0.001611	0.100179
6.3	-0.160373	-0.099323	-0.077451	-0.026586	0.000375	0.101290
6.4	-0.156667	-0.096356	-0.074712	-0.024401	0.002299	0.102334
6.5	-0.153083	-0.093483	-0.072070	-0.022295	0.004149	0.103340
6.6	-0.149635	-0.090704	-0.069549	-0.020284	0.005903	0.104284
6.7	-0.146298	-0.088027	-0.067097	-0.018334	0.007612	0.105187
6.8	-0.143068	-0.085447	-0.064730	-0.016452	0.009265	0.106051
6.9	-0.139926	-0.082934	-0.062453	-0.014642	0.010826	0.106886
7.0	-0.136863	-0.080531	-0.060243	-0.012900	0.012341	0.107663
7.1	-0.133958	-0.078188	-0.058108	-0.011215	0.013789	0.108414
7.2	-0.131101	-0.075920	-0.056037	-0.009603	0.015194	0.109131
7.3	-0.128363	-0.073745	-0.054056	-0.008027	0.016553	0.109804

Table 1A. Continued

7.4	-0.125676	-0.071620	-0.052122	-0.006516	0.017858	0.110468
7.5	-0.123042	-0.069554	-0.050240	-0.005032	0.019123	0.111083
7.6	-0.120535	-0.067574	-0.048431	-0.003635	0.020326	0.111686
7.7	-0.118106	-0.065650	-0.046692	-0.002276	0.021503	0.112249
7.8	-0.115726	-0.063777	-0.044988	-0.000954	0.022632	0.112797
7.9	-0.113417	-0.061967	-0.043348	0.000319	0.023724	0.113323
8.0	-0.111195	-0.060210	-0.041749	0.001560	0.024790	0.113809
8.1	-0.109002	-0.058498	-0.040200	0.002743	0.025790	0.114317
8.2	-0.106879	-0.056837	-0.038692	0.003900	0.026798	0.114763
8.3	-0.104818	-0.055219	-0.037241	0.005040	0.027749	0.115201
8.4	-0.102825	-0.053661	-0.035826	0.006123	0.028656	0.115608
8.5	-0.100859	-0.052133	-0.034446	0.007154	0.029559	0.116018
8.6	-0.098962	-0.050651	-0.033115	0.008175	0.030416	0.116405
8.7	-0.097105	-0.049204	-0.031813	0.009179	0.031252	0.116774
8.8	-0.095290	-0.047811	-0.030556	0.010149	0.032062	0.117126
8.9	-0.093523	-0.046447	-0.029317	0.011070	0.032860	0.117473
9.0	-0.091821	-0.045127	-0.028133	0.011977	0.033618	0.117783
9.1	-0.090159	-0.043830	-0.026962	0.012845	0.034346	0.118100
9.2	-0.088516	-0.042576	-0.025842	0.013698	0.035065	0.118397
9.3	-0.086938	-0.041349	-0.024734	0.014532	0.035755	0.118687
9.4	-0.085379	-0.040158	-0.023666	0.015333	0.036421	0.118960
9.5	-0.083854	-0.038984	-0.022616	0.016118	0.037080	0.119222
9.6	-0.082381	-0.037854	-0.021604	0.016868	0.037714	0.119475
9.7	-0.080931	-0.036748	-0.020609	0.017608	0.038328	0.119719
9.8	-0.079553	-0.035683	-0.019654	0.018317	0.038917	0.119954
9.9	-0.078172	-0.034629	-0.018706	0.019018	0.039480	0.120186
10.0	-0.076817	-0.033597	-0.017784	0.019709	0.040050	0.120411
10.5	-0.070540	-0.028825	-0.013538	0.022821	0.042624	0.121345
11.0	-0.064912	-0.024585	-0.009749	0.025563	0.044873	0.122133
$\lambda \downarrow \mid \alpha \rightarrow$	0.01	0.02	0.025	0.04	0.05	0.10
11.5	-0.059842	-0.020784	-0.006407	0.027966	0.046819	0.122770
12.0	-0.055255	-0.017392	-0.003412	0.030098	0.048531	0.123291
12.5	-0.051114	-0.014327	-0.000708	0.031980	0.050026	0.123727
13.0	-0.047339	-0.011552	0.001725	0.033664	0.051355	0.124079
13.5	-0.043889	-0.009055	0.003918	0.035157	0.052524	0.124365
14.0	-0.040724	-0.006747	0.005909	0.036500	0.053572	0.124596
14.5	-0.037822	-0.004656	0.007694	0.037699	0.054480	0.124805
15.0	-0.035151	-0.002762	0.009352	0.038784	0.055308	0.124958
15.5	-0.032685	-0.001014	0.010850	0.039764	0.056045	0.125093
16.0	-0.030383	0.000586	0.012218	0.040657	0.056694	0.125200
16.5	-0.028268	0.002073	0.013480	0.041444	0.057285	0.125281
17.0	-0.026302	0.003431	0.014648	0.042177	0.057812	0.125357
17.5	-0.024471	0.004693	0.015709	0.042825	0.058293	0.125411
18.0	-0.022750	0.005851	0.005851	0.016690	0.058715	0.125456

Table 1A. Continued

18.5	-0.021154	0.006931	0.017599	0.043967	0.059083	0.125477
19.0	-0.019658	0.007942	0.018437	0.044467	0.059422	0.125499
19.5	-0.018247	0.008911	0.019235	0.044918	0.059745	0.125521
20.0	-0.016935	0.009766	0.019949	0.045335	0.060028	0.125530
20.5	-0.015693	0.010567	0.020614	0.045729	0.060287	0.125547
21.0	-0.014518	0.011346	0.021244	0.046078	0.060514	0.125554
21.5	-0.013422	0.012052	0.021830	0.046386	0.060714	0.125561
22.0	-0.012383	0.012723	0.022374	0.046681	0.060905	0.125560
22.5	-0.011397	0.013344	0.022886	0.046948	0.061078	0.125566
23.0	-0.010471	0.013928	0.023348	0.047200	0.061233	0.125567
23.5	-0.009614	0.014482	0.023801	0.047428	0.061370	0.125573
24.0	-0.008769	0.014998	0.024219	0.047639	0.061500	0.125572
24.5	-0.007970	0.015477	0.024611	0.047831	0.061614	0.125574
25.0	-0.007208	0.015950	0.024969	0.048007	0.061716	0.125564
25.5	-0.006509	0.016391	0.025315	0.048170	0.061808	0.125573
26.0	-0.005832	0.016791	0.025634	0.048323	0.061892	0.125564
26.5	-0.005174	0.017176	0.025939	0.048463	0.061967	0.125552
27.0	-0.004559	0.017544	0.026223	0.048588	0.062036	0.125554
27.5	-0.003968	0.017898	0.026496	0.048705	0.062091	0.125542
28.0	-0.003411	0.018220	0.026747	0.048813	0.062153	0.125537
28.5	-0.002880	0.018530	0.026980	0.048919	0.062205	0.125533
29.0	-0.002364	0.018821	0.027205	0.049011	0.062245	0.125530
29.5	-0.001887	0.019102	0.027418	0.049094	0.062286	0.125524
30.0	-0.001412	0.019371	0.027628	0.049182	0.062326	0.125530
31.0	-0.000534	0.019867	0.028003	0.049323	0.062388	0.125514
32.0	-0.000270	0.020300	0.028323	0.049445	0.062445	0.125508
33.0	0.001017	0.020702	0.028611	0.049549	0.062486	0.125501
34.0	0.001689	0.021065	0.028878	0.049633	0.062532	0.125496
35.0	0.002328	0.021390	0.029116	0.049712	0.062557	0.125504
36.0	0.002914	0.021690	0.029332	0.049774	0.062578	0.125490
37.0	0.003455	0.021970	0.029521	0.049829	0.062591	0.125496
38.0	0.003950	0.022231	0.029697	0.049876	0.062610	0.125493
39.0	0.004439	0.022450	0.029851	0.049918	0.062614	0.125490
40.0	0.004874	0.022661	0.029990	0.049952	0.062626	0.125487
45.0	0.006671	0.023457	0.030504	0.050056	0.062642	0.125448
50.0	0.007977	0.023984	0.030829	0.050084	0.062631	0.125432

