

Combining Paired and Two-Sample Data Using a Permutation Test

Richard L. Einsporn* and Desale Habtzghi
University of Akron

Abstract: This paper presents a permutation test for the incomplete pairs setting. This situation arises in both observational and experimental studies when some of the data are in the form of a paired sample and the rest of the data comprise two independent samples. The proposed method uses the data from the two types of samples to test the difference between the mean responses. Our test statistic combines the observed mean difference for the complete pairs with the difference between the two means of the independent samples. The randomizations are carried out as is typically done with standard permutation tests for paired and independent samples. We show by a simulation study that our statistic performs well in comparison to other methods.

Key words: Equality of means, incomplete pairs, paired t -test, permutation test.

1. Introduction

In paired data situations it is often the case that some of pairs are missing one or the other piece of data. Consider, for example, a pre-test post-test study involving a quantitative dependent variable where some of the subjects are absent during the pre-test and others during the post-test. A common way to handle the resulting data is to discard the incomplete pairs and carry out a paired t -test using only the complete pairs. This approach is recommended in the textbooks by Motulsky (2010, p. 237), Ling (2012, Section 11.1), and Ha and Ha (2012, p. 173), but sacrifices potentially useful information contained in the discarded data. An alternative approach (Ha and Ha, 2012, p. 173) is to keep all the data and run the two-sample t -test, treating the pre-test and post-test data as two independent samples. However, this is an incorrect analysis, since the data from the complete pairs are clearly dependent.

*Corresponding author.

Since the 1970's various methods have been proposed for combining the data from the complete pairs and incomplete pairs in order to compare means. Early work in this area was done by Morrison (1973), Lin and Stivers (1974), Ekbohm (1976), Woolson, Leeper, Cole and Clark (1976), Bhoj (1978), and Hamdan, Khuri and Crews (1978). Bhoj (1989) introduced additional methods and provided a summary and an empirical power comparison of many of these approaches, all of which are parametric in nature. Dubnicka, Blair and Hettmansperger (2002) developed nonparametric procedures for the incomplete pairs situation and showed that their methods, which combine Wilcoxon one and two-sample procedures, compare favorably in terms of efficiency to the parametric approaches of Bhoj, Ekbohm, and Lin and Stivers. A permutation test for the incomplete pairs setting was given by Maritz (1995). However, that method treats the incomplete pairs in a non-standard fashion, and does not reduce to the usual permutation-based alternative to the two-sample t when all the pairs are incomplete.

In this paper we propose a permutation-based method in which the complete pairs and incomplete pairs are handled in the standard ways associated with randomization testing and then combined into one statistic by optimally weighting the complete and incomplete portions of the data set, taking into account the relative sample sizes and the correlation coefficient for the complete pairs. We begin by presenting our method and then show its application to an experimental study to compare two methods of DNA extraction. Next, we apply our method to compare the times of runners at two 5K races. Finally, we review some of the alternative statistics that have been proposed for the incomplete pairs situation, and show through simulations that our statistic performs well in comparison to the alternative methods.

2. Development of the Test Statistic

We consider the situation where two variables X and Y are both observed for some of the cases, but for other cases only X or only Y is observed. Letting F_X and F_Y be the distributions of X and Y , we test the null hypothesis that $F_X = F_Y$ against the alternative that the distributions differ by a location shift. (See Good, 2006; Efron and Tibshirani, 1993, for details.) Thus, we are testing the equality of the means of the X 's and Y 's under the assumption of constant variance and constant shape for the underlying distributions. Let n_p represent the number of complete pairs where both observations (x, y) are present. Let n_{ux} be the number of unpaired observations where only the first value (x) is present and n_{uy} be the number of unpaired observations where only the second value (y) is present. For the complete pairs portion of the data, the standard permutation test considers each of the 2^{n_p} possible interchanges of x and y within each of

the n_p (x, y) pairs. (See, for example, Moore and McCabe, 2005.) The mean difference statistic, $\bar{d}_p = \bar{x}_p - \bar{y}_p$, where \bar{x}_p and \bar{y}_p are the means of the x 's and y 's for the complete pairs only, is calculated for the actual data as well as for each of the possible (x, y) permutations. The p -value for the paired data test is the fraction of the 2^{n_p} \bar{d}_p 's that are greater to the value of \bar{d}_p for the actual data. In cases where n_p is large, a sample of the possible permutations is used rather than every possible permutation.

