Digital Commons @ Wayne State

Journal of Modern Applied Statistical Methods

Volume 1 | Issue 2 Article 40

11-1-2002

A Comparison Of The D'Agostino S_u Test To The Triples Test For Testing Of Symmetry Versus Asymmetry As A Preliminary Test To Testing The Equality Of Means

Kimberly T. Perry
Pharmacia Corporation, Kalamazoo, MI

Michael R. Stoline
Western Michigan University

Recommended Citation

Perry, Kimberly T. and Stoline, Michael R. (2002) "A Comparison Of The D'Agostino S_u Test To The Triples Test For Testing Of Symmetry Versus Asymmetry As A Preliminary Test To Testing The Equality Of Means," *Journal of Modern Applied Statistical Methods*: Vol. 1 : Iss. 2 , Article 40.
DOI: 10.22237/jmasm/1036109040
Available at: http://digitalcommons.wayne.edu/jmasm/vol1/iss2/40

Journal of Modern Applied Statistical Methods
Fall 2002, Vol. 1, No 2, 316-325

1538 – 9472/02/$30.00

A Comparison Of The D'Agostino S_u Test To The Triples Test For Testing Of Symmetry Versus Asymmetry As A Preliminary Test To Testing The Equality Of Means

Kimberly T. Perry
Pharmacia Corporation
Kalamazoo, Michigan

Michael R. Stoline
Department of Statistics
Western Michigan University

This paper evaluates the D'Agostino S_U test and the Triples test for testing symmetry versus asymmetry. These procedures are evaluated as preliminary tests in the selection of the most appropriate procedure for testing the equality of means with two independent samples under a variety of symmetric and asymmetric sampling situations.

Key words: symmetry; asymmetry; preliminary testing.

Introduction

The purpose of this paper is to evaluate the performance of two tests, the D'Agostino S_U test and the Triples test for the testing of symmetry versus asymmetry (or skewness) as a preliminary test using two levels of significance: $\alpha = 0.05$ and $\alpha = 0.25$. The results could be used to select a method for testing the equality of two means, Ho: $\mu_1 = \mu_2$, based on two classes of preliminary tests: (1) a test of variance homogeneity, and (2) a test of symmetry.

Procedures for the D'Agostino S_U test and the Triples test for symmetry are given below, as well as details of the four symmetric distributions and five asymmetric distributions used in the simulations. Results of a simulation study comparing the two tests for the one - sample cases and as well as preliminary tests in two sample contexts are presented below.

Kimberly T. Perry is the Director of Clinical Biostatistics, Pharmacia Corporation, Kalamazoo, Michigan. Her areas of interest are innovated clinical study designs, multiple endpoint analysis, and interim analysis. Email: kimberly.t.perry @pharmacia.com. Michael R. Stoline is Professor, Department of Statistics, Western Michigan University, Kalamazoo, Michigan. His areas of expertise are: multiple comparisons, analysis of variance, and regulatory environmental statistics. Email address: stoline@wmich.edu

Methodology

Testing of Symmetry Versus Skewness

The D'Agostino test and the Triples test of symmetry are described first for a general random sample $x_1, \ldots, x_n$ from some distribution $f\ (x;\ \mu, \sigma)$. It is convenient to let $\bar{x}$ denote the sample mean of $x_1, \ldots, x_n$ and to let the sample estimates of $\beta_1^{1/2}$, the third standardized moment, and β_2, the fourth standardized moment, be denoted as

$$b_1^{1/2} = m_3 / m_2^{3/2}, \qquad (1)$$

and $$b_2 = m_4 / m_2^2, \qquad (2)$$

where $m_k = \Sigma(x_i - \bar{x})^k / n$ for k = 2, 3, 4. (3)

D'Agostino's Skewness Test

D'Agostino's test is a test of normality versus non-normality, which is sensitive to skewed nonnormal alternatives. A sketch of this procedure is now described.

First, compute $b_1^{1/2}$ from the sample data. Secondly compute $Z(b_1^{1/2})$, where

$$Z(b_1^{1/2}) = \delta \ln(Y/a + [(Y/a)^2 + 1]^{½}), \qquad (4a)$$

$$Y = b_1^{1/2}\left[\frac{(n+1)(n+3)}{6(n-2)}\right], \quad (4b)$$

$$W^2 = -1 + [2(\beta_2(b_1^{1/2}) - 1)]^{1/2}, \quad (4c)$$

$$\beta_2\left(b_1^{1/2}\right) = \frac{3(n^2+27n-70)(n+1)}{(n-2)(n+5)(n+7)(n+9)}, \quad (4d)$$

$$\delta = 1/(\ln W)^{1/2} \text{ and } a = 2/(W^2-1)^{1/2}. \quad (4e)$$

The α-level D'Agostino test of skewness is:

$$Z(b_1^{1/2}) > z_\alpha, \quad (5)$$

where z_α is the upper α-point of the standard unit normal. $Z((b_1)^{1/2})$ is approximately $n(0, 1)$ under the null hypothesis of population normality for cases where $n > 8$ (D'Agostino, Belanger, & D'Agostino, Jr., 1990).

