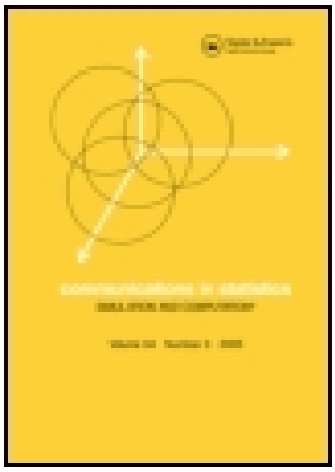
This article was downloaded by: [University of Georgia], [Mr Mohamad Hasan]
On: 30 June 2015, At: 14:06
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Communications in Statistics - Simulation and Computation

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Testing Equality of Two Normal Means Using Combined Samples of Paired and Unpaired Data

Nizam Uddin[a] & M. S. Hasan[b]

[a] Department of Statistics, University of Central Florida, Orlando, FL 32816

[b] Department of Statistics, University of Georgia, Athens, GA 30602

Accepted author version posted online: 30 Jun 2015.

Click for updates

To cite this article: Nizam Uddin & M. S. Hasan (2015): Testing Equality of Two Normal Means Using Combined Samples of Paired and Unpaired Data, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2015.1047527

To link to this article: http://dx.doi.org/10.1080/03610918.2015.1047527

Disclaimer: This is a version of an unedited manuscript that has been accepted for publication. As a service to authors and researchers we are providing this version of the accepted manuscript (AM). Copyediting, typesetting, and review of the resulting proof will be undertaken on this manuscript before final publication of the Version of Record (VoR). During production and pre-press, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal relate to this version also.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

ACCEPTED MANUSCRIPT

Testing equality of two normal means using combined samples of paired and unpaired data

Nizam Uddin[a] and M. S. Hasan[b]

[a]Department of Statistics, University of Central Florida, Orlando, FL 32816

[b]Department of Statistics, University of Georgia, Athens, GA 30602

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

Short Title: Testing equality of two normal means

Abstract: A test for equality of two normal means when the data consist of both paired and unpaired observations is proposed. The proposed test is compared with two other standard methods known in the literature with respect to the type I error rate and power using simulation results.

Key words: Normal distribution, paired and unpaired data, power, simulation, t-distribution, test of equality of means, type I error.

1. Introduction

We consider a situation in which two normal means μ_1 and μ_2 are to be compared when the data consist of two random samples $x_{11}, x_{12}, ...x_{1n_1}, u_{11}, u_{12}, ..., u_{1n}$, and $x_{21}, x_{22}, ..., x_{2n_2}, u_{21}, u_{22}, ..., u_{2n}$ of which (u_{11}, u_{21}), (u_{12}, u_{22}), ..., (u_{1n}, u_{2n}) may be regarded as a paired random sample from a bivariate normal distribution $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, and $x_{11}, x_{12}, ...x_{1n_1}$ and $x_{21}, x_{22}, ..., x_{2n_2}$ are independent random samples from two normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. For convenience, the later two independent samples will be referred to as unpaired data. We shall use $\bar{x_i}$ and s_{ii}^2 to denote, respectively, the sample mean and variance of n_i observations $x_{i1}, x_{i2}, ..., x_{in_i}$, and M_i and s_i^2 to denote, respectively, the sample mean and variance of $n_i + n$ observations $x_{i1}, x_{i2}, ..., x_{in_i}, u_{i1}, u_{i2}, ..., u_{in}$, i = 1, 2. Also, we use s_d^2 for the sample variance of differences $d_j = (u_{1j} - u_{2j})$, $j = 1, 2, \ldots, n$ and s_{12} for sample covariance of pairs (u_{1j}, u_{2j}), $j = 1, 2, \ldots, n$. Furthermore, for $i = 1, 2$, $\bar{u}_i$ stands for the mean of $u_{i1}, u_{i2}, \ldots, u_{in}$. Examples of various studies that require a test of $H_0 : \mu_1 - \mu_2 = 0$ in the above settings of both paired and unpaired data are cited in

[1]Corresponding author. email:nizam.uddin@ucf.edu

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

a number of papers (e.g., Mehrotra (2004), Looney and Jones (2003), Dimery *et al.* (1987), Steere *et al.* (1985), Nurnberger *et al.* (1982), Bhoj (1978), Ekbohm (1976), Lin and Stivers (1974)).

To test $H_0 : \mu_1 - \mu_2 = 0$ using data in the above settings, Bhoj (1978) suggested a linear combination of unpaired and paired t-test statistics and proposed the following test statistic, T_c, when variances are equal:

$$T_c = \lambda t_{f_e} + (1-\lambda) t_{f_p}, \text{ where } 0 \le \lambda \le 1, \; t_{f_e} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_{11}^2+(n_2-1)s_{22}^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, \; t_{f_p} = \frac{\bar{u}_1 - \bar{u}_2}{\sqrt{\frac{1}{n}s_d^2}}.$$

The test statistic T_c is a weighted average of the independent samples t-test statistic t_{f_e} with $f_e = n_1 + n_2 - 2$ degrees of freedom (df) based on the unpaired data assuming equal variances and the paired samples t-test statistic t_{f_p} with $f_p = n - 1$ df based on the paired data. With $\lambda = 0$ and $\lambda = 1$, T_c reduces, respectively, to the paired and two independent samples equal variance t-test statistics for the comparison of two means.

When $\sigma_1 \neq \sigma_2$, we replace t_{f_e} by the corresponding unequal variance test statistic

$$t_{f_u} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_{11}^2}{n_1} + \frac{s_{22}^2}{n_2}}}$$

with Satterthwaites (1946) approximation $f_u = (s_{11}^2/n_1 + s_{22}^2/n_2)^2 / (s_{11}^4/(n_1^2(n_1-1)) + s_{22}^4/(n_2^2(n_2-1)))$ for the degrees of freedom. The linear combination $\lambda' t_{f_u} + (1-\lambda') t_{f_p}$ will then be denoted by T_{cu} and is given by

$$T_{cu} = \frac{\lambda'(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_{11}^2}{n_1} + \frac{s_{22}^2}{n_2}}} + \frac{(1-\lambda')(\bar{u}_1 - \bar{u}_2)}{\sqrt{\frac{1}{n}s_d^2}}, \text{ where } 0 \le \lambda' \le 1.$$

With $\lambda' = 0$ and $\lambda' = 1$, T_{cu} reduces, respectively, to the paired and two independent samples unequal variance t-test statistic for comparing two means. The constants λ and λ' in T_c and T_{cu}

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

are chosen in our simulations so that the var(T_c) and var(T_{cu}) are minimized. Note that var(T_c) = λ^2var(t_{f_e}) + $(1-\lambda)^2$var(t_{f_p}) = $\lambda^2 f_e/(f_e-2) + (1-\lambda)^2 f_p/(f_p-2)$ which, for given sample sizes, is minimized when $\lambda = f_e(f_p-2)/((f_e-2)f_p + f_e(f_p-2))$. Similarly, var($T_{cu}$) is minimized when $\lambda' = f_u(f_p-2)/((f_u-2)f_p + f_u(f_p-2))$.