Table 1B. Percentiles of Skew-normal distributions

$\lambda \downarrow \mid \alpha \rightarrow$	0.90	0.95	0.96	0.975	0.98	0.99
0.01	1.293420	1.656375	1.762111	1.971312	2.064999	2.337648
0.02	1.301256	1.664239	1.769951	1.979043	2.072866	2.344827
0.03	1.309050	1.671955	1.777612	1.986665	2.080348	2.352550
0.04	1.316720	1.679481	1.785125	1.994172	2.087801	2.360183
0.05	1.324280	1.686977	1.792677	2.001513	2.095286	2.367557
0.06	1.331719	1.694337	1.799951	2.008842	2.102222	2.374510
0.07	1.339140	1.701540	1.807104	2.015848	2.109401	2.381502
0.08	1.346394	1.708666	1.814139	2.022961	2.116477	2.388371
0.09	1.353632	1.715712	1.821133	2.029747	2.123257	2.394928
0.10	1.360717	1.722577	1.828044	2.036431	2.129964	2.401439
0.15	1.394691	1.755185	1.860267	2.068003	2.160968	2.431694
0.20	1.426029	1.784750	1.889300	2.096069	2.188663	2.458094
0.25	1.454635	1.811324	1.915185	2.120902	2.213067	2.480934
0.30	1.480591	1.834861	1.938216	2.142349	2.233973	2.500336
0.35	1.503898	1.855531	1.958124	2.161018	2.251863	2.516574
0.40	1.524704	1.873564	1.975278	2.176768	2.267009	2.530021
0.45	1.543034	1.889094	1.990116	2.190004	2.279616	2.540934
0.50	1.559094	1.902265	2.002473	2.201200	2.290273	2.549641
0.55	1.573039	1.913487	2.013031	2.210150	2.298575	2.556615
0.60	1.585148	1.922783	2.021584	2.217448	2.305356	2.562057
0.65	1.595441	1.930570	2.028706	2.223336	2.310856	2.566115
0.70	1.604329	1.936988	2.034563	2.228120	2.315141	2.569253
0.75	1.611804	1.942238	2.039192	2.231836	2.318608	2.571621
0.80	1.618075	1.946476	2.042889	2.234565	2.320861	2.573240
0.85	1.623354	1.949947	2.045924	2.236780	2.322569	2.574507
0.90	1.627734	1.952547	2.048190	2.238430	2.324226	2.575317
0.95	1.631392	1.954670	2.049986	2.239752	2.325128	2.575999
1.0	1.634401	1.956423	2.051249	2.240476	2.326339	2.576275
1.1	1.638840	1.958590	2.053171	2.241585	2.327224	2.576799
1.2	1.641727	1.959986	2.054064	2.242350	2.327050	2.577051
1.3	1.643522	1.960541	2.054750	2.242413	2.327984	2.577070
1.4	1.644607	1.960917	2.054949	2.242456	2.327962	2.576978
1.5	1.645308	1.961144	2.059950	2.242557	2.327888	2.576880
1.6	1.645648	1.961166	2.055057	2.242478	2.327804	2.576944
1.7	1.645906	1.961200	2.054881	2.242736	2.327201	2.576940
1.8	1.645948	1.961158	2.055017	2.242478	2.327852	2.576950
1.9	1.645961	1.961119	2.054800	2.242613	2.327199	2.576926
2.0	1.645964	1.961075	2.054768	2.242509	2.327203	2.576712
2.1	1.645909	1.961014	2.054838	2.242351	2.327562	2.576704
2.2	1.645818	1.960899	2.054820	2.242209	2.327690	2.576740
2.3	1.645823	1.960930	2.054787	2.242234	2.327584	2.576644
2.4	1.645719	1.960825	2.054695	2.242143	2.327559	2.576768
2.5	1.645742	1.960853	2.054653	2.242216	2.327176	2.576606
2.6	1.645711	1.960833	2.054470	2.242426	2.326834	2.576642

Table 1B. Continued

2.7	1.645637	1.960791	2.054506	2.242231	2.326950	2.576564
2.8	1.645590	1.960666	2.054598	2.242026	2.327523	2.576594
2.9	1.645570	1.960694	2.054491	2.242087	2.326953	2.576518
3.0	1.645558	1.960707	2.054382	2.242221	2.326884	2.576578
3.1	1.645474	1.960613	2.054549	2.242039	2.327407	2.576532
3.2	1.645488	1.960613	2.054294	2.242132	2.326717	2.576504
$\lambda \downarrow \mid \alpha \rightarrow$	0.90	0.95	0.96	0.975	0.98	0.99
3.3	1.645415	1.960518	2.054481	2.241884	2.327434	2.576521
3.4	1.645466	1.960594	2.054265	2.242045	2.326620	2.576445
3.5	1.645430	1.960526	2.054435	2.241953	2.327321	2.576440
3.6	1.645399	1.960585	2.054206	2.242235	2.326654	2.576527
3.7	1.645338	1.960508	2.054390	2.241919	2.327152	2.576455
3.8	1.645341	1.960490	2.054281	2.241911	2.326835	2.576369
3.9	1.645379	1.960450	2.054197	2.242022	2.326782	2.576330
4.0	1.645231	1.960373	2.054328	2.241714	2.327218	2.576309
4.1	1.645315	1.960442	2.054154	2.241898	2.326652	2.576407
4.2	1.645232	1.960439	2.054178	2.241854	2.326687	2.576256
4.3	1.645216	1.960421	2.054332	2.241768	2.327075	2.576212
4.4	1.645209	1.960414	2.054117	2.242002	2.326591	2.576272
4.5	1.645218	1.960400	2.054053	2.241882	2.326596	2.576298
4.6	1.645162	1.960299	2.054246	2.241636	2.327101	2.576333
4.7	1.645162	1.960338	2.054202	2.241751	2.326936	2.576291
4.8	1.645208	1.960351	2.054063	2.241928	2.326533	2.576241
4.9	1.645163	1.960280	2.054060	2.241765	2.326623	2.576122
5.0	1.645114	1.960214	2.054205	2.241575	2.327085	2.576106
5.1	1.645086	1.960246	2.054112	2.241644	2.326876	2.576236
5.2	1.645165	1.960308	2.053991	2.241825	2.326466	2.576260
5.3	1.645135	1.960306	2.053968	2.241907	2.326391	2.576162
5.4	1.645122	1.960253	2.054074	2.241737	2.326709	2.576184
5.5	1.645047	1.960210	2.054136	2.241636	2.326965	2.576106
5.6	1.645074	1.960243	2.054071	2.241744	2.326695	2.576144
5.7	1.645095	1.960286	2.053924	2.241871	2.326429	2.576125
5.8	1.645071	1.960233	2.053909	2.241780	2.326384	2.576143
5.9	1.645000	1.960143	2.053952	2.241593	2.326621	2.576128
6.0	1.645002	1.960103	2.054066	2.241451	2.326907	2.576198
6.1	1.644974	1.960113	2.054026	2.241435	2.326910	2.576147
6.2	1.645003	1.960126	2.053954	2.241558	2.326640	2.576085
6.3	1.645052	1.960183	2.053896	2.241696	2.326429	2.576078
6.4	1.645045	1.960172	2.053872	2.241734	2.326339	2.576060
6.5	1.645009	1.960141	2.053902	2.241628	2.326459	2.576040
6.6	1.644982	1.960095	2.053987	2.241485	2.326643	2.576015
6.7	1.644965	1.960078	2.054038	2.241488	2.326882	2.576034
6.8	1.644951	1.960053	2.053990	2.241437	2.326824	2.576046
6.9	1.644988	1.960112	2.053900	2.241568	2.326596	2.576126
7.0	1.644949	1.960154	2.053835	2.241636	2.326317	2.576030
7.1	1.644964	1.960172	2.053805	2.241699	2.326186	2.575996
7.2	1.644953	1.960166	2.053807	2.241727	2.326238	2.576077
7.3	1.644914	1.960135	2.053827	2.241632	2.326301	2.575969
7.4	1.644974	1.960074	2.053883	2.241616	2.326524	2.576031