Results from D'Agostino's Monte Carlo simulations for $n < 25$ and checks with an existing table of Pearson and Hartley (1966) for $n \geq 25$ show that the accuracy of the transformation is very good. Therefore, due to its sensitivity to skewed nonnormal alternatives, the D'Agostino test was chosen as a possible preliminary test for symmetry/skewness.

Triples Test

The Triples test is described in a paper by Randles, Fligner, Policello, and Wolfe (1980). Let $x_i, \ldots, x_n$ denote a random sample from a continuous population where i, j, k are distinct integers such that $1 \leq i < j < k \leq n$. The Triples test is an asymptotically distribution-free procedure which examines each triple (x_i, x_j, x_k). If the middle observation is closer to the smaller observation than it is to the largest observation, then a "right triple" is formed (looks skewed to the right). If the middle observation is closer to the larger observation than it is to the smaller observation, then a "left triple" is formed (looks skewed to the left). The Triples test statistic is a function of the number of right triples and left triples.

The Triples test rejects H_o of symmetry if $|T_1| > t_{n, (\alpha/2)}$ where $t_{n, (\alpha/2)}$ is the upper α/2 point of a t distribution with n degrees of freedom,

$$T_1 = n^{1/2}\ \hat{\eta} / \hat{\sigma}_n, \quad (6a)$$

$$\hat{\eta} = \frac{\{(\text{number of right triples}) - (\text{number of left triples})\}}{\left[3\binom{n}{3}\right]} \quad (6b)$$

and $\hat{\sigma}_n$ is the standard deviation of $\hat{\eta}$. The statistic $\hat{\eta}$ is calculated as

$$\hat{\eta} = \binom{n}{3}^{-1} \sum_{i<j<k} f^*(x_i, x_j, x_k) \quad (7)$$

where $f^*(x_i, x_j, x_k) = \{\text{sign}(x_i + x_j - 2x_k) + \text{sign}(x_i + x_k - 2x_j) + \text{sign}(x_j + x_k - 2x_i)\}/3$, and sign($u$) = -1, 0, or 1 as u <, =, or > 0.

To compute var $(\hat{\eta}) = \hat{\sigma}_n^2$, let

$$\frac{\hat{\sigma}_n^2}{n} = \binom{n}{3}^{-1} \sum_{c=1}^{3} \binom{3}{c}\binom{n-3}{3-c}\hat{\xi}_c \quad (8a)$$

where $\hat{\xi}_c = \text{var}\,[f_c^*(x_1, \ldots, x_c)]$. (8b)

Then $\hat{\xi}_1 = var\,[\ f_1^*(x_1)\,]$, with $f_1^*(x) = \text{E}\,[f^*(x, x_2, x_3)]$, yields

$$\hat{\xi}_1 = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_1^*\ (x_i) - \hat{\eta})^2, where \quad (9a)$$

$$\hat{f}_1^*(x_i) = \frac{1}{\binom{n-1}{2}} \sum_{\substack{j<k \\ j\neq i\neq k}} f^*(x_i, x_j, x_k). \quad (9b)$$

Similarly,

$$\hat{\xi}_2 = \frac{1}{\binom{n}{2}} \sum_{j<k} (\hat{f}_2^*(x_j, x_k) - \hat{\eta})^2, where \quad (10a)$$

$$f_2^*(x_j, x_k) = \frac{1}{n-2} \sum_{\substack{i=1 \\ i\neq j\neq k \\ i\neq k}} f^*(x_i, x_j, x_k), \quad (10b)$$

and $\hat{\xi}_3 = \frac{1}{9} - \hat{\eta}^2$. (11)

Randles, et al. (1980) compared three procedures for testing whether a univariate population is symmetric about some unspecified value compared to an immense class of asymmetric distribution alternatives. The Triples test was compared to Gupta's skewness test (Gupta, 1967) and Gupta's nonparametric procedure (Gupta, 1967). Randles et al. (1980) show that the Triples Test is superior to either competitor, even for sample sizes as small as 20, while possessing good power for detecting asymmetric alternative distributions.

