

Journal of Statistical Theory and Applications Vol. **19(2)**, June 2020, pp. 238–247 DOI: https://doi.org/10.2991/jsta.d.200507.003; ISSN 1538-7887 https://www.atlantis-press.com/journals/jsta



Comparing Two Means Using Partially Paired and Unpaired Data

Nizam Uddin^{1,*}, Mohamad S. Hasan²

¹ Department of Statistics and Data Science, University of Central Florida, Orlando, FL 32816, USA ² Beckman Coulter Inc., 11800 SW 147th Ave, Miami, FL 33196, USA

ARTICLE INFO

ABSTRACT

Article History Received 10 Feb 2019 Accepted 25 Apr 2020

Keywords

Normal distribution Paired and unpaired data Power Simulation t-distribution Test of equality of means Type I error

AMS Subject Classification (2010): 62H15, 62J15

A new test statistic is proposed to test the equality of two normal means when the data is partially paired and partially unpaired. The test statistic is based on a linear combination of the differences of both paired and unpaired sample means. Using t-distribution as the approximate null distribution, the proposed method is evaluated against some other standard methods known in the literature. For samples from normal and logistic distributions with equal variances, the proposed method appears to perform better than other methods with respect to power while keeping the type I error rates very competitive.

© 2020 *The Authors*. Published by Atlantis Press SARL. This is an open access article distributed under the CC BY-NC 4.0 license (http://creativecommons.org/licenses/by-nc/4.0/).

1. INTRODUCTION

This paper presents a new test statistic for comparing two means μ_1 and μ_2 under the following two types of data:

- i. Type U Data: Data obtained from two independent random samples but, unknowingly or by accident, some units are used or found to be the same in both samples that resulted in some paired observations. The number of units common to both samples is expected to be small compared to unpaired sample sizes.
- ii. Type P Data: Data obtained from paired design, but some data values are missing completely at random that resulted in some unpaired observations. Pairs with both missing observations are deleted. The numbers of first and second missing values (or equivalently the unpaired sample sizes) are expected to be small compared to complete pairs.

Type U data may be visualized as two random samples $(X_1, U_1) = x_{11}, x_{12}, \dots, x_{1n_1}, u_{11}, u_{12}, \dots, u_{1n}$, and $(X_2, U_2) = x_{21}, x_{22}, \dots, x_{2n_2}, u_{21}, u_{22}, \dots, u_{2n}$ of which $(U_1, U_2) = (u_{11}, u_{21}), (u_{12}, u_{22}), \dots, (u_{1n}, u_{2n})$ are paired, and $X_1 = x_{11}, x_{12}, \dots, x_{1n_1}$ and $X_2 = x_{21}, x_{22}, \dots, x_{2n_2}$ are unpaired observations of the two independent samples.

Type P data is visualized as $n_1 + n_2 + n$ pairs $(X_1, .), (., X_2), (U_1, U_2)$ where n_1 pairs $(X_1, .)$ have only second observations missing, n_2 pairs $(., X_2)$ pairs have only first observations missing, and n complete pairs (U_1, U_2) with no missing observations.

We assume that the paired data (U_1, U_2) is a random sample from a bivariate normal distribution $BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, and the unpaired data X_1 and X_2 are independent random samples from normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Notice that the means and variance parameters are same in bivariate normal and normal distributions.

Some studies that require testing H_0 : $\mu_1 - \mu_2 = 0$ for Type U and Type P data described above are mentioned by several researchers (e.g., Uddin and Hasan [1], Samawi and Vogel [2], Rempala and Looney [3], Mehrotra [4], Looney and Jones [5], Hermann and Looney [6], Dimery *et al.* [7], Steere *et al.* [8], Nurnberger *et al.* [9], Bhoj [10], Ekbohm [11], Lin and Stivers [12]). A survey of various statistical methods proposed for use in this setting can be found in Guo and Yuan [13].

