Journal Of Modern Applied Statistical Methods
Winter 2002, Vol. 1, No. 1, 32-41

1538 - 9472/02/$30.00

REGULAR ARTICLES

Rank-Based Procedures for Mixed Paired and Two-Sample Designs

Suzanne R. Dubnicka
Department of Biostatistics
The University of North Carolina
at Chapel Hill

R. Clifford Blair
Department of Epidemiology and Biostatistics
College of Public Health, &
Jaeb Center For Health Research
University Of South Florida

Thomas P. Hettmansperger
Department of Statistics
The Pennsylvania State University

This paper presents a rank-based procedure for parameter estimation and hypothesis testing when the data are a mixture of paired observations and independent samples. Such a situation may arise when comparing two treatments. When both treatments can be applied to a subject, paired data will be generated. When it is not possible to apply both treatments, the subject will be randomly assigned to one of the treatment groups. Our rank-based procedure allows us to use the data from the paired sample and the independent samples to make inferences about the difference in the mean responses. The rank-based procedure uses both types of data by combining the Wilcoxon signed-rank statistic and the Wilcoxon-Mann-Whitney statistic. The exact and asymptotic distributions of the test statistic under the null hypothesis are determined as well as the limiting distribution of the point estimate. We also consider the Pitman efficacy of our rank-based procedure and its efficiency with respect to mean-based procedures.

Keywords: Wilcoxon signed-rank statistic, Wilcoxon-Mann-Whitney statistic, Pitman efficacy.

Introduction

The purpose of this paper is to introduce a simple robust, nonparametric approach to testing and estimation in mixed paired and two sample designs. The new methods are more robust and have excellent efficiency when compared to more traditional methods based on the means. In addition, they are easy to compute and apply to data. Mixtures of paired and unpaired data may be realized in a variety of research contexts. We consider measurement data in this article. For a discussion of categorical data see Thompson (1995) and references therein.

For example, in a clinical trial designed to compare two methods of laser surgery that are used to correct a certain eye condition, patients with both eyes eligible for study may have one eye assigned to one surgical method, but the other treated with the competing method. Visual acuity or other measures taken on the treated eyes will likely be correlated, thereby producing paired data. However, patients who have only one eye eligible for study will have the one eye randomly assigned to one of the competing treatments. This produces unpaired observations.

Other examples include designs in which subjects are to be observed at two points in time with some intervention between observations. Often in such situations, data may be missing at either the first or second observation times. Subjects with no missing data will thereby produce paired sets, while those with a missing data point will be unpaired. A special case of this example arises when pairwise comparisons are performed in repeated measures designs with missing observations.

Finally, in matched pair designs, some subjects in the pool may not have suitable matches. In such situations, the unmatched subjects can simply be randomly assigned to treatments, thereby taking advantage of all eligible study subjects.

As will be noted in one of the example analyses presented below, researchers are often uncertain as to how such data might be efficiently analyzed. Unfortunately, strategies wasteful of available information are often employed. When the underlying distributions are normal, there is a sufficient but not complete statistic. In this case, the maximum likelihood estimators may not behave very well and they are quite complex computationally. Consequently, we do not attempt to develop optimal methods. Instead, we suggest nonparametric methods that work well in a broad spectrum of models. Bhoj (1989) presented a

Suzanne R. Dubinicka is a Post-doctoral Fellow. This article was submitted while at the Department of Mathematical Sciences, Indiana University - Purdue University at Fort Wayne. Her research interests include nonparametric methods, survival analysis, measurement error, and permutation tests. R. Clifford is Professor, Department of Epidemiology and Biostatistics, University of South Florida. His research was partially supported by Jaeb Center for Health Research. Thomas P. Hettmansperger is Professor, Department of Statistics, The Pennsylvania State University.

thorough summary of tests based on means when the underlying distributions are normal. He proposed two new tests and compared various methods via Monte Carlo simulations. Perhaps the easiest version of such a test was proposed by Morrison (1973). See Ekbohm (1976, 1981); Hamdan, Khuri, and Crews (1978); and Woolson, Leeper, Cole and Clarke (1976) for additional early work based on means.

Simple Rank Methods

We consider the situation in which observations come from a bivariate distribution $f(x,y)$ with marginal densities $f_1(x)$ and $f_2(y)$; means θ_1 and θ_2; variances σ_1^2 and σ_2^2; and correlation ρ. We will assume that the marginal variances are equal and call this common variance σ^2. In addition, we will assume that the marginal densities of X and Y have the same shape and differ by at most a shift in location. We are concerned with making inferences about the difference in marginal means, $\Delta = \theta_1 - \theta_2$. In particular, we will be interested in comparing a new treatment with an old treatment or with a control.

Paired data will be generated when the new and old treatments can be applied to each of the subjects. When complete pairs are available, we collect $(X_1,Y_1),...,(X_n,Y_n)$. Let $D_i = X_i - Y_i, i = 1,...,n$. The D_i have a probability density function f_d with $E(D_i) = \Delta$ and $Var(D_i) = 2\sigma^2(1-\rho)$.

Complete pairs may not always be available. In some situations, it will only be possible to apply one of the treatments to a subject. In this case, the treatments will be randomly assigned to the subjects, and two independent samples will be generated. We denote these samples of sizes n_1 and n_2 as $T_1,...,T_{n_1}$ and $C_1,...,C_{n_2}$, where the T_i have density $f_1(t-\theta_1), E(T_i) = \theta_1, Var(T_i) = \sigma^2$, and the C_j have density $f_2(c-\theta_2), E(C_j) = \theta_2, Var(C_j) = \sigma^2$.

Combining the paired data and the independent samples, we would like to make inferences about the parameter Δ. In particular, we would like to determine if the new treatment is better than the old treatment, and to estimate the difference between the treatments. Thus, we construct a test $H_o : \Delta = 0$ vs $H_a : \Delta > 0$, an estimate of Δ, and a confidence interval for Δ.