Table 1B. Continued

7.5	1.644917	1.960047	2.053996	2.241583	2.326695	2.576053
7.6	1.644905	1.96007	2.053986	2.241468	2.326820	2.575998
7.7	1.644881	1.960032	2.053922	2.241447	2.326658	2.575943
7.8	1.644929	1.960073	2.053898	2.241500	2.326506	2.575923
7.9	1.644862	1.960086	2.053796	2.241544	2.326368	2.575921
8.0	1.644862	1.960086	2.053796	2.241544	2.326368	2.575921
8.1	1.644943	1.960109	2.053767	2.241632	2.326222	2.575927
8.2	1.644900	1.960052	2.053782	2.241569	2.326266	2.575962
8.3	1.644896	1.960029	2.053752	2.241511	2.326375	2.575969
8.4	1.644865	1.959982	2.053821	2.241426	2.326483	2.575973
$\lambda \downarrow \mid \alpha \rightarrow$	0.90	0.95	0.96	0.975	0.98	0.99
8.5	1.644835	1.959963	2.053842	2.241366	2.326642	2.576032
8.6	1.644836	1.959950	2.053918	2.241271	2.326747	2.575987
8.7	1.644781	1.959948	2.053914	2.241361	2.326842	2.576077
8.8	1.644840	1.959959	2.053904	2.241307	2.326759	2.576039
8.9	1.644839	1.959961	2.053845	2.241415	2.326647	2.575979
9.0	1.644861	1.959939	2.053783	2.241389	2.326466	2.575960
9.1	1.644858	1.959983	2.053723	2.241461	2.326388	2.575918
9.2	1.644885	1.960014	2.053756	2.241531	2.326252	2.575928
9.3	1.644882	1.960022	2.053730	2.241538	2.326224	2.575987
9.4	1.644870	1.960017	2.053703	2.241534	2.326170	2.575862
9.5	1.644853	1.960025	2.053738	2.241544	2.326254	2.576012
9.6	1.644877	1.960038	2.053720	2.241512	2.326289	2.575952
9.7	1.644810	1.959976	2.053759	2.241423	2.326382	2.575969
9.8	1.644811	1.959973	2.053800	2.241396	2.326444	2.575875
9.9	1.644769	1.959939	2.053802	2.241360	2.326500	2.575831
10.0	1.644781	1.959948	2.053859	2.241353	2.326638	2.575943
10.5	1.644804	1.959943	2.053758	2.241444	2.326391	2.575953
11.0	1.644804	1.960052	2.053653	2.241601	2.326104	2.575897
11.5	1.644781	1.959947	2.053681	2.241464	2.326186	2.575794
12.0	1.644741	1.959817	2.053787	2.241293	2.326544	2.575793
12.5	1.644716	1.959876	2.053757	2.241254	2.326530	2.575768
13.0	1.644757	1.959929	2.053645	2.241376	2.326241	2.575774
13.5	1.644741	1.959922	2.053582	2.241459	2.326090	2.575783
14.0	1.644732	1.959921	2.053618	2.241399	2.326186	2.575842
14.5	1.644684	1.959809	2.053642	2.241237	2.326385	2.575796
15.0	1.644638	1.959769	2.053717	2.241128	2.326577	2.575814
15.5	1.644644	1.959782	2.053737	2.241101	2.326580	2.575795
16.0	1.644663	1.959782	2.053644	2.241154	2.326394	2.575750
16.5	1.644685	1.959836	2.053550	2.241262	2.326212	2.575741
17.0	1.644674	1.959845	2.053551	2.241346	2.326080	2.575808
17.5	1.644698	1.959858	2.053542	2.241395	2.326020	2.575665
18.0	1.644710	1.959889	2.053565	2.241372	2.326083	2.575678
18.5	1.644645	1.959817	2.053607	2.241333	2.326142	2.575747
19.0	1.644663	1.959800	2.053619	2.241271	2.326221	2.575731
19.5	1.644634	1.959747	2.053661	2.241215	2.326354	2.575740
20.0	1.644617	1.959751	2.053672	2.241223	2.326401	2.575616
20.5	1.644623	1.959762	2.053698	2.241181	2.326468	2.575791

Table 1B. Continued

21.0	1.644602	1.959722	2.053689	2.241216	2.326337	2.575552
21.5	1.644638	1.959779	2.053610	2.241245	2.326266	2.575705
22.0	1.644612	1.959777	2.053609	2.241303	2.326148	2.575632
22.5	1.644621	1.959820	2.053550	2.241317	2.326071	2.575730
23.0	1.644659	1.959839	2.053537	2.241369	2.326041	2.575757
23.5	1.644647	1.959841	2.053499	2.241389	2.325941	2.575648
24.0	1.644614	1.959878	2.053466	2.241425	2.325935	2.575692
24.5	1.644658	1.959882	2.053480	2.241397	2.325907	2.575722
25.0	1.644639	1.959836	2.053463	2.241375	2.325944	2.575668
25.5	1.644635	1.959838	2.053506	2.241361	2.326009	2.575766
26.5	1.644627	1.959805	2.053518	2.241291	2.326085	2.575695
27.0	1.644580	1.959718	2.053498	2.241234	2.326143	2.575725
27.5	1.644571	1.959727	2.053531	2.241205	2.326167	2.575708
28.0	1.644597	1.959732	2.053560	2.241215	2.326225	2.575627
28.5	1.644562	1.959702	2.053602	2.241172	2.326275	2.575609
$\lambda \downarrow \mid \alpha \rightarrow$	0.90	0.95	0.96	0.975	0.98	0.99
29.0	1.644596	1.959695	2.053639	2.241164	2.326395	2.575679
29.5	1.644554	1.959687	2.053625	2.241102	2.326410	2.575647
30.0	1.644567	1.959746	2.053641	2.241064	2.326440	2.575569
31.0	1.644542	1.959689	2.053634	2.241111	2.326441	2.575682
32.0	1.644529	1.959683	2.053624	2.241048	2.326381	2.575607
33.0	1.644538	1.959699	2.053558	2.241121	2.326288	2.575529
34.0	1.644552	1.959725	2.053554	2.241144	2.326236	2.575658
35.0	1.644579	1.959725	2.053534	2.241180	2.326200	2.575612
36.0	1.644555	1.959697	2.053427	2.241170	2.326102	2.575533
37.0	1.644047	1.959700	2.053440	2.241219	2.326077	2.575647
38.0	1.644570	1.959698	2.053447	2.241282	2.326026	2.575629
39.0	1.644555	1.959742	2.053442	2.241288	2.325948	2.575530
40.0	1.644563	1.959744	2.053426	2.241311	2.325948	2.575642
45.0	1.644552	1.959688	2.053419	2.241203	2.326003	2.575680
50.0	1.644524	1.959636	2.053461	2.241076	2.326203	2.575640

Conclusion

For any number of runs, the program will calculate the percentiles of skew-normal distributions for a given value of λ . It is advisable that at least one million runs be used to find percentiles. The above table is created using 100 million runs, so these reported results are more precise. One can expect to have better results if the number of runs is very high. From the table it may be noted that for $\lambda = 1.1$, $\alpha = 0.05$, $Z'_{\alpha}(\lambda) = Z'_{0.05}(1.1) = -0.69448$.

However, if $\lambda = -1.1$, $\alpha = 0.05$, $Z'_{\alpha}(\lambda) = Z'_{0.05}(-1.1)$ is not available in the table. Now the value of $Z'_{\alpha}(\lambda) = Z'_{0.05}(-1.1)$ can be derived as $Z'_{0.05}(-1.1) = -Z'_{0.95}(1.1) = -1.95859$. Let $X' \sim SN(3.0, 1.5, 1.1)$. Then, $X'_{0.05}(1.1)$ can be obtained as $X'_{0.05}(1.1) = 3 + (1.5) Z'_{0.05}(1.1) = 3 + (1.5)(0.69448) = 1.95828$, and $X'_{0.05}(-1.1)$ is derived as $X'_{0.05}(-1.1) = 3 - (1.5) Z'_{0.95}(1.1) = 3 - (1.5)(1.95859) = 0.062115$.