In an attempt to offer a better alternative to T_c, Looney and Jones (2003) proposed the following test statistic (Z_{corr}):

$$Z_{\text{corr}} = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1+n} + \frac{s_2^2}{n_2+n} - \frac{2ns_{12}}{(n_1+n)(n_2+n)}}}.$$

However, Z_{corr} has a problem with regards to the estimation of the variance of the difference $M_1 - M_2$ between the two sample means. Note that this variance is expressed as

$$\text{var}(M_1 - M_2) = \frac{\sigma_1^2}{n_1+n} + \frac{\sigma_2^2}{n_2+n} - \frac{2n\sigma_{12}}{(n_1+n)(n_2+n)}.$$

where σ_{12} is the population covariance of the paired observations. Looney and Jones (2003) estimated this variance by $\frac{s_1^2}{n_1+n}+\frac{s_2^2}{n_2+n}-\frac{2ns_{12}}{(n_1+n)(n_2+n)}$ where all (n_i+n) data values $x_{i1}, x_{i2}, ..., x_{in_i}, u_{i1}, u_{i2}, ..., u_{in}$ are used to calculate the estimate s_i^2 of σ_i^2, but only n paired data are available and used to calculate the estimate s_{12} of σ_{12}. Since the variance and covariance parameters are estimated using different sets of data with different sample sizes, it is quite possible, especially for small sample sizes, to end up with an observed data covariance matrix that is not positive definite. This may result in a negative value for the estimated variance $\frac{s_1^2}{n_1+n} + \frac{s_2^2}{n_2+n} - \frac{2ns_{12}}{(n_1+n)(n_2+n)}$ that is used in the denominator of Z_{corr}. Here we offer an example in favor of this argument using samples of sizes $n_1 = n_2 = 20, n = 5$. For the unpaired data, the twenty X_1-observations are -0.04649, 0.07511, 0.01553, 0.08333, -0.05896, 0.04752, 0.04265, -0.0501, 0.04889, -0.08706, 0.04738, 0.02475, 0.04031, 0.02102, -0.04169, -0.05797, -0.08568, 0.02198, -0.04612, and -0.04226, and the twenty X_2-observations are 0.97027, 0.98132, 0.97722, 0.98784, 0.96520, 1.00969, 0.98689, 0.99289, 1.03771, 0.93423, 0.96841, 0.90209, 0.97670, 0.93800, 0.92953, 0.96941, 0.98554, 0.98112, 0.96386, and 1.02635. The five (u_1, u_2) paired observations are (-0.28480, 0.61651), (1.87131, 2.87115), (-2.05722, -

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

0.94238), (-0.04226, 1.02635), and (0.27421, 0.82283). These samples yield $s_1^2 = 0.3310641$, $s_2^2 = 0.3098131, s_{12} = 1.8820029$ for which $\frac{s_1^2}{n_1+n} + \frac{s_2^2}{n_2+n} - \frac{2ns_{12}}{(n_1+n)(n_2+n)} = -0.004476958 < 0$ and hence Z_{corr} is undefined. Here we offer a simple fix to this negative variance estimation problem. First transform each (u_{1j}, u_{2j}) pair to the weighted difference $d_j^* = nu_{1j}/(n_1+n) - nu_{2j}/(n_2+n)$ and write $\bar{d^*} = (1/n)\sum_{j=1}^{n} d_j^* = \frac{n\bar{u_1}}{(n_1+n)} - \frac{n\bar{u_2}}{(n_2+n)}$. Then the numerator of Z_{corr} can be rearranged as

$$M_1 - M_2 = \frac{n_1\bar{x_1}}{(n_1+n)} - \frac{n_2\bar{x_2}}{(n_2+n)} + \frac{n\bar{u_1}}{(n_1+n)} - \frac{n\bar{u_2}}{(n_2+n)} = \frac{n_1\bar{x_1}}{(n_1+n)} - \frac{n_2\bar{x_2}}{(n_2+n)} + \bar{d^*}.$$

The variance of $M_1 - M_2$ can now be expressed as

$$\text{var}(M_1 - M_2) = \frac{n_1^2\sigma^2_{\bar{x_1}}}{(n_1+n)^2} + \frac{n_2^2\sigma^2_{\bar{x_2}}}{(n_2+n)^2} + \sigma^2_{\bar{d^*}}.$$

where $\sigma^2_{\bar{x_1}}$, $\sigma^2_{\bar{x_2}}$ and $\sigma^2_{\bar{d^*}}$ are, respectively, the variance of the sampling distributions of $\bar{x_1}$, $\bar{x_2}$ and $\bar{d^*}$. Since $\sigma^2_{\bar{x_1}} = \sigma_1^2/n1$, $\sigma^2_{\bar{x_2}} = \sigma_2^2/n_2$, and $\sigma^2_{\bar{d^*}} = \sigma^2_{d^*}/n$, we estimate these variances by replacing σ_i^2 by the corresponding sample variance s_{ii}^2, $i = 1, 2$, and $\sigma^2_{d^*}$ by the sample variance $s^2_{d^*}$ of the weighted differences $d_j^* = nu_{1j}/(n_1+n) - nu_{2j}/(n_2+n))$, $j = 1, 2, \ldots, n$. Notice that the estimated variance is now the sum of the three positive terms and thus results in a positive value for the estimated value of var$(M_1 - M_2)$. Using this estimated variance of var$(M_1 - M_2)$, we modify Looney and Jone's (2003) Z_{corr} to Z_c as follows:

$$Z_c = \frac{M_1 - M_2}{\sqrt{\frac{n_1 s_{11}^2}{(n_1+n)^2} + \frac{n_2 s_{22}^2}{(n_2+n)^2} + \frac{s^2_{d^*}}{n}}}$$

Before we address the performance of the above Z_c with respect to the type I error rate and power, we would like to note that the above problem can also be viewed as a test of $H_0 : \mu_1 - \mu_2 = 0$ in the framework of a paired or repeated measurements experiment with missing data as follows: of the $n_1 + n_2 + n$ pairs of possible observations, n_1 pairs have data for X_1 only $(x_{11}, .), (x_{12}, .), ..., (x_{1n_1}, .)$, where '.' is used to denote the missing observation in each pair, n_2 pairs have data for X_2 only $(., x_{21}), (., x_{22}), ..., (., x_{2n_2})$, and n subjects have data for both X_1 and X_2 which are the paired obser-

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

vations (u_{11}, u_{21}), (u_{12}, u_{22}), ..., (u_{1n}, u_{2n}). In the framework of a block design with two correlated observations per block, a test of the above hypothesis can then be carried out using SAS Proc Mixed which utilizes all complete and incomplete pairs assuming missing data occurred at random. Details of this test, here referred to as *REML* can be found in Patterson and Thompson (1971) as well as in SAS documentation for Proc Mixed. Using simulation results, Mehrotra (2004) provided empirical support to conclude that REML results in higher power than T_c for all positive values of correlations and higher power than Z_{corr} for large correlations (≥ 0.7). However, as noted above, Z_{corr} can have a negative estimated variance and there is a problem with the way the critical values of the null distribution of T_c are used in Mehrotra (2004). In Section 2 of this paper, we approximate the null distributions of T_c and T_{cu}. Using the approximate critical values obtained in Section 2, a simulation study is conducted to compare the empirical type I error rates and powers of Z_c with that of T_c, T_{cu}, and *REML* mentioned above. This is done in Sections 4 and 5 following a discussion on the null distributions in Section 2 and a numerical example in Section 3. Note here that the *REML* values are calculated using Proc Mixed with default degrees of freedom when variances are equal and using Kenward-Roger's adjustment for the degrees of freedom when variances are unequal. Details of these methods can be found in SAS documentations for Proc Mixed.