An intuitively appealing statistic is the sum of the usual rank statistics for paired data, the Wilcoxon signed-rank statistic, and for two independent samples, the Wilcoxon-Mann-Whitney statistic. Let

$$T^+(\Delta) = S^+(\Delta) + U^+(\Delta)$$

$$= \sum_{i \le j} \sum I\left(\frac{D_i + D_j}{2} > \Delta\right) + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} I(T_i - C_j > \Delta)$$

$$= \sum_{i=1}^{n} R(|D_i|) I(D_i > 0) + \sum_{j=1}^{n_1} R_j - \frac{n_1(n_1+1)}{2} \qquad (1)$$

where $I(A) = 1$ if event A occurs and $= 0$ otherwise, $R(|D_i|)$ is the rank of $|D_i|$ among $|D_1|,...,|D_n|$ and R_j is the rank of T_j in the combined independent samples.

Distribution Theory

From Hettmansperger and McKean (1998, Section 1.7 & 2.4) it follows immediately that, under $H_o : \Delta = 0$,

$$E_0 T^+(0) = \frac{n(n+1)}{4} + \frac{n_1 n_2}{2} \qquad (2)$$

$$Var_0 T^+(0) = \frac{n(n+1)(2n+1)}{24} + \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}. \qquad (3)$$

Theorem 1

Under $H_o : \Delta = 0$, as $M \to \infty$,

$$\frac{T^+(0) - E_0 T^+(0)}{\sqrt{Var_0 T^+(0)}} \xrightarrow{D} Z \sim N(0,1)$$

where

$\frac{n}{m} \to \gamma$, $\frac{n_1+n_2}{m} \to 1-\gamma$, $\frac{n_1}{n_1+n_2} \to \lambda$, $\frac{n_2}{n_1+n_2} \to \lambda$, and $M = n + n_1 + n_2$, $0 < \gamma < 1$, and $0 < \lambda < 1$.

Thus, an asymptotic size α test will reject $H_o : \Delta = 0$ in favor of $H_a : \Delta > 0$ if

$$\frac{T^+(0) - E_0 T^+(0) - \frac{1}{2}}{\sqrt{Var_0 T^+(0)}} \ge z_\alpha \qquad (4)$$

where Z_α is the upper α quantile of the standard normal distribution.

A convolution can be used to compute the exact p-value. From the independence of $S^+(0)$ and $U^+(0)$,

$$P(T^+ \ge t) = \sum_s P(U^+ \ge t - s) P(S^+ = s)$$

$$= \sum_u P(S^+ \ge t - u) P(U^+ = u)$$

Hence, tables of probabilities and tail probabilities can be used to compute $P(T^+ \ge t)$. The exact distribution of T^+ can be constructed recursively as well.

Define $\bar{P}_{n,n_1,n_2}(k)$ as the number of data configurations such that $T^+ = k$. Because there are $2^n \binom{n_1+n_2}{n_1}$ such configurations,

$$P\left(T^+(0)=k\right)=\frac{\overline{P}_{n,n_1,n_2}(k)}{2^n\binom{n_1+n_2}{n_1}}$$

where $k = 0,1,...,[n(n+1)]/2 + n_1n_2$. Note that T^+ is distribution-free. A recursive formula can be used to compute $\overline{P}_{n,n_1,n_2}(k)$.

Theorem 2

$$\overline{P}_{n,n_1,n_2}(k)=\overline{P}_{n,n_1-1,n_2}(k-n_2)+\overline{P}_{n,n_1,n_2-1}(k)$$

where $\overline{P}_{h,i,j}(k)=0$ or 1 as $k<0$ or $=0$, and

$$\overline{P}_{h,i,0}(k)=\overline{P}_{h,0,j}(k)=\overline{P}_h(k)=\overline{P}_{h-1}(k-h)+\overline{P}_{h-1}(k)$$

with $\overline{P}_h(k)=0$ if $k<0$ and $\overline{P}_0(k)=1$ or 0 as $k=0$ or $\neq 0$.

This recursive formula can be used to construct an exact test for our problem. The exact test will reject $H_o:\Delta=0$ in favor of $H_a:\Delta>0$ if $T^+(0)\geq c$, where c is chosen as $P\left(T^+(0)\geq c\right)\doteq\alpha$. *S-plus* functions and Fortran programs have been written to compute the p-value for this test, as well as construct upper tail probabilities for the distribution of T^+ for given n, n_1, and n_2. Table 1 was constructed using the *S-plus* functions. In this table, we have assumed that $n_1 = n_2$ and computed selected upper quantiles for various values of n and n_1. Because T^+ is discrete, it is not usually possible to obtain the exact quantile desired. The values in the table give the values closest to the upper 5%, 2.5%, and 1% quantiles.

We can also compute these upper quantiles using the normal approximations. For example, when $n = n_1 = n_2 = 5$, $P\left(T^+\geq 30\right)\doteq 0.058$ using the normal approximations in (4). Compare this to 0.059 in Table 1 using the exact distribution. For the same sample sizes, $P\left(T^+\geq 32\right)\doteq 0.029$ and $P\left(T^+\geq 34\right)\doteq 0.013$. For larger sample sizes, the approximations are even better. For $n = n_1 = n_2 = 10$, $P\left(T^+\geq 105\right)\doteq 0.051$, $P\left(T^+\geq 110\right)\doteq 0.026$, and $P\left(T^+\geq 116\right)\doteq 0.011$ using the normal approximation. Notice the first two agree with exact probabilities to three decimal places and the third differs from the exact probability by only 0.001. We conclude that for most practical situations the normal approximation is adequate.

Estimation

To estimate Δ, we consider the following form of our estimating function:

$$T(\Delta)=\sum_{i\leq j}\sum sgn\left(\frac{D_i+D_j}{2}-\Delta\right)+\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}sgn\left(T_i-C_j-\Delta\right). \quad (5)$$

where sgn(x)=1 if $x > 0$, = 0 if $x = 0$, and = -1 if $x < 0$. Notice that $T(\Delta)=2T^+(\Delta)-[n(n+1)/2+n_1n_2]$ and $E_\Delta T(\Delta)=0$. We estimate Δ by $\hat{\Delta}_R$ so that $T\left(\hat{\Delta}_R\right)=0$. We find

$$\hat{\Delta}_R=med\left\{\frac{D_i+D_j}{2},T_k-C_l;1\leq i\leq j\leq n,1\leq k\leq n_1,1\leq l\leq n_2\right\}. \quad (6)$$