For the unpaired portion of the data, the standard permutation test considers each of the $\binom{n_{ux} + n_{uy}}{n_{ux}}$ possible ways the data can be partitioned into one set of n_{ux} x -values and another set of n_{uy} y -values. (See, for example, Moore and McCabe, 2005.) For each partition, the value of $\bar{d}_u = \bar{x}_u - \bar{y}_u$ is calculated, where \bar{x}_u and \bar{y}_u are the means of the x 's and the y 's for the unpaired data. The p -value for the standard two-sample permutation test is the fraction of the \bar{d}_u 's that are greater than or equal to the value of \bar{d}_u for the actual data. Again, when the number of partitions is extremely large, a random sample of partitions is used rather than enumerating every possible partition.

In order to utilize information contained in both the paired and unpaired data, we propose the following statistic, a linear combination of the standard paired and unpaired statistics:

$$T = w\bar{d}_p + (1 - w)\bar{d}_u,$$

where the weight $w \in [0, 1]$ is chosen to minimize the variance of T . Let μ_X , μ_Y , σ_X^2 and σ_Y^2 be the means and variances of the two variables X and Y , and ρ be the correlation. We assume that the study design is such that the paired subjects are independent of the unpaired subjects and that, among the unpaired subjects, those with x values only are independent of those with y values only. Then T is unbiased for $\mu_X - \mu_Y$ and the variance of T is

$$\text{var}(T) = w^2 \left(\frac{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y}{n_p} \right) + (1 - w)^2 \left(\frac{\sigma_X^2}{n_{ux}} + \frac{\sigma_Y^2}{n_{uy}} \right).$$

The variance of T is minimized when the weight is set to

$$w = (\sigma_X^2/n_{ux} + \sigma_Y^2/n_{uy}) / [(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)/n_p + (\sigma_X^2/n_{ux} + \sigma_Y^2/n_{uy})].$$

When $\sigma_X^2 = \sigma_Y^2$, which is the case under our null hypothesis, this reduces to

$$w = (1/n_{ux} + 1/n_{uy}) / [(2 - 2\rho)/n_p + (1/n_{ux} + 1/n_{uy})].$$

Note that when $\rho = 1$, then the weight is $w = 1$, so that in the case of perfect correlation, the entire weight will be given to the complete pairs. If $\rho = 0$ and

the number of unpaired x 's and y 's is the same, say $n_u = n_{ux} = n_{uy}$, then $w = n_p / (n_p + n_u)$. Thus, when there is no correlation between the complete pairs, the weights given to the complete and incomplete pairs are proportional to the sample sizes. In practice the true value of ρ would be unknown and replaced by the sample correlation. As the exact distribution of the statistic T is unknown, this is the type of situation where a permutation approach may prove beneficial. Our method is to compute T for the given data, using the sample to obtain the weight w . Then we randomly permute the complete pairs, and we randomly re-partition the incomplete pairs, as described above. There are $2^{n_p} \binom{n_{ux} + n_{uy}}{n_{ux}}$ ways to do this, each resulting in a value of T . For a right-tailed alternative, the p -value of our test is the proportion of these T 's that are equal to or greater than our observed value of T .

More formally, our algorithm goes as follows.

