

Regression: Comparing Predictors and Groups of Predictors Based on a Robust Measure of Association

Rand R. Wilcox
University of Southern California

Abstract: Let ρ_j be Pearson's correlation between Y and X_j ($j = 1, 2$). A problem that has received considerable attention is testing $H_0: \rho_1 = \rho_2$. A well-known concern, however, is that Pearson's correlation is not robust (e.g., Wilcox, 2005), and the usual estimate of ρ_j , r_j has a finite sample breakdown point of only $1/n$. The goal in this paper is to consider extensions to situations where Pearson's correlation is replaced by a particular robust measure of association. Included are results where there are $p > 2$ predictors and the goal to compare any two subsets of $m < p$ predictors.

Key words: Dependent correlations, generalized variances, halfspace depth, median ball algorithm, robust methods.

1. Introduction

Let $Y_i, X_{i1}, \dots, X_{ip}$ ($i = 1, \dots, n$) be a random sample from some unknown $p + 1$ -variate distribution, let ρ_{yj} be Pearson's correlation between Y and X_j , and let ρ_{jk} be the correlation between X_j and X_k ($j, k = 1, \dots, p; j \neq k$). A problem that has received considerable attention is testing

$$H_0 : \rho_{yj} = \rho_{yk} \tag{1.1}$$

(e.g., Hittner, May and Silver, 2003; Hotelling, 1940; Olkin, 1967; Dunn and Clark, 1969; Meng, Rosenthal and Rubin, 1992); Steiger, 1980; Williams, 1959; Wilcox and Tian, 2008; Hendrickson, Stanley and Hills, 1970). Most methods are now known to be unsatisfactory in terms of Type I errors (e.g., Hittner *et al.*, 2003; Wilcox and Tian, 2008). Indeed, Wilcox and Tian (2008) found only one method that performs well in simulations.

The goal in this paper is to extend methods for testing (1.1) in two ways. The first is to compare any two predictors based on a particular robust measure of association with Y . The second is to consider a generalization of the first method to situations where the strength of the association between m variables

and Y is of interest, $1 < m < p$. To be slightly more precise, consider two non-identical subsets of the integers $1, \dots, p$, both having cardinality m . Denote these subsets by M_1 and M_2 , let X_{M_j} represent the predictor variables indexed by M_j ($j = 1, 2$), and let τ_{yM_j} be some measure of the strength of the association between Y and X_{M_j} . The goal is to test

$$H_0 : \tau_{yj} = \tau_{yk}. \quad (1.2)$$

Several methods were considered for the case $m = 1$, only one of which performed reasonably well in simulations, in terms of Type I errors and power. The method extends in an obvious way to $m > 1$, but it does not control the probability of a Type I error reasonably well in simulations. The one method that did perform well in simulations, when $m > 1$, is based in part on Tukey's notion of halfspace depth. It can be used when $m = 1$, but in terms of power it is not recommended for this special case for reasons that will be made evident.

The suggested method when $m = 1$ is based on a particular robust covariance. As well known, there are many robust analogs of the usual covariance matrix that result in robust analogs of ρ , which include the minimum volume ellipsoid (MVE) estimator (Rousseeuw and Leroy, 1987), the (fast) minimum covariance determinant (MCD) estimator (e.g., Rousseeuw and van Driessse, 1999), an estimate based on the median ball algorithm proposed by Olive (2004), the OGK estimator suggested by Maronna and Zamar (2002), the TBS (translated biweight) estimator derived by Rocke (1996), and various skipped estimators (e.g., Wilcox, 2005). Here the focus is on Olive's estimator, primarily because it is relatively simple and easy to compute. It is not being suggested that other robust measures of scatter should be eliminated from consideration. However, some results are briefly noted using the MVE and the (fast) MCD estimators to illustrate that the choice of estimator has practical consequences regarding the approximation of an appropriate critical value.

2. Some Background

This section reviews some methods and results needed in this paper.

