psychometrika—vol. 1, no. 1, 1–0
march 1936

ON A TEST OF DIMENSIONALITY IN REDUNDANCY ANALYSIS

Yoshio Takane

mcgill university

Heungsun Hwang

hec montreal

The work reported in this paper was supported by Grant A6394 to the first author from the Natural Sciences and Engineering Research Council of Canada.

Requests for reprints should be sent to Yoshio Takane, Department of Psychology, McGill Universty, 1205 Dr. Penfield Ave., Montreal, QC H3A 1B1 CANADA.

Abstract

Lazraq and Cléroux (Psychometrika, 2002, 411–419) proposed a test for identifying the number of significant components in redundancy analysis. This test, however, is ill conceived. A major problem is that it regards each redundancy component as if it were a single observed predictor variable, which cannot be justified except for the rare situations in which there is only one predictor variable. Consequently, the proposed test leads to drastically biased results, particularly when the number of predictor variables is large, and it cannot be recommended for use. This is shown both theoretically and by Monte Carlo studies.

Key words: Reduced rank regression, PCA of instrumental variables, Parallel analysis, Permutation tests.

Introduction

Redundancy analysis (van den Wollenberg, 1977) is a popular method of multivariate analysis for analyzing the relationship between two sets of variables (e.g., Reinsel and Velu, 1998). Unlike canonical correlation analysis which maximizes the correlation between two sets of variables, redundancy analysis maximizes predictability of one set of variables from the other. Whereas in canonical correlation analysis it is sufficient to have one variable in one set having a large correlation with at least one variable in the other set to obtain a large canonical correlation, it is necessary in redundancy analysis that most, if not all, of the criterion variables are sufficiently predictable from the predictor set to obtain a large value of redundancy index. The latter thus captures the overall relationship between the two sets of variables more accurately (e.g., Lambert, Wildt, and Durand, 1988).

There are several different formulations of redundancy analysis, resulting in different names for essentially the same analytic technique. Redundancy analysis is called by at least two other names, reduced rank regression (Anderson, 1951) and principal components of instrumental variables (Rao, 1964). Given matrices of criterion variables and predictor variables, these techniques all extract a set of mutually orthogonal components (called redundancy components) from the set of predictor variables that can maximally predict the variability in the criterion variables. As in other techniques with similar objectives that extract components based on some optimality criteria (such as canonical correlation analysis), the question of "how many components to extract" is of paramount importance. Lazraq and Cléroux (2002) recently proposed a procedure for testing the number of significant components in redundancy analysis. This procedure, however, is ill-founded. A major problem stems from the fact that it treats each redundancy component as if it were a single observed predictor variable. This, however, can only be justified when the weights in the linear combination of predictor variables forming the redundancy components are fixed, which is strictly true only when the number of predictor variables is one. Consequently, the proposed test leads to drastically biased results, as the number of predictor variables increases, and it cannot be recommended for use. This will be shown in this paper both theoretically and by means of Monte-Carlo studies. Some alternative methods are suggested for determining the number of significant components in redundancy analysis.

A Summary of the Lazraq-Cléroux Procedure

In this section, we briefly overview the Lazraq-Cléroux procedure (2002). Let Y (p by 1) and X (q by 1) be random vectors of the criterion and predictor variables, respectively. It is assumed that $\begin{pmatrix} Y \\ X \end{pmatrix} \sim N(0, \Sigma)$, where the covariance matrix Σ is partitioned into

$$\Sigma = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}. \tag{1}$$

Let $s = \min(p, q)$, and let $T = [t_1, \ldots, t_s]'$ denote the s-component vector of population redundancy components defined by

$$T = A'X, \tag{2}$$

where $A = [\alpha_1, \ldots, \alpha_s]$ (q by s) is the matrix of generalized eigenvectors of $\Sigma_{XY}\Sigma_{YX}$ with respect to Σ_{XX} corresponding to its nonzero generalized eigenvalues. Each successive $t_h (h = 1, \ldots, s)$ maximizes the proportion of variability in Y that can be accounted for by t_h. It follows that $\begin{pmatrix} Y \\ T \end{pmatrix} \sim N(0, \Sigma^*)$ with

$$\Sigma^* = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YT} \\ \Sigma_{TY} & \Sigma_{TT} \end{pmatrix} = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX}A \\ A'\Sigma_{XY} & A'\Sigma_{XX}A \end{pmatrix}, \tag{3}$$

where it follows further that $A'\Sigma_{XX}A = I_s$, the identity matrix of order s. Let

$$\rho_I(Y, t_h) = \frac{\mathrm{tr}(\Sigma_{Yt_h}\sigma_{t_h t_h}^{-1}\Sigma_{t_h Y})}{\mathrm{tr}(\Sigma_{YY})} \tag{4}$$

denote the population redundancy index for predicting Y from t_h. This quantity represents the proportion of the total variance in Y that can be accounted for by t_h. We are interested in testing whether $\rho_I(Y, t_h) = 0$ for $h = 1, \ldots, s$. This is carried out separately (but sequentially) for each h.