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

2. Null distributions of T_c, T_{cu}, and Z_c

Let t_f be a statistic having a *t*-distribution with f df. Assuming equal variance, we begin with T_c which can be expressed as $T_c = \lambda t_{f_e} + (1 - \lambda) t_{f_p}$ where $f_e = n_1 + n_2 - 2$, $f_p = n - 1$. Following Patil (1965), the null distribution of hT_c is approximated by a t-distribution with f_c df where h and f_c are determined by equating the second and fourth moments of the distribution of hT_c to that of the distribution of t_{f_c}. Solving these two equations, we obtain

$$h = \sqrt{(2B - 3A^2)/AB} \quad \text{and} \quad f_c = (4B - 6A^2)/(B - 3A^2)$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

where

$$A = \frac{\lambda^2 f_p}{f_p - 2} + \frac{(1-\lambda)^2 f_e}{f_e - 2}, \qquad B = \frac{3\lambda^4 f_p^2}{(f_p - 2)(f_p - 4)} + \frac{6\lambda^2(1-\lambda)^2 f_p f_e}{(f_p - 2)(f_e - 2)} + \frac{3(1-\lambda)^4 f_e^2}{(f_e - 2)(f_e - 4)}.$$

The critical values of the null distribution of hT_c is approximated from the distribution of $t_{int(f_c)}$ or $t_{int(f_c)+1}$ if f_c is not an integer. Here int(.) is the integer function; int(f_c) is the largest integer not exceeding f_c. A better approximation would be to estimate the critical values interpolating from the distributions of $t_{int(f_c)}$ and $t_{int(f_c)+1}$. We have used linear interpolation to approximate the critical values of hT_c. It should be noted here that Mehrotra (2004) assumed that the null distribution of T_c is t-distribution with $(n_1 + n_2 + n - 3)$ df but we find no justification for such an assumption. When variances are unequal, the null distribution of $h'T_{cu}$ is derived in the above fashion by replacing $f_e = n_1 + n_2 - 2$ by Satterthwaites (1946) approximate degrees of freedom $f_u = (s_{11}^2/n_1 + s_{22}^2/n_2)^2/(s_{11}^4/(n_1^2(n_1 - 1)) + s_{22}^4/(n_2^2(n_2 - 1)))$ in the above expressions for *A* and *B*.

Following Looney and Jones (2003), the null distributions of Z_c is assumed to follow approximately standard normal distribution. The null distributions described here and their critical values are used in our simulation studies in Sections 4 and 5 below following an illustrative numerical example in Section 3.

3. A Numerical Example

In this section, the competing test statistics mentioned above are evaluated for the clinical study data reported by Rempala and Looney (2006) in their Table 3. The data consist of the Karnofsky Performance Status (KPS) scale scores obtained from 37 patients on the day before they died, from 32 patients on their last day of life, but nine patients were common on both days. This study thus results in two unpaired samples of sizes $n_1 = 28$, $n_2 = 23$ and one paired (next-to-last day, last day) sample of size n = 9. The summary measures are in Table 1, and the results of the statistical tests of the null hypothesis of equality of KPS mean ratings on the last day and next-to-last day of life are in Table 2. The reader is referred to Herman and Looney (2001) and Rempala and Looney

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

(2006) for details of this clinical study data on symptom management among hospice patients in the last days of life. All four methods give two-sided p-values that are much smaller than 0.05 indicating that these methods result in the conclusion that the KPS mean ratings on the last day and next-to-last day of life are significantly different. A closer look at the data indicates that the variances of the two groups are not significantly different. However, the data distributions do not appear to be normal.

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

4. Simulation Study Using Samples From Normal Distributions

For the purpose of calculating the type I error rate and power, we have used the nominal significance value $\alpha = 0.05$ and set our null and research hypothesis as $H_0 : \mu_1 - \mu_2 = 0$ and $H_1 : \mu_1 - \mu_2 = 0.35$, respectively. For each set of sample size values (n_1, n_2, n)= $(30, 30, 10)$, $(10, 10, 30)$, $(45, 45, 10)$, $(10, 10, 45)$, $(40, 25, 10)$, $(10, 15, 30)$ and correlations $\rho = -0.9, -0.8, ..., 0.9$, a total of 10,000 random samples are generated from the corresponding normal and bivariate normal distributions. These sample sizes are chosen to cover combinations of both moderate and small samples for unpaired and paired data. SAS IML (version 9.3) functions "rand('normal', μ, σ)" and "randnormal(n, mean, cov)" are used for generating these samples. Here "rand('normal', μ, σ)" gives a random observation from a normal distribution with mean μ and standard deviation σ and, "randnormal(n, mean, cov)" gives n random pairs of observations from bivariate normal distribution with mean vector specified by "mean" and covariance matrix specified by "cov". Further details of simulation methods and results are presented below.

4.1 Comparison of Z_c, T_c, T_{cu} and $REML$ when $\sigma_1 = \sigma_2$

Under the equal variance assumption, (σ_1, σ_2) are set to (1, 1) for all n_1, n_2, n, and ρ values listed above. For each set of simulation parameters n_1, n_2, n and ρ, the type I error rate when $H_0 : \mu_1 - \mu_2 = 0$ and power when $H_1 : \mu_1 - \mu_2 = 0.35$ were calculated for each of the four methods mentioned above based on 10,000 simulations. We chose $\mu_1 - \mu_2 = 0.35$ so that the results can be compared with those of Mehrotra (2004). For T_c and T_{cu}, we have used the null distributions described in Section 2. The test statistic T_{cu} defined previously for unequal variances

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

is included here under the equal variance assumption to see it's performance with respect to T_c. For *REML*, both the empirical type I error rate and power are determined by comparing significance probabilities obtained from the analysis carried out using SAS Proc Mixed with nominal $\alpha = 0.05$. The null distribution of Z_c is assumed to follow a N(0,1). The results of these simulations are presented in Figures 1-2.

It appears that the empirical type I error rates of all four test statistics are very close to the nominal $\alpha = 0.05$ except for *REML* which is more conservative showing a smaller empirical type I error rate than nominal $\alpha = 0.05$ whenever n is small relative to n_1, n_2. The test statistic Z_c, the corrected version of Looney and Jone's (2003) Z_{corr} statistic, performs better than T_c, T_{cu} and *REML* with respect to empirical power for all negative values of ρ as well as for weak to moderate positive values of ρ. For large positive values of ρ, the choice of a test statistic depends on the number of paired and unpaired observations; *REML* shows higher empirical power than both T_c and T_{cu} when n is large compared to n_1, n_2 whereas T_c and T_{cu} have higher empirical powers than *REML* when n is small compared to n_1, n_2. However, this small loss in empirical power of *REML* compared to T_c and T_{cu} for large positive values of ρ when n is small may have resulted from a smaller empirical type I error rate of *REML* (see Figure 1). Taking this into consideration and the availability of software, one may prefer *REML* to T_c and T_{cu} for large positive values of ρ, say $\rho > 0.7$, irrespective of the number of paired and unpaired observations. For other values of ρ, Z_c may be a clear choice with the standard normal as its null distribution. We like to note here that there is hardly any noticeable differences between T_c and T_{cu} with respect to both empirical type I error rates and powers for all sample sizes and correlations. Our simulation results are not quite in line with that obtained by Mehrotra (2004) where the author provided empirical support to conclude that *REML* results in higher empirical power than T_c for all positive values of ρ. When the number of paired observations is small, the empirical powers of T_c as well as of T_{cu} appear to be higher than *REML* for all ρ (see Figure 2). However, as noted above, Z_c performs better than others with respect to empirical power for all negative values of ρ as well as for weak to moderate

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

positive values irrespective of the number of paired and unpaired observations. Thus for samples from normal distributions with equal variance, our simulations provide empirical support in favor of $REML$ for large positive ρ and of Z_c for all other correlations irrespective of the number of paired and unpaired sample sizes.