2.1 The median ball algorithm

For completeness, the computational details associated with Olive's estimator, which is based on what he calls the reweighted median ball algorithm (RMBA), are outlined here. His approach begins with two estimates of location and scatter, $(T_{0,j}, \mathbf{C}_{0,j})$ ($j = 1, 2$). Compute all n Mahalanobis distances $D_i(T_{0,j}, \mathbf{C}_{0,j})$, $i = 1, \dots, n$, based on the j th measure of location and scatter. The next iteration consists of estimating the usual mean and covariance matrix based on the $c_n \approx$

$n/2$ cases corresponding to the smallest distances, yielding $(T_{1,j}, \mathbf{C}_{1,j})$. Repeating this process, based on $D_i(T_{1,j}, \mathbf{C}_{1,j})$, yields an updated measure of location and scatter, $(T_{2,j}, \mathbf{C}_{2,j})$. As done by Olive, $(T_{5,j}, \mathbf{C}_{5,j})$ is used here. The first of the two starting values used by Olive takes $(T_{0,1}, \mathbf{C}_{0,1})$ to be the usual mean and covariance matrix. The other starting value, $(T_{0,2}, \mathbf{C}_{0,2})$, is the usual mean and covariance based on the c_n cases that are closest to the coordinate wise median in Euclidean distance. Let $(T_A, \mathbf{C}_A) = (T_{5,j}, \mathbf{C}_{5,j})$, where $j = 1$ if the determinant $|\mathbf{C}_{5,1}| \leq |\mathbf{C}_{5,2}|$, otherwise $j = 2$. The RMBA estimate of location is T_A and the measure of scatter is

$$\mathbf{C}_{\text{RMBA}} = \frac{\text{MED}(D_i^2(T_A, \mathbf{C}_A))}{\chi_{p,0.5}^2} \mathbf{C}_A,$$

where MED indicates the sample median and $\chi_{p,0.5}^2$ is the .5 quantile of a chi-squared distribution with p degrees of freedom. Olive and Hawkins (2006) show that the RMBA estimate is \sqrt{n} consistent.

It is convenient to standardize X_{ij} yielding

$$Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j},$$

$i = 1, \dots, n; j = 1, \dots, p$, where μ_j and σ_j are the mean and standard deviation of X_{ij} . In the event the population mean and variance are not known, their usual estimates are used instead. Momentarily, focus on $p = 2$. The (sample) generalized variance associated with Y_i and Z_{ij} reflects how tightly clustered together the n pairs of points happen to be. From basic principles, the population analog is the determinant of the usual covariance matrix, namely

$$\sigma_y^2 - \rho_{yj}^2 \sigma_y^2,$$

where σ_y^2 is the variance of Y . And of course, if Y is standardized as well, the generalized variance reduces to $1 - \rho_{yj}^2$. So if we take τ_{yj} to be the generalized variance between Y and Z_j , testing (1.2) reduces to testing (1.1). However, a well-known practical concern is that the usual generalized variance is not robust, roughly meaning that a single outlier can give a distorted sense of the strength of the association among the bulk of the observations.

Standardizing Y , in addition to standardizing X , is perhaps the seemingly more natural way of measuring the strength of an association. But this is not done here because when both are standardized, it becomes more difficult to control the probability of a Type I error when comparing the strength of the association corresponding to any two predictors. In particular, a method was found that performs well in simulations when Y is not standardized, but when Y is standardized, the method performs poorly in simulations and alternative methods

that were considered also fail to perform well. Note, however, that even though Y is not standardized, the relative strength of the association corresponding to two subsets of variables can be measured with τ_{y1}/τ_{y2} .

Again consider two subsets of predictors indexed by M_j ($j = 1, 2$), both sets having cardinality m . One way of measuring the strength of the association between Y and the Z_{M_j} variables is to use some robust analog of the generalized variance and here, when $m = 1$, the strategy is to simply replace the usual covariance matrix with the RMBA estimate of scatter. Now the estimate of τ_{yj} , $\hat{\tau}_{yj}$, is taken to be the determinant of this matrix and testing (1.2) provides a way of comparing the strength of association between these two sets of variables. Of course, another possible way of measuring the strength of association between Y and two or more predictors is to fit some robust regression line and use some obvious analog of the usual squared multiple correlation coefficient. Perhaps there is some practical advantage associated with this alternative approach, but this issue is not pursued here.