In practical data analysis situations, the population covariance matrix is usually unknown, and it should be estimated from data. Suppose a random sample of size n, $\begin{pmatrix} Y_1 \\ X_1 \end{pmatrix}, \ldots, \begin{pmatrix} Y_n \\ X_n \end{pmatrix}$, were drawn from $N(0, \Sigma)$. Then, Σ is estimated by the unbiased sample covariance matrix, S, partitioned in the same way as Σ,

$$S = \begin{pmatrix} S_{YY} & S_{YX} \\ S_{XY} & S_{XX} \end{pmatrix}, \tag{5}$$

where $S_{YY} = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'/(n-1)$, $S_{YX} = S'_{XY} = \sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})'/(n-1)$, $S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'/(n-1)$, $\bar{Y} = \sum_{i=1}^{n} Y_i/n$, and $\bar{X} = \sum_{i=1}^{n} X_i/n$. The population redundancy index in (4) is estimated by the corresponding sample redundancy index

$$R_I(Y, t_h) = \frac{\mathrm{tr}(S_{Yt_h} s_{t_h t_h}^{-1} S_{t_h Y})}{\mathrm{tr}(S_{YY})}, \tag{6}$$

where S_{Yt_h}, and $s_{t_h t_h}$ are sample estimates of Σ_{Yt_h}, and $\sigma_{t_h t_h}$, respectively. Under the hypothesis that $\rho_I(Y, t_h) = 0$, the following statistic,

$$\phi_h \equiv \frac{R_I(Y, t_h)}{1 - R_I(Y, t_h)}, \tag{7}$$

can be expressed as the ratio of the trace of two independent Wishart variables generated from $N(0, \Sigma_{YY})$,

$$V_1 = (n-1) \cdot \mathrm{tr}(S_{Yt_h} s_{t_h t_h}^{-1} S_{t_h Y}) \tag{8}$$

over

$$V_2 = (n-1) \cdot \mathrm{tr}(S_{YY} - S_{Yt_h} s_{t_h t_h}^{-1} S_{t_h Y}). \tag{9}$$

Then, for any non-negative quantity v (Meaningful choices of the value of v will be discussed below.),

$$\mathrm{Prob}(\phi_h \leq v) = \mathrm{Prob}(\frac{V_1}{V_2} \leq v) = \mathrm{Prob}(V_1 - vV_2 \leq 0), \tag{10}$$

where $V_1 - vV_2$ is known to follow the distribution of a random variable defined as a weighted sum of squares of $p(n-1)$ independent standard normal variables (Lazraq and Cléroux, 1988), where the weights are obtained by the p eigenvalues of Σ_{YY} and the p eigenvalues of $-v\Sigma_{YY}$ repeated $n-2$ times. Note that these weights depend only on Σ_{YY}, and consequently the probability in (10) depends only on Σ_{YY} (and the value of v). A good approximation of this probability is obtained by substituting S_{YY} for Σ_{YY}.

The probability in (10) is evaluated (approximated) for a specific value of v, say, the observed value of ϕ_h to obtain 1 minus the p-value. The p-value may then be compared with a selected significance level α. Alternatively, the critical value, g_α, may be obtained such that $\mathrm{Prob}(R_I(Y, t_h)/(1 - R_I(Y, t_h)) \leq g_\alpha) = 1 - \alpha$. The value of g_α is then compared with the observed value of ϕ_h.

The Problem

The proposed procedure looks mathematically elegant. However, there is a pitfall in the above derivation. The most serious problem is that the population redundancy components are tacitly equated to sample redundancy components. The former is obtained by (2), but A in (2) is usually unknown and has to be estimated from a sample. Let $\hat{A}$ denote an estimate of A. It is obtained by the generalized eigenvectors of $S_{XY}S_{YX}$ with respect to S_{XX}. Sample redundancy components, $\hat{T}$, are then obtained by $\hat{T} = [\hat{t}_1, \ldots, \hat{t}_s]' = \hat{A}'X$. In the part of the above derivation pertaining to the sampling distribution of $R_I(Y, t_h)$ and ϕ_h, t_h should have been replaced by $\hat{t}_h$. To avoid confusion, we put a hat on these quantities to indicate that t_h is replaced by $\hat{t}_h$, namely,

$$\hat{\phi}_h = \frac{\hat{R}_I(Y, \hat{t}_h)}{1 - \hat{R}_I(Y, \hat{t}_h)}, \tag{11}$$

where

$$\hat{R}_I(Y, \hat{t}_h) = \frac{\mathrm{tr}(S_{Y\hat{t}_h} s^{-1}_{\hat{t}_h\hat{t}_h} S_{\hat{t}_h Y})}{\mathrm{tr}(S_{YY})}. \tag{12}$$