4.2 Comparison of Z_c, T_{cu} and $REML$ when $\sigma_1 \neq \sigma_2$

The simulation method is similar to that of the equal variance case described above except (σ_1, σ_2) are set at (1, 2) for all sample sizes considered here. For each set of simulation parameters, the type I error rates and powers were calculated for each of Z_c, T_{cu}, and $REML$ in the fashion of Section 4.1. We did not include T_c here since it was appropriate only when the variances are equal. The critical values of T_{cu} are obtained from the null distribution described in Section 2 with $\lambda' = f_u(n-3)/((n-1)(f_u-2)+(n-3)f_u)$ which minimizes the var(T_{cu}). The null distribution of Z_c is assumed to follow $N(0, 1)$. However, for $REML$, both the empirical type I error rates and powers are determined by comparing significance probabilities obtained from the analysis carried out by Proc Mixed with nominal $\alpha = 0.05$ using Kenward-Rogers adjustment for degrees of freedom. The results of these simulations are presented in Figures 3-4. It appears from these plots that T_{cu} and $REML$ have better control on type I error rates than Z_c; empirical type I error rates of both T_{cu} and $REML$ are about the same and are close to the nominal value $\alpha = 0.05$ for all sample sizes and ρ whereas Z_c show a bit higher type I error rates than the nominal value $\alpha = 0.05$ used in the simulation. The empirical power plots in Figure 4 show that the statistic $REML$ has higher powers than that of T_{cu} across all sample sizes and all correlation values. This simulation study thus provides empirical support in favor of using $REML$ for normally distributed data with unequal variances.

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

5. Simulation Results Using Samples From Non-normal Distributions

In our simulations in Section 4, random samples were generated from normal distributions. To investigate the robustness of these methods to other types of data, both symmetric and skewed non-normal distributions are included in the simulation study. These distributions as well as the parameter values used in the simulations are described below.

5.1. Symmetric non-normal distribution

In this category, we chose the logistic distribution which is symmetric similar to normal but with heavier tails than the normal distribution. For $i = 1, 2$, the i-th logistic distribution is described by the probability density function

$$f_i(x) = \frac{e^{-(x-\mu_i)/\sigma_i}}{\sigma_i(1 + e^{-(x-\mu_i)/\sigma_i})^2}$$

The mean $\mu_{i\ell}$ and variance $\sigma^2_{i\ell}$ of the i-th logistic distribution are $\mu_{i\ell} = \mu_i, \sigma^2_{i\ell} = \sigma^2_i\pi^2/3, i = 1, 2$. Similar to Section 4, the null and research hypotheses are set as $H_0 : \mu_{1\ell} - \mu_{2\ell} = 0$ and $H_1 : \mu_{1\ell} - \mu_{2\ell} = 0.35$, and standard deviations of the two distributions are set at $(\sigma_{1\ell}, \sigma_{2\ell}) = (1, 1)$ when they are assumed equal, and at $(\sigma_{1\ell}, \sigma_{2\ell}) = (1, 2)$ when assumed unequal. Note here that the bivariate logistic distribution with the above marginals yield only $\rho = 1/2$ as the correlation between the two logistic variables, see Kotz, Balakrishnan, and Johnson (2000). Instead of using the bivariate logistic distribution with this restricted ρ, we first randomly generated correlated pairs of cumulative probabilities and then utilize the cumulative distribution function of logistic distribution to generate the paired sample data. All four test statistics Z_c, T_c, T_{cu} and $REML$ are included in the calculations of empirical type I error rates and powers for all sample sizes considered in Section 4. The empirical type I error rates and powers are presented in Figures 5 - 8. For both equal and unequal variances, the simulation results appear to be in line with those obtained for normal data in Section 4 with the exception that the statistic Z_c shows a bit higher type I error rates for logistic data than normal data specially when n is large relative to the unpaired sample sizes. A closer look at the empirical type I error rates in Figures 5 and 7 in conjunction with

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

an analysis of power plots in Figures 6 and 8, one may argue that this simulation study provides empirical support in favor of *REML* for all ρ when n is large compared to unpaired sample sizes and for moderate to large ρ when n is small compared to the unpaired sample sizes. For all other parameter values, T_{cu} may be used if one is concerned about a minor inflation of empirical type I error rates shown by Z_c for data generated from logistic distributions.

5.2. Skewed non-normal distribution

In this category, we chose skewed log-normal distribution. We use $\text{LN}_i(\mu_i, \sigma_i^2)$ to refer to the i-th log-normal distribution with parameters μ_i and σ_i^2, i = 1, 2. The notation ρ is used to denote the correlation parameter in the bivariate log-normal distribution. With these parameters, the mean (μ_{iL}) and variance (σ_{iL}^2) of LN_i are $\mu_{iL} = e^{\mu_i + \frac{1}{2}\sigma_i^2}$, and $\sigma_{iL}^2 = (e^{\sigma_i^2} - 1)e^{2\mu_i + \sigma_i^2}, i = 1, 2$. Random samples of unpaired log-normal data are obtained by exponentiating the random sample values generated from the normal distributions, and the bivariate log-normal paired data are obtained by exponentiating the paired samples generated from bivariate normal distribution with ρ_L as its correlation parameter. Note that the correlation parameter ρ of the bivariate log-normal distribution is related to the parameters of the normal and bivariate normal distributions as

$$\rho = \frac{e^{\sigma_1 \sigma_2 \rho_L} - 1}{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}.$$

The values of ρ are thus restricted by the choice of these parameter values and are reflected in the plots of the empirical type I error rates in Figures 9 and 10. All four test statistics show higher type I error rates than the nominal value $\alpha = 0.05$ for both the equal and unequal variances assumptions across all sample sizes and correlation values considered here. However, Z_c appears to have better control on type I error rates compared to other three methods. Due to inflated type I error rates, empirical powers are not calculated for any of these methods.

6. Summary and Conclusions

The empirical results of Sections 4 and 5 indicate that the choice of a method for testing the

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

equality of two means in the presence of both paired and unpaired data depends on several factors including sample sizes (the numbers of unpaired and paired data), correlations of paired data, parent distributions and their variances.

For samples from normal distributions with equal variance, $REML$ and Z_c may be preferred to T_c and T_{cu} where the choice of $REML$ and Z_c depends on the correlation coefficient; choose $REML$ for large positive correlations between the paired data when n is large compared to n_1 and n_2, otherwise choose Z_c. When variances are unequal, the weighted t-statistic T_{cu} appears to have better control on type I error rates across all sample sizes and all values of ρ. However, given that the empirical type I error rates of both $REML$ and T_{cu} are very close to the nominal value $\alpha = 0.05$ when n is large compared to n_1 and n_2 and that $REML$ shows higher powers than that of T_{cu} for these sample sizes, one may choose $REML$ (T_{cu}) whenever n is large (small) compared to the unpaired sample sizes for all values of ρ.