^{*}Corresponding author. Email: nizam.uddin@ucf.edu

If the number (n) of paired data is too small compared to unpaired sample sizes n_1 and n_2 , one may simply ignore the paired data and use two independent samples methods (e.g., independent samples t-test). Similarly, paired sample methods (e.g., paired t-test) can be used if the unpaired sample sizes n_1 and n_2 are too small compared to n, the complete pairs. A third approach would be to ignore the pairing of data and run the two independent samples t-test. In each of the above three approaches, some information would be lost due to omitted observations or correlations between paired observations.

The present paper focuses only on tests of H_0 : $\mu_1 - \mu_2 = 0$ that use all available data on both samples as well as the correlation from paired observations. For the purpose of a brief literature review, first we use the following notations:

- \overline{x}_i = sample mean of n_i observations $x_{i1}, x_{i2}, \dots, x_{in_i}, i = 1, 2$.
- s_{ii}^2 = sample variance of n_i observations $x_{i1}, x_{i2}, \dots, x_{in_i}, i = 1, 2$.
- M_i = sample mean of $n_i + n$ observations $x_{i1}, x_{i2}, \dots, x_{in_i}, u_{i1}, u_{i2}, \dots, u_{in}, i = 1, 2$.
- s_i^2 = sample variance of $n_i + n$ observations $x_{i1}, x_{i2}, \dots, x_{in_i}, u_{i1}, u_{i2}, \dots, u_{in}, i = 1, 2$.
- \overline{d} = sample mean of *n* paired differences $u_{1i} u_{2i}$, j = 1, 2, ..., n.
- s_d^2 = sample variance of *n* paired differences $u_{1j} u_{2j}, j = 1, 2, ..., n$.
- $s_{d^*}^2$ = sample variance of *n* weighted paired differences $nu_{1j}/(n_1 + n) nu_{2j}/(n_2 + n)$, j = 1, 2, ..., n.

For Type U data under the assumption that $\sigma_1^2 = \sigma_2^2$, Bhoj [10] suggested the following test statistic T_c :

$$T_{c} = \lambda t_{f_{e}} + (1 - \lambda) t_{f_{p}} \text{ where}$$

$$t_{f_{e}} = \frac{\overline{x}_{1} - \overline{x}_{2}}{\sqrt{\frac{(n_{1} - 1)s_{11}^{2} + (n_{2} - 1)s_{22}^{2}}{n_{1} + n_{2} - 2}} \left(\frac{1}{n_{1}} + \frac{1}{n_{2}}\right)}, \quad t_{f_{p}} = \frac{\overline{d}}{\sqrt{\frac{1}{n}s_{d}^{2}}}, \quad 0 \le \lambda \le 1.$$

The above test statistic T_c is linear combination of the two independent samples t-statistic t_{f_c} for unpaired data with $f_e = n_1 + n_2 - 2$ degrees of freedom and paired t-test statistic t_{f_p} with $f_p = n - 1$ degrees of freedom. To accommodate for unequal variances, T_c is modified by Uddin and Hasan [1] to T_{cu} :

$$T_{cu} = \lambda' t_{f_u} + (1 - \lambda') t_{f_p}, \text{ where } t_{f_u} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_{11}^2}{n_1} + \frac{s_{22}^2}{n_2}}}$$

The sampling distribution of t_{fu} is approximated by t-distribution with Satterthwaites [14] approximate degrees of freedom f_u , where

$$f_u = \frac{\left(\frac{s_{11}^2}{n_1} + \frac{s_{22}^2}{n_2}\right)^2}{\frac{s_{11}^4}{n_1^2(n_1-1)} + \frac{s_{22}^4}{n_2^2(n_2-1)}}.$$

Looney and Jones [5] proposed the following alternative to T_c :

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

However, it is pointed out by Uddin and Hasan [1] that the test statistic Z_{corr} has a negative estimated variance problem and offered a solution to the problem. They suggest to modify Z_{corr} to Z_c as follows:

$$Z_{c} = \frac{M_{1} - M_{2}}{\sqrt{\frac{n_{1}s_{11}^{2}}{(n_{1} + n)^{2}} + \frac{n_{2}s_{22}^{2}}{(n_{2} + n)^{2}} + \frac{s_{d^{*}}^{2}}{n}}}$$

For the Type P data that results from a paired design with missing observations, Mehrotra [4] suggested that a test (here referred to as REML) of the above hypothesis can be carried out using SAS Proc Mixed.