Two peculiar things occur:

1) The Lazraq-Cléroux test does not depend on the number of predictor variables, q. The size of the largest generalized eigenvalue of $S_{XY}S_{YX}$ with respect to S_{XX}, however, tends to increase as q increases. The test statistic, $\hat{\phi}_1$, for testing the significance of the most dominant component is a simple monotonic function of the largest generalized eigenvalue.

2) The test does not depend on h, where h is the index (order) of redundancy components. A smaller value of h indicates a more dominant component corresponding to a larger generalized eigenvalue. The size of generalized eigenvalues naturally decreases as h increases.

These peculiar characteristics stem from the fact that Lazraq and Cléroux set $\hat{T} = T$. As indicated in the previous section, T follows the *iid* standard normal distribution. On the basis of this, Lazraq and Cléroux treat each $\hat{t}_h$ as if it were a single observed predictor variable that follows the *iid* standard normal distribution. However, this can only be justified for the population redundancy components, where the weights applied to X are fixed constants. This is not generally true for sample redundancy components, except for $q = 1$. The weights used to obtain sample redundancy components, being derived from a sample, are not fixed constants, but are functions of the random variables whose linear combinations are taken to obtain the sample redundancy components. (Linear combinations of X is taken to obtain $\hat{T}$, i.e., $\hat{T} = \hat{A}'X$, but $\hat{A}$ itself is a function of X.) Under these circumstances the linear composites (the sample redundancy components) do not follow the presumed normal distribution (Lancaster, 1963).

We formalize the above observation in the following theorem, and show that $\hat{\phi}_1$ is stochastically larger than any random variable defined analogously by a fixed linear combination of X, if $q > 1$. The proof is similar to that of Lancaster's (1963), who proved that n (the sample size) times the largest eigenvalue from canonical analysis of contingency tables was stochastically larger than a chi-square variable with specified degrees of freedom.

Theorem. Let Y and X be as introduced in the previous section. Let u be a linear combination of X with fixed weights (i.e., $u = b'X$ for some fixed vector b). Define

$$\psi(u) = \frac{\delta(u)}{\text{tr}(S_{YY}) - \delta(u)}, \tag{13}$$

where

$$\delta(u) = \text{tr}(S_{Yu} s_{uu}^{-1} S_{uY}). \tag{14}$$

Then, $\hat{\phi}_1 = \psi(\hat{t}_1) > \psi(u)$ with probability 1, if $q > 1$, where $\hat{t}_1$ is the most dominant sample redundancy component.

Proof. Let

$$\hat{\lambda}_1 = \max_a \frac{a' S_{XY} S_{YX} a}{a' S_{XX} a} \tag{15}$$

denote the largest generalized eigenvalue of $S_{XY} S_{YX}$ with respect to S_{XX}, and let $\hat{a}_1$ be the corresponding eigenvector. Then, $\hat{t}_1 = \hat{a}_1' X$, and thus, $\hat{\lambda}_1 = \delta(\hat{t}_1) \geq \delta(u) = \delta(b'X)$ for any fixed q-component vector b. Since $\hat{\lambda}_1$ is a continuous random variable, the probability of $\hat{\lambda}_1 = \delta(u)$ is infinitesimally small (i.e., the chance of $\hat{a}_1 = b$ is nil) if $q > 1$, and $\hat{\lambda}_1 > \delta(u)$ with probability 1. Since $\psi(u)$ is a strictly monotonic function of $\delta(u)$, $\hat{\phi}_1 = \psi(\hat{t}_1) > \psi(u)$ with probability 1, if $q > 1$. When $q = 1$, b is a scalar, and no matter what its value is, $u = \hat{t}_1$. QED

Corollary. $\hat{\phi}_1 > \psi(x_i)$ with probability 1 for any single observed predictor variable x_i, if $q > 1$.

Proof. Set $b = e_i$, where e_i is the q-component vector with a one in the i^{th} position and zeros elsewhere. (i.e., $u = e_i'X = x_i$). QED

The above theorem and corollary suggest that the Lazraq-Cléroux test is too liberal for the test of significance of the most dominant redundancy component.