For samples from logistic distributions with equal and unequal variances, $REML$ and Z_c appear to be competitors of each other; choose $REML$ for all ρ when n is large compared to unpaired sample sizes and for moderate to large ρ when n is small compared to the unpaired sample sizes. However, when variances are unequal, one may use T_{cu} instead of Z_c if minor inflation of empirical type I error rates of Z_c is of great concern. Note, however, that the test statistic Z_c with standard normal as its null distribution is much easier to use in practice than T_{cu}.

All four methods are in general found to be liberal with respect to the type I error rates when samples are drawn from the skewed log-normal distributions irrespective of the variances being equal or not. In general, our simulation study does not provide sufficient empirical support in favor of recommending any of these methods for use in practice for non-normal skewed data. However, Z_c may be a good option only when the number of paired data is small compared to the unpaired sample sizes.

It should be noted here that the empirical comparisons of these methods are limited in that the simulations are carried out only for some selected combinations of n_1, n_2, n, σ_1, σ_2, and ρ with

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

$H_1 : \mu_1 - \mu_2 = 0.35$ using samples from normal, logistic and log-normal distributions.

Acknowledgments.

The authors are grateful to a anonymous referee for valuable comments and suggestions that greatly improved the paper. We are also grateful to Dr. David Nickerson, Professor and Chair of the Department of Statistics here at the University of Central Florida for thoroughly reading this revised version of the paper.

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

5. References

Bhoj, D. S. (1978). Testing equality of correlated variates with missing observations on both responses. Biometrika, 65, 225-228.

Dimery, I. W., Nishioka, K., Grossic, B., Ota, D., Schantz, S. P., Byers, R., Robbins, K. T. and Hong, W. K. (1987). Polyamine metabolism in carcinoma of the oral cavity compared with adjacent and normal oral mucosa. American Journal of Surgery, 154, 429-433.

Ekgohm, G. (1976). Comparing means in the paired case with incomplete data on both responses. Biometrika, 63, 299-304.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). Continuous Multivariate Distributions, Volume I: Models and Applications. John Wiley & Sons, Inc., 551-574.

Lin, P. E. and Stivers, L. E. (1974). On the difference of means with incomplete data. Biometrika, 61, 325-334.

Looney, S. W. and Jones, P. W. (2003). A method of comparing two normal means using combined samples of correlated and uncorrelated data. Statistics in Medicine, 22, 1601-1610.

Mehrotra, D. Y. (2004). Letter to the Editor. Statistics in Medicine, 23, 1179-1180

Nurnberger, J., Jimerson, D. C., Allen, J. R., Simmons, S. and Gershon, E. (1982). Red cell oabain-sensitive Na^+ -K^+-adenosine triphosphatase: a state marker in affective disorder inversely related to plasma cortisol. Biological Psychiatry, 17, 981-992.

Patil, V. H. (1965). Approximation to the Behrens-Fisher distributions. Biometrika, 52, 267-271.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.

Rempala, G. A. and Looney, S. W. (2006). Asymptotic properties of a two sample randomized test for partially dependent data. J. Statistit. Plann. Infer., 136, 68-89.

Satterthwaite, F.W. (1946). An Approximate Distribution of Estimates of Variance Components, Biometrics Bulletin, 2, 110114.

Steere, A. C., Green, J., Schoen, R. T., Taylor, E., Hutchinson, G. J., Rahn, D. W., Malawista, S. E. (1985). Successful parenteral penicillin therapy of established Lyme arthritis. New England Journal of Medicine, 312, 869-874.

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1: Summary Measures of KPS Data Mentioned in Example 1.

Group	Description	Sample I	Sample II	
Unpaired	Sample Size	28	23	
	Mean	20.89	17.39	
	Variance	42.69	51.98	
Unpaired/Paired	Sample Size	37	32	
Combined	Mean	21.08	17.34	
	Variance	41.85	48.36	
	n	Mean	Variance	
Paired	Difference (d)	9	4.44	34.03
	Weighted Difference (d*)	9	0.43	2.40

Table 2: Table 2. Results of Competing Methods for KPS Data

Methods	Observed Value	Two-tailed p-value
Corrected Z	Z_c =2.46	0.0139
Weighted t (equal variance)	T_c = 2.02	0.0111
	(λ =0.4388, h = 1.3707, df=22.14)	
Weighted $t_{(Unequalvariances)}$	T_{cu} = 2.01	0.0115
	(λ' =0.4397, h′ = 1.3702, df=21.86)	
$REML_{(EqualVariance)}$	F (1, 29) = 7.68	0.0096
$REML_{(UnequalVariances)}$	F (1, 28.3) = 7.53	0.0104

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

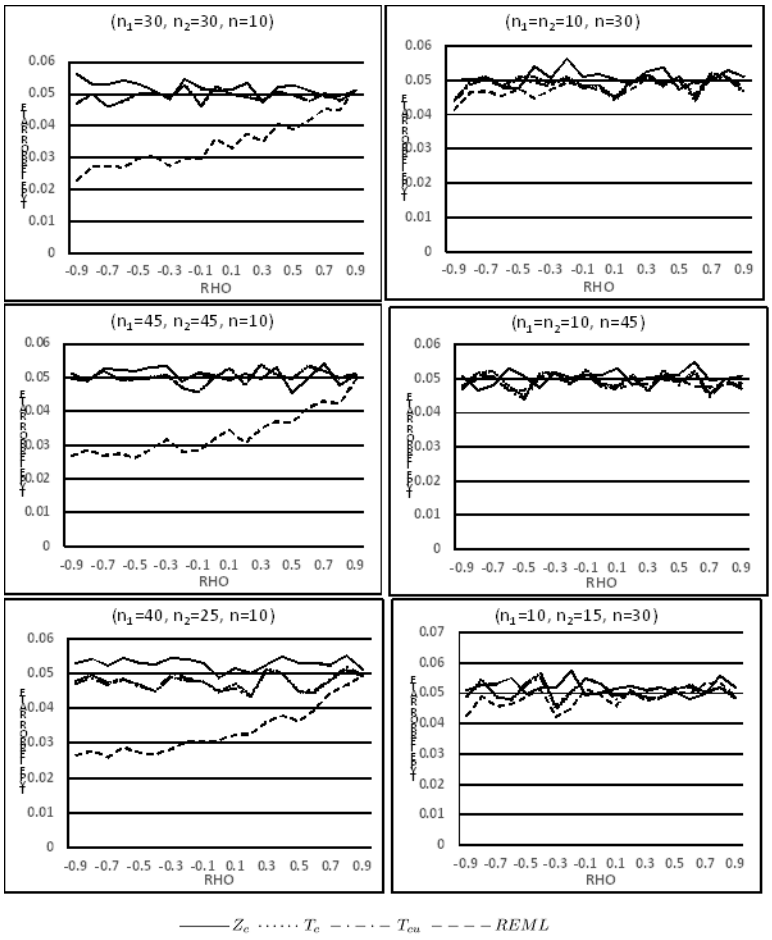


Figure 1: Empirical type I error rates for samples from normal distributions with equal variance.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