3. Methods Considered for Testing (1.2) When $m = 1$

Momentarily, attention is focused on the case $p = 2$, $m = 1$. The first method for testing (1.2) that was considered here was a basic percentile bootstrap method (e.g., Efron and Tibshirani, 1993; Liu and Singh, 1997). It performs relatively well for a variety of situations where the goal is to compare robust measures of location or scatter (e.g., Wilcox, 2005). However, for the problem at hand, this approach was found to be unsatisfactory. A basic concern is that the actual level of the test is too sensitive to the distribution used to generate data in simulations. So in particular, if some adjustment is made so that, under normality, this approach has a Type I error probability equal to some specified value, the actual level of the test when generating data from a non-normal distribution can exceed the nominal level by a considerable amount. Consequently, this approach was abandoned.

The second strategy considered was to use an obvious test statistic based in part on a bootstrap estimate of the standard error of $\hat{\tau}_{yj}$. To elaborate, let $Y_i^*, X_{i1}^*, \dots, X_{ip}^*$ be a bootstrap sample obtained by randomly sampling with replacement n vectors of observations from $Y_i, X_{i1}, \dots, X_{ip}$, ($i = 1, \dots, n$). Let $\hat{\tau}_{yj}^*$ be the estimate of τ_{yj} based on this bootstrap sample. Repeat this process B times yielding the B bootstrap estimates $\hat{\tau}_{yjb}^*$, $b = 1, \dots, B$. Then an estimate of the squared standard error of $\hat{\tau}_{y1} - \hat{\tau}_{y2}$ is

$$\frac{1}{B-1} \sum_{b=1}^B (\hat{\tau}_{y1b}^* - \hat{\tau}_{y2b}^* - \bar{\tau}_{y1}^* + \bar{\tau}_{y2}^*)^2,$$

where $\bar{\tau}_{yj}^* = \sum \hat{\tau}_{yjb}^*/B$. This performed considerably better in simulations, but

the estimated probability of a Type I error was still judged to be a bit too unstable as a function of the distributions generating the data.

Note that in the bootstrap world, $\bar{\tau}_{yj}^* = \sum \hat{\tau}_{yjb}^*/B$ is attempting to estimate $\hat{\tau}_{yj}$, which is known. This suggests estimating the squared standard error of $\hat{\tau}_{y1} - \hat{\tau}_{y2}$ with

$$\hat{\omega}^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{\tau}_{y1b}^* - \hat{\tau}_{y2b}^* - \hat{\tau}_{y1} + \hat{\tau}_{y2})^2,$$

this provided more stability in simulations (estimated Type I errors are less affected by the distribution used to generate the data) and so the test statistic used here is

$$T = \frac{\hat{\tau}_{y1} - \hat{\tau}_{y2}}{\hat{\omega}}. \quad (3.1)$$

There remains the problem of approximating the null distribution of T . Still focusing on the case $p = 2$, $m = 1$, consider the situation where sampling is from a trivariate normal distribution with correlations $\rho_{y1} = \rho_{y2} = \rho_{12} = 0$. The initial strategy was to use simulations to determine an appropriate critical value for this special case and then use this critical value under non-normality. But this was not quite satisfactory. When sampling from a skewed distribution with all of the correlations set equal to .5, the actual Type I error probability, when testing at the .05 level, exceeded .075 in some cases.

To elaborate, let Z be a standard normal distribution, in which case

$$X = \begin{cases} \frac{\exp(gZ)-1}{g} \exp(hZ^2/2), & \text{if } g > 0 \\ Z \exp(hZ^2/2), & \text{if } g = 0 \end{cases}$$

has a g-and-h distribution where g and h are parameters that determine the first four moments. When $g = h = 0$, X has a standard normal distribution. With $g = 0$ this distribution is symmetric and it becomes increasingly skewed as g gets large. As h gets large, the g-and-h distribution becomes more heavy-tailed. Table 1 shows the skewness (κ_1) and kurtosis (κ_2) for each distribution considered in the simulations used here. They correspond to a standard normal ($g = h = 0$), a symmetric heavy-tailed distribution ($h = 0.2$, $g = 0.0$), an asymmetric distribution with relatively light tails ($h = 0.0$, $g = 0.2$), and an asymmetric distribution with heavy tails ($g = h = 0.2$). (Additional properties of the g-and-h distribution are summarized by Hoaglin, 1985.) It was found that altering h , with g fixed, has little or no effect on the probability of a Type I error. But altering g (increasing skewness) did have some effect. Also, increasing the correlation increased somewhat the probability of a Type I error, but as ρ approaches 1, the reverse was found to be true. Based on these results, it was decided to determine critical values when sampling from a g-and-h distribution with $h = 0.0$, $g = 0.2$ and a common correlation $\rho = .5$.