One may still argue that the above results are only for finite n, and that the difference between $\psi(\hat{t}_1)$ and $\psi(u)$ diminishes as n goes to infinity. In particular, $S_{XX}^{-1}S_{XY}S_{YX}$ converges in probability to $\Sigma_{XX}^{-1}\Sigma_{XY}\Sigma_{YX}$, $\hat{A}$ to A, and $\hat{T}$ to T. That is, $\hat{T}$ may be replaced by T for n sufficiently large, and the usual asymptotic theory holds. Unfortunately, this is not so due to a violation of one of the regularity conditions typically assumed in the asymptotic theory. Under the null hypothesis that $\rho_I(Y, t_1) = 0$, A is not uniquely determined (being eigenvectors of a zero matrix), and hence T is not identifiable. The usual asymptotic theory does not hold in this case (e.g., Wilks, 1938). This is similar to the "singularity" problem associated with the likelihood ratio tests involving some hierarchical models, where some of the model parameters postulated under the alternative hypothesis are not identifiable under the null hypothesis. Such models include not only redundancy analysis but also virtually all models that require dimensionality selection (Takane, van der Heijden, and Browne, 2003).

Monte-Carlo Studies

To confirm our theoretical assertions in the previous section, and to show the severity of the problem with the Lazraq-Cléroux procedure (which was not indicated by the theory), we have conducted a series of small Monte-Carlo studies. The scope of the studies is limited to the demonstrations of the seriousness of the problem with the Lazraq-Cléroux procedure. There is no intention to develop alternative criteria or correction formula for the Lazraq-Cléroux procedure. The first study intends to show that Lazraq and Cléroux' result holds for $q = 1$. However, the second study shows that their result is incorrect for $q > 1$, and that the bias of the test gets larger as q gets larger. The third study investigates the behavior of non-dominant redundancy components, and the fourth study questions the adequacy of the complete independence assumption between Y and X in investigating the behavior of non-dominant redundancy components.

The design of the studies is similar to that of "parallel analysis" (Horn, 1965) often used to determine the number of significant components in principal component analysis. In our adaptation of parallel analysis, we repeatedly generated data according to the assumption of multivariate normality and under the hypothesis that $\rho_I(Y, t_1) = 0$ (except in the fourth study) and examined the distribution of the test statistic, $\hat{\phi}_h$, as functions of various factors. We specifically look for the critical values, g_α, for the significance level α as functions of n (the sample size), p (the number of criterion variables), and h (index of redundancy components). We also look at the probability of Type I error committed by the Lazraq-Cléroux test. All the computations were done by MATLAB 6.1. In all cases, reported results are based on 10,000 Monte Carlo samples.

The first study was designed to see that Lazraq and Cléroux (2002)' theory holds for $q = 1$. We generated data with $q = 1$, but varying the values of n and p in exactly the same way as in Lazraq and Cléroux. That is, $n = 25$, 50, 75, 100 and 200, and $p = 2$, 3, 4 and 5. We generated each data set, $\{Y_i, x_i\}$ for $i = 1, \ldots, n$, according to x_i following *iid* standard normal, Y_i *iid* multivariate normal with zero means and prescribed variances, and x_i and Y_i independent of each other. The variances of Y were identical to those assumed by Lazraq and Cléroux, i.e., $\sigma_{y_i y_i} = i^2$ for the i^{th} criterion variable, y_i. (Covariances can all be assumed zero without loss of generality.) Redundancy analysis was applied to each contrived data set to obtain the distribution of the test statistic, $\hat{\phi}_1$. Table 1 reports the critical value, g_α, of $\hat{\phi}_1$ for selected values of the significance level α obtained by the Monte Carlo study. (The combinations of the values of p, n and α for which g_α is reported in Table 1 coincide with those of Lazraq and Cléroux.) For comparison, the exact critical value, as reported by Lazraq and Cléroux, is also provided in parentheses. In all cases, the reported critical values agree very well with the exact critical value, suggesting that the Lazraq-Cléroux procedure is more or less correct when $q = 1$. However, this is not the typical situation in which redundancy analysis is applied.