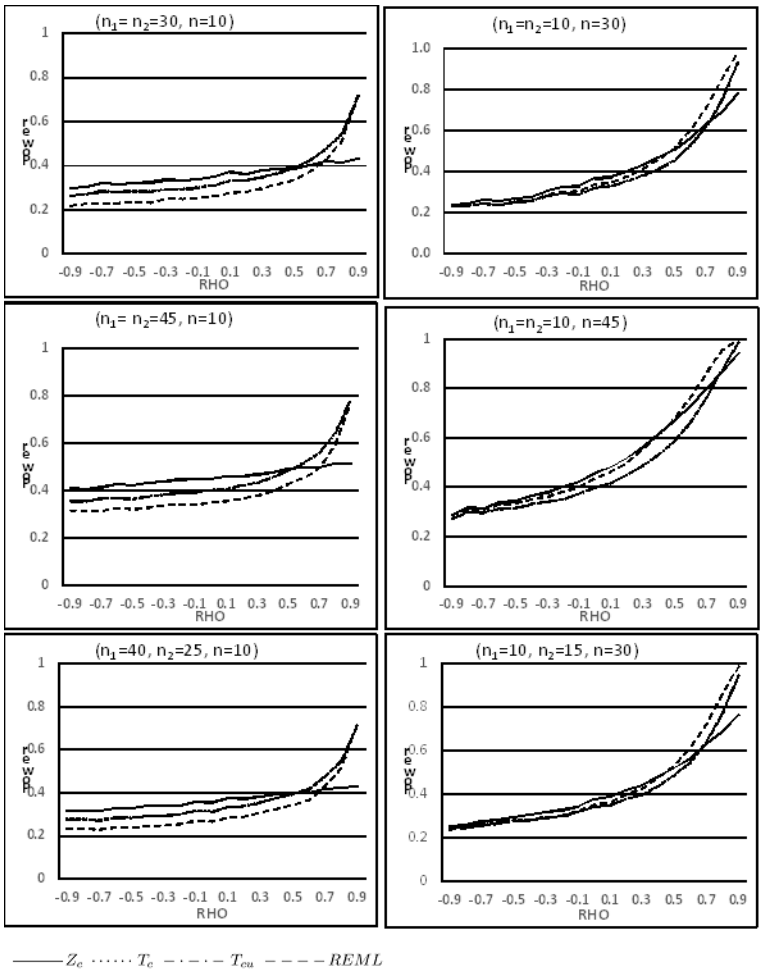


Figure 2: Empirical Powers for samples from normal distributions with equal variance.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

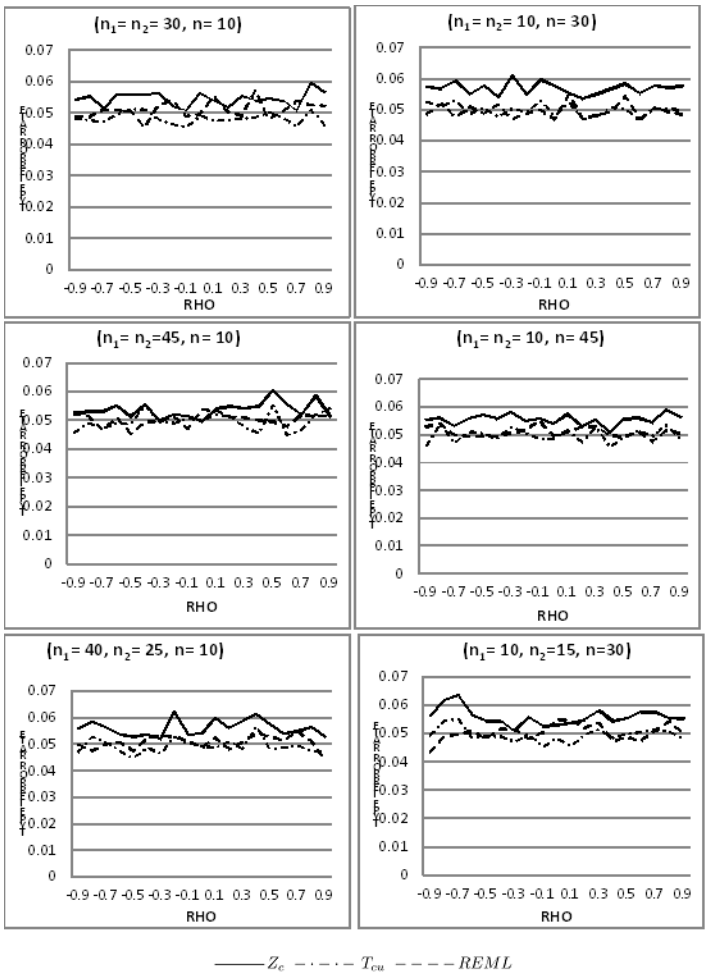


Figure 3: Empirical type I error rates for samples from normal distributions with unequal variances.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

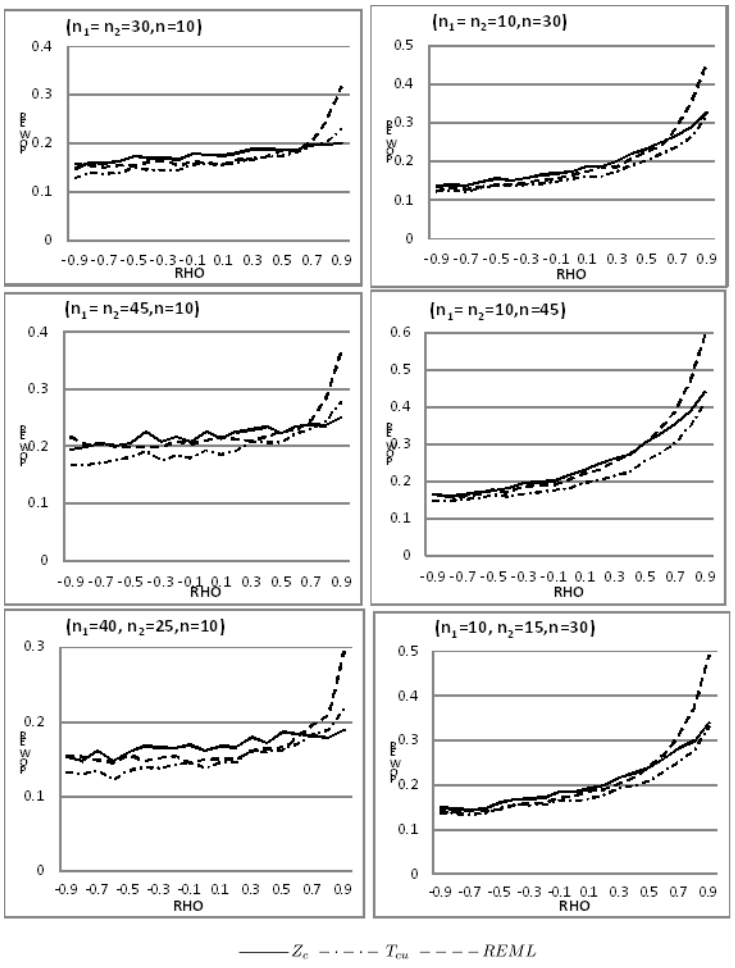


Figure 4: Empirical powers for samples from normal distributions with unequal variances.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