1. Compute the observed value of the statistic T using the original data; call it T_{obs} .
2. For the paired observations generate n_p uniform random variables u_i . Set $(x_i^* = x_i, y_i^* = y_i, \text{ if } u_i \leq 0.5)$; otherwise, $(x_i^* = y_i, y_i^* = x_i, \text{ if } u_i > 0.5)$, for $i = 1, \dots, n_p$. For each permutation of the data calculate \bar{d}_p . Calculate the correlation coefficient r for the permuted paired data and use this to determine the weight w .
3. Let $N = n_{ux} + n_{uy}$ be the combined unpaired sample size and let $Z = (z_1, \dots, z_N)$ be the combined vector of the N unpaired data values, where, for convenience, the first n_{ux} of the z_i 's are observed x 's in the original data and the remaining z_i 's are observed y 's in the original data. For a random partition, choose n_{ux} of the N z 's at random and without replacement to be x 's, with the remaining n_{uy} z 's designated as y 's. Calculate \bar{d}_u .
4. Repeat Steps 2 and 3 a large number of times, say B , and compute the test statistic T each time. Denote the results by T_1^*, \dots, T_B^* . (Following Manly, 1997, we use $B = 5000$ permutations. See the Appendix for a justification.)
5. The approximate p -value for a two-tailed test is calculated as $P_{H_0}(|T^*| \geq |T_{obs}|) \approx [\sum_{j=1}^B I(|T_j^*| \geq |T_{obs}|)] / B$. Here, $I(\cdot) = 1$ if the argument is true; 0 otherwise. The test rejects H_0 for small p -values.

This algorithm is implemented in R as a permutation test function, which may be obtained at <http://gozips.uakron.edu/dh52/>.

3. Examples

One of the goals of an honors research project conducted by Riordan (2012) was to compare two methods for extracting DNA from coyote blood samples. A total of 30 different coyotes were used for the study. One of the methods was the QIAGEN DNeasy® Blood and Tissue Kit and the other was the more traditional chloroform isoamyl alcohol method. The researcher wanted to determine whether these extraction methods differ with respect to the mean concentration of DNA. Due to time and cost considerations, DNA was measured using both methods for only 6 of these coyotes, selected at random from the 30. Riordan decided to use the kit method only for 8 of the coyotes and the chloroform method only for the remaining 16 coyotes. These were randomly assigned by the researcher. (The researcher chose twice as many for chloroform analysis in order to gain more experience using this more widely used technique.)

In this example, $n_p = 6$, $n_{ux} = 8$, and $n_{uy} = 16$. The data are the obtained concentrations of DNA, and are shown in Table 1, where the first two lines are for the six complete pairs. The mean difference statistic for the paired data values is $\bar{d}_p = -1.03$ and the mean difference for the unpaired portion is $\bar{d}_u = \bar{x}_u - \bar{y}_u = 2.53 - 2.82 = -0.29$. As the permutation test requires equal variances, we note that the standard deviations are 1.16 and 1.07 for the unpaired Kit and Chloroform samples, respectively. Using only the unpaired data, Levene's test for equality of variances for the two DNA extraction methods yields a p -value of 0.467, suggesting that the equal variance assumption is tenable. We chose not to use Levene's test for the combined paired and unpaired data, as the paired portion would violate the assumption of independent samples. Ideally, we could test equality of variances using a method that is applicable for a combination of complete and incomplete pairs. Bhoj (1979) and Ekbohm (1982) have proposed tests for this situation, but their methods require normality, and are therefore not applicable for the DNA data. We are unaware of the existence of any robust test for equality of variance when some of the data are paired and the remaining data are unpaired. Further research is needed to address this issue.

Table 1: DNA extraction concentrations ($ng/\mu L$) for Kit and Chloroform methods

Kit	7.27	4.17	2.21	3.68	12.07	5.06		
Chlor.	7.60	0.16	3.60	0.04	6.16	23.08		
Kit only	2.00	2.83	0.95	1.25	4.15	3.85	2.08	3.10
Chlor. only	3.24	1.80	2.72	2.36	3.23	3.80	3.20	4.80
	2.96	3.00	2.56	0.24	1.16	3.16	3.68	3.12

With a correlation of $r = 0.144$ for the complete pairs, the resulting weight is $w = 0.397$. The combined permutation statistic is $T = -0.587$, which has a corresponding p -value of approximately 0.823 based on 5000 permutations. Thus,

there is not a statistically significant difference between the two techniques with respect to DNA extraction concentration.

