Notes on two sample tests for partially correlated (paired) data

Article *in* Journal of Applied Statistics · January 2014

DOI: 10.1080/02664763.2013.830285

2 authors:

Hani Samawi
Georgia Southern University

Robert L Vogel
Georgia Southern University

Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Notes on two sample tests for partially correlated (paired) data

Hani M. Samawi[a] & Robert Vogel[a]
[a] Jiann-Ping Hsu College of Public Health, PO Box 8015, Georgia Southern University, Statesboro, GA 30460, USA
Published online: 19 Aug 2013.

Journal of Applied Statistics, 2014
Vol. 41, No. 1, 109–117, http://dx.doi.org/10.1080/02664763.2013.830285

Notes on two sample tests for partially correlated (paired) data

Hani M. Samawi* and Robert Vogel

Jiann-Ping Hsu College of Public Health, PO Box 8015, Georgia Southern University, Statesboro, GA 30460, USA

(*Received 1 February 2013; accepted 28 July 2013*)

We provide several methods to compare two Gaussian distributed means in the two sample location problems under the assumption of partially dependent observations. Simulation studies indicate that our test procedure is frequently more powerful than other methods depending on the ratio of the unpaired data and the strength and direction of the correlation between the two variables. The tests used in our comparative study are illustrated with an example based on data from a small gynecological study.

Keywords: partially dependent observation; corrected z-test, weighted t-tests; weighted z-test; power of the test

1. Introduction

Many study designs used in applied sciences such as epidemiology, public health and the biomedical sciences give rise to correlated and partially correlated data. As such, the importance in developing new statistical inference methods to treat partially correlated data and new approaches to model partially correlated data has grown over the past few decades. These methods attempt to account for the special nature of partially correlated data.

Correlated data arise under various situations such as 'matched-pairs' or repeated measures. However, due to the difficulties in matching observational or experimental study subjects, researchers are faced with situations where the observed sample consists of a combination of correlated and uncorrected data, or what is called 'partially correlated data'. Partially correlated data are the result of a matched pair which is missing one of the two correlated values or when a subject in a repeated measures design is missing at least one of their repeated observations, but not all observations. When we have partial data on a matched-pair or repeated measure, we must address the nature of the mechanism that caused the data to be missing. One assumption is to consider the observations are missing completely at random (MCAR), see for example [2,12]. An alternative and more viable assumption is missing at random (MAR), which is considered by Akritas *et al*. [1]. For quantitative responses, statistical methods are well established for analyzing

*Corresponding author. Email: hsamawi@georgiasouthern.edu; samawi.hani2@gmail.com

correlated data. However, for partially correlated data, there are concerns that need to be addressed due to the complexity of the analysis in particular for small sample sizes.

As an example of partially correlated data for the MCAR design, consider the case where the researcher compares two different treatment regiments for eye redness. The researcher randomly assigns one treatment to each eye for each experimental subject. Some patients may drop out after the first treatment, while other patients may refuse to receive the first treatment, but decide to receive the second treatment. Finally, the remaining patients receive both treatments as assigned. In this situation, we may have two groups of patients. The first group of patients received both treatments as assigned and is considered matched-pair data. The second group of patients received only the first treatment or the second treatment and is considered unmatched data.

Additional examples for partially correlated data can be found in the literature [3,15,20]. Several authors have presented various tests that consider the problem of estimating the difference of means from a bivariate normal distribution when some observations corresponding to both variables are missing. Under the assumption of bivariate normality and MCAR, Ekbohm [5] summarized five procedures for testing the equality of two means. Using Monte Carlo results, Ekbohm [5] indicated that the two tests based on a modified maximum-likelihood estimator are preferred. One of those tests was developed by Lin and Stivers [11] and it is preferred when the number of complete pairs is large. The other test was proposed in Ekbohm's paper and is preferred when the variances of the two responses do not differ substantially. When the correlation coefficient between the two responses is low, two other tests have been suggested. Of these two other tests, one is a test proposed by Ekbohm and is preferred when the homoscedasticity assumption is not strongly violated. The other is a Welch-type statistic suggested by Lin and Stivers and preferred when there is a low correlation [11].