Table 1: Some properties of the g-and-h distribution

g	h	κ_1	κ_2
0.0	0.0	0.00	3.0
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

In particular, .05 critical values were estimated using the method just described for $n = 30, 50, 100, 200$ and 400 and it was found that the .05 critical value is given approximately by

$$2.06 - \frac{5.596}{\sqrt{n}}.$$

Note that if asymptotic normality is assumed, the intercept 2.06 would be 1.96. When using this critical value, situations are avoided where in the simulations to be described, the estimated Type I error probability exceeds .075 when the nominal level is .05. Simultaneously, for the normal and other distributions considered, the estimated probability generally does not drop below .025. One exception is when $n = 30, h = 0.0, g = 0.2$ and the common correlation is $\rho = .8$, in which case the estimate was .024. This will be called method R1.

4. Methods Considered for Testing (1.2), $m > 1$

Now consider the case $p > 2$ with $m = 2$. The strategy for determining a critical value, just outlined, extends in an obvious way to the situation at hand. But now simulations indicate that the actual Type I error probability is much more sensitive to the distribution used to generate data. If adjustments are made to avoid Type I errors that are well above the nominal level, there are situations (including normality) where the method is too conservative, meaning that when testing at the .05 level, the actual level is estimated to be less than .01. Moreover, power can be poor relative to the method suggested here.

Currently, only one method has been found that controls Type I error probabilities reasonably well, which is based in part on Tukey's notion of halfspace depth. Roughly, the idea is that if Y and X_{M_1} have a stronger association than Y and X_{M_2} , this will be characterized by the points (Y, Z_{M_1}) being more tightly clustered together than the points (Y, Z_{M_2}) . And the extent to which this is true can be investigated with a slight extension of results in Liu and Singh (1993), which is described in Wilcox (2005, p. 256).

Tukey's halfspace depth represents an extension of the notion of ranks to multivariate data. It reflects how deeply a point is nested within a multivariate data cloud without reliance on any type of covariance matrix. To describe its

formal definition, consider any point \mathbf{x} where \mathbf{x} is a column vector having length p , let \mathcal{H} be any closed halfspace containing the point \mathbf{x} , and let $P(\mathcal{H})$ be the probability associated with \mathcal{H} . Then Tukey's halfspace depth is

$$T_D = \inf_{\mathcal{H}} \{P(\mathcal{H}) : \mathcal{H} \text{ is a closed halfspace containing } \mathbf{x}\}. \quad (4.1)$$

There are relatively simple and reasonably accurate methods for approximating the halfspace depth of point and here the method given by Wilcox (2005, p. 207) will be used.

Momentarily consider the case of two independent, multivariate distributions, F and G . Let $\mathbf{U}_1, \dots, \mathbf{U}_{n_1}$ and $\mathbf{V}_1, \dots, \mathbf{V}_{n_2}$ be random samples from F and G , respectively, and let

$$R(\mathbf{v}; F) = P_F(T_D(\mathbf{U}; F) \leq T_D(\mathbf{v}; F)).$$

That is, $R(\mathbf{v}; F)$ is the probability that the depth of a randomly sampled \mathbf{U} , relative to F , is less than or equal to the depth of some particular point, \mathbf{v} , again relative to F . Said another way, $R(\mathbf{v}; F)$ is the fraction of the F population that is less central than the value \mathbf{v} . For \mathbf{V}_i , $R(\mathbf{V}_i)$ is estimated with the proportion of points among $\mathbf{U}_1, \dots, \mathbf{U}_{n_1}$ that have a smaller halfspace depth than \mathbf{V}_i , relative to F . The quality index proposed by Liu and Singh (1993) is