For a second example, we compare the runners' times for two local 5K races. As one course (Kent) tends to have more hills, we expect that runners in that race will take about one minute longer to finish, on average, than for the race with a flatter 5K course (Tallmadge). There were 32 runners who competed in both of these races, 478 who competed in only the Kent race, and 541 who competed in only the Tallmadge race. The runners' times were found online at the races' websites. We merged the data using the runners' names and converted their times to seconds. The resulting data set is provided at <http://gozips.uakron.edu/dh52/>. We wish to test $H_0 : \mu_X - \mu_Y = 60$ sec. versus a two-tailed alternative. For the 32 complete pairs, the mean difference $\bar{d}_p = 89.1$ sec. and the paired t -test has a p -value of 0.183. For the incomplete pairs the mean difference is $\bar{d}_u = 105.5$. The standard deviations are 382.6 and 429.0 for the unpaired Kent and Tallmadge data, respectively. Levene's test for equality of variance gives a p -value of 0.216, suggesting that the equal variance assumption may be valid. The pooled two-sample t -test performed on the incomplete pairs yields a p -value of 0.076. If the paired nature of the 32 complete pairs is ignored and a two-sample t -test is used for all of the data, the resulting p -value for this (improper) approach is 0.083. Thus, none of these methods produce statistically significant results at the $\alpha = 0.05$ level. However, our permutation test gives an approximate p -value of 0.044 based on 5000 permutations. Therefore, at the 5% significance level we conclude that times for runners at the Kent 5K tend to be more than a minute slower than for the Tallmadge 5K.

4. Alternative Approaches

Bhoj (1978) proposed using a weighted average of the paired t -statistic computed for the complete pairs and the pooled two-sample t -statistic computed for the incomplete pairs. However, the distribution of that statistic and other straight-forward statistics, such as our own, is unknown for small sample sizes. Therefore, much of the early research consisted of attempts to derive related statistics that have known distributions. For example, Bhoj (1984) came up with a method for transforming the paired t and the two-sample t such that a combination of those transformed statistics is approximately normal. Lin and Stivers (1974) and Bhoj (1989, p. 283) developed statistics of the form $(\bar{x} - \bar{y})/D$, where \bar{x} and \bar{y} are based on all of the paired and unpaired data without distinction. Thus, as pointed out in Dubnicka *et al.* (2002), these statistics ignore the actual design of the study. The denominator D in each case is a complex expression, leading to statistics with approximate t -distributions under the assumptions of normality and equal variances. Details for both of these statistics are provided

in Bhoj (1989). Another statistic proposed by Bhoj (1989, p. 282), referred to as Z_B , performed well in a simulation study. Dubnicka *et al.* (2002) showed that Z_B is equivalent to a statistic based on a linear model approach.

As the early research into this problem took place prior to the resurgence of permutation methods, it seems that a permutation approach was overlooked in the 1970's and 1980's. Because, straight-forward statistics based on combining \bar{d}_p and \bar{d}_u do not have known distributions, these were bypassed in favor of complex and non-intuitive expressions developed in order to have approximate t or Z distributions. Instead, we believe that this is exactly the type of setting for which a permutation test may be an attractive alternative.

In 1995, Maritz proposed a permutation test for the incomplete pairs problem. For that test, the complete pairs are permuted in the standard fashion. That is, the complete (x, y) pairs are randomly interchanged. However, the incomplete pairs are not handled in the standard way which fixes n_{ux} and n_{uy} for all randomizations. (See, for example, Good, 2006, p. 37.) Maritz' approach does not maintain fixed sample sizes for the unpaired data for the randomizations. Denoting the incomplete pairs as either (x, \cdot) or (\cdot, y) , Maritz' method randomly interchanges the x or y with the missing value. Thus, the set of permutations includes the extreme cases where all of the incomplete data are x 's or all are y 's. Maritz' combined statistic is $\bar{d} = \bar{x} - \bar{y}$, where \bar{x} and \bar{y} are based on the entire set of x 's and y 's. Note that there are $2^{n_p+n_{ux}+n_{uy}}$ total permutations. We believe that this approach is especially inappropriate for experimental studies, such as the Coyote DNA extraction example, where the values of n_{ux} and n_{uy} are fixed by the experimental design.

A rank-based approach was developed by Dubnicka *et al.* (2002). They suggested using a combination of the Wilcoxon Signed Rank Statistic for the complete pairs and the Mann-Whitney U Statistic for the incomplete pairs. They gave a simple version that is just the sum of the Signed Rank and U statistics as well as a more complex version that is a weighted sum of these two nonparametric statistics.