Theorem 3

Under $H_o:\Delta=0$, as $M\to\infty$ provided $0<\int f_d^2<\infty$

Table 1: Upper Quantiles for the Distribution of T^+: (P_t,t) where $P_t = P(T^+ \geq t)$

	$n_1 = n_2$					
n	5	6	7	8	9	10
	(0.059,30)	(0.050,38)	(0.047,47)	(0.048,57)	(0.051,68)	(0 .047,81)
5	(0.028,32)	(0.026,40)	(0.028,49)	(0.024,60)	(0.023,72)	(0.024,85)
	(0.011,34)	(0.012,42)	(0.011,52)	(0.011,63)	(0.012,83)	(0.011,89)
	(0.044,35)	(0.049,42)	(0.046,51)	(0.046,61)	(0.048,72)	(0 .051,84)
6	(0.022,37)	(0.027,44)	(0.027,53)	(0.023,64)	(0.027,75)	(0 .027,88)
	(0.014,38)	(0.010,47)	(0.011,56)	(0.011,67)	(0.009,80)	(0.010,93)
	(0.044,40)	(0.046,47)	(0.052,55)	(0.050,65)	(0.050,76)	(0 .053,88)
7	(0.023,42)	(0.027,49)	(0.026,58)	(0.027,68)	(0.025,80)	(0 .024,93)
	(0.011,44)	(0.010,52)	(0.011,61)	(0.010,72)	(0.011,84)	(0.009,98)
	(0.052,45)	(0.052,52)	(0.055,60)	(0.051,70)	(0.050,81)	(0 .052,93)
8	(0.023,48)	(0.025,55)	(0.023,64)	(0.023,74)	(0.025,85)	(0 .024,98)
	(0.012,50)	(0.010,58)	(0.011,67)	(0.009,78)	(0.009,90)	(0.010,103)
	(0.045,52)	(0.053,58)	(0.055,66)	(0.050,76)	(0.048,87)	(0 .049,99)
9	(0.027,54)	(0.028,61)	(0.025,70)	(0.024,80)	(0.025,91)	(0 .024,104)
	(0.008,58)	(0.010,65)	(0.009,74)	(0.010,84)	(0.010,96)	(0.010,109)
	(0.045,59)	(0.052,65)	(0.052,73)	(0.046,83)	(0.051,93)	(0.051,105)
10	(0.023,62)	(0.023,69)	(0.025,77)	(0.023,87)	(0.024,98)	(0.026,110)
	(0.011,65)	(0.011,72)	(0.010,81)	(0.011,91)	(0.010,103)	(0.010,116)

and $0 < \int f_1^2 < \infty$, and the additional sample size limits in Theorem 1 are satisfied,

$$\sqrt{M}(\hat{\Delta}_R - \Delta) \xrightarrow{D} Z \sim N\left(0, \frac{1}{c_{T^+}^2}\right) \tag{7}$$

where

$$c_{T^+} = \sqrt{12}\frac{\gamma^2 \int f_d^2 + \lambda(1-\lambda)(1-\gamma)^2 \int f_1^2}{\sqrt{\gamma^3 + \lambda(1-\lambda)(1-\gamma)^3}} \tag{8}$$

is the Pitman efficiency of T^+. See (26) in the appendix.

A Confidence Interval and the Standard Error of $\hat{\Delta}$

To construct a confidence interval for Δ, let $Y_{(1)} \le \ldots \le Y_{(N^*)}$ be the ordered values of the combined Walsh averages, $(D_i + D_j)/2$, and pairwise differences, $T_k - C_l$, where $N^* = n(n+1)/2 + n_1 n_2$. Then $\left[Y_{(k+1)}, Y_{(N^*-k)}\right)$ is a $(1-\alpha)$ 100% confidence interval for Δ where $P(T^+ \le k) = P(T^+ \ge N^* - k) = \alpha/2$. Using a normal approximation with continuity correction

$$k \doteq \left(\frac{n(n+1)}{4} + \frac{n_1 n_2}{2}\right) - \frac{1}{2} - z_{\frac{\alpha}{2}}\sqrt{\frac{n(n+1)(2n+1)}{24} + \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}.$$

This confidence interval can be used to estimate the asymptotic variance of $\hat{\Delta}_R$. Note that

$$\frac{\sqrt{M}\left[Y_{(N^*-k)} - Y_{(k+1)}\right]}{2z_{\alpha/2}} \xrightarrow{P} \frac{1}{c_{T^+}}. \tag{9}$$

See Hettmansperger and McKean (1998, Section 1.5 & 2.4). Hence, we say $\hat{\Delta}_R$ is approximately normally distributed with mean Δ and variance $M^{-1}\hat{c}_{T^+}^{-1}$, where $\hat{c}_{T^+}^{-1}$ is defined by (9). In practice we would choose α around 0.10.

Example 1

The Krypton Argon Regression of Neovascularization Study (KARNS) [The Krypton Argon Regression of Neovascularization Research Study Group (1993)] was designed to compare the efficacy of red krypton versus blue-green argon laser photocoagualation for the management of high-risk proliferative diabetic retinopathy. To this end a randomized prospective clinical trial was performed on patients with diabetes and neovascularization of the optic disc. Patients were randomized in the following manner: Patients with both eyes eligible for study had the right eye randomly assigned to one of the treatments, with the left eye then being assigned to the competing treatment. Patients who had only one eye eligible for study had that eye randomly assigned to one of the two treatments. As a result of this assignment scheme, part of the outcome data were paired and part were unpaired.

A variety of outcomes were measured 3 months after surgery. One outcome of interest was visual acuity as measured by the number of letters correctly read from the *Early Treatment Diabetic Retinopathy Visual Acuity Chart.* As noted by the authors, these measures had a substantial negative skew. Although the KARNS authors do not indicate exactly how these acuity scores were analyzed, they do note for other outcomes that paired and unpaired data were analyzed separately, thereby complicating the assessment of a possible treatment effect.