$$Q(F, G) = E_G(R(\mathbf{V}; F)),$$

the average of all $R(\mathbf{v}; F)$ values with respect to the distribution G . Put another way, for a randomly sampled \mathbf{U} and \mathbf{V} ,

$$Q(F, G) = P(D(\mathbf{U}; F) \leq D(\mathbf{V}; F))$$

is the probability that the depth of V is greater than or equal to depth of U . Liu and Singh show that the range of Q is $[0, 1]$ and when $F = G$, $Q(F, G) = 1/2$. An estimate of $Q(F, G)$ is

$$\hat{Q}(F, G) = \frac{1}{n_2} \sum_{i=1}^{n_2} R(\mathbf{V}_i; F_{n_1}),$$

where F_{n_1} is the usual empirical distribution associated with the first group, and an estimate of $Q(G, F)$ is

$$\hat{Q}(G, F) = \frac{1}{n_2} \sum_{i=1}^{n_2} R(\mathbf{U}_i; G_{n_2}).$$

To determine whether the distributions differ in scatter, one can test

$$H_0 : Q(F, G) = Q(G, F), \quad (4.2)$$

and the basic percentile bootstrap method appears to perform well in terms of Type I errors (Wilcox, 2005, p. 256). That is, randomly sample with replacement n_j vectors of observations from the j th group, estimate $Q(F, G)$ and $Q(G, F)$, yielding say $Q^*(F, G)$ and $Q^*(G, F)$, and let $D^* = Q^*(F, G) - Q^*(G, F)$. Repeat this process B times and put the results in ascending order yielding $D^*_{(1.1)} \leq \dots \leq D^*_{(B)}$. Then a $1 - \alpha$ confidence interval for $Q(F, G) - Q(G, F)$ is simply $(D^*_{(\ell+1)}, D^*_{(u)})$, where $\ell = \alpha B/2$, rounded to the nearest integer, and $u = B - \ell$.

Returning to the problem at hand, simple modifications of the method just outlined were considered, with all but one performing poorly in simulations. For example, taking bootstrap samples of all n vectors and computing D^* was considered. Another variation was to use independent bootstrap samples when computing $Q^*(F, G)$ and $Q^*(G, F)$. Another strategy is to split the data into two independent components and use the method outlined in the previous paragraph. Certainly, this approach is not ideal, but in terms of controlling the Type I error probability, it is only approach found that performs reasonably well in simulations.

More formally, let $n_1 = n/2$, rounded down to the nearest integer. Then, under random sampling, (Z_{i,M_1}, Y_i) ($i = 1, \dots, n_1$) is independent of (Z_{i,M_2}, Y_i) ($i = n_1 + 1, \dots, n$), and the issue is whether the (population) amount of scatter corresponding to these two groups differs. And here the hypothesis that they do not differ is tested with the method just outlined. This will be called method R2.

A variation of this method also controls Type I errors in the simulations considered here. First test the hypothesis using method R2 as just described, and then test it again, only now using (Z_{i,M_2}, Y_i) ($i = 1, \dots, n_1$) and (Z_{i,M_1}, Y_i) ($i = n_1 + 1, \dots, n$). Denote the corresponding p-values with p_1 and p_2 . Then apply a sequentially rejective method designed to control the probability of at least one Type I error (e.g., Rom, 1990). That is, reject at the α level if both p_1 and p_2 are less than or equal to α , or if either one is less than or equal to $\alpha/2$. However, this approach was found to have less power than method R2, so further details are omitted.

5. Two Illustrations

Data for \mathbf{X} were generated from a trivariate normal distribution with $\rho_{12} = .5$, and Y was taken to be $Y = X_1 + .5X_2 + .1X_3 + \epsilon$ with ϵ standard normal. Here the sample size is $n = 200$. Figure 1 shows a scatterplot of Y versus Z_1 and Z_3 . The points marked by * correspond to the first predictor and points marked by o are for the third predictor. Notice how the first set of points tend to be nested within the points marked by an o, indicating that the first predictor has a stronger association with Y . And the extent to which this is true can

be quantified with halfspace as previously described. The polygons denote the central half of the data as measured by halfspace depth. The .95 confidence interval for $Q(F, G) - Q(G, F)$ is (0.15, 0.46). If instead the goal is to compare the strength of association based of the first two predictors versus predictors 2 and 3, the .95 confidence interval is (0.176, 0.486), indicating that the first two predictors, taken together, have a stronger association with Y .