References

- Adcock, C., & Shutes, K. (1999). Portfolio selection based on the multivariate skewed normal distribution, *Financial Modeling*, (A Skulimowski edited), Progress and Business Publishers, Krakow, 35-60.
- Aigner, D. J., & Lovell, C. A. K. (1977). Formulation and estimation of stochastic frontier production function model. *Econometrics*, 12, 21-37.
- Arnold, B. C., & Lin, G. D. (2004). Characterization of the skew-normal and generalized Chi distributions. *Sankhya*, 66(4), 593-606.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12, 171-178.
- Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica*, 46, 199-208.
- Azzalini, A., & Dalla valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4), 715-726.
- Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, B*, 61(3), 579-602.
- Bartolucci, F., De Luca G., & Loperfido, N. (2000). A generalization for skewness in the basic stochastic volatility model, *Proceedings of the 15th International Workshop on Statistical Modelling*, 140-145.
- Box, G. E. P. & Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29, 610-611.
- Gupta, A. K., Nguyen, T. T., & Sanqui, J. A. T. (2004a). Characterization of the skew-normal distribution. *Annals of the Institute of Statistical Mathematics*, 56(2), 351-360.
- Gupta, A. K., Gonzalez-Farias, G., & Dominguez-Molina, J. A. (2004b). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89, 181-190.
- O'Hagan, A., & Leonard, T. (1982). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63, 201-203.
- Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association*, 61(316), 1184-1190.
- Rubinstein, R. Y. (1981). *Simulation and the monte carlo method*. New York, N.Y.: Wiley & Sons.

Appendix

```
Imports System.Math
Public Class Form1
    Inherits System.Windows.Forms.Form

    Dim u1, u2, z1, z2, delta, lamda, zScore, num As Single
    Dim p1, p2, p3, p4, p5, p6, p7, p8, p9, p10, p11, p12 As Single
    Dim i, n As Integer
    Dim file1 As System.IO.StreamWriter

    Private Sub Button1_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button1.Click
        n = Val(TextBox1.Text)
        lamda = Val(TextBox2.Text)
        file1 = System.IO.File.CreateText("C:\file1.txt")

        Dim zScoreArray(n) As Single

        For i = 1 To n
            u1 = Rnd()
            u2 = Rnd()
            z1 = (Sqrt((-2) * Log(u1))) * Cos((360) * u2)
            z2 = (Sqrt((-2) * Log(u1))) * Sin((360) * u2)

            delta = (lamda / (Sqrt(1 + (lamda * lamda))))
            zScore = (delta * Abs(z1) + Sqrt(1 - (delta * delta)) * z2)
            zScoreArray(i) = zScore
        Next
        Array.Sort(zScoreArray)
        p1 = 0.01 * n
        p2 = 0.02 * n
        p3 = 0.025 * n
        p4 = 0.04 * n
        p5 = 0.05 * n
        p6 = 0.1 * n
        p7 = 0.9 * n
        p8 = 0.95 * n
        p9 = 0.96 * n
        p10 = 0.975 * n
        p11 = 0.98 * n
        p12 = 0.99 * n

        TextBox3.Text = zScoreArray(p1)
        TextBox4.Text = zScoreArray(p2)
        TextBox5.Text = zScoreArray(p3)
        TextBox7.Text = zScoreArray(p4)
        TextBox8.Text = zScoreArray(p5)
        TextBox9.Text = zScoreArray(p6)
        TextBox10.Text = zScoreArray(p7)
    End Sub
End Class
```

```

TextBox11.Text = zScoreArray(p8)
    TextBox12.Text = zScoreArray(p9)
    TextBox13.Text = zScoreArray(p10)
    TextBox14.Text = zScoreArray(p11)
    TextBox15.Text = zScoreArray(p12)

    file1.WriteLine("Number of Runs = " & n & vbCrLf & vbCrLf _
        & "Lamda = " & lamda & vbCrLf & vbCrLf _
        & "0.01 = " & zScoreArray(p1) & vbCrLf _
        & "0.02 = " & zScoreArray(p2) & vbCrLf _
        & "0.025 = " & zScoreArray(p3) & vbCrLf _
        & "0.04 = " & zScoreArray(p4) & vbCrLf _
        & "0.05 = " & zScoreArray(p5) & vbCrLf _
        & "0.1 = " & zScoreArray(p6) & vbCrLf _
        & "0.9 = " & zScoreArray(p7) & vbCrLf _
        & "0.95 = " & zScoreArray(p8) & vbCrLf _
        & "0.96 = " & zScoreArray(p9) & vbCrLf _
        & "0.975 = " & zScoreArray(p10) & vbCrLf _
        & "0.98 = " & zScoreArray(p11) & vbCrLf _
        & "0.99 = " & zScoreArray(p12) & vbCrLf & _
        vbCrLf _ & "Date and Time is " & _
        System.DateTime.Now)

    file1.Close()

End Sub

Private Sub Form1_Load(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles MyBase.Load
    Randomize() 'Calling the random number generator
End Sub
End Class

```

Translations, Ephemerals, & Biographies Joseph Liouville's 'Mathematical Works Of Évariste Galois'

Shlomo S. Sawilowsky
Wayne State University

John L. Cuzzocrea
University of Akron

Liouville's 1846 introduction to the mathematical works of Galois is translated from French to flowing (American) English. It gave an overview of the tragic circumstances of the undergraduate mathematician whose originality led to major advances in abstract Algebra.

Key words: Galois, Liouville, *Journal de Mathématiques Pures et Appliquées*

Introduction

The mathematics of Évariste Galois (1811 – 1832) flourished throughout the 20th century. It centered on groups of permutations, currently known in group theory as Galois connections or Galois representations. Its application has greatly expanded, emerging from the bowels of abstract Algebra, proliferating even as a key component of Wiles' (1995) solution to one of the most famous problems in mathematics, the 17th century Fermat's Last Theorem.

In the 1820s, Galois became interested in mathematics at the Lycée of Louis-le-Grand. Although he excelled, he was denied admission to the more prestigious École Polytechnique because his other coursework was wanting. Subsequently, he returned to the Lycée and became a student of Louis Paul Émile Richard (1795 – 1849). Richard (who never published) was also the mentor of other illustrious students: Joseph Le Verrier (1811 – 1877), Joseph Alfred Serret (1819 – 1885), and Charles Hermite (1822 – 1901).

Galois' tragic demise abruptly truncated a promising academic career to about a half dozen presentations and publications. (Approximately a dozen personal letters exist, www.galois-group.net/gtp/index//EN/.) Either inadvertently or deliberately, development of his scholarly work was hindered by Sylvestre François Lacroix (1765 – 1843), Jean Baptiste Joseph Fourier (1768 – 1830), and Augustin Louis Cauchy (1789 – 1857), but was facilitated by Marie-Sophie Germain (1776 – 1831), Siméon Denis Poisson (1781 – 1840), and Jacques Charles François Sturm (1803 – 1855).

Galois' work may have been forever lost were it not for the dedication of Joseph Liouville (1809 – 1882), one of the most influential mathematicians of the middle of the 19th century. In 1836, he became the Founding Editor of *Journal de Mathématiques Pures et Appliquées* (extant today as an Elsevier Science journal), known fondly as *Liouville's Journal*.

Liouville's (1846) introduction to Galois' work was written in French. We translate it in flowing (American) English. It is followed by a scan of the first page (Figure 1) as it originally appeared in *Liouville's Journal*.

Shlomo Sawilowsky is Wayne State University Distinguished Faculty Fellow, Professor of Evaluation and Research, and "data analyst." Contact him at shlomo@wayne.edu. John Cuzzocrea is an Assistant Professor of Educational Research. His research interests are in Monte Carlo methods, measurement theory, nonparametric, and exact statistics. He is also the Editorial Assistant of *JMASM*. Contact him at jcuzzocrea@cogeco.ca.