Cabilio & Masaro (1996) compared their symmetry test, S_K , to several other tests of symmetry including the Triples test. The Triples test again performed well and therefore, is selected as a second possible preliminary test of symmetry/skewness.

Generation of Random Realizations From Six Distributions

This section contains details of how the random realizations are generated for each specified distribution among members of the normal, uniform, double exponential, logistic, lognormal, and gamma families of random variables used in the simulations. The normal, uniform, double exponential, and logistic are symmetric; the lognormal and gamma are asymmetic.

For one-sample cases, it is convenient to let $x_1, \ldots, x_n$ be a random sample of size n from $f(x)$. Let the sample mean and sample standard deviation be denoted as $\bar{x}$ and s, respectively.

The IMSL random number generator RNSET, which initializes the seed, is used in all of the simulations.

Normal Distribution

In the case of the normal distribution, population means are set to zero, $\mu = 0$ with unit standard deviations, $\sigma = 1$. The distribution $f(x)$ is normal (0, 1). The FORTRAN function RNNOF was used to generate the normal (0, 1) random numbers.

Uniform Distribution

Let x be uniform (a, b) with mean $\mu = (a + b)/2$ and standard deviation $\sigma = (b - a) / \sqrt{12}$. The uniform distribution $f(x)$ used in the simulations is a uniform (-1/2, 1/2) distribution yielding a mean μ= 0 and standard deviation $\sigma = 1/\sqrt{12}$.

The random numbers u_i from a uniform (0,1) distribution are first generated using the FORTRAN function RNUN. The uniform (-1/2, 1/2) random realizations are then generated using the transformation:

$$x_i = (u_i - ½) \quad (12)$$

Double Exponential Distribution

Let x have the double exponential probability density function $f(x)$ where

$$f(x) = \frac{\exp[-|x|]}{2}, \; -\infty < x < \infty . \quad (13)$$

The mean and variance are

$$\mu = 0 \text{ and} \quad (14)$$

$$\sigma^2 = 2. \quad (15)$$

To simulate x for this double exponential distribution, we use the following transformation:

$$x = (y_1 - y_2)/2 \quad (16)$$

where y_1 and y_2 are two independent chi-square random variables, each with two degrees of freedom. The two degree of freedom chi-squared random number y is generated as

$$y = -2 \ln (u) \quad (17)$$

where u is an independent random number from a uniform (0,1) distribution (see Uniform Distribution subsection).

Logistic Distribution

Let $f(x)$ represent the probability density function for a logistic distribution

$$f(x) = \frac{e^x}{\left(1+e^x\right)^2} \quad where\ -\infty \le x \le \infty . \qquad (18)$$

The mean and variance are

$$\mu = 0 \text{ and} \qquad (19)$$

$$\sigma^2 = 3/\pi^2 \qquad (20)$$

The random numbers x_i for this logistic distribution are generated using the transformation

$$x_i = \frac{\sqrt{3}}{\pi}\log\left(\frac{u_i}{1-u_i}\right) \qquad (21)$$

where u_i is uniform (0,1).

Lognormal Distribution

The probability density function for the lognormal distribution with parameters a and b is:

$$f(x) = \ln(x;a,b) = \frac{1}{b\,x\,(2\pi)^{1/2}}\exp\left(-\frac{1}{2b^2}(\ln x - a)^2\right)$$
$$\text{for } x > 0. \qquad (22)$$

The mean μ, variance σ^2, and coefficient of skewness are

$$\mu = \exp\left(a + \frac{b^2}{2}\right) \qquad (23)$$

$$\sigma^2 = w(w-1)\exp(2a) \text{, and} \qquad (24)$$

$$\text{coefficient of skewness} = (w+2)\,(w-1)^{½} \qquad (25)$$

where $w = \exp(b^2)$. Let y be $n(a, b)$, which designates a normally distributed variable with mean a and standard deviation b, then $x = e^y$ has the lognormal probability density function $ln(x; a, b)$ in (22).

Three lognormal distributions are selected due to their varying degrees of skewness. In each of the three cases, the sample from the lognormal distribution $ln(x; a, b)$, denoted as lognormal (a, b), has a set to zero. The three b parameter values chosen are: (1) $b = 0.4$, (2) $b = 1.0$, and (3) $b = 1.75$. The coefficient of skewness for these cases are 1.3, 6.2, and 105.6, respectively. The case of $b = 0.4$ is denoted as slight skewness, $b = 1.0$ as moderate skewness, and $b = 1.75$ as heavy skewness.