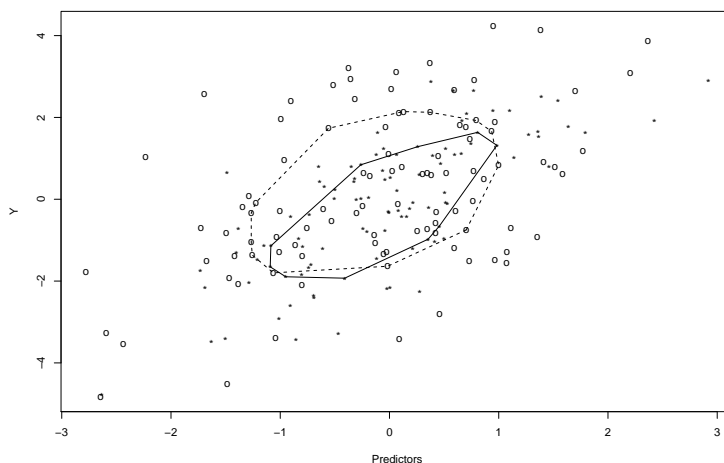


Figure 1: Scatterplot corresponding to two predictors

The following illustrates that the choice of method can make a practical difference when analyzing actual data. In an unpublished study by L. Doi, of interest were predictors of reading ability among children. One portion of the study dealt with how well measures of accuracy and speed, when identifying lower case letters, predicts scores on a reading comprehension test. Figure 2 shows a scatterplot of the points after the predictors have been standardized. As is evident, the points indicated by an *, which correspond to a measure of accuracy, are more tightly clustered together except for a few points in the lower right portion of the plot. If Pearson correlations are compared using the method in Zou (2007), no difference is found at the .05 level. The sample size is 81 and the estimates of Pearson's correlation for accuracy and speed are $-.11$ and $-.33$, respectively. Note that Figure 2 suggests that, there is a seemingly strong positive association. Pearson's correlation is negative due to the outliers in the lower right corner. The correlation based on the RMBA covariance matrix is $.59$.

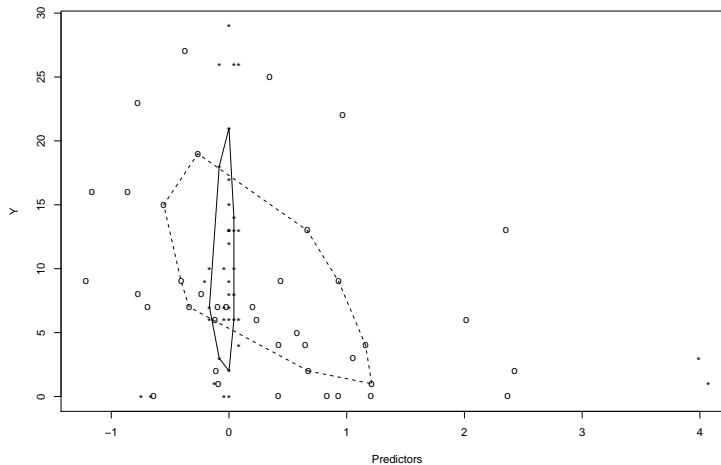


Figure 2: Scatterplot corresponding the reading data.

Table 2: Estimated type I errors, $\alpha = .05$, $n = 40$

g	h	ρ_{12}	$p = 2, m = 1$	$p = 3, m = 2$
0.0	0.0	0.0	.030	.026
0.0	0.2	0.0	.030	.026
0.2	0.0	0.0	.032	.051
0.2	0.2	0.0	.032	.051
0.0	0.0	0.5	.034	.032
0.0	0.2	0.5	.034	.032
0.2	0.0	0.5	.042	.047
0.2	0.2	0.5	.042	.047

6. Simulation Results

Simulations were used to check the ability of methods R1 and R2 to control the probability of a Type I error when the marginal distributions have one of the four g -and- h distributions previously described. The predictors were taken to have a common correlation of $\rho = 0$ or $.5$. Correlations among the predictors were formed as follows. Let \mathbf{R} be the desired correlation matrix and form the Cholesky decomposition $\mathbf{U}'\mathbf{U} = \mathbf{R}$, where \mathbf{U}' is the transpose of \mathbf{U} . If \mathbf{X} represents an $n \times p$ matrix of data where the marginal distributions are independent, then \mathbf{XU} produces an $n \times p$ matrix of data that has population correlation matrix \mathbf{R} .