We have selected a subset of the data from this study to illustrate the rank-based methods of this paper; see Table 2. For this example, we have $n = 20$ pairs of visual acuity measurements for those patients with both eyes eligible for study. In addition, we selected 20 patients with one eye eligible for study, with $n_1 = 10$ in the group receiving blue-green argon laser photocoagulation and $n_2 = 10$ in the group receiving red krypton. The rank based estimate for the difference in mean visual acuities for the two groups is the median of the combined 310 Walsh averages and pairwise differences: $\hat{\Delta}_R = 4.00$. In addition, to compute a 95% confidence interval for Δ_R, we need to find k such that $P(T^+ \le k) = P(T^+ \ge 310 - k) = 0.025$. Using the exact distribution for T^+, $P(T^+ \le 96) = P(T^+ \ge 214) = 0.0246$. Thus, a 95% confidence interval for Δ_R is $\left[Y_{(97)}, Y_{(214)}\right) = [-3, 9.5)$. Also, suppose we are interested in testing that the mean visual acuity score is higher for the blue-green argon group. In this case, $S^+ = 135$ and $U^+ = 189$. Thus, our observed test statistic $T^+ = 189$ with an exact p-value of 0.126. Using normal approximation, the approximate p-value for this test is 0.124. Note that there are 210 Walsh averages and 100 pairwise differences.

Example 2

Another situation in which these procedures may be used is a repeated measures design in which data are not available for all the time points. For example, the Optic Neuritis Treatment Trial (ONIT)/Longitudinal Optic Neuritis Study (LONS) [The Optic Neuritis Study Group (1991)] was conducted to assess the relative benefit to visual function of two corticosteroid treatments as compared with placebo for patients suffering from optic neuritis. This

Table 2: Visual Acuity Measurements

Paired Data

X_i	4	69	87	35	39	79	31	79	65	95
Y_i	62	80	82	83	0	81	28	69	48	90
D_i	-58	-11	5	-48	39	-2	3	10	17	5
X_i	68	62	70	80	84	79	66	75	59	77
Y_i	63	77	0	55	83	85	54	72	58	68
D_i	5	-15	70	25	1	-6	12	3	1	9

Unpaired Data

T_i	36	86	39	85	74	72	69	85	85	72
C_j	88	83	78	30	58	45	78	64	87	65

disease is characterized by the fact that patients usually experience a rapid loss of visual function over a period of about one week, with most of this function subsequently restored over a period of some months.

One of the important vision function measures used in this trial was the *Farnsworth Munsell 100-Hue Test* which assesses the patient's ability to make fine color discrimination. This test was administered to patients as baseline, six months, one year, and annually thereafter until the eighth year after disease diagnosis. A question arises as to when along this temporal course changes in color discrimination occur and when such changes no longer occur. It is of interest then to perform pairwise comparisons to make these assessments. In this case we use data from the placebo group, as interest lies in the natural course of the disease.

Problematic is the fact that, as expected, some patients miss their testing appointments for various reasons, thereby producing missing observations in the longitudinal data set. The testing procedures discussed in this paper can be used to test pairwise hypotheses in the face of missing data.

Weighted Rank Methods

Above, we based our inferences on simple sums of one and two sample methods. It should be possible with some added complexity to increase the efficiency by considering weighted sums. The weights may be constructed to depend on sample size proportions and correlation in the bivariate portion of the data. Of course, the efficiency depends on the underlying model distribution. In this section, we explore various ways to construct weights, always trying to keep in mind the practical side of computing and applying the methods.

Consider

$$T_a = a\frac{2}{n(n+1)}S^+ + b\frac{1}{n_1 n_2}U^+ \qquad (10)$$

where a+b=1. Then, T_a is an estimate of $\delta = aP((D_1 + D_2)/2 > 0) + bP(T - C > 0)$ and under $H_0 : \Delta = 0, \delta = 1/2$.

Define, with $N = n_1 + n_2$ and $a = nN/(nN + 2n_1 n_2)$,

$$T^* = \frac{2N}{(nN + 2n_1 n_2)(n+1)}S^+ + \frac{2}{nN + 2n_1 n_2}U^+ \qquad (11)$$

for testing $H_0 : \Delta = 0$. Under H_0, $E_0 T^* = 1/2$ and

$$Var_0 T^* = \left(\frac{nN}{nN + 2n_1 n_2}\right)^2 \frac{n(n+1)(2n+1)}{24} + \left(\frac{2n_1 n_2}{nN + 2n_1 n_2}\right)^2 \frac{n_1 n_2 (N+1)}{12}.$$

Also, under H_0, T^* has an asymptotic normal distribution. Thus, one can construct an asymptotic size α test based on T^*. This test is close to the optimal test derived in the appendix when the underlying distribution is bivariate normal.

We have, from (26) in the appendix,

$$c_{T^*} = \sqrt{12}\frac{\gamma \int f_d^2 + \lambda(1-\lambda)(1-\gamma)\int f_1^2}{\sqrt{\gamma + \lambda(1-\lambda)(1-\gamma)}}. \qquad (12)$$

For a bivariate normal with $\lambda = \frac{1}{2}$, Table 4 gives the efficiency of T^* relative to $T_{opt,normal}$ and relative to the simpler T^+ from above. Note that T^* loses very little efficiency and is not below .95 for the range of γ and ρ values considered when compared to $T_{opt,norm}$. On the other hand, it is quite a bit better than T^+ for values of γ below

Table 3: Efficiency of T^* (a) with respect to $T_{opt,norm}$ and (b) with respect to T^+

γ	$\rho = .2$		$\rho = .5$		$\rho = .8$	
	(a)	(b)	(a)	(b)	(a)	(b)
.2	0.987	1.187	1.000	1.360	0.952	1.818
.5	0.990	1.000	1.000	1.000	0.975	1.000
.8	0.996	1.058	1.000	1.034	0.992	1.001

0.5. It remains to show that T^* has robust efficiency against its natural normal theory competition. This is shown later. We complete this section with discussion of the corresponding estimation and confidence interval.

Computing the rank estimate is not as straightforward as in the simple sum case. Again, we estimate Δ by $\hat{\Delta}_R^*$ such that $T^*(\hat{\Delta}_R^*) = 1/2$. In this case, $T^*(\Delta)$ is a nondecreasing step function which steps down $2N/[(nN + 2n_1n_2)(n+1)]$ units at each Walsh average and $2/(nN + 2n_1n_2)$ units at each pairwise difference. Because the order of the Walsh averages and pairwise differences depends on the actual data, this estimate cannot be written in closed form. The estimate can be easily found in *S-plus* or *Minitab*. First, list the Walsh averages and pairwise differences along with the corresponding weights: $2N/[(nN + 2n_1n_2)(n+1)]$ or $2/(nN + 2n_1n_2)$. Then, sort the combined Walsh averages and pairwise differences with the associated weights in descending order. Starting with the largest item in the list, accumulate the weights until 1/2 is reached. The Walsh average or pairwise difference corresponding to the weight required to reach 1/2 is $\hat{\Delta}_R^*$.