The FORTRAN function RNLNL is used to create the random realizations for the ln $(x; a, b)$ distributions using the transformation $x = e^y$, where y is $n(a, b)$ (IMSL, STAT/Library, 1989).

Gamma Distribution

The probability density function for the gamma distribution with shape parameter α and scale parameter β is

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right) \qquad (26)$$
$$where\ x > 0, \alpha > 0, \beta > 0$$

with mean $\alpha\beta$, variance $\alpha\beta^2$ and coefficient of skewness $2(\alpha)^{-1/2}$.

Two gamma distributions are selected, one with shape parameter equal to 3 and unit scale parameter (denoted as G(3,1)), and the other with shape parameter equal to 2 and unit scale parameter (denoted as G(2,1)). The G(3,1) distribution is only slightly skewed (coefficient of skewness = 1.15), whereas the skewness is more pronounced in the G(2,1) distribution (coefficient of skewness = 1.41).

The gamma random realizations are generated using RNGAM (IMSL Routine) which yields random numbers with shape parameter α and unit scale parameter ($\beta = 1$).

Results

Results For Testing of Symmetry Versus Asymmetry For One Sample Cases

The robustness and the power of the D'Agostino S_U test for skewness at significance levels of $\alpha = 0.05$ and 0.25, denoted D(α), and the Triples test for symmetry at significance levels of $\alpha = 0.05$ and 0.25, denoted as T(α), are examined in this section for the one sample cases.

To assess the Type I error, the simulated null rejection rates are examined for the four symmetric distributions (normal, uniform, double exponential, and logistic). The Type I error simulated results for the two procedures are presented below. The five asymmetric distributions (lognormal (0,0.4), lognormal (0,1), lognormal (0,1.75), gamma (3,1) and gamma (2,1)) are used to investigate the power. The simulated power results for the two tests, and discussion of the one sample results also appear below.

Type I Error Comparisons in One Sample Case

For the one sample cases, n random realizations are generated from each of the four symmetric distributions for each of three samples: n = 10, 20, or 40. The hypothesis of symmetry is tested using the D'Agostino S_U test and the Triples test.

The two procedures are compared for control of significance level at two levels: $\alpha = 0.05$ and $\alpha = 0.25$ using the four symmetric distributions. A total of 10,000 simulation runs are obtained for each of the three sample sizes for each of the four symmetric distributions. Hence, twelve simulated Type I error p-values are obtained for the Triples test for the $\alpha = 0.05$ cases, and twelve simulated p-values are also obtained for the $\alpha = 0.25$ cases. Likewise, twelve simulated Type I error p-values are obtained for the D'Agostino S_U test for each of these two levels.

Significant Level Testing at 5%

For the 5% significance-level testing cases, the simulated Type I error rates (expressed as percentages) are categorized into one of the following five 5% significance level categories:

1. $x \leq 2.5$ (extremely conservative) (27)
2. $2.5 < x \leq 4.0$ (slightly conservative)
3. $4.0 < x \leq 6.0$ (robust)
4. $6.0 < x \leq 10.0$ (slightly liberal)
5. $x > 10.0$ (extremely liberal)

The value "x" represents the percentage of rejections for testing H_o: symmetry based on the 10,000 simulations. A value "x" is obtained for each sample size and symmetric distribution combination for each procedure. Hence, twelve x values were obtained for the T(.05) cases, and twelve for the D(.05) cases.

The five 5% significance-level testing categories in (27) are labeled as robust, conservative (slightly or extremely), and liberal (slightly or extremely). These five Type I error categories are now further defined.

The outcome of the D(.05) test and the T(.05) test for a particular symmetric case is defined to be robust if the simulated null rejection rate is > 4.0 and ≤ 6.0. The outcome of the D(.05) and the T(.05) test is defined to be slightly conservative if the simulated null rejection rate is > 2.5 and ≤ 4.0; and extremely conservative if the simulated null rejection rate is ≤ 2.5. Likewise, the test is categorized as slightly liberal if the simulated null rejection rate is > 6.0 and ≤ 10.0; and extremely liberal if the simulated rejection rate is > 10.0.

The frequency and percentage of simulated Type I error rates observed in each of the five categories: $a < x \leq b$ (given in (27)) is presented in Table 1 for the D(.05) and T(.05) tests.