Insert Table 1 about here

In the above study, the number of predictor variables, q, was set to unity. The second Monte Carlo study examines the effect of q on the critical value for selected values of n and p ($n = 75$, 200, and 5000, and $p = 2$ and 5). (Some of the values of n and p used by Lazraq and Cléroux were not included in this study. The included ones were, however, deemed sufficient to reveal general tendencies.) The value of q was varied over five levels: 1, 2, 5, 10, and 20. The data were generated in a manner similar to the first study except that when $q > 1$, elements of X_i were also assumed independent. Table 2 reports the critical value of $\hat{\phi}_1$ associated with $\alpha = .05$. As before, the critical value tends to decrease as both n and p increase. More importantly, however, the critical value gets larger in all cases, as q gets larger. This is understandable because as q gets larger, the chance of any of the predictor variables, which happens to be correlated with any of the criterion variables by chance, gets larger, inflating the critical value. This is thus as expected but contrary to Lazraq and Cléroux' theory. Note that this tendency does not disappear even for the sample size as large as 5000. Table 3 reports the probability of Type I error by the Lazraq-Cléroux procedure in testing the most dominant redundancy component. Results for $n = 5000$ use approximate critical values obtained by the Monte Carlo study, since Lazraq and Cléroux do not provide the "exact" critical value for this case. This probability goes up rather quickly, as q goes up for all the sample sizes examined, suggesting that the Lazraq-Cléroux test is seriously biased for large q. In particular, note that the probability of Type I error becomes nearly one when $q = 10$ even for $n = 5000$.

Insert Tables 2 and 3 about here

The third Monte Carlo study examines the behavior of $\hat{\phi}_h$ for $h > 1$. The sample size (n) and the number of criterion variables (p) were varied in the same way as in the second study. The number of predictor variables was also varied in the same way as in the second study, but excluding the case in which $q = 1$. Otherwise, the data were generated in exactly the same way as in the second study. Table 4 reports the critical value of $\hat{\phi}_h(h = 1, ..., s)$ for $\alpha = .05$ as functions of n, p, and q. It is observed that for all combinations of n, p and q, the critical value decreases as h increases.

This is quite natural, since components are ordered by size, but is contrary to Lazraq and Cléroux' assumption that the distribution of $\hat{\phi}_h$ is independent of h. Again, this tendency does not disappear even for $n = 5000$. Table 5 reports the probability of Type I error by the Lazraq-Cléroux procedure to test the significance of non-dominant redundancy components. As in the test of the most dominant component (Table 3), the probability of Type I error increases as q increases. This tendency is partially offset by the reverse tendency observed as h increases. However, it does not seem practical to figure out for what combinations of n, p, q, and h the Lazraq-Cléroux test gives correct significance levels.

Insert Tables 4 and 5 about here

In the third Monte Carlo study (as well as in all other simulation studies reported above), data were generated under the hypothesis that Y and X were completely independent (i.e., $\rho_I(Y, t_1) = 0$). This is consistent with the Lazraq-Cléroux procedure in which the distributions of ϕ_h for different components are all derived under the assumption of complete independence between Y and X. However, with the Lazraq-Cléroux procedure the second component is never tested unless the first component is significant. (It is impossible to find a significant second component unless the first component is significant. The observed value of $\hat{\phi}_2$ is necessarily smaller than that of $\hat{\phi}_1$, and an identical critical value is used for testing both.) This means that by the time we test the significance of the second component, we know that the first component is significant. Then, it might be more appropriate to derive the distributions of $\hat{\phi}_h$ for non-dominant components under non-null values of $\rho_I(Y, t_1)$. The fourth simulation study investigates the effects of non-null values of $\rho_I(Y, t_1)$ on the distribution of the test statistics for subsequent components. The data for the criterion variables were generated in the same way as before. The data on the predictor side were generated according to

$$X_i = (1-c)Z_i + cY_{i1}1_q, \tag{16}$$

where Z_i was assumed *iid* standard normal (independent also across elements of X_i; Z_i is like X_i in the second and third Monte Carlo studies), Y_{i1} is the first element in Y_i, and 1_q is the q-component vector of ones. This incorporates varying degrees of rank-one dependency between Y and X, the size of which are modulated by the value of c, without affecting the variance-covariance structure of Y. The value of c was varied over four levels: .05, .1, .2, and .5. Stronger dependency was effected by larger values of c. In this study, p was fixed at 5, although q and n were varied in the same way as in the third study. (The cases of $n = 5000$ were not reported in Table 6, but basically the same tendency holds.) Table 6 reports the 5% critical value, $g_{.05}$, of the test statistic, $\hat{\phi}_h$, for $h = 1$, 2, and 3. Quite naturally, we find larger values of $\hat{\phi}_1$ than in the corresponding cases of complete independence between X and Y. More importantly, however, the critical value of $\hat{\phi}_2$ also tends to get larger as c (and consequently, the strength of dependency) gets larger. (Compare the third and the sixth columns of Table 6 against the fourth column in the second and the third row blocks of Table 4, respectively, which give the results for $c = 0$.) This increasing trend levels off beyond certain values of c, more slowly for smaller n, and more quickly for larger n. The distribution of $\hat{\phi}_3$, on the other hand, is much less affected by the non-null rank-one dependency between Y and X. This study indicates that non-null values of $\rho_I(Y, t_1)$ do affect the distributions of subsequent sample redundancy indices, and that the assumption of complete independence among variables in traditional parallel analysis is at best controversial in testing the significance of non-dominant components. Table 7 gives the power ($h = 1$), and the probability of Type I error ($h = 2, 3$) of the Lazraq-Cléroux procedure in the non-null cases.