References

- Liouville, J. (1846). Oeuvres mathématiques d'Évariste Galois. *Journal de Mathématiques Pures et Appliquées*, 11, 381-384.
- Wiles, A. (1995). Modular elliptic curves and Fermat's last theorem. *Annals of Mathematics*, 141(3), 443-551.
- www.galois-group.net/gtp/index//EN/

Notice: Mathematical Works of Évariste Galois
Toul, October 30, 1846
J. Liouville

The ingenious and profound geometrician, whose works are presented here, died being scarcely twenty years old. Even so, he managed to squander fruitlessly the greater part of the last two years of such a short a life, in political activism, nightclubs, or imprisoned under lock and key at Sainte-Pélagie. He was born October 26, 1811. In the month of May, 1832, he succumbed to a fatal duel, undoubtedly the culmination of some frivolous quarrel. It removed him from the mathematical sciences, of which he would have cultivated with considerable brilliance!

The primary work of Évariste Galois has as a goal the conditions of resolving equations by using radicals. The author posed the basis of a general theory which he applied in detail to the equations of which the degree is a prime number. At the early age of sixteen, and on the benches of the Louis-le-Grand College, Galois enrolled in this difficult subject matter, where his contented studies were encouraged by the excellent Professor and gentleman, Mr. Richard [1]. He presented several memoirs to the Academy containing the results of his work. Apart from a few fragments and some notes, however, nothing remains extant with the exception of the article he submitted to them on January 17, 1831.

The reviewers [2], in their referees' report, reproached the young analyst for his obscure draft. Their reproach, which had already been addressed (Galois himself tells us) in his preceding communications, was well-founded, as he must have admitted. Perhaps, his exaggerated desire for conciseness was the cause of this defect, and is something which one must endeavor to refrain from when dealing with the abstract and mysterious matters of pure Algebra. Clarity is, indeed, all the more necessary when one has intention of leading the reader away from the beaten roads into the desert.

"When dealing with transcendental questions," said Descartes, "strive to be transcendently clear." Galois had frequently neglected this precept; and we understand that famous geometricians had deemed it suitable to

attempt to try to lead this genius novice, inexperienced albeit on the right path, by the severity of their wise council. The author that they censured was in front of them, passionate, active, and he would benefit from their advice.

But at present all that has changed. Alas, Galois is no more! Let us cease carrying on with useless criticisms; let us leave its defects, and instead see its qualities.

While bowing to the wishes of Évariste's friends, I engaged myself, under the watchful eyes of his brother [3], so to speak, to the attentive study of all the publications and manuscripts he left. I believed it my duty and primary goal to research and arrange them, in order to bring forth to the best of my abilities that which was novel in his works.

My zeal was soon rewarded. I experienced great pleasure the moment when, after having filled in the minor gaps, I recognized both the scope and precision of the method that Galois proved. In particular, this is his beautiful theorem: In order for an irreducible equation of a primary degree to be solvable by radicals, it is necessary and sufficient that all the roots are rational functions of any two among them. This method, truly worthy of the attention of the geometricians, would be enough by itself to ensure our compatriot being ranked with the small number of the scientists who have deserved the title of 'Inventor'.

We will initially reproduce the material Galois published from 1828 – 1830 which appeared in the *Annales* of Mr. Gergonne and in the *Bulletin des Sciences* of Mr. Ferussac. Then, his unedited writings follow. Finally, we propose some commentary to complete certain passages and clarify a few delicate points.

On the eve of his death, and in anticipation of the disastrous fate which awaited him, Galois quickly traced the summary of the broad ideas which occupied him, and addressed it, in the form of a letter, to his best friend, Mr. Auguste Chevalier. This last writing, a sort of scientific testament, which we will place as a preface to his posthumous works, is impossible to read without emotion while contemplating the circumstances under which it was composed.

This letter appeared in 1832, in the *Revue Encyclopédique* (September issue, page 568). An obituary notice on Mr. Galois appeared

in the same issue (page 744). We did not believe it necessary to include it in our collection to follow. It contains interesting details, but, for the most part, does not pertain to science. Moreover, certain extreme assertions and opinions concerning persons and things would only motivate others to rise to contradict him. It is true that even in the eyes of one who was furthest removed from his opinions, the author of this notice excused him in advance due to professed tender friendship in unity with Galois. As for those who neither knew nor had seen this unfortunate young man, we will contain ourselves in our role of geometrician (and the observations which we will allow ourselves in publishing his works under the inspiration of his family) to only portray his mathematics.

Appendix

The ordinary requirements of our publication and the extent of Galois' works prevent us from presenting it in a single issue. We shall present his material in two parts, in this volume and in the next.

[1] Mr. Le Verrier, Mr. Hermite, and other distinguished scholars were in Mr. Richard's class. These excellent students glorify their teacher.

[2] Misters Lacroix and Poisson were signatories to the reviewer's report. We can see their conclusions in the report were expressed somewhat dryly by the way in which Mr. Lacroix expressed himself in the sixth edition of his *Complements of the Elements of Algebra*, page 345: "In 1831, a young Frenchman, Évariste Galois, dead the following year, had announced, in a memoir presented to the Academy of Sciences, that in order for an irreducible equation of the first degree to be solvable by radicals, it is necessary and sufficient that any two of the roots be known and the others could be deduced rationally; but this memoir appeared almost unintelligible to the reviewers responsible for examining it."

[3] Mr. Alfred Galois.

Figure 1. Scan of the first page of Liouville's (1846) introduction to the works of Galois, from the private collection of Shlomo S. Sawilowsky.

Pietro Paoli, Italian Algebraist

John L. Cuzzocrea
University of Akron

Shlomo S. Sawilowsky
Wayne State University

Pietro Paoli was a leading Italian mathematician in the late 18th century. His signed letter pertaining to the death of astronomer Giuseppe Antonio Slop is translated from Italian to flowing (American) English.

Key words: *Paoli, Slop, Elementi di Algebra*

Introduction

The first edition of Paoli's 1780 *Opuscola Analytica* was dedicated to the Duke of Toscana (Tuscany), Peter Leopold II Joseph (1747 – 1792, who was to become the penultimate Roman Emperor the last two years of his life). *Analytical Works* was a common title for scientific anthologies (see. e.g., the 1783/1785 compendium with the same title by Leonhard Euler (1707 – 1783) published posthumously). Another edition appeared in 1783.

Paoli's more important two volume *Elementi di Algebra* was published in 1794. The expanded 3rd and final edition of *di Algebra* was published in two volumes in 1803 with a third volume in 1804. The mathematics (and slide rule) historian Cajori (1929) cited the first edition of *di Algebra* as helping to establish various 17th and 18th century mathematics symbols in Italy. Some examples include trigonometric, inverse, and powers of trigonometric functions; and Euler's *e* as the base of natural logarithms. Extant ephemera include thirty-five letters written by Paoli held in the David Bierens de Haan (1822 - 1895) mathematics collection, University of Leiden.

Paoli's students included Vincenzo Brunacci (1768 – 1818), who published the two volume *Elementi di Algebra e di Geometria* in 1809, and Antonio Bordoni.

The following letter written by Paoli pertained to the death of the astronomer Giuseppe Antonio Slop (Josephi Slopilii de Cadenberg, 1740 – 1808). He was a professor at the University of Pisa and Director of their observatory. Among his many works (including letters held by the University Library of Pisa), he was the editor of six volumes pertaining to the observatory that were published in 1769, 1774, 1778, 1779, 1789, and 1795.

The translation, scan (Figure 1), and title page of *Elementi di Algebra* (Figure 2) follow.

Dear Sir,
Pisa, the 20th of May, 1808

I hope you have received my reply to your circular from March 18 conveyed to me on May 2nd. As I have already written to you, I have searched futilely for a picture of the deceased Slop, but I was not able to find one. However, at my request, Mr. Giarrini Ajuto, as the chair of Astronomy and his probable successor, was the person placed in charge of compiling Slop's eulogy. When it is finished, I will send it your way. Please give my very best to the very deserving Signore Cavalieri President, your uncle.

With distinct esteem I am, Pietro Paoli.

References

Cajori, F. (1929). *A history of mathematical notations. Volume II: Notations mainly in higher mathematics*. Chicago, IL: The Open Court Publishing Co.