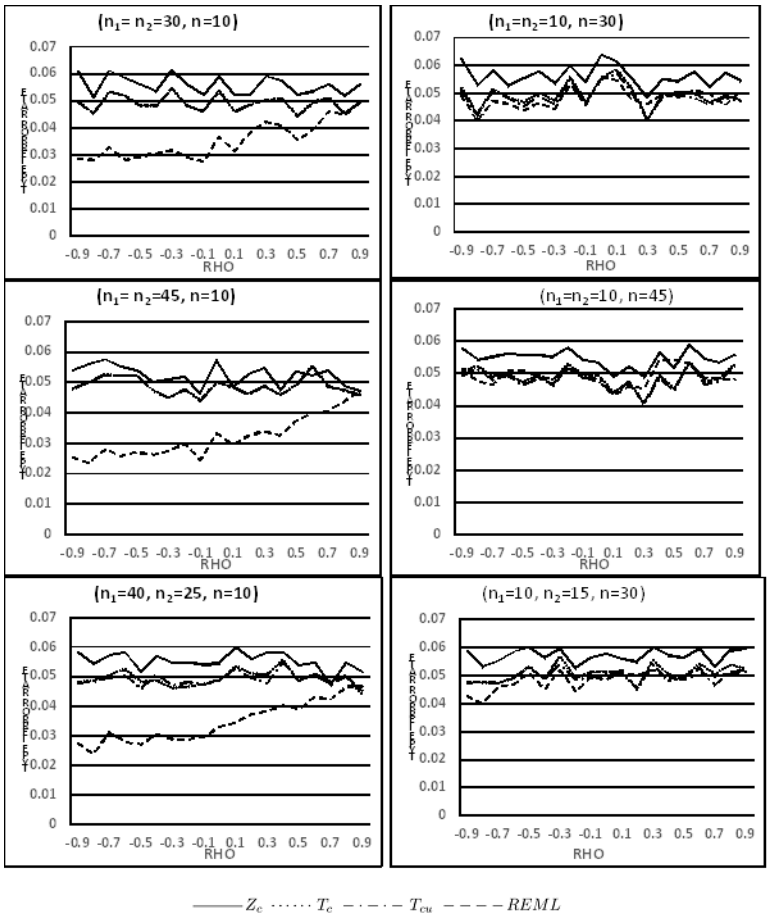


Figure 5: Empirical type I error rates for samples from logistic distributions with equal variance.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

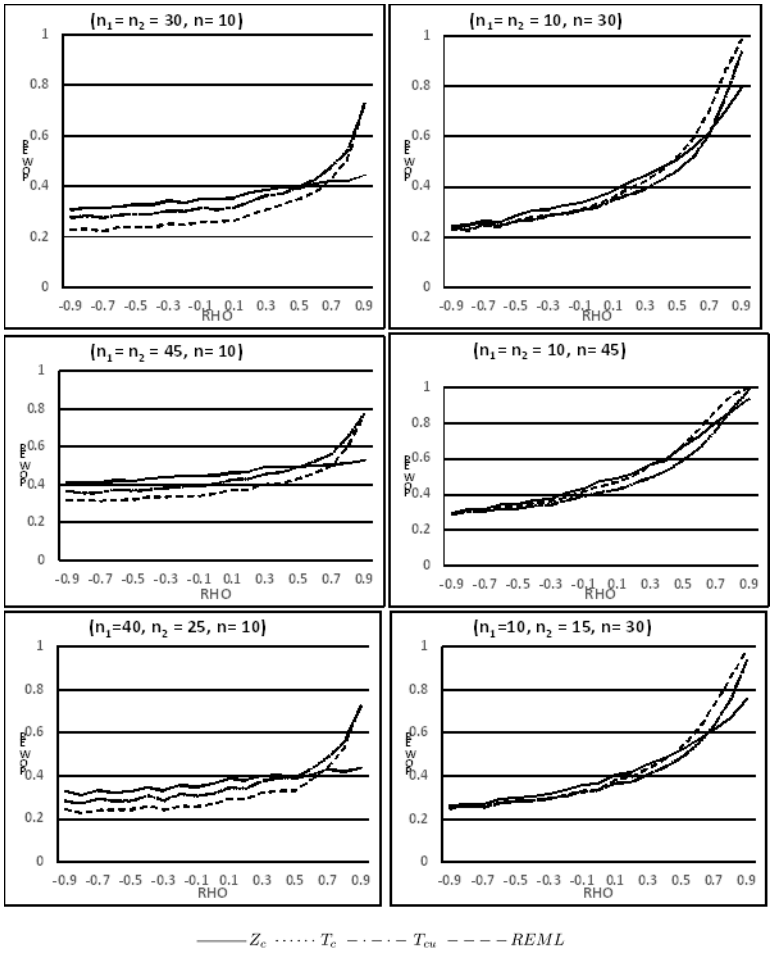


Figure 6: Empirical powers for samples from logistic distributions with equal variance.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

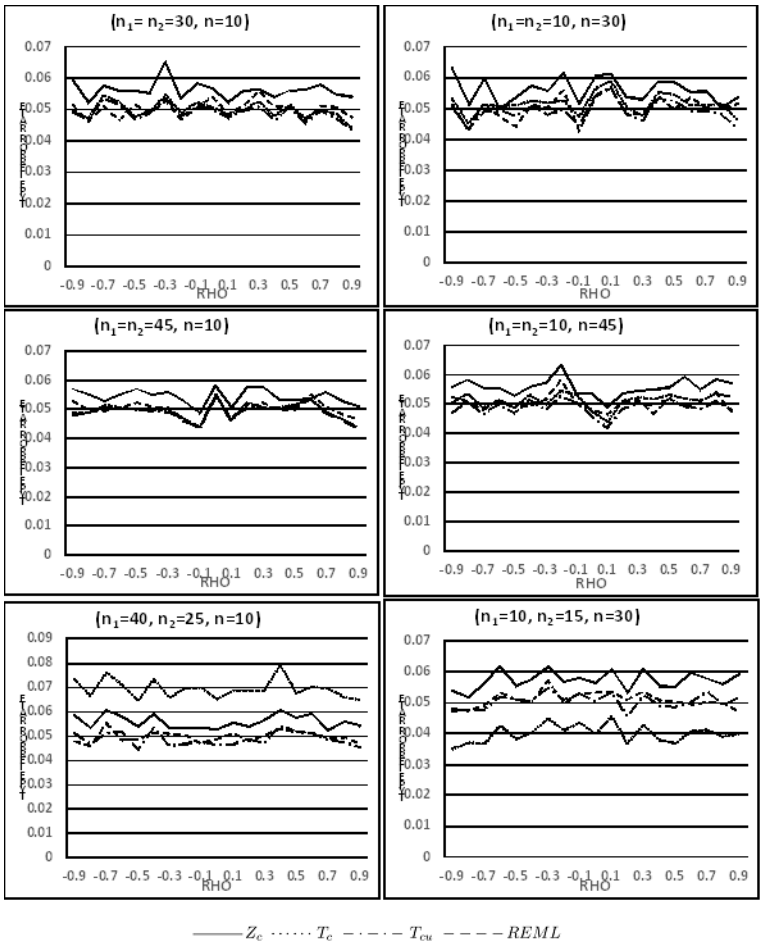


Figure 7: Empirical Type I error rates for samples from logistic distributions with unequal variances.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

