

5-22-2020

On Statistical Significance of Discriminant Function Coefficients

Tolulope T. Sajobi

University of Calgary, tolu.sajobi@ucalgary.ca

Gordon H. Fick

University of Calgary, ghfick@ucalgary.ca

Lisa M. Lix

University of Manitoba, lisa.lix@umanitoba.ca

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Sajobi, Tolulope T.; Fick, Gordon H.; and Lix, Lisa M. (2020) "On Statistical Significance of Discriminant Function Coefficients," *Journal of Modern Applied Statistical Methods*: Vol. 18 : Iss. 1 , Article 27.

DOI: [10.22237/jmasm/1571745720](https://doi.org/10.22237/jmasm/1571745720)

Available at: <https://digitalcommons.wayne.edu/jmasm/vol18/iss1/27>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

On Statistical Significance of Discriminant Function Coefficients

Cover Page Footnote

This research was supported by the National Science and Engineering Research Council Discovery Grant to the first author.

On Statistical Significance of Discriminant Function Coefficients

Tolulope T. Sajobi
University of Calgary,
Calgary, Alberta, Canada

Gordon H. Fick
University of Calgary,
Calgary, Alberta, Canada

Lisa M. Lix
University of Manitoba,
Winnipeg, Manitoba, Canada

Discriminant function coefficients are useful for describing group differences and identifying variables that distinguish between groups. Test procedures were compared based on asymptotically approximations, empirical, and exact distributions for testing hypotheses about discriminant function coefficients. These tests are useful for assessing variable importance in multivariate group designs.

Keywords: discriminant function coefficients, relative importance, multivariate group designs, statistical inference

Introduction

Linear discriminant analysis (DA; Fisher, 1936) is a multivariate procedure for predicting group membership (predictive discriminant analysis; PDA) and describing group separation (descriptive discriminant analysis; DDA). PDA focuses on the development of efficient predictive models with minimal misclassification error and is used in many fields including biomedical, behavioral, and engineering sciences for predicting group membership (McLachlan, 1992; Sherry, 2006; Onur, et al., 2007). DDA focuses on describing group differences and identifying variables that maximize group separation. In DDA, measures of discriminant function coefficients are used to rank order a set of variables according to their contribution to group separation (Rencher, 2002).

The concept of variable importance in multivariate group designs was studied using DDA and multivariate analysis of variance. (Huberty & Wisenbaker, 1992; Huberty & Olejnik, 2006; Sherry, 2006). Huberty and Wisenbaker (1992) identified three major perspectives for defining relative importance of outcome variables in multivariate group designs. These include (a) relative contribution of a variable to a latent

construct, (b) relative contribution of variable to linear discriminant function scores, and (c) relative contribution of a variable to a grouping variable. Several relative importance measures have been developed, based on these perspectives, for evaluating the relative importance of variables in multivariate group designs.

Huberty and Smith (1982) and Huberty (1984) suggested an index originally derived by Urbakh (1971) for evaluating relative importance of variables in two-group multivariate designs. It measures the decrease in the inter-group Mahalanobis distance due to the removal of each of the criterion variables in turn. It was shown to be equivalent to the F -to-remove statistic.

Huberty and Wisenbaker (1992) proposed the use of the magnitude of standardized discriminant function coefficient associated with each variable as a measure of its relative contribution to the linear discriminant function score. The rationale for using the standardized weight as a variable-ordering index is that a large weight, in an absolute value sense, indicates that the associated variable contributes to the linear discriminant function score for each experimental unit. However, this measure was criticized for its instability and large standard errors especially in high correlated multivariate outcomes.

The structure coefficient, which is the within-groups correlations of the outcome variables with the linear discriminant function scores, was also recommended for assessing the relative importance of variables (Bray & Maxwell, 1985; Huberty & Morris, 1989). Its use has been limited by criticisms about its appropriateness for evaluating the relative importance of outcome variables. Bargmann (1970) showed that structure coefficients are equivalent to univariate F -statistic, which ignores the correlation with other outcome variables, and is not suitable for evaluating the relative importance of outcome variables in multivariate group designs. Thomas (1992) proposed discriminant ratio coefficient, which can be estimated as the product of the corresponding standardized discriminant function coefficients and the structure coefficients, for evaluating the relative importance of outcome variables. The discriminant ratio coefficient is the multivariate equivalent of the Pratt index, a measure of relative importance for linear regression earlier developed by Pratt (1987). However, the discriminant ratio coefficient is sensitive to strong between-variable correlation and suppression effects.