The above test statistics are empirically compared with respect to type I error rates and powers for some selected values of sample sizes and correlations (see Uddin and Hasan [1], Mehrotra [4]). Under the assumption of equal variances, Mehrotra [4] used simulations to compare

type I error rates and powers of REML, Z_{corr} and T_c and concluded that REML results in higher power than both T_c and Z_{corr} for large positive correlations.

In a recent paper, Uddin and Hasan [1] modified Z_{corr} to Z_c to correct for negative estimated variance problem, T_c to T_{cu} to account for unequal variances, and carried out a simulation study to compare REML, Z_c and T_{cu} . For normally distributed data with equal variances, their simulation results provided empirical support to conclude that REML results in higher powers than Z_c and T_{cu} for large positive correlations. For all negative and small to moderate positive values of correlation, Z_c appears to have higher powers than both REML and T_{cu} . For samples from normal distributions with unequal variances, Z_c is outperformed jointly by T_{cu} and REML, where REML is preferred to T_{cu} for large positive correlations (ρ) and T_{cu} for all other correlations. The choice of a test statistic thus depends on ρ which will often be unknown in practice. Viewing this as a limitation of the above approaches, this paper is devoted to finding a single test statistic that would outperform others for all ρ with respect to statistical powers while keeping the type I error rates very competitive.

In Section 2 below, we describe such a test statistic and offer approximate null sampling distribution of the proposed test statistic. Following an example in Section 3 from Rempala and Looney [3], we use simulation to show that the proposed test statistic performs very well in comparison to the alternative methods available in the literature for samples from normal and logistic distributions with equal variances.

2. PROPOSED TEST STATISTIC AND APPROXIMATE NULL SAMPLING DISTRIBUTION

As noted above both T_c and T_{cu} are defined as weighted average of unpaired and paired t-statistic. An alternative test statistic considered here is based on the weighted average $L(\omega) = \omega(\overline{x}_1 - \overline{x}_2) + (1 - \omega)(\overline{u}_1 - \overline{u}_2)$ of unpaired and paired mean differences with $\omega \in (0, 1)$ chosen so that $var(L(\omega)) = \omega^2 \sigma_{(\overline{x}_1 - \overline{x}_2)}^2 + (1 - \omega)^2 \sigma_{\overline{d}}^2$ is minimized. It follows that $\omega = \frac{\sigma_{\overline{d}}^2}{\sigma_{(\overline{x}_1 - \overline{x}_2)}^2 + \sigma_{\overline{d}}^2}$ minimizes this variance where $\sigma_{\overline{d}}^2$ is the variance of the sampling distribution of paired mean difference \overline{d} and $\sigma_{(\overline{x}_1 - \overline{x}_2)}^2$ is the variance of the sampling distribution of $(\overline{x}_1 - \overline{x}_2)$. If the null hypothesis H_0 is true and variances are known, then

$$\frac{\omega(\overline{x}_1 - \overline{x}_2) + (1 - \omega)\overline{d}}{\sqrt{\omega^2 \sigma_{(\overline{x}_1 - \overline{x}_2)}^2 + (1 - \omega)^2 \sigma_{\overline{d}}^2}}$$

is distributed as standard normal. In practice, the variances $\sigma_{(\overline{x}_1 - \overline{x}_2)}^2$ and $\sigma_{\overline{d}}^2$ and hence ω will be unknown and must be estimated from data. Since $\sigma_{(\overline{x}_1 - \overline{x}_2)}^2$ is the variance of the sampling distribution of $(\overline{x}_1 - \overline{x}_2)$, a function of only unpaired observations, it is sensible to estimate this variance from unpaired data only. Similarly, $\sigma_{\overline{d}}^2$ can be estimated using paired observations only. Analogous to paired and independent samples test of H_0 : $\mu_1 = \mu_2$, we suggest the following estimators:

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}^2 = \frac{s_{11}^2}{n_1} + \frac{s_{22}^2}{n_2}$$
$$\hat{\sigma}_{\bar{d}}^2 = \frac{1}{n} s_d^2$$

Replacing the unknown variances and hence ω by their estimates as shown above, and using a correction factor for the purpose of sampling distribution, we define the following test statistic:

$$\begin{split} T_{u} &= \theta \cdot \frac{\hat{\omega} \left(\bar{x}_{1} - \bar{x}_{2}\right) + (1 - \hat{\omega})\bar{d}}{\sqrt{\hat{\omega}^{2} \hat{\sigma}_{x_{1} - \bar{x}_{2}}^{2} + (1 - \hat{\omega})^{2} \hat{\sigma}_{d}^{2}}} \quad where \; \theta = \sqrt{\frac{2D - 3C^{2}}{CD}} \cdot \frac{\sqrt{\hat{\sigma}_{d}^{2} + \hat{\sigma}_{x_{1} - \bar{x}_{2}}^{2}}}{\left(\hat{\sigma}_{d} + \hat{\sigma}_{x_{1} - \bar{x}_{2}}\right)} \\ \hat{\omega} &= \frac{\hat{\sigma}_{d}^{2}}{\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})}^{2} + \hat{\sigma}_{d}^{2}}, \quad C = \frac{(n - 1)\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})}^{2}}{(n - 3)\left(\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})} + \hat{\sigma}_{d}\right)^{2}} + \frac{f_{u}\hat{\sigma}_{d}^{2}}{(f_{u} - 2)\left(\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})} + \hat{\sigma}_{d}\right)^{2}}, \\ D &= \frac{3(n - 1)^{2}\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})}^{4}}{(n - 3)(n - 5)\left(\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})} + \hat{\sigma}_{d}\right)^{4}} + \frac{6(n - 1)f_{u}\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})}^{2}\hat{\sigma}_{d}^{2}}{(n - 3)\left(f_{u} - 2\right)\left(\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})} + \hat{\sigma}_{d}\right)^{4}}. \\ &+ \frac{3f_{u}^{2}\hat{\sigma}_{d}^{4}}{(f_{u} - 2)\left(f_{u} - 4\right)\left(\hat{\sigma}_{(\bar{x}_{1} - \bar{x}_{2})} + \hat{\sigma}_{d}\right)^{4}}. \end{split}$$

It is noted here that the above T_u without the correction factor θ reduces to T_{New} of Samawi and Vogel [2] when $\hat{\omega} = 1/2$ and the common variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is estimated by the pooled variance of the unpaired samples. Here, we do not know the exact sampling distribution of T_u under the null hypothesis. Following the procedure described in Uddin and Hasan [1], we approximate the null sampling distribution of T_u by *t*-distribution with $(4D - 6C^2)/(D - 3C^2)$ degrees of freedom. This approximate null distribution is used in our simulation in Section 4 for the calculation of empirical type I error rates and powers of T_u for some selected values of sample sizes and correlation values of paired observations. We have an illustrative example in Section 3 below.

3. A NUMERICAL EXAMPLE

In this example, we use the data from a clinical study on symptom management among hospice patients in the last days of life (see Hermann and Looney [6], Rempala and Looney [3]). The data values reported in Table 3 of Rempala and Looney [3] are Karnofsky Performance Status (KPS) scale scores. These values are 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. The reader may review Hermann and Looney [6], Rempala and Looney [3] for more details of this clinical study on symptom management among hospice patients in the last days of life. This study results in data in the above setting with unpaired samples of sizes $n_1 = 28$, $n_2 = 23$ and n = 9 paired (next-to-last day, last day) data values (see Table 3 in Rempala and Looney [3] for all data values). Summary results obtained from these samples are reported by Uddin and Hasan [1] in their Table 1. These samples yield $T_u = 2.83$ with 16.01 degrees of freedom and two-sided p-value = 0.012 which is small enough to conclude that the KPS mean ratings on the last day and next-to-last day of life are significantly different. This conclusion is consistent with that reported by Uddin and Hasan [1] in their Table 2 under five other test statistics considered in that paper. An *R* program is displayed in Table 1 that can be used to do the calculations where the user needs only to enter the data and run the program.