Insert Tables 6 and 7 about here

Conclusion

It is clear from the above that the Lazraq-Cléroux procedure is ill conceived and cannot be recommended for use. Monte-Carlo studies indicate how serious the problem is. One natural question to ask is: What alternative procedures are there?

It seems difficult to obtain general distributional results on $\hat{\phi}_h$ from the kind of "parallel" analysis employed in the present study, since it presupposes a complete specification of Σ_{YY}. Although this matrix can be taken as a diagonal matrix without loss of generality (so that only its diagonal elements have to be specified), quite a bit of freedom is still left to one's discretion. The choice is non-trivial, and it is almost impossible to cover the range of all possible variance-covariance structures of interest. Of course, in the specific contexts of redundancy analysis where Σ_{YY} is known reasonably accurately, parallel analysis is always feasible and presents an attractive procedure for testing the significance of the most dominant redundancy component. As has been noted earlier, it is extremely difficult to develop a general-purpose procedure for testing non-dominant components based on parallel analysis.

There is a well established procedure for dimensionality selection in redundancy analysis when parameters in the model are estimated by the maximum likelihood method (Reinsel and Velu, 1998) based on the multivariate normality assumption on Y, but with fixed X (as in the standard multiple regression analysis). This case is equivalent to canonical correlation analysis under the assumption of joint multivariate normality on Y and X. The number of significant components in redundancy analysis can be tested by the same procedure used in canonical correlation analysis to determine the number of significant canonical variates. The likelihood ratio tests (Wilks' λ test and Bartlett's (1951) approximation thereof) are available for testing the significance of the difference between the model with specific dimensionality (the number of components) and the saturated model. The logic underlying this procedure is described in detail by Reinsel and Velu (1998). The significance of non-dominant components can also be tested sequentially using a step-down technique (Bartlett, 1951). These tests are also based on an asymptotic rationale. However, they avoid the "singularity" problem (like the one encountered by the Lazraq-Cléroux procedure) by forming a likelihood ratio between the model of specific dimensionality (H_0: $\rho_I(Y, t_h) = 0$) and the saturated model (H_1: Σ_{XY} unconstrained). (One gets into the singularity problem, if one forms the likelihood ratio between H_0: $\rho_I(Y, t_h) = 0$ and H_1: $\rho_I(Y, t_h) \neq 0$ (Takane, et al., 2003).)

A permutation test similar to the one used by Takane and Hwang (2002; see also Legendre and Legendre (1998) and ter Braak and Šmilauer (1998)) in generalized constrained canonical correlation analysis, can be easily adapted to redundancy analysis when the multivariate normality assumption fails. The permutation test can also be applied to successively test the significance of non-dominant redundancy components by eliminating the effects of previous components, and reapplying the permutation tests on the residual data matrices. Legendre and ter Braak (in preparation) have recently conducted extensive numerical studies that demonstrate the validity and usefulness of this approach. They have also succeeded in finding a correction formula for possible biases in testing non-dominant components based on residuals from the sample estimates of more dominant components. The permutation test presupposes a specific data set to be analyzed. It is straightforward to apply the method, however, once the specific data set is at hand.

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics, 22*, 327–351.

Bartlett, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Annals of Eugenics, 16*, 199–214.

Horn, J. L. (1965). A rationale and the test of the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

Lambert, Z. V., Wildt, A. R., and Durand, R. M. (1988). Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring interset association. *Psychological Bulletin, 104*, 282–289.

Lancaster, H. O. (1963). Canonical correlations and partitions of χ^2. *Quarterly Journal of Mathematics, 14*, 220–224.

Lazraq, A., and Cléroux, R. (1988). Un algorithme pas à pas de sélection de variables en régression linéaire multivariée [A stepwise variable selection algorithm in multivariate linear regression]. *Statistique et Analyse des donnes, 13*, 39–58.

Lazraq, A., and Cléroux, R. (2002). Testing the significance of the successive components in redundancy analysis. *Psychometrika, 67*, 411–419.

Legendre, P., and Legendre, L. (1998). *Numerical ecology.* Second English edition. Oxford: Elsevier.