Another common approach to deal with partially correlated data is to ignore some of the data, either the correlated data or the uncorrelated data depending on the size of each subset. Looney and Jones [13] argued that ignoring some of the correlated observations would bias the estimation of the variance of the difference in treatment means. Ignoring correlated observations may dramatically affect the performance of the statistical test in terms of type I error rates and statistical power (see [19]). As a result, Looney and Jones [13] proposed a corrected z-test method to overcome the challenges created by ignoring some of the correlated observations. Our preliminary investigations indicate that the method of Looney and Jones [13] pertains to large samples and it is not the most powerful test procedure. Furthermore, Rempala and Looney [16] studied asymptotic properties of a two-sample randomized test for partially dependent data. They indicated that a linear combination of randomized t-tests is asymptotically valid and can be used for non-normal data. However, the large sample permutation tests are difficult to perform and only have some optimal asymptotic properties in the Gaussian family of distributions when the correlation between the paired observations is positive.

Other researchers, such as Xu and Harrar [22] and Konietschke *et al*. [9] also discuss the problem for continuous variables including the normal distribution by using weighted statistics. However, the procedure suggested by Xu and Harrar [22] is a functional smoothing to the Looney and Jones [13] procedure. As such, the Xu and Harrar procedure is not a practical alternative for the non-statistician researcher. The procedure suggested by Konietschke *et al*. [9], is a non-parametric procedure based on ranking.

In this paper, we propose using a weighted t-test and a weighted z-test to combine a two independent sample test with a matched pair's test. Section 2 outlines various competing methods found in the literature. Our proposed weighted tests and their null distributions are provided in Section 3. Section 4 provides the detail for our proposed test statistic. In Section 5, we will provide a bootstrap method to estimate the P-value of the weighted test in case of small samples. Simulation studies are provided in Section 6. Section 7 illustrates with an example, while Section 8 concludes with discussion and final remarks.

2. Competing methods used in the literature

Methods that are most commonly used to analyze a combination of correlated and uncorrelated data are as follows:

(1) Using all data with a t-test for two independent samples assuming no correlation among the observations in the two treatments.
(2) Ignoring the paired observations and perform the usual t-test of two independent samples after deleting the correlated data.
(3) Ignore the independent observations of treatment 1 and 2 and perform the usual paired t-test on the correlated data.
(4) The corrected z-test by Looney and Jones [13].

To perform the Looney and Jones test, let $\{X_1, X_2, \ldots, X_{n_1}\}$ and $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ denote two independent random samples of subjects receiving either treatment 1 or treatment 2, respectively. Suppose there are n paired subjects in which one member of the pair receives treatment 1 and the other paired member receives treatment 2. Let $\{(U_1, V_1), (U_2, V_2), \ldots, (U_n, V_n)\}$ denote the observed values of the paired (correlated) subjects. Looney and Jones assumed that x-and$u$-observations come from a common normal parent population and y-and v-observations come from a common normal parent population, which may be different from x and u observations. Let M_1 denotes the sample mean for all treatment 1 subjects; that is the mean of all x- and u-values combined, and let M_2 denote the sample mean for all treatment 2 subjects; that is, the mean of all $y-$ and $v-$values combined. Let S_1^2 denote the sample variance for all treatment 1 subjects and let S_2^2 denote the sample variance for all treatment 2 subjects. The Looney and Jones proposed test statistic is

$$Z_{\text{Corr}} = \frac{M_1 - M_2}{\sqrt{S_1^2/(n_1+n) + S_2^2/(n_2+n) - 2nS_{uv}^2/(n_1+n)(n_2+n)}},$$

where S_{uv}^2 is the sample covariance of the paired observations.

Under the null hypothesis, Z_{Corr} has asymptotic $N(0,1)$ distribution. However, this test works only for a large sample size. In case of small sample sizes, the exact distribution is not clear. An approximation to the exact distribution critical values is needed. Bootstrap methods to find the p-value of the test may also work. In addition, under the assumption of a large sample size, this test is not a uniformly powerful test. Its power depends on the correlation between the correlated observations. As an alternative, we propose the following test procedure.