4. SIMULATION: DISTRIBUTIONS AND PARAMETERS

To make it comparable to other simulation studies reported in the literature (e.g., Uddin and Hasan [1], Mehrotra [4]), we 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 some sample sizes and correlations, 10,000 random samples are generated from the corresponding normal and bivariate normal distributions with $\sigma_1^2 = \sigma_2^2 = 1$ under both null and research hypothesis. SAS IML (version 9.22) functions "rand('normal', μ, σ)" and "randnormal(n, mean, cov)" are used for generating these samples. For each set of simulation parameters, type I error rates when H_0 : $\mu_1 - \mu_2 = 0$ and powers when H_0 : $\mu_1 - \mu_2 = 0.35$ were calculated for each of Z_c , REML and T_u based on 10,000 samples mentioned above. Note that T_{cu} is not included here in our simulation with sampling from normal distributions with equal variances since Z_c and REML are found by Uddin and Hasan [1] to jointly outperform T_{cu} for all correlation values. Our goal here is to compare the proposed test statistic T_u to Z_c and *REML* with respect to type I error rate and power. In the calculation of type I error rates and powers, we have used the null sampling distribution of T_u from Section 2 above, but for Z_c , the null

Table 1Calculation of T_U for example 1 using R.

```
(User needs to insert data in X1, X2, and U1U2. Assumptions n > 3, Satterthwaite's fractional df > 2)
#sample one- enter the X1 data
x1 <- c(10,20,25,30,20,30,15,20,30,15,15,20,10,25,30,20,20,30,25,30,20,20,10,25,20,10,20,20);
#sample two- enter the X2 data x2 <- c(15,25,30,20,10,20,10,30,10,10,10,25,15,20,20,20,20,10,10,10,20,30,10);
# sample of n pairs -enter the (u1, u2) data pairs in the order (u11, u21, u12, u22, .....)
u1u2 =matrix(c(20, 10, 30, 20, 25, 10, 20, 20, 25, 20, 10, 10, 15, 15, 20, 20, 30, 30), ncol=2, byrow=TRUE);
n1 <- length(x1); n2 <- length(x2); n <- nrow(u1u2);
diff = u1u2[, 1] - u1u2[, 2]; omega = (var(diff)/n)/(var(x1)/n1 + var(x2)/n2 + var(diff)/n);
theta = sqrt((var(x1)/n1 + var(x2)/n2)/(var(diff)/n)); A = theta/(1 + theta);
fdf = (((var(x1))/n1+(var(x2))/n2)^{*}2)/((var(x1)^{*}var(x1))/(n1^{*}n1^{*}(n1-1))+(var(x2)^{*}var(x2))/(n2^{*}n2^{*}(n2-1)));
C = A^{2} * (n - 1)/(n - 3) + (1 - A)^{2} * (fdf)/(fdf - 2);
D = (A^{**}4)^{*}3^{*}(n-1)^{*}(n-1)/((n-3)^{*}(n-5)) + ((1-A)^{**}4)^{*}3^{*}((fdf)^{**}2)/((fdf-2)^{*}(fdf-4)) + ((1-A)^{*}4)^{*}3^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}(n-1)^{*}
     6*(A**2)*((1-A)**2)*(n-1)*(fdf)/((n-3)*(fdf-2));
df = ((4^{*}D - 6^{*}C^{*}C)/(D - 3^{*}C^{*}C)); f1 <- sqrt((2^{*}D - 3^{*}C^{2})/(C * D));
f_2 <- \operatorname{sqrt}(\operatorname{var}(\operatorname{diff})/n + \operatorname{var}(x_1)/n_1 + \operatorname{var}(x_2)/n_2)/(\operatorname{sqrt}(\operatorname{var}(\operatorname{diff})/n) + \operatorname{sqrt}(\operatorname{var}(x_1)/n_1 + \operatorname{var}(x_2)/n_2));
f3 <- (omega^*(mean(x1) - mean(x2)) + (1 - mega)^*mean(diff))/(sqrt(omega^2 * (var(x1)/n1 + var(x2)/n2) + (1 - mega)^*mean(diff))/(sqrt(omega^2 + var(x1)/n1 + var(x2)/n2))/(sqrt(omega^2 + var(x1)/n1 + var(x2)/n2))/(sqrt(omega^2 + var(x1)/n1 + var(x2)/n2))/(sqrt(omega^2 + var(x1)/n1 + var(x1)/n1))/(sqrt(omega^2 + var(x1)/n1 + var(x1)/n1))/(sqrt(omega^2 + var(x1)/n1 + var(x1)/n1))/(sqrt(omega^2 + var(x1)/n1 + var(x1)/n1))/(sqrt(omega^2 + va
      (1-\text{omega})^2 * var(diff)/n));
TU = f1*f2*f3; pvalue = 2*(1-pt(abs(TU), df));cat("");
Summary <- data.frame(Group=c(1,2,'Pair Diff'), N=c(n1, n2, n), Mean=c(mean(x1), mean (x2), mean(diff)),
      Variance=c(var(x1), var(x2), var(diff)));
print(Summary, print.gap=5, row.names=FALSE); cat(""); cat("");
 Result <- c(TU,df, pvalue); names(Result) <- c("TU Value ", "DF", "Prob > |TU|"); print(Result);
```