Significance Level Testing at 25%

For the D(.25) test and the T(.25) test, the percentages of rejections (%) is tabulated for the five categories listed below:

1. $x \leq 12.5$ (extremely conservative) (28)
2. $12.5 < x \leq 17.5$ (slightly conservative)
3. $17.5 < x \leq 32.5$ (robust)
4. $32.5 < x \leq 37.5$ (slightly liberal)
5. $x > 37.5$ (extremely liberal)

The outcome of the D(.25) test and the T(.25) test for the symmetric cases is defined to be robust if the simulated null rejection rate is > 17.5 and ≤ 32.5. The definitions for the conservative and liberal classifications in (28) for the D(.25) and T(.25) tests are similar to those defined in (27) for the D(.05) and T(.05) cases.

The frequency and percentage of simulated Type I error rates observed in each of the categories: $a < x \leq b$ (given in (28)) are also presented in Table 2 for the D(.25) and T(.25) tests.

Discussion of Robustness for Symmetric Cases

Tables 1 and 2 show that the Triples test is more robust than the D'Agostino S_U test for symmetric cases, especially for $\alpha = 0.25$ testing. The T(.25) test is robust in 91.7% (11 of 12) of the cases compared to 33.3% (4 of 12) of the cases for the D(.25) test. The T(.05) test is robust in 41.6% (5 of 12) of the cases compared to 25.0% (3 of 12) for the D(.05) test.

Table 1. Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) for Symmetry Versus Asymmetry Tests With Nominal 5% Level--One Sample Cases.

Test	Extremely Conservative ≤2.5	Slightly Conservative >2.5, ≤4.0	Robust >4.0, ≤6.0	Slightly Liberal >6.0, ≤10	Extremely Liberal >10
D(.05)	3 (25.0%)	0 (0.0%)	3 (25.0%)	0 (0.0%)	6 (50.0%)
T (.05)	3 (25.0%)	2 (16.7%)	5 (41.6%)	2 (16.7%)	0 (0.0%)

Note: Table 1 results are based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and three sample sizes (n = 10, 20 and 40).

Table 2. Summary of Symmetric Distributions: Frequency of Simulated Null Rejection Rate (%) for Symmetry Versus Asymmetry Tests With Nominal 25% Level--One Sample Cases.

Test	Extremely Conservative ≤12.5	Slightly Conservative >12.5, ≤17.5	Robust >17.5, ≤32.5	Slightly Liberal >32.5, ≤37.5	Extremely Liberal >37.5
D(.25)	2 (16.7%)	1 (8.3%)	4 (33.3%)	1 (8.3%)	4 (33.3%)
T (.25)	0 (0.0%)	1 (8.3%)	11 (91.7%)	0 (0.0%)	0 (0.0%)

Note: Table 2 results are based on the four symmetric distributions (normal, uniform, double exponential, and logistic) and three sample sizes (n = 10, 20 and 40).

Tables 1 and 2 also show that the D'Agostino S_U test is appreciably more liberal than the Triples test for symmetric cases. The D(.05) test is observed to be liberal in 50.0% (6 of 12) of the cases compared to 16.7% (2 of 12) for the T(.05) test. Also, the D(.25) test is observed to be liberal in 41.6% (5 of 12) of the cases compared to 0.0% (0 of 12) of the T(.25) cases.

On the basis of the results presented in Tables 1 and 2, it is concluded that the Triples Test is superior to the D'Agostino S_U test for controlling Type I error. It is also concluded that the D'Agostino S_U test does not control the Type I error rate for symmetric cases since it fails to maintain the Type I error rate at or below the stated level of significance.

Results of Power Analysis in One Sample Cases

The results of a power comparison of the D'Agostino S_U test and the Triples test is now reported. A total of 10,000 simulation runs are obtained for each of the three sample sizes n = 10, 20, and 40 for each of the five asymmetric distributions. Hence, fifteen simulated power p-values are obtained for the Triples test for each of the T(.05), T(.25), and D(.05), and D(.25) cases.

Definition of Power Categories

The results of the simulation for the five asymmetric distributions are combined in Table 3 over all sample sizes for the four power categories defined below:

1. $x \leq 50.0$ (low power) (29)
2. $50.0 < x \leq 75.0$ (moderate power)
3. $75.0 < x \leq 90.0$ (high power)
4. $x > 90.0$ (extremely high power)

The value "x" represents the power to detect asymmetry based on 10,000 simulations for each sample size configuration. Each entry in Table 3 denotes both the frequency and percentage at which $a < x \leq b$ occurs, as in Table 1.

The four power categories in (29) are conveniently labeled in order of increasing power: low power (power <50%), moderate power (50% < power ≤ 75%), high power (75% < power ≤ 90%), and extremely high power (power > 90%).

Table 3. Summary of Asymmetric Distributions: Frequency of Simulated Power Rates (%) for Symmetry Versus Asymmetry Tests With Nominal 5% and 25% Levels, One Sample Cases.