3. Proposed weighted tests

As in Looney and Jones [13], let $\{X_1, X_2, \ldots, X_{n_1}\}$ and $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ denote two independent random samples of subjects receiving either treatment 1 or treatment 2, respectively. Suppose there are n paired subjects in which one member of the pair receives treatment 1 and the other paired member receives treatment 2. Let $\{(U_1, V_1), (U_2, V_2), \ldots, (U_n, V_n)\}$ denote the observed values of the paired subjects. Assume that x-and u-observations come from a common normal parent population $N(\mu_1, \sigma_1^2)$ and y-and$v$-observations come from a common normal parent population $N(\mu_2, \sigma_2^2)$. Let $D_i = U_i - V_i, i = 1, 2, \ldots, n$. D_i is $N(\mu_1 - \mu_2, \sigma_D^2)$, where ρ is the correlation coefficient between U and V and $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$ and $\sigma_D^2 = 2\sigma^2(1-\rho)$, when $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Let, $\bar{X}$, $\bar{Y}$ and $\bar{D}$ denote the sample means of x-observations, y-observations, and d-observations, respectively. Also, let S_x^2, S_y^2 and S_d^2 denote the sample variances of x-observations, y-observations, and d-differences between u and v observations, respectively. Let $N = n_1 + n_2 + n$

and $\gamma = (n_1 + n_2)/N$, which represents the proportion of the independent observations. We propose the following test procedure for testing the null hypothesis H_0: $\mu_1 = \mu_2$, where μ_1 and μ_2 are the respective response means of treatment 1 and treatment 2:

$$T_0 = \sqrt{\gamma}\frac{\bar{X} - \bar{Y}}{\sqrt{S_x^2/n_1 + S_y^2/n_2}} + \sqrt{1-\gamma}\frac{\bar{D}}{S_d/\sqrt{n}}. \tag{1}$$

When $\gamma = 1$, this test reduces to the two-sample t-test. Also, when $\gamma = 0$, this test is the matched paired t-test.

Case 1. Large sample sizes will generally mean both a large number of matched paired observations and a large number of two independent samples from treatment 1 and treatment 2. By applying Slutsky's Theorem, under the null hypothesis, T_0 has an approximate $N(0,1)$ distribution. The p-value of the test can, therefore, be directly calculated from the standard normal distribution.

Case 2. Without loss of generality, we will consider that the paired data have a small sample size while the independent two-samples from the two treatments have large sample sizes. To find the distribution of the weighted test, under the null hypothesis, let $T_0 = \sqrt{\gamma}X + \sqrt{1-\gamma}T_k$ where X has $N(0,1)$ and T_k has t-distribution with k degrees of freedom. Then

$$f_{T_0}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\gamma}}\phi\left(\frac{t-\sqrt{\gamma}x}{\sqrt{1-\gamma}}\right) t_k(x)\,\mathrm{d}x. \tag{2}$$

Nason [14] in an unpublished report found the distribution of T_0 when the degrees of freedom are odd. The distribution provided by Nason [14] is very complicated and cannot be used directly to find percentiles from this distribution. To find the p-value of T_0 you need to use a package published by Nason [14]. Therefore, we provide a simple bootstrap algorithm to find the $p-$value of test procedure based on the distribution of T_0. A similar approach may be taken when the paired data has a large sample size and the independent data have a small sample size.

Case 3. Both data sets, the independent samples and the matched paired data, have small sample sizes. Under the null hypothesis T_0 has weighted t-distribution. Let $T_0 = \gamma T_{k_1} + (1-\gamma)T_{k_2}$ where T_{k_1} and T_{k_2} are two independent t-variates with k_1 (in case of unequal variances Satherwaite's method can be used to estimate the degrees of freedom, see [10]) and k_2 degrees of freedom, respectively. Walker and Saw [21] derived the distribution of a linear combination of t-variates when all degrees of freedom are odd. In our case, since we have only two t-variates with k_1 and k_2 degrees of freedoms, we need to assume that k_1 and k_2 are odd numbers or we can manipulate the data to have both numbers to be odd. Using Walker and Saw [21] results, one can find the p-value of the suggested test statistics T_0. However, the Walker and Saw [21] method still needs an extensive amount of computation. Therefore, a bootstrap algorithm also may be used to find the p-value of T_0.

4. New test procedure

Under the assumption of MCAR, to test the null hypothesis $H_0 : \mu_1 = \mu_2$, we introduce the following notation: $D_i = U_i - V_i$, $i = 1, 2, \ldots, n$, and $\bar{D} = \sum_{i=1}^{n} D_i/n$.