We compare the performance of our statistic to some of these competing statistics. Specifically, we consider the Z_B statistic Bhoj (1989, p. 282), the statistic T_{LS} by Lin and Stivers (1974), and Maritz (1995) permutation statistic T_M . Applying these to 5K data, we obtain p -values: 0.035 for Z_B , 0.074 for T_{LS} , and 0.067 for T_M . We also consider the simple and weighted rank-based statistics of Dubnicka *et al.* denoted R and R_W , respectively. For the 5K data, these give p -values of 0.162, and 0.240. Thus, only our method and Bhoj's Z_B statistic achieve significance at the 5% level.

Tables 2-5 show the results of simulation studies to compare these statistics. We tested $H_0 : \mu_X - \mu_Y = 0$ against $H_0 : \mu_X - \mu_Y > 0$, where the true difference

$\delta = \mu_X - \mu_Y$ was varied among 0, 0.5, and 1.0, and the correlation was varied among $\rho = 0.1, 0.5,$ and 0.9 . Each empirical size and power calculation is based on 1000 data sets, using a significance level of $\alpha = 0.05$. For Tables 2 and 3, the x 's and y 's were generated from normal distributions; for Tables 4 and 5 we use exponential distributions. In all cases, we assume equal variances with $\sigma_X^2 = \sigma_Y^2 = 1.0$. For our statistic T and Maritz' statistic T_M , 5000 permutations were used for each data set. The R program used to perform these simulations is available from the authors.

Table 2: Empirical size and power of our test (T) and several competing methods. Normal distribution; $n_p = 10$, $n_{ux} = 5$, and $n_{uy} = 10$

δ	ρ	T	T_M	T_{LS}	Z_B	R	R_W
0	0.1	0.052	0.050	0.049	0.047	0.051	0.052
0	0.5	0.048	0.060	0.052	0.051	0.051	0.052
0	0.9	0.058	0.052	0.050	0.051	0.051	0.052
0.5	0.1	0.312	0.308	0.315	0.298	0.298	0.292
0.5	0.5	0.577	0.475	0.456	0.552	0.517	0.557
0.5	0.9	0.999	0.582	0.583	1.000	0.883	0.943
1.0	0.1	0.825	0.835	0.825	0.803	0.793	0.785
1.0	0.5	0.979	0.943	0.951	0.986	0.951	0.974
1.0	0.9	1.000	0.977	0.993	1.000	0.980	0.993

Table 3: Empirical size and power of our test (T) and several competing methods. Normal distribution; $n_p = 20$, $n_{ux} = 30$, and $n_{uy} = 10$

δ	ρ	T	T_M	T_{LS}	Z_B	R	R_W
0	0.1	0.047	0.056	0.042	0.043	0.041	0.045
0	0.5	0.045	0.040	0.033	0.035	0.041	0.045
0	0.9	0.046	0.047	0.041	0.037	0.041	0.045
0.5	0.1	0.587	0.629	0.586	0.602	0.538	0.554
0.5	0.5	0.903	0.806	0.772	0.922	0.798	0.896
0.5	0.9	1.000	0.853	0.857	1.000	0.984	1.000
1.0	0.1	0.995	0.998	0.996	0.997	0.988	0.993
1.0	0.5	1.000	1.000	1.000	1.000	1.000	1.000
1.0	0.9	1.000	1.000	1.000	1.000	1.000	1.000

In each of the four settings, our statistic T maintained its size close to the nominal 0.05 level, as did all of the other tests. With smaller samples from

Table 4: Empirical size and power of our test (T) and several competing methods. Exponential distribution with scale parameter = 1; $n_p = 10$, $n_{ux} = 5$, and $n_{uy} = 10$