Legendre, P., and ter Braak, C. J. F. (in preparation). Partial canonical analysis, variation partitioning, and tests of significance of eigenvalues in redundancy analysis.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya, A, 26*, 329–358.

Reinsel, G. C., and Velu, R. P. (1998). *Multivariate reduced-rank regression.* New

York: Springer.

Takane, Y., and Hwang, H. (2002). Generalized constrained canonical correlation analysis. *Multivariate Behavioral Research, 37*, 163–195.

Takane, Y., van der Heijden, P. G. M., and Browne, M. W. (2003). On likelihood ratio tests for dimensionality selection. In Higuchi, T., Iba, Y., and Ishiguro, M. (Eds.), *Proceedings of Science of Modeling: The 30^{t}h Anniversary Meeting of the Information Criterion (AIC)*, (pp. 348-349).

ter Braak, C. J. F., and Šmilauer, P. (1998). *CANOCO reference manual and user's guide to Canoco for windows: Software for canonical community ordination (version 4)*. Ithaka, N. Y.: Microcomputer Power.

van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical correlation analysis. *Psychometrika, 42*, 207–219.

Wilks, S. S. (1938). The large-sample distriution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics, 9*, 60–62.

Table 1.

Approximate (and exact) critical values, g_α, of $\hat{\phi}_1$ for $q = 1$ as functions of p, n, and α.

	$p = 2$	$p = 3$	$p = 4$	$p = 5$
	$\alpha = .05$	$\alpha = .01$	$\alpha = .10$	$\alpha = .05$
$n = 25$	.1443	.1899	.0908	.1063
	(.144)	(.195)	(.091)	(.107)
$n = 50$	.0668	.0911	.0439	.0512
	(.066)	(.090)	(.043)	(.050)
$n = 75$	.0438	.0564	.0282	.0330
	(.043)	(.058)	(.028)	(.033)
$n = 100$	.0317	.0433	.0213	.0240
	(.032)	(.043)	(.021)	(.024)
$n = 200$	.0158	.0211	.0102	.0120
	(.016)	(.021)	(.010)	(.012)

TABLE 2.

Approximate critical values, $g_{.05}$, of $\hat{\phi}_1$ as functions of p, q, and n.

	$q = 1$	$q = 2$	$q = 5$	$q = 10$	$q = 20$
			$n = 75$		
$p = 2$	.0434	.0653	.1158	.1976	.3794
$p = 5$	.0322	.0448	.0712	.1102	.1875
			$n = 200$		
$p = 2$	.0153	.0233	.0408	.0670	.1178
$p = 5$	.0117	.0164	.0259	.0389	.0634
			$n = 5000$		
$p = 2$	.0006	.0009	.0016	.0025	.0043
$p = 5$	.0005	.0006	.0010	.0015	.0024

Table 3.
Probabilities of Type I error by the Lazraq-Cléroux test for the first redundancy component as functions of p, q, and n. The "*" indicates that the critical value is not exact, but is a numerical approximation obtained by the Monte Carlo study reported in Table 2.

	$q = 1$	$q = 2$	$q = 5$	$q = 10$	$q = 20$	"exact" $g_{.05}$
			$n = 75$			
$p = 2$	.0515	.1490	.6240	.9816	1.0	.043
$p = 5$	.0500	.1500	.6727	.9930	1.0	.033
			$n = 200$			
$p = 2$	.0498	.1469	.6062	.9748	1.0	.016
$p = 5$	.0504	.1532	.6660	.9917	1.0	.012
			$n = 5000$			
$p = 2$	.0500	.1434	.6051	.9733	1.0	.00063*
$p = 5$	.0500	.1504	.6620	.9819	1.0	.00047*

Table 4.

Approximate critical values, $g_{.05}$, of $\phi_h (h = 1, \ldots, s)$ as functions of p, q, and n.

	$p=2, n=75$		$p=2, n=200$		$p=2, n=5000$	
	$h=1$	$h=2$	$h=1$	$h=2$	$h=1$	$h=2$
$q=2$	.0628	.0104	.0238	.0038	.0009	.0001
$q=5$	.1162	.0323	.0416	.0121	.0016	.0005
$q=10$	.1964	.0694	.0679	.0248	.0025	.0010
$q=20$	.3801	.1417	.1187	.0483	.0043	.0019
		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
			$p=5$, $n=75$			
$q=2$		.0440	.0136			
$q=5$		.0717	.0323	.0147	.0055	.0012
$q=10$		.1111	.0580	.0323	.0167	.0072
$q=20$		.1880	.1079	.0667	.0398	.0201
			$p=5$, $n=200$			
$q=2$		.0163	.0053			
$q=5$		.0255	.0117	.0054	.0021	.0004
$q=10$		.0388	.0207	.0117	.0062	.0027
$q=20$		.0636	.0383	.0240	.0143	.0073
			$p=5$, $n=5000$			
$q=2$		.0006	.0002			
$q=5$		.0010	.0005	.0002	.0001	.0000
$q=10$		.0015	.0008	.0005	.0002	.0001
$q=20$		.0024	.0015	.0009	.0006	.0003

Table 5.