We propose the following test statistics to test $H_0 : \mu_1 = \mu_2$ as follows:

$$T_{\text{New}} = \frac{\overline{D} + (\bar{X} - \bar{Y})}{\sqrt{\theta_1 S_d^2 + \theta_2 S_P^2}}, \tag{3}$$

where

$$S_d^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}, \quad S_P^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_1 + n_2 - 2},$$

$$\theta_1 = \frac{1}{n} \quad \text{and} \quad \theta_2 = \frac{1}{n_1} + \frac{1}{n_2}.$$

Note that $\mathrm{Var}(\overline{D} + \bar{X} - \bar{Y}) = \theta_1\sigma_d^2 + \theta_2\sigma^2$, under the normality assumption and $H_0 : \mu_1 = \mu_2$, $\frac{(n-1)S_d^2}{\sigma_d^2} \xrightarrow{L} \chi^2_{(n-1)}$ and $\frac{\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2}{\sigma^2} \xrightarrow{L} \chi^2_{(n_1+n_2-2)}$.

Therefore, under the null hypothesis and using Satherwaite's method (see[11]),

$$T_{\mathrm{New}} \xrightarrow{L} t(\mathrm{d}f_s);\ \mathrm{d}f_s \approx \frac{(\theta_1 S_d^2 + \theta_2 S_P^2)^2}{(\theta_1 S_d^2)^2/(n-1) + (\theta_2 S_p^2)^2/(n_1+n_2-2)}. \tag{4}$$

5. Bootstrap method to estimate the p-value of T_0 in case II and III

Uniform bootstrap resampling was introduced by Efron [4]. The uniform resampling for the two independent sample case is discussed by Ibrahim [6] and Samawi *et al.* [17,18]. We suggest applying uniform bootstrap resampling as a means of obtaining p-values for our test procedure. However, since our test procedure involves t-statistic, there are some conditions discussed by Janssen and Pauls [8] and Janssen [7] that need to be verified to insure that the test statistic under consideration will have a proper convergence rate. They indicated that the bootstrap works if and only if the so-called central limit theorem holds for the test statistics.

In our case $\{X_1, X_2, \ldots, X_{n_1}\}$ and $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ and $\{(U_1, V_1), (U_2, V_2), \ldots, (U_n, V_n)\}$ are independent samples, thus the bootstrap p-value can be calculated as follows:

(1) Use the original sample sets $\{X_1, X_2, \ldots, X_{n_1}\}$ and $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ and $\{(U_1, V_1), (U_2, V_2), \ldots, (U_n, V_n)\}$ to calculate T_0.
(2) Let $\{X_1^*, X_2^*, \ldots, X_{n_1}^*\}$ and $\{Y_1^*, Y_2^*, \ldots, Y_{n_2}^*\}$ and $\{(U_1^*, V_1^*), (U_2^*, V_2^*), \ldots, (U_n^*, V_n^*)\}$ be the centered samples by subtracting the sampling means $\bar{X}$, $\bar{Y}$ and $(\bar{U}, \bar{V})$ respectively.
(3) With placed probabilities $((1/n_1, 1/n_2, 1/n)$ on the samples in step (2), respectively, generate independently bootstrap samples namely $\{X_{i1}^{**}, X_{i2}^{**}, \ldots, X_{in_1}^{**}\}$ and $\{Y_{i1}^{**}, Y_{i2}^{**}, \ldots, Y_{in_2}^{**}\}$ and $\{(U_{i1}^{**}, V_{i1}^{**}), (U_{i2}^{**}, V_{i2}^{**}), \ldots, (U_{in}^{**}, V_{in}^{i*})\}$, $i = 1, 2, \ldots, B$.
(4) For each $i = 1, 2, \ldots, B$ set of samples in step 3, compute the corresponding bootstrap version of T_0 statistics, namely, $T_1^{**}, T_2^{**}, \ldots, T_B^{**}$.
(5) Bootstrap p-value is computed as

$$P^{**} = \frac{\sum_{i=1}^{B} I(|T_i^{**}| \geq |T_0)}{B},$$

where

$$I(|T_i^{**}| \geq |T_0|) = \begin{cases} 1 & \text{if } |T_i^{**}| \geq |T_0 \\ 0 & \text{Otherwise.} \end{cases}$$

6. Simulation study

To investigate the performance of our proposed test procedures for partially correlated data, a simulation study was conducted. We used 0.05 level of significance and 1000 simulated samples from normal distributions for each power estimation. It was compared to the test of

Table 1. Empirical comparison with the size of the tests with Looney and Jones [13] test for a size ($\alpha = 0.05$) and $\sigma_1 = \sigma_2 = 1$.