Test	Low Power ≤50.0	Moderate Power >50.0, ≤75.0	High Power >75.0, ≤90.0	Extremely High Power >90.0
		Nominal 5% Level		
D(.05)	6 (40.0%)	3 (20.0%)	3 (20.0%)	3 (20.0%)
T(.05)	7 (46.7%)	3 (20.0%)	2 (13.3%)	3 (20.0%)
		Nominal 25% Level		
D(.25)	2 (13.3%)	3 (20.0%)	3 (20.0%)	7 (46.7%)
T(.25)	3 (20.0%)	4 (26.7%)	2 (13.3%)	6 (40.0%)

Note: Table 3 results are based on the asymmetric distributions [lognormal (0, 0.40), lognormal (0, 1.0), lognormal (0, 1.75), G(3,1), and G(2,1)] and three sample sizes (n = 10, 20 and 40).

These four power categories are used in Table 3 for both 5% and 25% results.

Discussion of Power for Asymmetric Cases

Table 3 shows that both the T(.05) and D(.05) tests lack power. The power is ≤ 0.75 for 60% of the cases when using the D(.05) test, and is ≤ 0.75 for 66.7% of the cases when using the T(.05) test. The D(.05) test is generally more powerful then the T(.05) test for asymmetric cases.

The D(.25) test tends to be somewhat more powerful than the T(.25) test. The power is > .90 for approximately 47% of the cases when using the D(.25) test compared to 40% of the cases when using the T(.25) test. In addition, the power is ≤ 0.50 for 20% of the cases when using the T(.25) test compared to approximately 13% when using the D(.25) test.

It is concluded that the D'Agostino S_U test is somewhat more powerful than the Triples test for detecting asymmetric distributions.

Discussion of One Sample Simulation Results

Table 4 contains summary statistics describing the mean, standard deviation (denoted as *s*), minimum, and maximum of the four sets of twelve simulated p-values obtained by using the D(.05), T(.05), D(.25), and T(.25) procedures for the symmetric cases. The symmetric case summary statistics can be used to characterize the Type I error properties of these test procedures. The symmetric mean p-value is denoted as $\bar{p}_s$ in Table 4.

Table 5 also contains the corresponding summary statistics of the four sets of fifteen simulated p-values obtained by the same four test procedures for the asymmetric cases. The asymmetric case summary statistics can be used to characterize the power properties of these procedures. The asymmetric mean p-value is denoted as $\bar{p}_a$ in Table 5.

For the symmetric cases summarized in Table 4, the average Type I error rates for the T(.05) and T(.25) procedures are $\bar{p}_s$ = 4.1% and $\bar{p}_s$ = 21.5%, respectively, compared to $\bar{p}_s$ = 11.2% and $\bar{p}_s$ = 31.0% for the D(.05) and D(.25) procedures, respectively. The average Type I error rates for the Triples test are observed to be closer to the stated significance levels of 5% and 25% then are those for the D'Agostino S_U test.

For the symmetric cases summarized in Table 4, the standard deviations *s* and ranges of the p-values for the T(.05) and the T(.25) procedures are appreciably smaller than the comparable standard deviations and ranges for the D(.05) and the D(.25) procedures.

Table 4. Descriptive Statistics of the Simulated p-values for Four Test Procedures: D(.05), T(.05), D(.25), and T(.25) for Symmetric Cases (Summary statistics displayed as percentages)

Type I Error Summary Statsitics	Significance level 5%		Significance level 25%	
	D(.05)	T(.05)	D(.25)	T(.25)
$\overline{p_a}$	11.2	4.1	31.0	21.5
s	10.5	1.6	16.0	2.6
minimum	0.2	1.6	8.0	16.3
maximum	33.2	6.3	58.0	25.0
n	12	12	12	12

Table 5 contains the corresponding summary statistics of the four sets of fifteen simulated p-values obtained by the same four test procedures for the asymmetric cases. The asymmetric case summary statistics can be used to characterize the power properties of these procedures. The asymmetric mean p-value is denoted as $\overline{p}_a$ in Table 5.

Table 5. Descriptive Statistics of the Simulated p-values for Four Test Procedures: D(.05), T(.05), D(.25), and T(.25) for Asymmetric Cases (Summary statistics displayed as percentages).