distribution described in Uddin and Hasan [1] is used. However, for REML, both type I error rate and power are determined by comparing significance probabilities obtained from the analysis carried out by SAS Proc Mixed with nominal $\alpha = 0.05$. The empirical type I error rates and powers obtained from our simulations are presented graphically in Figures 1 and 2 for some combinations of sample sizes and parameter values.

To study the robustness of the proposed test statistic, the above simulation is repeated for samples from logistic distribution, a non-normal symmetric distribution similar to normal but with heavier tails than the normal. The i-th logistic probability density function is given by

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

for which the mean and variance are μ_i and $\sigma_i^2 \pi^2/3$. For notational convenience, we shall use $\mu_{i\ell}$ and $\sigma_{i\ell}^2$ to refer to the mean and variance of the i-th logistic distribution, i = 1, 2. The null and research hypotheses are set as H_0 : $\mu_{1\ell} - \mu_{2\ell} = 0$, H_1 : $\mu_{1\ell} - \mu_{2\ell} = 0.35$, and standard deviations of the two distributions are set at $(\sigma_{1\ell}, \sigma_{2\ell}) = (1, 1)$. Since the bivariate logistic distribution with the above marginal logistic distributions yield only $\rho = 1/2$, we randomly generated correlated pairs of cumulative probabilities, and then utilize the cumulative distribution function of the above logistic distribution to generate the paired sample data. This technique was previously used by Uddin and Hasan [1] in their simulation. Using samples from logistic distributions, Uddin and Hasan [1] provided empirical support in favor of using *REML* or Z_c depending on sample sizes and correlation values. Here the empirical type I error rates and powers of the proposed test statistic T_u , displayed in Figures 3 and 4, are compared to that of *REML* and Z_c using samples from logistic distributions described above.

5. DISCUSSION AND CONCLUSION

For sampling from normal distributions with equal variances, the test statistic T_u shows higher power than *REML* for all ρ , and higher powers than Z_c for moderate to large ρ considered here. However, due to the higher type I error rates of T_u , particularly for negative ρ and type U data (see Figure 1) compared to that of Z_c and *REML*, one may argue that such power gains by T_u over Z_c and *REML* is the result of its inflated type I error rates. However, for type P data with positive ρ (see Figure 2), the competitive empirical type I error rates and power curves provide empirical support in favor of recommending T_u over Z_c and *REML*. Note that for type P data, we expect small unpaired sample sizes n_1 , n_2 compared to paired sample size n whereas for type U data only a small number of complete pairs is expected compared to unpaired sample sizes. Note further that the correlation parameter ρ is expected to be positive in practice for paired data.