ρ	N	n_1	n_2	T_{New}	Z_{Corr}	T_0-Normal approximation
0.5	10	5	5	0.024	0.034	0.047
	10	10	5	0.018	0.038	0.054
	10	5	10	0.025	0.041	0.051
−0.5	10	5	5	0.027	0.044	0.043
	10	10	5	0.026	0.049	0.042
	10	5	10	0.020	0.035	0.037
0.9	10	5	5	0.027	0.029	0.050
	10	10	5	0.036	0.047	0.069
	10	5	10	0.023	0.019	0.054
−0.9	10	5	5	0.019	0.043	0.033
	10	10	5	0.031	0.050	0.040
	10	5	10	0.031	0.051	0.045

Table 2. Empirical power comparison with Looney and Jones [13] for $\sigma_1 = \sigma_2 = 1$.

E=true effects	ρ	n	n_1	n_2	T_{New}	Z_{Corr}	T_0-Normal approximation
0.20	0.50	20	10	20	0.110	0.149	0.154
		20	10	10	0.057	0.120	0.128
		20	20	10	0.097	0.144	0.144
		20	20	20	0.119	0.127	0.133
	0.90	20	10	20	0.167	0.235	0.414
		20	10	10	0.148	0.254	0.431
		20	20	10	0.151	0.231	0.385
		20	20	20	0.200	0.216	0.411
0.50	0.50	20	10	20	0.587	0.671	0.681
		20	10	10	0.484	0.660	0.660
		20	20	10	0.577	0.677	0.671
		20	20	20	0.710	0.710	0.735
	0.90	20	10	20	0.706	0.836	0.983
		20	10	10	0.526	0.844	0.987
		20	20	10	0.669	0.833	0.983
		20	20	20	0.832	0.844	0.988

Looney and Jones [13]. Tables 1– 4 present the simulated probability of the type I error and the power comparison between our weighted test T_0 (Normal approximation), T_{New} and the corrected z-test Z_{Corr} by Looney and Jones [13]. The sets of parameters used in the simulation are $\{(\sigma_1 = \sigma_2 = 1), (\sigma_1 = 1, \sigma_2 = 3)$ and $\rho = \pm 0.1 \pm 0.5, \pm 0.9\}$, $\{n_1 = 5, 10, 20, n_2 = 5, 10, 20,$ and $n = 10, 20, 30\}$ and true effects size E=0, 0.2, and 0.5. For the sake of comparison with a recent approach discussed by Konietschke *et al.* [9], the parametric test T_θ^{para} is presented in Table 5. We used the following parameters in our simulation: total sample size N=30, the percentage of MAR r=10% from each arm, the true effects size E=0, 0.5, 0.55, 0.60, 0.70, and 0.80, and the $\{\sigma_1 = \sigma_2 = 1$ and $\rho = 0.5\}$. Table 5 shows our test is more powerful than the Looney and Jones [13] test in this case. Also, our test and the Looney and Jones [13] test are more powerful than the Konietschke *et al.* [9] test except when the power of the test get close to 100%.

Our simulation shows that our weighted test procedure is more powerful than Looney and Jones [13] in most of the presented cases. Also, our simulation shows that the new test T_{New} is a conservative test procedure. In general, T_{New} is not more powerful than Looney and Jones [13]. However, Looney and Jones [13] test procedure is an asymptotic procedure and valid only for large sample cases.

Table 3. Empirical power comparison with Looney and Jones [13] for $\sigma_1 = \sigma_2 = 1$.

E=true effects	ρ	n	n_1	n_2	T_{New}	Z_{Corr}	$T_{0\text{-Normal approximation}}$
0.20	0.10	20	10	20	0.135	0.151	0.163
		20	10	10	0.118	0.147	0.158
		20	20	10	0.115	0.143	0.143
		20	20	20	0.151	0.160	0.166
	−0.10	20	10	20	0.103	0.123	0.129
		20	10	10	0.104	0.126	0.122
		20	20	10	0.122	0.141	0.132
		20	20	20	0.129	0.139	0.152
0.50	0.10	20	10	20	0.516	0.576	0.557
		20	10	10	0.439	0.542	0.537
		20	20	10	0.490	0.560	0.533
		20	20	20	0.637	0.648	0.651
	−0.10	20	10	20	0.448	0.507	0.492
		20	10	10	0.421	0.491	0.499
		20	20	10	0.492	0.551	0.532
		20	20	20	0.559	0.580	0.590

Table 4. Empirical power comparison with Looney and Jones [13] for $\sigma_1 = 1, \sigma_2 = 3$.