Despite these developments, the existing measures of variable importance in multivariate group designs may not always result in similar conclusions about the relative importance of variables in discriminating between independent groups. These measures are based on different statistical models with different motivations and goals. For example, for measures of relative importance based on discriminant analysis analysis, a variable's importance is interpreted as its contribution to

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

grouping variable effects and discriminant function scores (Huberty & Wisenbaker, 1992; Thomas & Zumbo, 1996). However, measures based on multivariate analysis of variance evaluate a variable's importance in terms of the additional information that it contributes to group discrimination (Thomas, 1997).

Also, these measures are not equally sensitive to different data-analytic conditions. The discriminant ratio coefficient, for example, is sensitive to strong variable correlations while the *F*-to-remove statistic is sensitive to the number of study variables (Thomas & Zumbo, 1996). As a result, they may not always result in consistent rankings of variables. Hence, no single measure of relative importance has been recommended for use and there may be confusion amongst researchers about which measure to adopt in practice.

The rank ordering of variables according to their relative importance is typically based on sample estimates of the relative importance measure. However, the rank ordering of variables according to sample estimates of relative importance indices might not correspond to population-level importance of these variables (Thomas, 1992). Therefore, stating that one variable is more important than the other, based on sample estimates of relative importance indices, might be misleading. Nevertheless, there are rules of thumb and descriptive analyses of relative importance being used to evaluate if the magnitude of the relative importance measure for a variable is large enough to ascertain its importance. For example, Thomas (1992) suggested variables with discriminant ratio coefficient values greater than $1/2p$, p being the number of outcome variables, can be considered important while variables with negative DRC values are suggested to be suppressor variables with no direct contribution to group separation. Dagleish (1994) suggested variables with structure coefficients above 0.4 should be considered as important. But these guidelines ignore error variation in estimates of relative importance measures and are sensitive to sample size and the number of variables in the study

More formal parametric and non-parametric test procedures were developed to test the statistical significance of discriminant function coefficients. These procedures are based on asymptotic approximations to a distribution, empirical distributions, or exact distributions. Das Gupta (1968) proposed an asymptotically normal *z*-test. Rao (1970) proposed an *F*-test based on the Mahalanobis distance (Timms, 2002). Computationally-intensive resampling-based methods such as bootstrap have been employed to approximate the distribution of differences in discriminant function coefficients in small samples (Huberty & Wisenbaker, 1992). Dagleish (1994) proposed the bootstrap method to test the statistical significance of a structure coefficient. Huberty and Wisenbaker (1992) proposed the use of bootstrap for estimating the variability in the rank ordering of the relative importance of the *F*-to-remove values

for the outcome variables. More recently, Bodnar and Okhrin (2011) developed a test statistic for the discriminant function coefficient vector based on t distribution.

Despite the availability of these tests, they have not been used for to evaluate the relative importance of correlated outcome variables in multivariate group designs. In addition, there have been no comparisons of the performance of these test procedures when their underlying derivational assumptions were satisfied or violated. For example, the assumption of normality of outcome variables may not be tenable in many research studies such as health-related quality of life studies where patient-reported outcomes frequently exhibit heavy tails and/or skewed distributions (Beaumont, Lix, Hahn, Yost, 2006). Choosing an appropriate test procedure for evaluating statistical significance of the discriminant function coefficients when describing the relative importance of variables in multivariate group designs is not straightforward. There are no guidelines about the choice of a test procedure that controls the family-wise error rate (FWER) while affording maximum power to correctly determine the relative importance of variables when the outcome variables are non-normal and group covariances are heterogeneous.

The purpose of this study is to propose these tests can be used as supplementary information to aid the assessment of the relative importance of variables in multivariate group designs. Specifically, the Type I error and statistical power of the four test procedures are compared to evaluate statistical significance of discriminant function coefficients for assessing relative importance of variables under a variety of data analytic conditions.