δ	ρ	T	T_M	T_{LS}	Z_B	R	R_W
0	0.1	0.051	0.064	0.064	0.039	0.051	0.049
0	0.5	0.050	0.074	0.063	0.044	0.051	0.049
0	0.9	0.048	0.070	0.051	0.039	0.051	0.049
0.5	0.1	0.384	0.377	0.372	0.337	0.471	0.447
0.5	0.5	0.571	0.520	0.488	0.611	0.692	0.709
0.5	0.9	0.993	0.636	0.618	1.000	0.915	0.964
1.0	0.1	0.827	0.843	0.818	0.811	0.876	0.871
1.0	0.5	0.965	0.923	0.925	0.965	0.964	0.972
1.0	0.9	1.000	0.966	0.971	1.000	0.983	0.994

Table 5: Empirical size and power of our test (T) and several competing methods. Exponential distribution with scale parameter = 1; $n_p = 20$, $n_{ux} = 30$, and $n_{uy} = 10$

δ	ρ	T	T_M	T_{LS}	Z_B	R	R_W
0	0.1	0.057	0.059	0.052	0.047	0.051	0.054
0	0.5	0.061	0.038	0.031	0.043	0.051	0.054
0	0.9	0.058	0.044	0.033	0.046	0.051	0.054
0.5	0.1	0.625	0.634	0.594	0.623	0.792	0.774
0.5	0.5	0.882	0.778	0.751	0.915	0.953	0.968
0.5	0.9	1.000	0.837	0.839	1.000	1.000	1.000
1.0	0.1	0.987	0.994	0.990	0.995	1.000	0.999
1.0	0.5	1.000	1.000	1.000	1.000	1.000	1.000
1.0	0.9	1.000	0.998	0.999	1.000	1.000	1.000

normal distributions, our test showed higher power than the competitors when the correlation for the complete pairs was moderate to high, as would typically be the case in practice. As expected, the rank-based procedures tended to have higher power than the other methods when the underlying distributions were exponential. Among the parametric procedures, the Z_B statistic tended to outperform T_{LS} . Comparing the two permutation approaches, our statistic demonstrated higher power than Maritz' statistic for all cases where the correlation for the complete pairs was moderate (0.5) to high (0.9).

5. Application and Summary

For many actual situations where a paired t would apply, we find that some of the pairs are incomplete. Rather than discarding the incomplete data, it makes sense to use that information in the analysis. Compared to other methods, the permutation test we propose is both straightforward and maintains a competitive degree of power across a variety of settings. It is easily implemented in *R*. However, caution must be taken when our statistic or one of the competing statistics is used with observational data. Conclusions may be misleading if there is a lurking factor related to the response variable that is also linked whether subjects have missing x or y values. In an experimental setting such as the coyote DNA example, this is not a problem, because the assignment of subjects to complete (x, y) pairs, only x 's, or only y 's, was completely randomized.

Appendix: Number of Permutations Needed

When the null hypothesis is true the two sided significance level of the permutation test is

$$\alpha_p = P_{H_0}(|T^*| \geq |T_{obs}|) = \frac{\sum_{i=1}^A I(|T_i^*| \geq |T_{obs}|)}{A}, \quad \text{where } A = 2^{n_p} \binom{n_{ux} + n_{uy}}{n_{ux}}.$$

For small sample sizes it is feasible to compute T^* for every possible permutation and thereby obtain α_p exactly. For large sample sizes T^* is calculated for only $B < A$ of these permutations. Following Efron and Tibshirani (1993, pp. 209-211), the number of permutations B is estimated as follows. Let $\hat{\alpha}_p$ be the estimated p -value based on the selected value of B . Then $B\hat{\alpha}_p$ equals the number of absolute values of T_B^* exceeding the absolute value of T_{obs} for the actual data, where T_B^* is the value of test statistic based on the B permuted samples. Then

$$B\hat{\alpha}_p \sim \text{Bin}(B, \alpha_p), \quad \text{and} \quad \text{var}(\hat{\alpha}_p) = \frac{\alpha_p(1 - \alpha_p)}{B}.$$

The coefficient of variation of $\hat{\alpha}_p$ is therefore

$$cv_B(\hat{\alpha}_p) = \sqrt{\left(\frac{(1 - \alpha_p)/\alpha_p}{B}\right)}.$$

Now suppose we want Monte Carlo error to affect the estimated coefficient of variation of α_p by no more than 0.10 or 0.05, that is, $cv_B(\hat{\alpha}_p) \leq 0.10$ or $cv_B(\hat{\alpha}_p) \leq 0.05$. When $\alpha_p = 0.05$, the number of permutation replications required for the above two cases are 1900 and 7900, respectively. Based on these two values, we

use $B = 5000$, which is close to the average of the two. This value was also supported by Manly (1997).