Probabilites of Type I error by the Lazraq-Cléroux test for non-dominant redundancy components as functions of p, q, and n.

		$p=2, n=75$		$p=2, n=200$	
		$g_{.05}=.043$		$g_{.05}=.016$	
		$h=1$	$h=2$	$h=1$	$h=2$
$q=2$		.1560	.0001	.1429	.0001
$q=5$		.6285	.0144	.5884	.0119
$q=10$		.9816	.3126	.9741	.2974
$q=20$		1.000	.9648	1.000	.9567

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
$p=5,\ n=75,\ g_{.05}=.033$					
$q=2$	.1531	.0007			
$q=5$	.6685	.0436	0.0	0.0	0.0
$q=10$	.9922	.6296	.0409	0.0	0.0
$q=20$	1.0	1.0	.9134	.1837	.0001
$p=5,\ n=200,\ g_{.05}=.012$					
$q=2$	.1539	.0004			
$q=5$	.6648	.0429	.0001	0.0	0.0
$q=10$	.9912	.6188	.0426	.0001	0.0
$q=20$	1.0	.9998	.9124	.1818	.0001

Table 6.

Approximate critical values, $g_{.05}$, of $\phi_h(h = 1, \ldots, 3)$ when the first component is non-null ($p = 5$).

	$n = 75$			$n = 200$		
	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
			$c = .05$			
$q = 2$	.0634	.0169		.0325	.0072	
$q = 5$	.1093	.0378	.0161	.0589	.0154	.0063
$q = 10$	.1698	.0669	.0347	.0970	.0264	.0132
$q = 20$	.2769	.1217	.0707	.1596	.0449	.0261
			$c = .1$			
$q = 2$	.1127	.0207		.0752	.0082	
$q = 5$	.1987	.0435	.0171	.1406	.0167	.0065
$q = 10$	.3013	.0740	.0363	.2208	.0271	.0134
$q = 20$	.4341	.1303	.0723	.3261	.0466	.0262
			$c = .2$			
$q = 2$	.2634	.0221		.2075	.0086	
$q = 5$	.4106	.0457	.0174	.3376	.0168	.0065
$q = 10$	.5223	.0748	.0363	.4350	.0275	.0134
$q = 20$	.6090	.1322	.0732	.5087	.0462	.0263
			$c = .5$			
$q = 2$	.6196	.0231		.5360	.0084	
$q = 5$	.6631	.0464	.0177	.5769	.0168	.0066
$q = 10$	.6830	.0753	.0369	.5925	.0273	.0135
$q = 20$	.6927	.1320	.0732	.5984	.0465	.0265

Table 7.

Powers ($h = 1$) and actual significance levels ($h = 2, 3$) of the Lazraq-Cléroux test when the first component is non-null ($p = 5$).

	$n = 75$, $g_{.05} = .033$			$n = 200$, $g_{.05} = .012$		
	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
$c = .05$						
$q = 2$	.3559	.0014		.0465	0.0	
$q = 5$	.8850	.1002	0.0	.5077	.0001	0.0
$q = 10$	.9999	.7537	.0712	.9691	.0088	0.0
$q = 20$	1.0	.9997	.9300	1.0	.3723	.0037
$c = .1$						
$q = 2$	.7945	.0051		.7528	0.0	
$q = 5$	.9975	.1707	.0004	.9996	0.0	0.0
$q = 10$	1.0	.7990	.0870	1.0	.0113	0.0
$q = 20$	1.0	1.0	.9365	1.0	.3816	.0035
$c = .2$						
$q = 2$	.9997	.0085		1.0	0.0	
$q = 5$	1.0	.1938	.0008	1.0	.0002	0.0
$q = 10$	1.0	.8139	.0972	1.0	.0124	0.0
$q = 20$	1.0	.9999	.9388	1.0	.3980	.0037
$c = .5$						
$q = 2$	1.0	.0139		1.0	0.0	
$q = 5$	1.0	.1919	.0010	1.0	.0002	0.0
$q = 10$	1.0	.8057	.0978	1.0	.0109	0.0
$q = 20$	1.0	1.0	.9471	1.0	.3998	.0040