Description of Methods

Let \mathbf{Y}_{ij} be the $p \times 1$ vector of random variables estimated by \mathbf{y}_{ij} for the i th study participant ($i = 1, \dots, n_j$) in the j th group ($j = 1, 2; n = n_1 + n_2$) and assume $\mathbf{Y}_{ij} \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ are the j th group mean vector and covariance matrix, respectively. The focus is on the two-group case, although all of the procedures can be generalized to multi-group designs. For the linear DA procedure, the discriminant function coefficients vector is

$$\mathbf{a} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (1)$$

where $\boldsymbol{\mu}_j$ is estimated by

$$\hat{\boldsymbol{\mu}}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \mathbf{y}_{ij},$$

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

and Σ is the pooled variance-covariance matrix and is estimated by

$$\hat{\Sigma} = \frac{(n_1 - 1)\hat{\Sigma}_1 + (n_2 - 1)\hat{\Sigma}_2}{n - 2} \quad (2)$$

Let $\mathbf{H}_0 = \{H_{0k}\}_{k=1}^p$ and $\mathbf{H}_A = \{H_{Ak}\}_{k=1}^p$ denote the family of null and alternative hypotheses for p outcome variables, respectively, where

$$\mathbf{H}_0 : \mathbf{a} = \mathbf{0} \quad \text{vs.} \quad \mathbf{H}_A : \mathbf{a} \neq \mathbf{0},$$

\mathbf{a} is vector of discriminant function coefficients defined in equation (1), and $\mathbf{0}$ is the null vector. Let $H_{0k} : a_k = 0$ ($k = 1, \dots, p$), and the corresponding alternative hypotheses $H_{Ak} : a_k \neq 0$, where a_k is the population standardized discriminant function coefficients for the k th outcome variable. We consider a family of p hypotheses about these coefficients for p outcomes. The outcome specific null and alternative hypotheses are defined as

$$H_{0k} : \mathbf{C}_k^T \mathbf{a} = 0 \quad \text{vs.} \quad H_{Ak} : \mathbf{C}_k^T \mathbf{a} \neq 0$$

respectively, where \mathbf{C}_k is a $p \times 1$ vector defining the hypothesis to be tested.

Das Gupta (1968) proposed an asymptotic test procedure for evaluating statistical significance of discriminant function coefficients. The test statistic for the k th discriminant function coefficients is

$$z_{k*}^* = \frac{n^{1/2} \mathbf{C}_k^T \mathbf{a}}{\mathbf{C}_k^T \mathbf{M} \mathbf{C}_k} \quad (3)$$

where

$$\mathbf{M} = \Psi_1 + \frac{n(n-1)}{n_1 n_2} \hat{\Sigma}^{-1} + \frac{n-p+1}{n-p-1} \Psi_2,$$

$$\Psi_1 = q(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^T \hat{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2) \hat{\Sigma}^{-1}, \quad (4)$$

$$\Psi_2 = \hat{\Sigma}^{-1} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2) (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^T \hat{\Sigma}^{-1}, \quad (5)$$

and

$$q = \frac{(n-1)^2}{(n-p)(n-p-1)(n-p-3)}. \quad (6)$$

Statistical significance of the k th variable's discriminant function coefficient is evaluated by comparing z_k^* to the $100(1 - \alpha/2)$ percentile of the normal distribution.

Rao (1970) suggested the importance of a variable is directly related to its contribution to the grouping effect, and proposed a test procedure based on the Mahalanobis distance to examine statistical significance of a variable's discriminant function coefficient,

$$F_k^* = \frac{k_3(n-p)(\delta^2 - \delta_{-k}^2)}{(n-p-1) + k_3\delta_{-k}^2}, \quad (7)$$

where

$$\delta^2 = (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^T \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2), \quad (8)$$

$k_3 = n/n_1n_2$, and δ_{-k}^2 is the value of δ^2 when the k th variable is omitted. Statistical significance of the k th variable's discriminant function coefficient is evaluated by comparing F_k^* to $F(1, n, p)$.

Bodnar and Okhrin (2011) derived the multivariate distribution of the discriminant function coefficients vector for two independent groups and proposed a t statistic for evaluating the statistical significance of sample discriminant function coefficients. This statistic is

$$t_k^* = \frac{\sqrt{n-p+1} \mathbf{C}_k \hat{\mathbf{a}}}{\sqrt{\mathbf{C}_k^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{C}_k} \sqrt{1 + (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^T \mathbf{R}_k (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)}}, \quad (9)$$

where $\mathbf{R}_k = \hat{\boldsymbol{\Sigma}}^{-1} - \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{C}_k^T \mathbf{C}_k \hat{\boldsymbol{\Sigma}}^{-1} / \mathbf{C}_k^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{C}_k$. Statistical significance of the k th variable's discriminant function coefficient is evaluated by comparing t_k^* to the $100(1 - \alpha/2)$ percentile of the t distribution.