Power Summary Statsitics	Significance level 5%		Significance level 25%	
	D(.05)	T(.05)	D(.25)	T(.25)
$\overline{p_a}$	60.4	52.6	80.0	75.2
s	30.0	34.2	20.1	24.0
minimum	16.8	5.7	44.3	33.0
maximum	100.0	100.0	100.0	100.0
n	15	15	15	15

Summary

For symmetric cases summarized in Tables 1,2, and 4, it is concluded that the Triples test is superior to the D'Agostino S_U test for the control of Type I error. The Triples test tends to hold to the stated level of significance. The D'Agostino S_U test does not hold to the stated level of significance and often tends to be liberal.

For the asymmetric cases summarized in Table 5, the average powers of the T(.05) and the T(.25) procedures are $\overline{p_a}$ = 52.6% and $\overline{p_a}$ = 75.2%, respectively, compared to $\overline{p_a}$ = 60.4% and $\overline{p_a}$ = 80.0%, respectively for the D(.05) the D(.25) procedures. The D'Agostino S_U test is observed to be slightly more powerful than the corresponding Triples test. The D'Agostino S_U test may be more powerful for asymmetric alternatives because the D'Agostino S_U test tends to be liberal with respect to Type I error control.

Testing Symmetry Versus Asymmetry In Preliminary Testing For Two Sample Cases

A purpose of this study is to select a preliminary test of testing symmetry versus asymmetry, and using the preliminary test to select the most appropriate method for testing the equality of two independent means H_o: $\mu_1 = \mu_2$. A two sample *t* procedure is commonly used if the underlying distributions are symmetric, and a Mann-Whitney-Wilcoxon (MWW) procedure may be more appropriate if the underlying distributions are asymmetric. The decision to use the *t* or the MWW procedure is often based on the personal preference of the investigator, or an examination of descriptive and graphical comparative statistics between the two samples.

Little evidence exists in the statistical literature of the use of tests of symmetry versus asymmetry as a preliminary test to select the *t* or MWW methods prior to testing H_o: $\mu_1 = \mu_2$. In these situations, the *t* procedure would be used if the preliminary test for skewness is non-significant; otherwise, the MWW procedure is used.

Two Sample Preliminary Testing Strategies

Assume there are two independent samples of sizes n_1 and n_2 from two distributions $f_1(x_1; \mu_1, \sigma_1)$ and $f_2(x_2; \mu_2, \sigma_2)$, respectively. Let us assume that the same skewness test is applied to the data from the two samples separately where the same significance level α is used for both tests.

Two preliminary testing protocols are defined. One utilizes the MWW test of H_o: $\mu_1 = \mu_2$ if at least one (ALO) of the two preliminary skewness tests is significant. The other utilizes the MWW test if both (BOTH) preliminary tests are significant. There two preliminary testing strategies are conveniently labeled: ALO and BOTH.

Selection of a Preliminary Testing Strategy

The one-sample simulation results summarized in Tables 4 and 5 are used to select a preliminary testing method between the BOTH and ALO protocols. For this purpose, it is convenient to utilize the average p-values: $\overline{p_s}$ and $\overline{p_a}$ p-values of the twelve symmetric and fifteen asymmetric distributions, respectively, summarized in Tables 4 and 5 for the D(.05), T(.05), D(.25), and T(.25) one-sample skewness test procedures.

Assuming symmetry (SYM) is true, the probability of correct selection of the *t* method for testing H_o: $\mu_1 = \mu_2$ is approximately given as:

$$1 - \overline{p_s}^2 \text{ for the BOTH method, and} \quad (30a)$$

$$(1 - \overline{p_s})^2 \text{ for the ALO method.} \quad (30b)$$

Assuming asymmetry (ASY) is true, the probability of correct selection of the MWW method for testing H_o: $\mu_1 = \mu_2$ is approximately given as:

$$\overline{p_a}^2 \text{ for the BOTH method, and} \quad (31a)$$

$$1 - (1 - \overline{p_a})^2 \text{ for the ALO method.} \quad (31b)$$

Table 6 contains the probabilities of correct preliminary test selection of the *t* or MWW method for testing H_o: $\mu_1 = \mu_2$ depending on whether the underlying distribution in symmetric (SYM) or asymmetry (ASY), and whether the BOTH or ALO preliminary test strategy is used.

For SYM cases, the BOTH method has the higher probabilities of correct selection of the *t* test since: $1 - \overline{p}_s^2 > (1 - \overline{p}_s)^2$. Whereas for ASY cases, the ALO method has the higher probabilities of correct selection of the MWW test since: $1 - (1 - \overline{p}_a)^2 > \overline{p}_a^2$.