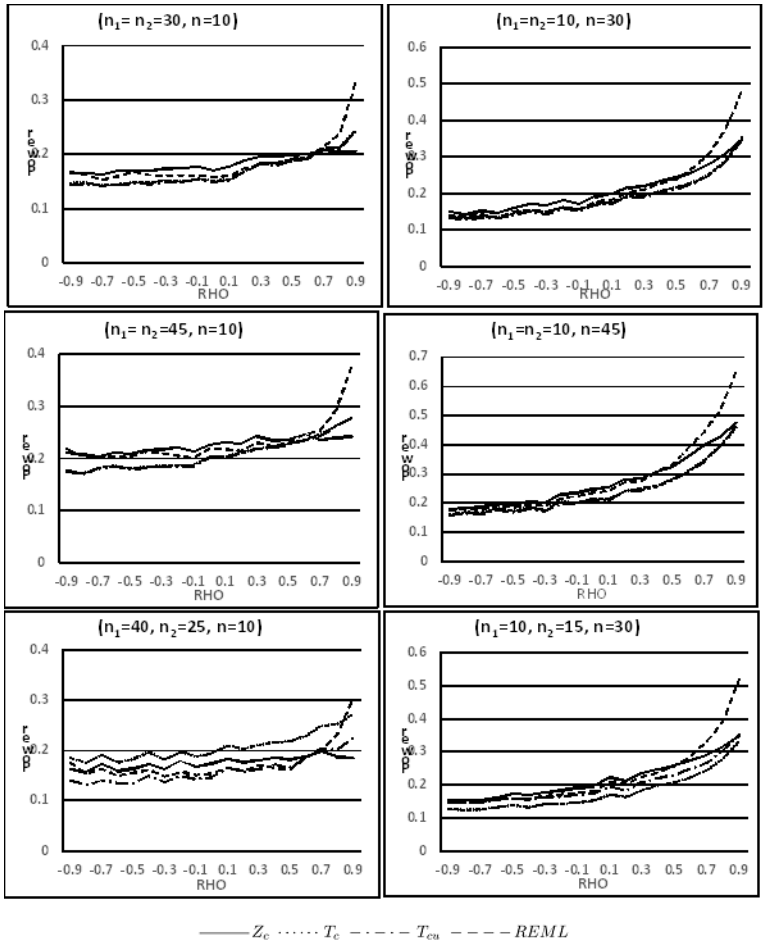


Figure 8: Empirical powers for samples from logistic distributions with unequal variances.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

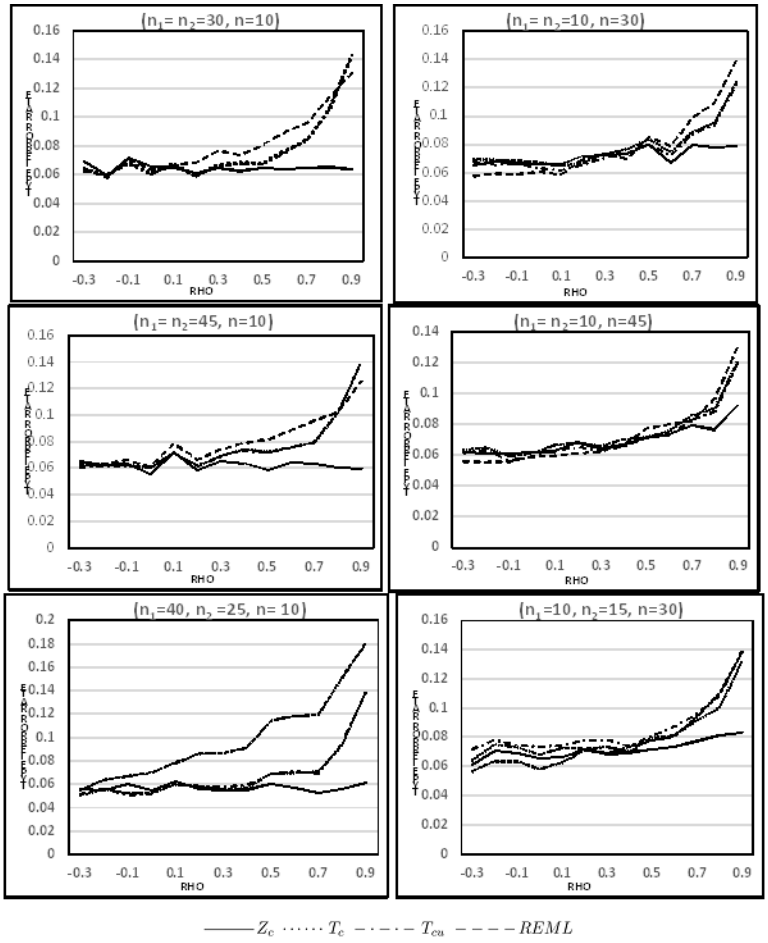


Figure 9: Empirical type I error rates for samples from log-normal distributions with equal variance.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Downloaded by [University of Georgia], [Mr Mohamad Hasan] at 14:06 30 June 2015

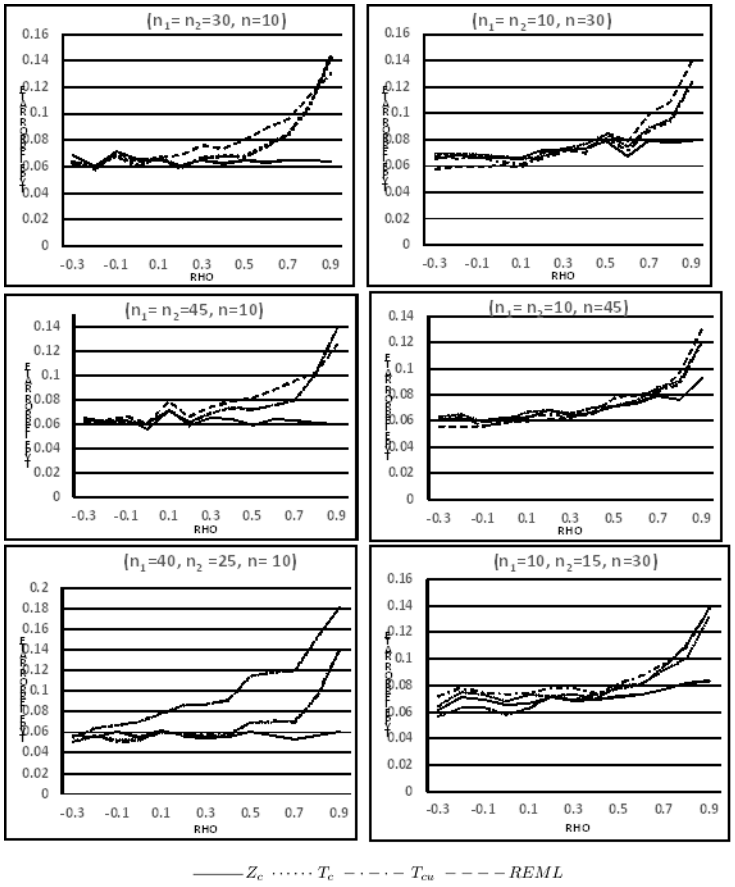


Figure 10: Empirical type I error rates for samples from log-normal distributions with unequal variances.

ACCEPTED MANUSCRIPT