John Cuzzocrea is Assistant Professor of Educational Research. Contact him at jcuzzocrea@cogeco.ca. Shlomo Sawilowsky is Wayne State University Distinguished Faculty Fellow. Contact him at shlomo@wayne.edu.

Chia.^{mo} Sig.^{re}
 Pisa Li 20. Maggio 1808
 Spero che Ella avrà ricevuta la mia responsiva alla sua
 circolare del 18. Marzo ripetutami in data del 2. Maggio.
 Come già Le scrissi, ho cercato inutilmente un ritratto
 del defunto Laplace, che non mi è riuscito di rinvenirlo. Ma
 l'elogio di lui si è incaricato a mia richiesta di com-
 pilare il sig.^{re} Piazzini Ajuto di Laplace nella cattedra
 di Astronomia, e probabilmente suo successore. Quando
 sarà terminato, Lo manderò alla sua direzione.
 La prego di porgermi i miei ossequj al degnissimo sig.^{re} Cav.^{re}
 Presidente suo Lio, e con distinta stima mi rassegno

Ino. Dio. mo. S. mo. L.
 Pietro Paoli

Figure 1. Letter of 1808 Signed By Pietro Paoli.
 From the private collection of Shlomo S. Sawilowsky

Figure 2. Pietro Paoli's Title Page Of Volume 1 of the 3rd edition (1803).
From the private collection of Shlomo S. Sawilowsky

Two Years in the Making...

Intel® Visual Fortran 8.0

The next generation of Visual Fortran is here! Intel Visual Fortran 8.0 was developed jointly by Intel and the former DEC/Compaq Fortran engineering team.

Visual Fortran Timeline

- 1997** DEC releases
Digital Visual Fortran 5.0
- 1998** Compaq acquires DEC
and releases DVF 6.0
- 1999** Compaq ships CVF 6.1
- 2001** Compaq ships CVF 6.6
- 2001** Intel acquires CVF
engineering team
- 2003** Intel releases
Intel Visual Fortran 8.0

Intel Visual Fortran 8.0

- CVF front-end +
Intel back-end
- Better performance
- OpenMP Support
- Real*16

Performance

Outstanding performance on Intel architecture including Intel® Pentium® 4, Intel® Xeon™ and Intel Itanium® 2 processors, as well as support for Hyper-Threading Technology.

Compatibility

- Plugs into Microsoft Visual Studio® .NET
- Microsoft PowerStation4 language and library support
- Strong compatibility with Compaq® Visual Fortran

Support

1 year of free product upgrades and Intel Premier Support

"The Intel Fortran Compiler 7.0 was first-rate, and Intel Visual Fortran 8.0 is even better. Intel has made a giant leap forward in combining the best features of Compaq Visual Fortran and Intel Fortran. This compiler... continues to be a 'must-have' tool for any Twenty-First Century Fortran migration or software development project."

—Dr. Robert R. Trippi
Professor Computational Finance
University of California, San Diego

FREE trials available at:
programmersparadise.com/intel

Programmer's Paradise®

To order or request additional information call:
800-423-9990
Email: intel@programmers.com

NCSS

329 North 1000 East
Kaysville, Utah 84037

Announcing NCSS 2004 Seventeen New Procedures

NCSS 2004 is a new edition of our popular statistical **NCSS** package that adds seventeen new procedures.

New Procedures

Two Independent Proportions
Two Correlated Proportions
One-Sample Binary Diagnostic Tests
Two-Sample Binary Diagnostic Tests
Paired-Sample Binary Diagnostic Tests
Cluster Sample Binary Diagnostic Tests
Meta-Analysis of Proportions
Meta-Analysis of Correlated Proportions
Meta-Analysis of Means
Meta-Analysis of Hazard Ratios
Curve Fitting
Tolerance Intervals
Comparative Histograms
ROC Curves
Elapsed Time Calculator
T-Test from Means and SD's
Hybrid Appraisal (Feedback) Model

Documentation

The printed, 330-page manual, called *NCSS User's Guide V*, is available for \$29.95. An electronic (pdf) version of the manual is included on the distribution CD and in the Help system.

Two Proportions

Several new exact and asymptotic techniques were added for hypothesis testing (null, noninferiority, equivalence) and calculating confidence intervals for the difference, ratio, and odds ratio. Designs may be independent or paired. Methods include: Farrington & Manning, Gart & Nam, Conditional & Unconditional Exact, Wilson's Score, Miettinen & Nurminen, and Chen.

Meta-Analysis

Procedures for combining studies measuring paired proportions, means, independent proportions, and hazard ratios are available. Plots include the forest plot, radial plot, and L'Abbe plot. Both fixed and random effects models are available for combining the results.

Curve Fitting

This procedure combines several of our curve fitting programs into one module. It adds many new models such as Michaelis-Menten. It analyzes curves from several groups. It compares fitted models across groups using computer-intensive randomization tests. It computes bootstrap confidence intervals.

Tolerance Intervals

This procedure calculates one and two sided tolerance intervals using both distribution-free (nonparametric) methods and normal distribution (parametric) methods. Tolerance intervals are bounds between which a given percentage of a population falls.

Comparative Histogram

This procedure displays a comparative histogram created by interspersing or overlaying the individual histograms of two or more groups or variables. This allows the direct comparison of the distributions of several groups.

Random Number Generator

Matsumoto's Mersenne Twister random number generator (cycle length $> 10^{**}6000$) has been implemented.

Binary Diagnostic Tests

Four new procedures provide the specialized analysis necessary for diagnostic testing with binary outcome data. These provide appropriate specificity and sensitivity output. Four experimental designs can be analyzed including independent or paired groups, comparison with a gold standard, and cluster randomized.

ROC Curves

This procedure generates both binormal and empirical (nonparametric) ROC curves. It computes comparative measures such as the whole, and partial, area under the ROC curve. It provides statistical tests comparing the AUC's and partial AUC's for paired and independent sample designs.

Hybrid (Feedback) Model

This new edition of our hybrid appraisal model fitting program includes several new optimization methods for calibrating parameters including a new genetic algorithm. Model specification is easier. Binary variables are automatically generated from class variables.

Statistical Innovations Products

Through a *special arrangement* with Statistical Innovations (S.I.), NCSS customers will receive \$100 discounts on:

Latent GOLD[®] - latent class modeling

SI-CHAID[®] - segmentation trees

GOLDMineR[®] - ordinal regression

For demos and other info visit:

www.statisticalinnovations.com

Please rush me the following products:

Qty _____

_____ **NCSS 2004 CD upgrade from NCSS 2001**, \$149.95 \$_____

_____ **NCSS 2004 User's Guide V**, \$29.95 \$_____

_____ **NCSS 2004 CD, upgrade from earlier versions**, \$249.95 \$_____

_____ **NCSS 2004 Deluxe (CD and Printed Manuals)**, \$599.95 \$_____

_____ **PASS 2002 Deluxe**, \$499.95 \$_____

_____ **Latent Gold® from S.I.**, \$995 - \$100 NCSS Discount = \$895..... \$_____

_____ **GoldMineR® from S.I.**, \$695 - \$100 NCSS Discount = \$595 \$_____

_____ **CHAID® Plus from S.I.**, \$695 - \$100 NCSS Discount = \$595.... \$_____

Approximate shipping--depends on which manuals are ordered (U.S: \$10 ground, \$18 2-day, or \$33 overnight) (Canada \$24) (All other countries \$10) (Add \$5 U.S. or \$40 International for any S.I. product) \$_____

Total..... \$_____

TO PLACE YOUR ORDER

CALL: (800) 898-6109 **FAX:** (801) 546-3907

ONLINE: www.ncss.com

MAIL: NCSS, 329 North 1000 East, Kaysville, UT 84037

My Payment Option:

_____ Check enclosed

_____ Please charge my: ☐ VISA ☐ MasterCard ☐ Amex

_____ Purchase order attached _____

Card Number _____ Exp _____

Signature _____

Telephone:

() _____

Email:

Ship to:

NAME _____

ADDRESS _____

ADDRESS _____

ADDRESS _____

CITY _____ STATE _____

ZIP/POSTAL CODE _____ COUNTRY _____

Forest Plot of Odds Ratio

Statistical and Graphics Procedures Available in NCSS 2004

Analysis of Variance / T-Tests

Analysis of Covariance
Analysis of Variance
Bartlett Variance Test
Crossover Design Analysis
Factorial Design Analysis
Friedman Test
Geiser-Greenhouse Correction
General Linear Models
Mann-Whitney Test
MANOVA
Multiple Comparison Tests
One-Way ANOVA
Paired T-Tests
Power Calculations
Repeated Measures ANOVA
T-Tests – One or Two Groups
T-Tests – From Means & SD's
Wilcoxon Test

Time Series Analysis

ARIMA / Box - Jenkins
Decomposition
Exponential Smoothing
Harmonic Analysis
Holt - Winters
Seasonal Analysis
Spectral Analysis
Trend Analysis

Plots / Graphs

Bar Charts
Box Plots
Contour Plot
Dot Plots
Error Bar Charts
Histograms
Histograms: Combined*
Percentile Plots
Pie Charts
Probability Plots
ROC Curves*
Scatter Plots
Scatter Plot Matrix
Surface Plots
Violin Plots

Experimental Designs

Balanced Inc. Block
Box-Behnken
Central Composite
D-Optimal Designs
Fractional Factorial
Latin Squares
Plackett-Burman
Response Surface
Screening
Taguchi

Regression / Correlation

All-Possible Search
Canonical Correlation
Correlation Matrices
Cox Regression
Kendall's Tau Correlation
Linear Regression
Logistic Regression
Multiple Regression
Nonlinear Regression
PC Regression
Poisson Regression
Response-Surface
Ridge Regression
Robust Regression
Stepwise Regression
Spearman Correlation
Variable Selection

Quality Control

Xbar-R Chart
C, P, NP, U Charts
Capability Analysis
Cusum, EWMA Chart
Individuals Chart
Moving Average Chart
Pareto Chart
R & R Studies

Survival / Reliability

Accelerated Life Tests
Cox Regression
Cumulative Incidence
Exponential Fitting
Extreme-Value Fitting
Hazard Rates
Kaplan-Meier Curves
Life-Table Analysis
Lognormal Fitting
Log-Rank Tests
Probit Analysis
Proportional-Hazards
Reliability Analysis
Survival Distributions
Time Calculator*
Weibull Analysis

Multivariate Analysis

Cluster Analysis
Correspondence Analysis
Discriminant Analysis
Factor Analysis
Hotelling's T-Squared
Item Analysis
Item Response Analysis
Loglinear Models
MANOVA
Multi-Way Tables
Multidimensional Scaling
Principal Components

Curve Fitting

Bootstrap C.I.'s*
Built-In Models
Group Fitting and Testing*
Model Searching
Nonlinear Regression
Randomization Tests*
Ratio of Polynomials
User-Specified Models

Miscellaneous

Area Under Curve
Bootstrapping
Chi-Square Test
Confidence Limits
Cross Tabulation
Data Screening
Fisher's Exact Test
Frequency Distributions
Mantel-Haenszel Test
Nonparametric Tests
Normality Tests
Probability Calculator
Proportion Tests
Randomization Tests
Tables of Means, Etc.
Trimmed Means
Univariate Statistics

Meta-Analysis*

Independent Proportions*
Correlated Proportions*
Hazard Ratios*
Means*

Binary Diagnostic Tests*

One Sample*
Two Samples*
Paired Samples*
Clustered Samples*

Proportions

Tolerance Intervals*
Two Independent*
Two Correlated*
Exact Tests*
Exact Confidence Intervals*
Farrington-Manning*
Fisher Exact Test
Gart-Nam* Method
McNemar Test
Miettinen-Nurminen*
Wilson's Score* Method
Equivalence Tests*
Noninferiority Tests*

Mass Appraisal

Comparables Reports
Hybrid (Feedback) Model*
Nonlinear Regression
Sales Ratios

***New Edition in 2004**

"Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away."

- Antoine de Saint Exupery

F is a carefully crafted subset of the most recent version of Fortran, the world's most powerful numeric language.

Using F has some very significant advantages:

- Programs written in F will compile with any Fortran compiler
- F is easier to use than other popular programming languages
- ***F compilers are free*** and available for Linux, Windows, and Solaris
- Several books on F are available
- F programs may be linked with C, Fortran 95, or older Fortran 77 programs

F retains the modern features of Fortran—modules and data abstraction, for example—but discards older error-prone facilities of Fortran.

It is a safe and portable programming language.

F encourages Module-Oriented Programming.

It is ideal for teaching a programming language in science, engineering, mathematics, and finance.

It is ideal for new numerically intensive programs.

The Fortran Company

11155 E. Mountain Gate Place, Tucson, AZ 85749 USA

+1-520-256-1455 +1-520-760-1397 (fax)

<http://www.fortran.com> info@fortran.com

Introducing GGUM2004

Item Response Theory Models for Unfolding

The new GGUM2004 software system estimates parameters in a family of item response theory (IRT) models that unfold polytomous responses to questionnaire items. These models assume that persons and items can be jointly represented as locations on a latent unidimensional continuum. A single-peaked, nonmonotonic response function is the key feature that distinguishes unfolding IRT models from traditional, "cumulative" IRT models. This response function suggests

that a higher item score is more likely to the extent that an individual is located close to a given item on the underlying continuum. Such single-peaked functions are appropriate in many situations including attitude measurement with Likert or Thurstone scales, and preference measurement with stimulus rating scales. This family of models can also be used to determine the locations of respondents in particular developmental processes that occur in stages.

The GGUM2004 system estimates item parameters using marginal maximum likelihood, and person parameters are estimated using an expected *a posteriori* (EAP) technique. The program allows for up to 100 items with 2-10 response categories per item, and up to 2000 respondents. GGUM2004 is compatible with computers running updated versions of Windows 98 SE, Windows 2000, and Windows XP. The software is accompanied by a detailed technical reference manual and a new Windows user's guide. **GGUM2004 is free** and can be downloaded from:

<http://www.education.umd.edu/EDMS/tutorials>

GGUM2004 improves upon its predecessor (GGUM2000) in several important ways:

- It has a user-friendly graphical interface for running commands and displaying output.
- It offers real-time graphics that characterize the performance of a given model.
- It provides new item fit indices with desirable statistical characteristics.
- It allows for missing item responses assuming the data are missing at random.
- It allows the number of response categories to vary across items.
- It estimates model parameters more quickly.

Start putting the power of unfolding IRT models to work in your attitude and preference measurement endeavors. Download your free copy of GGUM2004 today!

JOIN DIVISION 5 OF APA!

The Division of Evaluation, Measurement, and Statistics of the American Psychological Association draws together individuals whose professional activities and/or interests include assessment, evaluation, measurement, and statistics. The disciplinary affiliation of division membership reaches well beyond psychology, includes both members and non-members of APA, and welcomes graduate students.

Benefits of membership include:

- subscription to *Psychological Methods* or *Psychological Assessment* (student members, who pay a reduced fee, do not automatically receive a journal, but may do so for an additional \$18)
- *The Score* – the division's quarterly newsletter
- Division's Listservs, which provide an opportunity for substantive discussions as well as the dissemination of important information (e.g., job openings, grant information, workshops)

Cost of membership: \$38 (**APA membership not required**); student membership is only \$8

For further information, please contact the Division's Membership Chair, Yossef Ben-Porath (ybenpora@kent.edu) or check out the Division's website:

<http://www.apa.org/divisions/div5/>

ARE YOU INTERESTED IN AN ORGANIZATION DEVOTED TO EDUCATIONAL AND BEHAVIORAL STATISTICS?

Become a member of the **Special Interest Group - Educational Statisticians** of the American Educational Research Association (SIG-ES of AERA)!

The mission of SIG-ES is to increase the interaction among educational researchers interested in the theory, applications, and teaching of statistics in the social sciences.

Each Spring, as part of the overall AERA annual meeting, there are seven sessions sponsored by SIG-ES devoted to educational statistics and statistics education.
We also publish a twice-yearly electronic newsletter.