Resampling-based methods, such as the bootstrap, were employed for approximating the distribution of a test statistic when the exact distribution is not readily available. Under this method, the original data in each group are randomly sampled with replacement B times. The discriminant function coefficients for the p outcomes are computed based on the bootstrap data. Statistical significance of each variable's discriminant function coefficient is obtained by comparing the B bootstrapped coefficients to the discriminant function coefficients estimated from the original data and estimating the proportion of discriminant function coefficients outside the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of the distribution. While $B = 500$ bootstrap samples have been shown to produce acceptable results for hypothesis testing, precision will increase as B increases (Mackinnon, 2006).

Simulation Study

A Monte Carlo study was used to evaluate the properties of test procedures for evaluating statistical significance of discriminant function coefficients, including: (a) Gupta's (1968) test (GUP), (b) Rao's (1970) F -test (RAO), (c) bootstrap test (BOT), and (d) Bodnar-Okhrin's (2011) test (BOD). All the procedures were used to test for statistical significance of discriminant function coefficients from a p -variate two-group design.

The simulation conditions investigated include: (a) population distribution, (b) number of outcome variables, (c) total sample size and equality/inequality of group sizes, (d) magnitude and pattern of variable correlations, (e) mean configuration, and (f) group covariance heterogeneity. Based on previous research, the number of outcome variables was set at $p = 4$ and 7 , representing small to moderate sets of outcome variables. Previous studies have investigated numbers of outcome variables ranging from $p = 4$ to 10 (Lebreton, Polyhart, & Ladd, 2004). Total sample sizes of $n = 60$, 140 , and 200 were investigated. Although previous simulation studies on discriminant function coefficients have primarily focused on equal group sizes (Roy et al., 2005a, 2005b; Finch & Laking, 2008), unequal group sizes have also been shown to influence their size (Baron, 1991), and may therefore influence tests of statistical significance. Both equal and unequal group size conditions were investigated. For $n = 60$, the unequal group size condition was $(n_1, n_2) = (24, 36)$. For $n = 140$, $(n_1, n_2) = (56, 84)$ was investigated, and for $n = 200$, $(n_1, n_2) = (80, 120)$. These group size conditions were chosen based on previous research (Baron, 1991; Sajobi, Lix, Laverty, & Li, 2011). For the unequal group size cases, the coefficient of group size variation (Lix & Fouladi, 2007) was 0.2 .

Discriminant function coefficients are sensitive to the magnitude of correlation among the outcome variables (Thomas & Zumbo, 1996). Therefore, three correlation structures and three magnitudes of correlation (ρ) were investigated: (a) \mathbf{Q}_1 : compound symmetric (CS) structure with $\rho = 0.3$; (b) \mathbf{Q}_2 : CS structure with $\rho = 0.6$, (c) \mathbf{Q}_3 : unstructured correlation matrix with average off diagonal elements of 0.3 , (e) \mathbf{Q}_4 : unstructured correlation matrix with average off-diagonal elements of 0.7 , and (f) \mathbf{Q}_5 simplex correlation matrix with off-diagonal elements of 0.3 and 0.7 . These structures were investigated with variable correlations ranging between 0.1 and 0.8 (Sajobi, Lix, Laverty & Li, 2011; Tonidandel, Lebreton, & Johnson, 2009). Details about these types of correlation structures are provided in the Appendix.

Conditions were investigated in which the group covariances were equal and unequal. In the latter case, both moderate and large degrees of covariance inequality were examined. In the moderate inequality case, the group covariances were in the

ratio of 1:3. In the large covariance inequality case, the group covariances were in a ratio of 1:7. Positive and negative pairings of group sizes and covariance matrices were investigated, in addition to cases in which equal group sizes were paired with equal or unequal group covariances. A positive pairing refers to the case in which the largest n_j is associated with the covariance matrix containing the largest element values; a negative pairing refers to the case in which the largest n_j is associated with the covariance matrix with the smallest element values. The pairing of group sizes and covariances is known to influence Type I error performance of univariate and multivariate procedures that assume covariance homogeneity (Harwell, Rubinstein, Hayes, & Olds, 1992).