Acknowledgements

The authors thank the referee for his or her careful and prompt review of the paper and for a number of helpful comments. We also thank Brittney Riordan for allowing us to use the coyote DNA data.

References

- Bhoj, D. S. (1978). Testing equality of means of correlated variates with missing data on both responses. *Biometrika* **65**, 225-228.
- Bhoj, D. S. (1979). Testing equality of variances of correlated variates with incomplete data on both responses. *Biometrika* **66**, 681-683.
- Bhoj, D. S. (1984). On difference of means of correlated variates with incomplete data on both responses. *Journal of Statistical Computation and Simulation* **19**, 275-289.
- Bhoj, D. S. (1989). On comparing correlated means in the presence of incomplete data. *Biometrical Journal* **31**, 279-288.
- Bhoj, D. S. (1991). Testing equality of means in the presence of correlated and missing data. *Biometrical Journal* **33**, 63-72.
- Dubnicka, S. R., Blair, R. C. and Hettmansperger, T. P. (2002). Rank-based procedures for mixed paired and two-sample designs. *Journal of Modern Applied Statistical Methods* **1**, 32-41.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, New York.
- Ekbohm, G. (1976). On comparing means in the paired case with incomplete data on both responses. *Biometrika* **63**, 299-304.
- Ekbohm, G. (1982). On comparing variances in the paired case with incomplete data. *Biometrika* **69**, 670-673.
- Good, P. (2006). *Resampling Methods: A Practical Guide to Data Analysis*, 3rd edition. Birkhäuser, Boston.
- Ha, R. R. and Ha, J. C. (2012). *Integrative Statistics for the Social and Behavioral Sciences*. Thousand Oaks, Sage Publications, California.

- Hamdan, M. A., Khuri, A. I. and Crews, S. L. (1978). A test for equality of means of two correlated normal variates with missing data on both responses. *Biometrical Journal* **20**, 667-674.
- Lin, P. E and Stivers, L. E. (1974). On the difference of means with incomplete data. *Biometrika* **61**, 325-334.
- Ling, D. (2012). *Introduction to Statistics Using LibreOffice.org Calc*, 5th edition. <http://www.comfsm.fm/~dleeling/statistics/text.html>.
- Manly, B. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 3rd edition. Chapman & Hall/CRC, London.
- Maritz, J. S. (1995). A permutation paired test allowing for missing values. *Australian Journal of Statistics* **37**, 153-159.
- Moore, D. S. and McCabe, G. P. (2005). *Introduction to the Practice of Statistics*, 5th edition. Edited by W. H. Freeman and Company, Chapter 14, 46-69. Supplementary Material, New York. http://bcs.whfreeman.com/ips5e/content/cat_080/pdf/moore14.pdf.
- Morrison, D. F. (1973). A test for equality of means of correlated variates with missing data on one response. *Biometrika* **60**, 101-105.
- Oden, N. L. (1991). Allocation of effort in Monte Carlo simulation for power of permutation tests. *Journal of the American Statistical Association* **86**, 1074-1076.
- Motulsky, H. (2010). *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*, 2nd edition. Oxford University Press, New York.
- Riordan, B. (2012). *Northeastern Ohio Coyote Hybridization with Wolves*. Honors research project, University of Akron, Akron.
- Woolson, R., Leeper, J., Cole, J. and Clarke, W. (1976). A Monte Carlo investigation of a statistic for a bivariate missing data problem. *Communications in Statistics - Theory and Methods* **A5**, 681-688.

Received October 16, 2012; accepted April 10, 2013.

Richard L. Einsporn
Department of Statistics
University of Akron
302 Buchtel Common. Akron, OH 44325, USA
rle@uakron.edu

Desale Habtzghi
Department of Statistics
University of Akron
302 Buchtel Common. Akron, OH 44325, USA
dh52@uakron.edu