To investigate the robustness of T_u to non-normal distributions, we repeated our simulation using samples from logistic distributions described above. The test statistic Z_c was found by Uddin and Hasan [1] to be more liberal with respect to type I error rates in the sense that Z_c results in higher empirical type I error rates than the nominal α value. They argued in favor of choosing T_{cu} or *REML* depending on sample sizes and correlations for comparing means of logistic distributions. We thus excluded Z_c and compared the proposed T_u with T_{cu} and *REML*. The empirical type I error rates and powers of T_u , *REML*, and T_{cu} are displayed in Figures 3 and 4. A closer look at the plots in these two figures reveal that T_{cu} results in higher type I error rate than the nominal α whereas *REML* is more conservative with smaller type I error rate than the assumed nominal α value for type U data. However, for type P data with positive correlations, the three test statistics appear to be very competitive with respect to type I error rates and T_u performs better with respect to power.

The present simulation study thus provided empirical support for choosing T_u for testing the equality of two means for type P data from normal distributions as well as from logistic distributions with equal variances when correlations of paired observations are positive. This is in contrast to findings in other simulations where, in the absence of T_u , the choice of a test statistic depends on paired/unpaired sample sizes and correlations of paired data.

We have expanded our simulation to investigate the robustness of T_u for unequal variances for all sample sizes and correlations shown in Figures 1–4 for both normal and logistic distribution. We followed the same procedure described in Uddin and Hasan [1] for these expanded simulations. The proposed statistic T_u does not appear to perform well when compared to others when variances are unequal. We like to note that the empirical comparisons of these methods are limited in that the simulations are carried out only for some selected combinations of $n_1, n_2, n, \sigma_1, \sigma_2$, and ρ with $H_1 : \mu_1 - \mu_2 = 0.35$ using samples from normal and logistic distributions.

CONFLICT OF INTEREST

There is no conflict of interest.

AUTHORS' CONTRIBUTIONS

The two authors contributed equally to prepare this paper for publication.



Figure 1 | Type I error rates and powers for samples from normal distributions with equal variance.



Figure 2 | Type I error rates and powers for samples from normal distributions with equal variance.



Figure 3 Type I error rates and powers for samples from logistic distributions with equal variance.



Figure 4 | Type I error rates and powers for samples from logistic distributions with equal variance.

Funding Statement

The work was not funded by any funding agency.

ACKNOWLEDGMENTS

The authors thank the Editor for carefully reading the manuscript and for his guidance on format in preparing this final version of the paper.

REFERENCES

- 1. N. Uddin, M.S. Hasan, Commun. Stat. Simul. Comput. 46 (2017), 2430-2446.
- 2. H.M. Samawi, R. Vogel, J. Appl. Stat. 41 (2014), 109-117.
- 3. G.A. Rempala, S.W. Looney, J. Stat. Plan. Inference. 136 (2006), 68-89.
- 4. D.Y. Mehrotra, Stat. Med. 23 (2004), 1179-1180.
- 5. S.W. Looney, P.W. Jones, Stat. Med. 22 (2003), 1601–1610.
- 6. C. Hermann, S. Looney, J. Hosp. Palliat. Care. 3 (2001), 88-96.
- 7. I.W. Dimery, K. Nishioka, B. Grossic, D. Ota, S.P. Schantz, R. Byers, K.T. Robbins, W.K. Hong, Am. J. Surg. 154 (1987), 429-433.
- 8. A.C. Steere, J. Green, R.T. Schoen, E. Taylor, G.J. Hutchinson, D.W. Rahn, S.E. Malawista, N. Engl. J. Med. 312 (1985), 869–874.
- 9. J. Nurnberger, D.C. Jimerson, J.R. Allen, S. Simmons, E. Gershon, Biol. Psychiat. 17 (1982), 981–992.
- **10.** D.S. Bhoj, Biometrika. 65 (1978), 225–228.
- 11. G. Ekbohm, Biometrika. 63 (1976), 299-304.
- 12. P.E. Lin, L.E. Stivers, Biometrika. 61 (1974), 325–334.
- 13. B. Guo, Y. Yuan, Stat. Methods Med. Res. 26 (2017), 1323–1340.
- 14. F.W. Satterthwaite, Biometrics Bull. 2 (1946), 110–114.