Both null and non-null mean configurations were considered. These configurations differed in the magnitude and pattern of the differences between the groups. The four non-null configurations (i.e., Configurations II through IV in Table 1) were chosen to represent small, moderate, and large magnitudes of group differences.

Pseudo-random observation vectors \mathbf{y}_{ij} from a multivariate normal distribution with mean vector $\boldsymbol{\mu}_j$ and correlation matrix $\mathbf{Q}_{s_j} = \mathbf{Q}_s$ ($s = 1, 2, \dots, 5$) were generated using the RANNOR function in SAS (SAS Institute Inc., 2011). A vector of p standard normal deviates, \mathbf{R}_{ij} , was transformed to a vector of multivariate observations via

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \mathbf{L}\mathbf{R}_{ij}. \tag{10}$$

The Cholesky decomposition was used to obtain \mathbf{L} , an upper triangular matrix of dimension m satisfying the equality $\mathbf{L}^T\mathbf{L} = \mathbf{Q}_j$. Then \mathbf{y}_{ij} was multiplied by \mathbf{V}_j ,

Table 1. Values of $\boldsymbol{\mu}_1$ Selected for the Monte Carlo Study

# Variables	Configuration	$\boldsymbol{\mu}_1$	D^2
4	I	(0 0 0 0)	0.0
	II	(1 1 1 1)	4
	III	(1 0.5 1 0.5)	2.5
	IV	(2 2 2 2)	16.0
	V	(2 1 2 1)	10.0
7	I	(0 0 0 0 0 0 0)	0.0
	II	(1 1 1 1 1 1 1)	7.0
	III	(1 0.5 1 0.5 1 0.5 1)	4.65
	IV	(2 2 2 2 2 2 2)	28
	V	(2 1 2 1 2 1 2)	20

Note: For all configurations, $\boldsymbol{\mu}_2$ is the null vector; D^2 = Euclidean distance between population mean vector

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

a diagonal matrix with elements σ , to obtain multivariate observations with the desired variances and covariances such that $\Sigma_j = \mathbf{V}_j \boldsymbol{\rho}_j \mathbf{V}_j^T$, where $\boldsymbol{\rho}_j$ is the correlation matrix for the j th group. For both groups, $\sigma^2 = 1$ for all investigated conditions. To assess departures from multivariate normality, the data were generated from skewed, heavy-tailed, and platykurtic multivariate distributions. Two skewed distributions were instigated: skewness (γ_1) and kurtosis (γ_2) values of $\gamma_1 = 1.8$ and $\gamma_2 = 5.9$ (SK-I) and $\gamma_1 = 13.2$ and $\gamma_2 = 42892.9$ (SK-II); for a normal distribution, $\gamma_1 = 0$ and $\gamma_2 = 0$. Two heavy-tailed distributions were investigated as well. One had skewness (γ_1) and kurtosis (γ_2) values of $\gamma_1 = 0$ and $\gamma_2 = 33$. The second was similar to a heavy-tailed Cauchy distribution, for which γ_1 and γ_2 are undefined.

A platykurtic distribution (PK) with $\gamma_1 = 5.9$ and $\gamma_2 = -4.0$ was also investigated. Field and Genton (2006) described a flexible family of multivariate non-normal distributions obtained by modifying their quantiles. The variables g and h , which control the magnitude of γ_1 and γ_2 , are used to transform a standard normal random variate, C , as follows:

$$X_{ijk} = \left(\frac{\exp(gR_{ijk}) - 1}{g} \right) \left(\exp \frac{h}{2} R_{ijk}^2 \right) \quad (11)$$

When $g = 0$, this equation reduces to

$$X_{ijk} = \left(\exp \frac{h}{2} R_{ijk}^2 \right) R_{ijk}. \quad (12)$$

A total of 5000 simulations were conducted for each combination of conditions; there were a total of 900 combination of conditions for the null means, and 3600 combination of conditions for the non-null mean configurations. For the BOT procedure, $B = 1000$ bootstrap samples were selected.