Table 6. Probabilities of Correct Preliminary Test Selection of the Method to Test H_o: $\mu_1 = \mu_2$

Correct Selection Probability	Preliminary Test Protocol	Underlying Distribution	Correct Methods
$1 - \overline{p}_s^2$	BOTH	SYM	*t*
$(1 - \overline{p}_s)^2$	ALO	SYM	*t*
$\overline{p}_a^2$	BOTH	ASY	MWW
$1 - (1 - \overline{p}_a)^2$	ALO	ASY	MWW

Table 7 contains the estimated probabilities of correct preliminary test method selection described in Table 6 for the various methods. The estimated probabilities in Table 7 are calculated utilizing the average $\overline{p}_s$ and $\overline{p}_a$ values tabled in Tables 4 and 5.

Table 7. Estimated Preliminary Test Probabilities of Correct Selection of the Method to Test H_o: $\mu_1 = \mu_2$

			BOTH		ALO	
Method	$\overline{p_s}$	$\overline{p_a}$	SYM	ASY	SYM	ASY
D(.05)	.112	.604	.987	.365	.789	.843
T(.05)	.041	.526	.998	.277	.920	.775
D(.25)	.310	.800	.904	.640	.476	.960
T(.25)	.215	.752	.954	.556	.616	.938

Discussion

Preliminary testing methods are recommended that maximize the Table 7 probabilities of correct selection for the SYM and ASY cases. Using this criterion, the BOTH method is preferred for correct *t* test selection for SYM cases, and the ALO method is preferred for correct MWW test selection for ASY cases. Also, the 5% significance level is preferred for SYM cases, and the 25% level is preferred for ASY cases. Furthermore, the Triples tests are preferred for SYM cases, and the D'Agostino S_U tests are preferred for ASY cases.

How then can a single preliminary testing strategy be selected if different strategies, significance levels, and methods are preferred for SYM versus ASY cases?

To resolve this question another preliminary test comparison criterion is introduced.

Preliminary testing methods are recommended that tend to provide equal or nearly equal probabilities of correct method selection for both SYM and ASY cases. Using this criterion with the results in Table 7, two methods are recommended for preliminary test usage. These are the T(.05) and D(.05) procedures, where both use the ALO method.

The probabilities of correct method selection are 0.920 for SYM cases and 0.775 for ASY cases using the T(.05) ALO method. The corresponding probabilities are 0.789 and 0.843, respectively, for the D(.05) ALO method. No other procedures in Table 7 have this high degree of balance between the equality of probabilities of correct model selection for typical SYM and ASY cases. The T(.05) method is preferred if more emphasis is needed for correct method selection for SYM cases, whereas, the D(.05) method is preferred if more emphasis is needed for correct method selection for ASY cases.

Conclusion

One Sample Symmetry Versus Asymmetry Tests

The one sample Triples test is superior to the D'Agostino S_U test for the control of Type I error for symmetric cases, whereas, the one sample D'Agostino S_U test is slightly more powerful than the Triples tests for asymmetric alternatives.

Preliminary Test Of Symmetry Versus Asymmetry Prior To A Test Of Equality Of Means

The Triples test using a 5% level of significance is preferred if more emphasis is needed for correct method selection for symmetric cases, whereas, the D'Agostino S_U test using a 5% level of significance level is preferred if more emphasis is needed for correct method selection for asymmetric cases.

Recommendations

A simulation study examining the characteristics of the use of a preliminary test of skewness versus asymmetry prior to testing H_o: $\mu_1 = \mu_2$ would be of interest. On the basis of the analyses reported here, the Triples test or the D'Agostino S_U test with a 5% level of significance is recommended over the Triples test or the D'Agostino S_U test with a 25% level of significance as a preliminary test of skewness versus asymmetry prior to testing H_o: $\mu_1 = \mu_2$.

References

Cabilio, P. & Masaro, J. (1996). A simple test of symmetry about an unknown median. *The Canadian Journal of Statistics*, *24*(3), 349-361.

D'Agostino, R. B., Belanger, A. & D'Agostino Jr., R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, *44* (4), 316-321.

Gupta, M. K. (1967). An asymptotically nonparametric test of symmetry. *Annals of Mathematical Statistics*, *38*, 849-866.

IMSL. (1989, December). *Math/Library User's Manual (Version 1.1)*. Houston, Texas: Author.

IMSL. (1989, January). *Stat/Library User's Manual (Version 1.1)*. Houston, Texas: Author.

Pearson, E. S. & Hartley, H. O. (1966). *Biometrika Tables for Statisticians*. *Vol. I*, 3rd edition. Cambridge University Press.

Randles, R. H., Fligner, M. A., Policello II, G. E., & Wolfe, D. A. (1980, March). An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association*, *75* (369), 168-172.