E=true effects	ρ	n	n_1	n_2	T_{New}	Z_{Corr}	$T_{0\text{-Normal approximation}}$
0.5	0.50	20	10	20	0.232	0.164	0.182
		20	10	10	0.134	0.172	0.182
		20	20	10	0.223	0.162	0.182
		20	20	20	0.200	0.207	0.217
	0.90	20	10	20	0.248	0.177	0.200
		20	10	10	0.172	0.217	0.234
		20	20	10	0.270	0.205	0.232
		20	20	20	0.232	0.244	0.256

Table 5. Empirical power comparison with Looney and Jones [13] and Konietschke *et al*. [9] test for total sample size of $N = 30$ and $\sigma_1 = \sigma_2 = 1$.

R	ρ	d=true effect	T_{New}	Z_{Corr}	$T_{0\text{-Normal approximation}}$	${}^{a}T_{\theta}^{\text{para}}$
10%	0.50	0.00	0.029	0.052	0.060	0.050
		0.50	0.179	0.677	0.729	0.052
		0.55	0.226	0.788	0.810	0.136
		0.60	0.283	0.871	0.894	0.420
		0.70	0.320	0.942	0.951	0.942
		0.80	0.360	0.982	0.983	1.000

Note: [a]Extracted from Konietschke *et al*. [9] Tables 2 and 3 for comparison purposes.

7. Illustration: A vaginal pessary satisfaction data

A vaginal pessary is a removable device placed into the vagina. It is designed to support areas of pelvic organ prolapse. Table 6 contains part of an unpublished study by Dr Catherine Bagley and Dr Robert Vogel on the value of estrogen therapy for certain types of patients currently using vaginal pessaries. The data come from a satisfaction survey of women aged 45 or older. The total score is 135 which would be interpreted as complete satisfaction. The data also include the age of the woman and the number of years she used a pessary (at time of first survey) as well as the date of the second survey.

Table 6. Vaginal pessary satisfaction data.

Patient	Score 1	Score 2	Patient	Score 1	Score 2	Patient	Score 1	Score 2
1	11	6	22	19	13	43	23	.
2	10	4	23	15	11	44	16	.
3	17	14	24	11	8	45	12	.
4	16	22	25	15	12	46	.	21
5	18	15	26	10	11	47	.	11
6	12	9	27	18	12	48	.	14
7	21	19	28	12	11	49	.	21
8	13	11	29	21	13	50	.	10
9	30	29	30	24	21	51	.	13
10	11	7	31	16	13	52	.	8
11	12	13	32	18	.	53	.	14
12	10	7	33	11	.	54	.	21
13	21	12	34	15	.	55	.	10
14	19	11	35	18	.	56	.	11
15	17	15	36	22	.	57	.	23
16	36	30	37	24	.	59	.	11
17	16	16	38	14	.	60	.	12
18	11	9	39	17	.	61	.	13
19	9	7	40	16	.	62	.	20
20	21	14	41	17	.			
21	13	16	42	24	.			

Table 7. Summary inference for vaginal pessary satisfaction data.

Method	Test	p-Value	Type
Z_{Corr}	4.512	0.00000322	Normal-approximation
T_0	5.529	0.00000000	Bootstrap method
T_{New}	3.435	0.00077690	T-approximation

Table 7 indicates that all of the tests discussed in this paper show strong statistical evidence, that on average, the satisfaction scores are lower on the second survey than on the first survey.

8. Discussion and final remarks

Occasionally, we end up with partial correlated data on a matched-pair or repeated measure studies due to missing observations and we must address these issues in our analysis. For quantitative responses, statistical methods are well established for analyzing correlated data. For partially correlated data, there are concerns that need to be addressed due to the complexity of the analysis in particular for small sample sizes.

Looney and Jones [13] and others introduced different method to deal with partially correlated data. In this paper, we introduced two different approaches, the weighted tests procedure and the new procedure to combine between the two data sets. Our simulation indicated that our two approaches are competing with those in the literature and weighted approach is more powerful than the others in most cases. However, our new proposed test is a conservative test and valid for a small and moderate sample size.

