Journal of Modern Applied Statistical Methods

Volume 5 | Issue 2 Article 8

11-1-2005

Determining Parallel Analysis Criteria

Marley W. Watkins
The Pennsylvania State University, mww10@psu.edu

Journal of Modern Applied Statistical Methods
November, 2006, Vol. 5, No. 2, 344-346

1538 – 9472/06/$95.00

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).

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.

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.

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 Jocobi 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.