A similar problem exists in constructing a confidence interval for Δ. We use the asymptotic normality of T^*. Thus, an approximate $(1-\alpha)100\%$ confidence interval for Δ is $(\hat{\Delta}_L, \hat{\Delta}_U)$ where

$$T^*(\hat{\Delta}_L) \doteq \frac{1}{2} + z_{\alpha/2}\sqrt{Var_0 T^*} \tag{13}$$

$$T^*(\hat{\Delta}_U) \doteq \frac{1}{2} - z_{\alpha/2}\sqrt{Var_0 T^*} \tag{14}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. Then, use the sorting and accumulation for the estimate to determine the $\hat{\Delta}_L$ and $\hat{\Delta}_U$ values.

Methodology

Comparison To Competitors

In this section, we compare T^* (or $\hat{\Delta}_R^*$) with various parametric and nonparametric competitors. We will use the Pitman efficiency as described in the appendix. Briefly, the Pitman efficiency of T^* (or $\hat{\Delta}_R^*$) with respect to T(or $\hat{\Delta}$) is the ratio $e = (c^*/c)^2$, where c is called the efficacy. Efficiency values greater than one favor T^* (or $\hat{\Delta}_R^*$). Pitman efficiency provides a single number comparison and has been shown to reflect small sample size as well as asymptotic behavior; see Hettmansperger (1984).

We discuss the comparisons in terms of estimates. The corresponding tests are often standardized versions of these estimates. In all cases, the estimation and test efficiency are identical and given by e. The competitors fall into two groups. One group is based on a linear models approach and includes a plug-in maximum likelihood estimator and an estimator derived from a test proposed by Bhoj (1989). The other group includes estimates for which the design is initially ignored and all the treatment observations are compared with all the control observations.

Linear Model

Consider the following linear model:

$$\boldsymbol{Z} = \boldsymbol{X}\beta + \varepsilon \tag{15}$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{T} \\ \mathbf{C} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \tag{16}$$

and

$$\text{Cov}\ \boldsymbol{\varepsilon} = \boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} \mathbf{I} & \rho\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \rho\mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{17}$$

Note that Z is a $(2n + n_1 + n_2)\times 1$ matrix, X is $(2n + n_1 + n_2)\times 2$, β is 2×1, Σ is $(2n + n_1 + n_2)\times(2n + n_1 + n_2)$. Then $\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\underline{Z}$ and $\text{Var}\hat{\beta} = (X'\Sigma^{-1}X)^{-1}$. Note that $\Delta = 2\beta_2$. To compute this estimate of Δ, one must know the correlation ρ.

Theorem 4

If ρ is known and $M = n + n_1 + n_2$,

$$\hat{\Delta}_{LM}(\rho) = \frac{1}{(n+n_1)(n+n_2) - n_1n_2\rho^2}\{nM(\bar{X} - \bar{Y}) + n_1n_2(1-\rho^2)(\bar{T} - \bar{C}) - n(1-\rho)[(n_1\bar{X} - n_2\bar{Y}) - (n_1\bar{T} - n_2\bar{C})]\} \tag{18}$$

$$Var\hat{\Delta}_{LM}(\rho) = \frac{2n(1-\rho) + (n_1 + n_2)(1-\rho^2)}{(n+n_1)(n+n_2) - n_1n_2\rho^2}\sigma^2 \tag{19}$$

and as $M \to \infty$, provided $0 < \sigma^2 < \infty$ and the sample size limits of Theorem 1 are satisfied,

$\sqrt{M}(\hat{\Delta}_{LM}(\rho)-\Delta)\xrightarrow{D}N(0,\xi^2)$ where

$$\xi^2=\frac{2\gamma(1-\rho)+(1-\gamma)(1-\rho^2)}{[\gamma+\lambda(1-\gamma)]\,[\gamma+(1-\lambda)(1-\gamma)]-\lambda(1-\lambda)(1-\gamma)^2\rho^2}\sigma^2. \quad (20)$$

Note that $c_{LM}=\xi^{-1}$. In practice we would use $\hat{\Delta}(r)$ where r is the correlation coefficient in the paired data. This estimate of Δ is also obtained using the plug-in method. For the assumed model, we derive the maximum likelihood estimator of Δ assuming that ρ and σ are known. Because ρ and σ are usually not known, we substitute, or plug-in, estimates of these unknown parameters into the expression for the estimate of Δ. Using the plug-in method, we get $\hat{\Delta}_{pmle}=\hat{\Delta}_{LM}(r)$ where r is the sample correlation coefficient for the paired data. Ekbohm (1976) used this approach to construct test statistics and compared various early approaches in a Monte Carlo study. We do not recommend full maximum likelihood because it is computationally quite complex and may not be stable because there is no complete statistic for the model with all parameters treated as unknown.

Bhoj (1989) compared six different test statistics based on means in a Monte Carlo power study. The different approaches entail refinements that yield approximate t distributions for the test statistics. In addition, some approaches consider the unequal variances case. See Bhoj's paper and references for a survey of parametric methods. We will concentrate on a new test proposed by Bhoj denoted Z_b which emerged as generally superior to the other tests based on sample averages. The corresponding estimate was not explicitly developed by Bhoj, so we include a brief description here.