References

[1] M.G. Akritas, J. Kuha, and D.W. Osgood, *A nonparametric approach to matched pairs with missing data*, Sociol. Methods Res. 30 (2002), pp. 425–462.

[2] E. Brunner and M.L. Puri, *Nonparametric methods in design and analysis of experiments*, Handbook of Statist. 13 (1996), pp 631–703.
[3] I.W. Dimery, K. Nishioka, B. Grossie, D.M. Ota, S.P. Schantz, K.T. Robbins, and W.K. Hong, *Polyamine metabolism in carcinoma of oral cavity compared with adjacent and normal oral mucosa*, Am. J. Surg. 154 (1987), 429–433.
[4] B. Efron, *Bootstrap methods: Another look at the jackknife*, Ann. Statist. 7 (1979), pp. 1–26.
[5] G. Ekbohm, *Comparing means in the paired case with missing data on one response*, Biometrika 63(1) (1976), pp. 169–172.
[6] H.I. Ibrahim, *Evaluating the power of the Mann–Whitney test using the bootstrap method*, Comm. Statist. Theory Methods 20 (1991), pp. 2919–2931.
[7] A. Janssen, *Resampling Student's t-type statistics*, Ann. Inst. Statist. Math. 57 (2005), pp. 507–529.
[8] A. Janssen and T. Pauls, *How do bootstrap and permutation tests work*? Ann. Statist. 31 (2003), pp. 768–806.
[9] F. Konietschke, S.W. Harrar, K. Lange, and E. Brunner, *Ranking procedures for matched pairs with missing values – asymptotic theory and a small sample approximation*, Comput. Statist. Data Anal. 56 (2012), pp. 1090–1102.
[10] M. Larsen, *An Introduction to Mathematical Statistics and its Applications*, 3rd ed., Prentice-Hall, Inc., Upper Saddle River, NJ 07458, 2001, p. 455.
[11] P. Lin and L.E. Stivers, *On difference of means with incomplete data*, Biometrika 61(2) (1974), pp. 325–334.
[12] R.J. Little, *A test of missing completely at random for multivariate data with missing values*, J. Amer. Statist. Assoc. 83 (1988), pp. 1198–1202.
[13] S.W. Looney and P.W. Jones, *A method for comparing two normal means using combined samples of correlated and uncorrelated data*, Stat. Med. 22 (2003), pp. 1601–1610.
[14] G.P. Nason, *On the sum of t and Gaussian random variables*, 2005. Available at http://www.maths.bris.ac.uk/∼guy/Research/papers/SumTGauss.pdf
[15] J. Nurnberger, D. Jimerson, J.R. Allen, S. Simmons, and E. Gershon, *Red cellouabain-sensitive* $Na^{+} - K^{+}$ *-adenosine triphosphatase: A state marker in affective disorder inversely related to plasma cortisol*, Biol. Psychiatry 17(9) (1982), pp. 981–992.
[16] G. Rempala and S. Looney, *Asymptotic properties of a two-sample randomized test for partially dependent data*, J. Statist. Plann. Inference 136 (2006), pp. 68–89.
[17] H.M. Samawi, G.G. Woodworth, and M.F. Al-Saleh, *Two-sample importance resampling for the bootstrap*, Metron LIV (1996), pp. 3–4.
[18] H.M. Samawi, G.G. Woodworth, and J. Lemke, *Power estimation for two-sample tests using importance and antithetic resampling*, Biom. J. 40(3) (1998), pp. 341–354.
[19] G.W. Snedecor and W.G. Cochran, *Statistical Methods*, 7th ed., Iowa State University Press, Ames, IA, 1980.
[20] A.C. Steere, J. Green, R.T. Schoen, E. Taylor, G.J. Hutchinson, D.W. Rahn, and S.E. Malawista, *Successful parenteral penicillin therapy of established Lyme arthritis*, New Engl. J. Med. 312(14) (1985), pp. 8699–8874.
[21] G.A. Walker and J.G. Saw, *The distribution of linear combinations of t-variables*, J. Amer. Statist. Assoc. 73(364) (1978), pp. 876–878.
[22] J. Xu and S.W. Harrar, *Accurate mean comparisons for paired samples with missing data: An application to a smoking-cessation trial*, Biom. J. 54 (2012), pp. 281–295.