Table 2 reports the estimated Type I error probabilities for methods R1 and R2 when $n = 40$. The results for method R2 are for the case $p = 3$, $M_1 = \{1, 2\}$

and $M_2 = \{2, 3\}$. That is, the goal is to test the hypothesis that the strength of association based on predictors 1 and 2 is the same as for predictors 2 and 3. As is evident, Type I error probabilities are controlled reasonably well among all of the situations considered.

7. Concluding Remarks

There are numerous variations of method R2 and perhaps some of them have practical value. For example, there are a variety of ways of measuring the depth of a point in a multivariate data cloud other than the approach used here (e.g., Wilcox, chapter 6). The main point is that many methods have been ruled out and two methods were found that perform reasonably well in terms of Type I errors. Method R1 is best, in terms of power, when $m = 1$, but for $m > 1$, only method R2 was found to be reasonably satisfactory in simulations.

Regarding method R1, of interest is whether some alternative robust covariance can be used with the same .05 critical value used here. The simulations were run again using the MVE and MCD estimators and it was found that now Type I error probabilities are not controlled well. Perhaps other robust covariances give satisfactory results, but this remains to be determined.

It is noted that a few simulations were run using method R2 with $p = 4$ and $m = 2$ and 3. The results were very similar to those in Table 2. Of course, perhaps method R2 breaks down with p sufficiently large, but this possibility is left for future studies,

Finally, both R and S-PLUS software are available from the author for applying both methods R1 and R2; ask for the functions `regpord` and `regpordv2`.

References

- Dunn, O. J. and Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association* **64**, 366-377.
- Dunn, O. J. and Clark, V. A. (1971). Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association* **66**, 904-908.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
- Hendrickson, G. F., Stanley, J. C. and Hills, J. R. (1970). Olkin's new formula for significance levels of r_{13} vs. r_{23} compared with Hotelling's method. *American Educational Research Journal* **7**, 189-195.
- Hittner, J. B., May, K., and Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology* **130**, 149-168.

- Hoaglin, D. C. (1985) Summarizing shape numerically: The g-and-h distributions. In *Exploring data tables, trends, and shapes*. (Edited by D. Hoaglin, F. Mosteller and J. Tukey, 461-515). Wiley.
- Hotelling, H. (1940). The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters. *Annals of Statistics*, **11**, 271-283.
- Liu, R. Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association* **88**, 252-260.
- Maronna, R. A. and Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* **44**, 307-317.
- Meng, X.-L., Rosenthal, R. and Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin* **111**, 172-175.
- Olkin, I. (1967). Correlations revisited. In *Improving Experimental Design and Statistical Analysis* (Edited by J. C. Stanley, 102-128). Rand McNally.
- Olive, D. J. (2004). A resistant estimator of multivariate location and dispersion *Computational Statistics and Data Analysis* **46**, 93-102.
- Olive, D. J. and Hawkins, D. M. (2006). Robustifying robust estimators. Preprint (www.math.siu.edu/olive/preprints.htm).
- Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics* **24**, 1327-1345.
- Rom, D. M. (1990). A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77**, 663-666.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley.
- Rousseeuw, P. J., Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212-223.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix *Psychological Bulletin* **87**, 245-251.
- Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*, 2nd Ed., Academic Press.
- Wilcox, R. R. and Tian, T. (2008) Comparing dependent correlations. *Journal of General Psychology* **135**, 105-112.
- Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B* **21**, 396-399.
- Yu, M. C. and Dunn, O. J. (1982). Robust tests for the equality of two correlation coefficients: A Monte Carlo study. *Educational and Psychological Measurement* **42**, 987-1004.

Received September 12, 2008; accepted October 29, 2008.

Rand R. Wilcox
Dept. of Psychology
University of Southern California