We follow the notation of Bhoj (1989). Let $\hat{\delta}=\omega\hat{\delta}_2+(1-\omega)\hat{\delta}_3$ where $\hat{\delta}_2=2\overline{T}-\overline{X}-\overline{Y}$ and $\hat{\delta}_3=\overline{X}+\overline{Y}-2\overline{C}$. The value of ω depends on ρ and was chosen to minimize $Var(\hat{\delta})$; it is given by Formula (2.2) in Bhoj (1989). Next, note that under normality $\overline{D}=\overline{X}-\overline{Y}$ and $\hat{\delta}$ are independent. Hence, the linear combination of $\overline{D}$ and $\hat{\delta}$ with minimum variance is

$$\hat{\Delta}_b=\frac{\sigma^2_{\hat{\delta}}}{\sigma^2_{\hat{\delta}}+\sigma^2_{\overline{D}}}\overline{D}+\frac{\sigma^2_{\overline{D}}}{\sigma^2_{\hat{\delta}}+\sigma^2_{\overline{D}}}\hat{\delta} \quad (21)$$

where, in practice, unknown variances must be estimated. Finally, the quantity relevant for the efficiency calculations is $\sigma^2_{\hat{\Delta}_b}=\left(\sigma^{-2}_{\hat{\delta}}+\sigma^{-2}_{\overline{D}}\right)^{-1}$. A formula for $\sigma^2_{\hat{\delta}}$ is given by (2.3) in Bhoj (1989). Although this estimate appears to be quite different from the linear model estimate, in fact the variance can be shown to be identical with the variance of $\hat{\Delta}_{LM}(r)$ defined in Theorem 4.

Hence, the linear model estimate includes the plug-in MLE and the Bhoj estimator. Our efficiency comparison then will compare $\hat{\Delta}^*_R$ to $\hat{\Delta}_{LM}(r)$.

The efficiency of $\hat{\Delta}^*_R$ with respect to the linear model estimate is

$$eff\left(\hat{\Delta}^*_R,\hat{\Delta}_{LM}(\rho)\right)=\xi^2 c^2_{T^*}. \quad (22)$$

In the following discussion, we compare $\hat{\Delta}^*_R$ to $\hat{\Delta}_{LM}(\rho)$ via (22). However, this is also the Pitman efficiency of the rank test T^* relative to a test based on $\hat{\Delta}_{LM}(\rho)$. Table 4 gives the efficiencies of $\hat{\Delta}^*_R$ with respect to the linear model estimate assuming an underlying bivariate normal distribution and $\lambda=1/2$.

Results

In Table 4, we see that when the underlying distribution is bivariate normal, the linear model estimate with ρ known is more efficient than the rank-based estimate when a^* is used regardless of the correlation, ρ, or the proportion of paired data, γ. The efficiencies are close to one, indicating that too much efficiency is not lost using the rank method.

Table 4: Efficiency of $\hat{\Delta}^*_R$ with respect to $\hat{\Delta}_{LM}(\rho)$

γ	$\rho=.2$	$\rho=.5$	$\rho=.8$
.2	0.942	0.955	0.909
.5	0.945	0.955	0.931
.8	0.951	0.955	0.948

In practice, ρ would be estimated from the data. When a Pearson correlation coefficient for the paired data is used, this introduces more variability into the linear model estimate of Δ. Simulations show that the linear model estimate in which ρ is estimated is still more efficient than the rank-based estimate using a^* under bivariate normality. In general, the gain in efficiency using this linear model estimate over the rank-based estimate is not great.

Instead of an underlying bivariate normal distribution, suppose the data are selected from a contaminated bivariate normal distribution. Let ε represent the proportion of contamination. Then with probability $1-\varepsilon$, the data will be selected from our usual bivariate normal distribution with equal marginal variances $\sigma^2_1=\sigma^2_2=\sigma^2$ and

correlation ρ. The rest of the data will be selected from a bivariate normal distribution with equal but larger marginal variances and possibly different correlation. It follows that the marginal distribution and distribution of differences are univariate contaminated normal distributions. Furthermore, if we mix a bivariate normal $(\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = 1, \rho)$ with a bivariate normal $(\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = \sigma^2, \rho_c)$ then

$$\int f_{d,\varepsilon}^2 = \frac{(1-\varepsilon)^2}{2\sqrt{\pi}\sqrt{2(1-\rho)}} + \frac{\varepsilon^2}{2\sqrt{\pi}\sqrt{2\sigma^2(1-\rho_c)}} + \frac{\sqrt{2}\varepsilon(1-\varepsilon)}{\sqrt{\pi}\sqrt{2(1-\rho)+2\sigma^2(1-\rho_c)}}$$

$$\int f_{1,\varepsilon}^2 = \frac{(1-\varepsilon)^2}{2\sqrt{\pi}} + \frac{\varepsilon^2}{2\sqrt{\pi}\sigma} + \frac{\sqrt{2}\varepsilon(1-\varepsilon)}{\sqrt{\pi}\sqrt{1+\sigma^2}}.$$

Now c_{T^*} is compared using a^* and (12).

When the data are a sample from a contaminated bivariate normal distribution, the asymptotic variance of $\hat{\Delta}_{LM}$ is given by

$$\xi_c^2 = \frac{u^2(u-w)[2\gamma u^2 + (1-\gamma)(u+w)]}{(u\gamma + \lambda(1-\gamma))(u\gamma + (1-\lambda)(1-\gamma))u^2 - \lambda(1-\lambda)(1-\gamma)^2 w^2} \qquad (23)$$

where $u = (1-\varepsilon) + \varepsilon\sigma^2$ and $w = (1-\varepsilon)\rho + \varepsilon\rho_c\sigma^2$. Recall ε is the proportion of contamination and ρ_c and σ^2 are the correlation and the marginal variance, respectively, in the contaminated portion of the distribution. Thus, the efficiency of the T^* methods with respect to the linear model methods under a contaminated bivariate normal distribution is given by $\xi_c^2 c_{T^*}^2$.

Suppose the data are sampled from the contaminated bivariate normal distribution with $\rho_c = 0$. This simulates the situation in which the components of the data are contaminated independently, as in the case of gross errors. Already for contamination of the order 5%, $\hat{\Delta}_R^*$ is more efficient than $\hat{\Delta}_{LM}$. See Table 5 for efficiencies of $\hat{\Delta}_R^*$ with respect to $\hat{\Delta}_{LM}$ when $\varepsilon = 0.05$ and $\sigma = 2$. We find similar efficiencies for other levels of contamination and values of $\lambda \neq 1/2$. We conclude that T^* (and $\hat{\Delta}_R^*$) are often more efficient than the linear based methods.