Past issues of the SIG-ES newsletter and other information regarding SIG-ES can be found at <http://orme.uark.edu/edstatsig.htm>

To join SIG-ES you must be a member of AERA. Dues are \$5.00 per year.

For more information, contact Joan Garfield, President of the SIG-ES, at jbg@umn.edu.

SOFTWARE SOLUTIONS
for Science & Engineering

Lahey/Fujitsu Fortran

The standard for Fortran programming
from the leader in Fortran language systems

LF95 Fortran for Linux and Windows

Full Fortran 95/90/77 support
Unsurpassed diagnostics
Intel and AMD optimizations

IMSL compatible
Fujitsu SSL2 math library
Wisk graphics package

LF Fortran for the Microsoft® .NET Framework - Coming Soon !

Visual Studio integration
Windows / Web Forms designer
Project and code templates

On-line integrated help
XML Web services
ADO.NET support

Visit www.lahey.com for more information

```
ELSE
  poly_coefficient
END IF
ELSE
  poly_coefficient
END IF
END FUNCTION poly_c
SUBROUTINE poly_init
TYPE(poly), INTENT
REAL(fpkind), INTE
IF ( .NOT. PRESENT
  NULLIFY ( p%coef
ELSE
  m = UBOUND(v,l)
  IF ( max_degree
  ALLOCATE ( p%
  p%coeffs
ELSE
  ALLOCATE ( p%
  p%coeffs
END IF
END IF
```

Lahey Computer Systems, Inc.
865 Tahoe Blvd - P.O. Box 6091
Incline Village, NV 89450 USA
1-775-831-2500
www.lahey.com

Instructions For Authors

Follow these guidelines when submitting a manuscript:

1. *JMASM* uses a modified American Psychological Association style guideline.
2. Submissions are accepted via e-mail only. Send them to the Editorial Assistant at ea@edstat.coe.wayne.edu. Provide name, affiliation, address, e-mail address, and 30 word biographical statements for all authors in the body of the email message.
3. There should be no material identifying authorship except on the title page. A statement should be included in the body of the e-mail that, where applicable, indicating proper human subjects protocols were followed, including informed consent. A statement should be included in the body of the e-mail indicating the manuscript is not under consideration at another journal.
4. Provide the manuscript as an external e-mail attachment in MS Word for the PC format only. (Wordperfect and .rtf formats may be acceptable - please inquire.) Please note that Tex (in its various versions), Exp, and Adobe .pdf formats are designed to produce the final presentation of text. They are not amenable to the editing process, and are **NOT** acceptable for manuscript submission.
5. The text maximum is 20 pages double spaced, not including tables, figures, graphs, and references. Use 11 point Times Roman font.
6. Create tables without boxes or vertical lines. Place tables, figures, and graphs “in-line”, not at the end of the manuscript. Figures may be in .jpg, .tif, .png, and other formats readable by Adobe Illustrator or Photoshop.
7. The manuscript should contain an Abstract with a 50 word maximum, following by a list of key words or phrases. Major headings are Introduction, Methodology, Results, Conclusion, and References. Center headings. Subheadings are left justified; capitalize only the first letter of each word. Sub-subheadings are left-justified, indent optional.
8. Do not use underlining in the manuscript. Do not use bold, except for (a) matrices, or (b) emphasis within a table, figure, or graph. Do not number sections. Number all formulas, tables, figures, and graphs, but do not use italics, bold, or underline. Do not number references. Do not use footnotes or endnotes.
9. In the References section, do not put quotation marks around titles of articles or books. Capitalize only the first letter of books. Italicize journal or book titles, and volume numbers. Use “&” instead of “and” in multiple author listings.
10. *Suggestions for style:* Instead of “I drew a sample of 40” write “A sample of 40 was selected”. Use “although” instead of “while”, unless the meaning is “at the same time”. Use “because” instead of “since”, unless the meaning is “after”. Instead of “Smith (1990) notes” write “Smith (1990) noted”. Do not strike spacebar twice after a period.

Print Subscriptions

Print subscriptions including postage for professionals are US \$95 per year; for graduate students are US \$47.50 per year; and for libraries, universities, and corporations are US \$195 per year. Subscribers outside of the US and Canada pay a US \$10 surcharge for additional postage. Online access is currently free at <http://tbf.coe.wayne.edu/jmasm>. Mail subscription requests with remittances to JMASM, P. O. Box 48023, Oak Park, MI, 48237. Email journal correspondence, other than manuscript submissions, to jmasm@edstat.coe.wayne.edu.

Notice To Advertisers

Send requests for advertising information to jmasm@edstat.coe.wayne.edu.

STATISTICIANS

HAVE YOU VISITED THE

Mathematics Genealogy Project?

The Mathematics Genealogy Project is an ongoing research project tracing the intellectual history of all the mathematical arts and sciences through an individual's Ph.D. advisor and Ph.D. students. Currently we have over 80,000 records in our database. We welcome and encourage all statisticians to join us in this endeavor.

Please visit our web site

<http://genealogy.math.ndsu.nodak.edu>

The information which we collect is the following:

The full name of the individual, the school where he/she earned a Ph.D., the year of the degree, the title of the dissertation, and, MOST IMPORTANTLY, the full name of the advisor(s). E.g., Fuller, Wayne Arthur; Iowa State University; 1959; *A Non-Static Model of the Beef and Pork Economy*; Shepherd, Geoffrey Seddon

For additions or corrections for one or two people a link is available on the site. For contributions of large sets of names, e.g., all graduates of a given university, it is better to send the data in a text file or an MS Word file or an MS Excel file, etc. Send such information to:

harry.coonce@ndsu.nodak.edu

The genealogy project is a not-for-profit endeavor supported by donations from individuals and sales of posters and t-shirts. If you would like to help this cause please send your tax-deductible contribution to: Mathematics Genealogy Project, 300 Minard Hall, P. O. Box 5075, Fargo, North Dakota 58105-5075E

Devereux Early Childhood Assessment for Infants and Toddlers DECA-I/T

What is the Devereux Early Childhood Assessment for Infants and Toddlers?

The DECA-I/T is a premier tool for assessing protective factors and screening for potential risks in the social and emotional development of very young children.

The DECA-I/T is a standardized, reliable and valid assessment.

Available
Spring 2007

There are two instruments
that make up the DECA-I/T:

- ~ An assessment for Infants from 1 month up to 18 months
- ~ An assessment for Toddlers from 18 months up to 36 months

#14110	DECA-I/T kit	\$149.95
#13844	DECA-I/T Infant Record Forms, Set of 20 (for children ages 1 month up to 18 months)	\$19.95
#13845	DECA-I/T Toddler Record Forms, Set of 30 (for children ages 18 months up to 36 months)	\$29.95
#13846	DECA-I/T User's Guide	\$39.95
#14016	DECA-I/T Set of 5 Parent/Teacher Profile Masters (4 infant and 1 toddler)	\$12.95
#14112	DECA-I/T CD Scoring Assistant and Technical Manual	\$99.95

EARLY LEARNING COMPANY

1-800-334-2014 • www.kaplanco.com

The easy way to find open access journals

www.doaj.org

The Directory of Open Access Journals covers free, full text, quality controlled scientific and scholarly journals. It aims to cover all subjects and languages.

Aims

- Increase visibility of open access journals
- Simplify use
- Promote increased usage leading to higher impact

Scope

The Directory aims to be comprehensive and cover all open access scientific and scholarly journals that use a quality control system to guarantee the content. All subject areas and languages will be covered.

In DOAJ browse by subject

Agriculture and Food Sciences
Biology and Life Sciences
Chemistry
General Works
History and Archaeology
Law and Political Science
Philosophy and Religion
Social Sciences

Arts and Architecture
Business and Economics
Earth and Environmental Sciences
Health Sciences
Languages and Literatures
Mathematics and statistics
Physics and Astronomy
Technology and Engineering

Contact

Lotte Jørgensen, Project Coordinator
Lund University Libraries, Head Office
E-mail: lotte.jorgensen@lub.lu.se
Tel: +46 46 222 34 31

Funded by

www.soros.org

Hosted by

LUND
UNIVERSITY
www.lu.se