Table 5: Efficiency of $\hat{\Delta}_R^*$ with respect to $\hat{\Delta}_{LM}$ with $\in = 0.05, \sigma = 2$, $\rho_c = 0$, $\lambda = 1/2$

γ	$\rho = .2$	$\rho = .5$	$\rho = .8$
.2	1.235	1.232	1.138
.5	1.039	1.182	1.137
.8	1.161	1.160	1.146

Without Regard To Design

In addition to the mean-based linear combination methods, one might consider a test statistic which treats the data as if it only consists of two samples. Let $T_1^*, \ldots, T_{n+n_1}^*$ and $C_1^*, \ldots, C_{n+n_2}^*$ represent observations from subjects receiving the treatment and the control, respectively. Then T_i^* is either an observation from the paired data, X_j, or one from the independent samples, T_k. Similarly, for C_i^*. Let

$$\bar{L}^* = \frac{1}{n+n_1}\sum_{i=1}^{n+n_1} T_i^* - \frac{1}{n+n_2}\sum_{j=1}^{n+n_2} C_j^* \qquad (24)$$

Thus, the computation of the statistic ignores the structure of the data. This was first investigated by Lin and Stivers (1974) and Ekbohm (1976). Bhoj (1989) included a version of this statistic in his simulation study. Usual computations reveal that the Pitman efficacy of T_M^* is

$$c_{\bar{L}^*} = \frac{1}{\sigma}\sqrt{\frac{[\gamma + \lambda(1-\gamma)]\,[\gamma + (1-\lambda)(1-\gamma)]}{2(1-\rho)\gamma + 1 - \gamma}} \qquad (25)$$

The efficiency of T^* with respect to $\bar{L}^*$ is given under bivariate normality in Table 6 where it can be seen that for moderate to high correlation T^* is superior even under normality.

Table 6: Efficiency of T^* with respect to $\bar{L}^*$, $\lambda = 1/2$

γ	$\rho = .2$	$\rho = .5$	$\rho = .8$
.2	0.95	1.06	1.56
.5	0.96	1.06	1.59
.8	0.93	1.00	1.24

The nonparametric counterpart to this approach is based on the Wilcoxon-Mann-Whitney statistic. The test consists of simply conducting the two sample rank test on the treatment versus the control observations irrespective of which part of the design they come from. The test will not have the usual Wilcoxon-Mann-Whitney permutation distribution. A normal approximation can be used to determine the p-value. However, certain quantities must be estimated. This approach was worked out in detail by Hollander, Pledger, and Lin (1974). The corresponding

estimator $\hat{\Delta}_{HPL}$ is simply the median of all the pairwise differences between treatment and control observations. In the case of bivariate normality, formulas from the paper can be used to compute the efficiency of $\hat{\Delta}^*_R$ with respect to $\hat{\Delta}_{HPL}$. However, we wish to make the comparison for contaminated bivariate normal distributions as well. Hence, we report simulated efficiencies in Table 7. They check with the formula values for the bivariate normal case. Note that T^* (and $\hat{\Delta}^*_R$) is generally superior to the rank test and $\hat{\Delta}_{HPL}$.

Table 7: Efficiency of $\hat{\Delta}^*_R$ with respect to $\hat{\Delta}_{HPL}$, $\lambda = 1/2$

γ	$\rho = .2$		$\rho = .5$		$\rho = .8$	
	Norm	CN	Norm	CN	Norm	CN
.2	1.001	1.013	1.119	1.096	1.639	1.626
.5	1.008	1.014	1.130	1.033	1.703	1.524
.8	1.011	0.970	1.079	1.114	1.362	1.579

Conclusion

A simple and efficient method for testing and estimation when the data collected is a mixture of paired and unpaired data is based on $T^+(\Delta)$, (1). The test statistic, point estimate, and confidence interval can be computed quickly and easily. The exact p-value is easily computed with an *S-plus* function for small sample sizes. The normal approximation works well for large sample sizes.

The $T^*(11)$ rank statistic is generally more efficient than T^+. The efficiency of T^* compares very favorably with the optimal weighted rank statistic. The linear model mean based estimate of Δ is a bit more efficient than the R-estimate based on T^* for a bivariate normal distribution, but we still estimate the correlation coefficient to use it. With a small amount of contamination the rank methods are superior. In addition, the rank methods, both test and estimate, are more robust in the sense that outliers will have less effect on them than on the mean based methods. In general, we recommend using T^*. However, if $n \geq n_1 + n_2$, Table 3 suggests that T^+ is almost as efficient as T^*. Hence, under those circumstances, we recommend using the simple methods based on T^+.

References

Bhoj, D. S (1989). On comparing correlated means in the presence of incomplete data. *Biometrics Journal, 3*, 279-288.

Dubnicka, S. R. (1998). Rank-based procedures for combined paired and unpaired data. Ph. D Thesis, The Pennsylvania State University.

Ekbohm, G. (1976). On comparing means in the paired case with incomplete data on both responses. *Biometrika, 63*, 299-304.

Ekbohm, G. (1981). On testing equality of means in the paired case with incomplete data on both responses. *Biometrical Journal, 23*, 251-259.

Hamdan, M., Khuri, A. & Crews, S. (1978). A test for equality of means of two correlated normal variates with missing data on both responses. *Biometrical Journal, 20*, 667-674.

Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. NY: John Wiley and Sons, Inc.

Hettmansperger, T. P., & McKean, J. W. (1998). *Robust nonparametric statistical methods*. NY: Arnold.

Hollander, M., Pledger, G., & Lin, P. -E. (1974). Robustness of the Wilcoxon test to a certain dependency between samples. *The Annals of Statistics, 2(1)*, 177-181.

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

Morrison, D. (1973). A test for equality of means of correlated variates with missing data on one response. *Biometrika, 61*, 101-105.

The Krypton Argon Regression Neovacularization Study Research Group (1993, November). Randomized comparison of krypton versus argon scatter photocoagulation for diabetic disc neovacularization. *Opthalmology, 100(11)*, 1655-1664.

The Optic Neuritis Study Group (1991). The clinical profile of optic neuritis: Experience of the optic neuritis treatment trial. *Arch. Opthal., 109*, 1673-1678.

Thompson, P. (1995). A hybrid paired and unpaired analysis for the comparison of proportions. *Statistics in Medicine, 14*, 1463-1470.

Woolson, R., Leeper, J., Cole, J. & .Clarke W. (1976). A Monte Carlo investigation of a statistic for a bivariate missing data problem. *Communication in Statistics - Theory and Methods, A5*, 681-658.

Appendix

Proof of Theorem 1: This follows immediately from the limiting normality of the stochastically independent signed rank and rank sum statistics. See Dubnicka (1998) for full details.

Proof of Theorem 2: Consider sequences that consists of an arrangement of n + and - signs followed by an ordering on n_1 T's and n_2 C's. A + in the i^{th} position indicates that the difference D_j is positive and $|D_j|$ is ranked i^{th} among $|D_1|,\ldots,|D_n|$. A - in the i^{th} positive means the D_j is negative and $|D_j|$ is ranked i^{th} among the absolute values. Also, a T(C) in the j^{th} position indicates that the j^{th} smallest response was obtained from a subject receiving the treatment (control). Then T^+ is simply the sum of the ranks

(positions) of the + signs plus the sum of the ranks of the T's minus $[n_1(n_1+1)]/2$.

When either $i = 0$ or $j = 0$, the independent samples portion only contains observations from either the control group or the treatment group. In these cases, the independent samples do not contribute to the value of $T^+(0)$. Thus, the value of $T^+(0)$ depends only on the observations from the paired data. Hence, when either $i = 0$ or $j = 0$, we can ignore the subscripts i and j. This completes the proof.

Pitman Efficiency

The Pitman efficacy of a test statistic T is defined as

$$c_T = \lim_{M\to\infty} \frac{\frac{d}{d\Delta}E_\Delta T(0)|_{\Delta=0}}{\sqrt{M}\sqrt{Var_0 T(0)}}. \qquad (26)$$

This quantity measures the standardized rate of change of the expected value of the statistic at the true value, 0 in this case. Relatively large values of c_T indicate high sensitivity of the test and hence large c_T is desirable. The computation of c_T depends on knowing the mean of T for general Δ and the variance under the true value.

Suppose further than $\hat{\Delta}$ is the estimate of Δ derived from T; then the asymptotic variance of $\sqrt{n}(\hat{\Delta}-\Delta)$ is c_T^{-2}. The Pitman testing and estimation efficiency of T_1 (and $\hat{\Delta}_1$) relative to T_2 (and $\hat{\Delta}_2$) is $e_{12} = (c_{T1}/c_{T2})^2$. See Hettmansperger and McKean (1998) for details.

Proof of Theorem 3: The finiteness of the integrals is sufficient to guarantee the existence of c_{T^+}. $Var_0 T^+(0)$ is given in (3). Also,

$$E_\Delta T^+(0) = n[1-F_d(-\Delta)] + \frac{n(n-1)}{2}\int[1-F_d(-z-2\Delta)]f_d(z)dz + n_1 n_2 \int[1-F_1(x-\Delta)]f_1(x)dx$$

where F_d represents the cumulative distribution function of the D_i in the paired data and F_1 is the common cumulative distribution function of the T_i and C_j in the unpaired data. By symmetry,

$$\frac{d}{d\Delta}E_\Delta T^+(0)|_{\Delta=0} = nf_d(0) + n(n-1)\int f_d^2(z)dz + n_1 n_2 \int f_1^2(x)dx.$$

Then

$$c_{T^+} = \lim_{M\to\infty} \frac{nf_d(0) + n(n-1)\int f_d^2(z)dz + n_1 n_2 \int f_1^2(x)dx}{\sqrt{M}\sqrt{\frac{n(n+1)(2n+1)}{24} + \frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

and the result immediately follows.

The Pitman efficiency (10) is given by

$$c_{Ta} = \sqrt{12}\frac{a\,2\int f_d^2 + b\int f_1^2}{\sqrt{a^2\frac{4}{\gamma} + b^2\frac{1}{\lambda(1-\lambda)(1-\gamma)}}}. \qquad (27)$$

We consider the weights that optimize the efficacy. The following lemma gives a formula for determining these weights and the maximum efficacy when the efficacy is of a particular form.

Lemma 1: Let $c^* = (ak_1 + bk_2)/(a^2k_3 + b^2k_4)^{1/2}$, $a+b=1$. Then c^* is maximized by $a = k_1k_4/(k_1k_4 + k_2k_3)$ and

$$c^*_{max} = \left(\frac{k_1^2}{k_3} + \frac{k_2^2}{k_4}\right)^{1/2}.$$

Applying Lemma 1 to c_{Ta}, we find that $a_{opt} = \gamma\int f_d^2 / \left[\gamma\int f_d^2 + 2\lambda(1-\lambda)\int f_1^2\right]$ maximizes the efficacy of T_a, and the maximum efficacy is

$$c_{Ta,max} = \sqrt{12}\sqrt{\gamma\left(\int f_d^2\right)^2 + \lambda(1-\lambda)(1-\gamma)\left(\int f_1^2\right)^2}.$$

In order to develop an efficient and, at the same time, practical statistic, consider a_{opt} when sampling from a bivariate normal distribution. We find, using $\int f_d^2 = 1/\left(2\sqrt{2\pi\sigma^2(1-\rho)}\right)$ and $\int f_1^2 = 1/\left(2\sqrt{\pi\sigma^2}\right)$, that $a_{opt,norm} = \gamma/\left[\gamma + 2\lambda(1-\lambda)(1-\gamma)\sqrt{2(1-\rho)}\right]$ where $\lambda \doteq n_1/(n_1+n_2)$ and $\gamma \doteq n/(n+n_1+n_2)$. We could estimate ρ from the data; however, the efficiency is fairly flat around $\rho = 1/2$, so we make that substitution to get

$$a^* = \frac{\gamma}{\gamma + 2\lambda(1-\lambda)(1-\gamma)} \doteq \frac{n(n_1+n_2)}{n(n_1+n_2)+2n_1n_2}$$

This is the optimal weighted rank statistic when sampling from a bivariate normal distribution with $\rho = 1/2$. T^*, (11), is based on a^*.

Proof of Theorem 4: The expressions for the estimator $\hat{\Delta}_{LM}(\rho)$ and its variance follow from routine but tedious calculations of $(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Z$ and $(X'\Sigma^{-1}X)^{-1}$. The limiting normality also follows immediately.