Empirical family-wise error rates (FWER) were obtained when the mean vectors for both groups were null (i.e., Configuration I). The FWERs were evaluated with Bradley's (1978) liberal criteria; FWERs outside the 0.05 ± 0.025 interval were considered as non-robust at $\alpha = .05$. Procedures with FWER below the lower bound of the Bradley's (1978) liberal criteria are considered as conservative in controlling the FWER, while those above the upper bound are considered liberal. For each condition, a Bonferroni correction was used to adjust for multiple correlated tests of significance (Dunn, 1961). For non-null mean configurations, the any-variable power rate, the power of a test procedure to reject at least one non-null hypothesis in the family of p non-null hypotheses, was used to assess the performance of the investigated procedures. Differences in statistical power less than 5% were considered negligible, while differences greater than 10% were considered substantial

(Lix, Deering, Fouladi, & Manivong, 2009). All analyses were conducted in SAS/IML (SAS Institute Inc., 2011).

Results

FWER

Results are presented for conditions when the data were sampled from multivariate normal distributions and group covariances were equal (Table 2) and also for conditions where the data were sampled from multivariate non-normal distributions and group covariances were unequal (Table 3). Overall, both Bodnar and Rao's *F*-test

Table 2. Mean Family-Wise Error Rates (%) of Discriminant Function Coefficient Test Procedures by Number of Outcome Variables (p), Total Sample Size (n), and Correlation Structures for the Multivariate Normal Distribution and Equal Group Covariances

# Outcome Variables	Total Sample	Correlation (ρ)	GUP	RAO	BOT	BOD
4	60	\mathbf{Q}_1	0.98	3.90	1.80	4.52
		\mathbf{Q}_2	0.94	3.88	1.60	4.36
		\mathbf{Q}_3	0.88	3.78	1.74	4.44
		\mathbf{Q}_4	0.90	4.10	1.72	4.38
		\mathbf{Q}_5	0.94	4.12	1.82	4.58
	200	\mathbf{Q}_1	3.98	5.08	4.56	5.22
		\mathbf{Q}_2	3.86	5.06	4.30	5.22
		\mathbf{Q}_3	3.58	4.84	4.42	5.02
		\mathbf{Q}_4	3.46	4.44	4.08	4.62
		\mathbf{Q}_5	3.16	4.10	3.80	4.36
7	60	\mathbf{Q}_1	0.28	4.12	0.86	4.72
		\mathbf{Q}_2	0.26	3.92	0.84	4.52
		\mathbf{Q}_3	0.32	4.04	0.92	4.60
		\mathbf{Q}_4	0.22	3.94	0.78	4.64
		\mathbf{Q}_5	0.18	3.82	0.76	4.28
	200	\mathbf{Q}_1	3.00	4.54	3.72	4.68
		\mathbf{Q}_2	2.90	4.65	3.58	4.79
		\mathbf{Q}_3	3.02	4.60	3.82	4.72
		\mathbf{Q}_4	2.97	4.58	3.60	4.74
		\mathbf{Q}_5	2.68	4.42	3.24	4.58

NB: GUP = Gupta's (1968) procedure; RAO = Rao's (1970) procedure; BOT = Bootstrap procedure; BOD = Bodnar-Okhrin's (2011) procedure. ρ = variable correlation; \mathbf{Q}_1 : compound symmetric (CS) structure with $\rho = 0.3$; (b) \mathbf{Q}_2 : CS structure with $\rho = 0.6$, (c) \mathbf{Q}_3 : unstructured correlation matrix with average off diagonal elements of 0.3, (e) \mathbf{Q}_4 : unstructured correlation matrix with average off-diagonal elements of 0.7, and (f) \mathbf{Q}_5 simplex correlation matrix with off-diagonal elements of 0.3 and 0.7;

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

Table 3. Mean Family-Wise Error Rates (%) of Discriminant Function Coefficient Test Procedures by Population Distribution and Covariance/Sample Size Pairing when $p = 4$

Distribution	Group Size/ Covariance Pairing	GUP	RAO	BOT	BOD
Normal	Equal Group Size/Equal Covariances	2.49	4.49	3.16	4.81
	Equal Group Size/Unequal Covariances	2.75	4.91	2.65	5.21
	Negative Pairing	6.60	10.44	0.77	10.97
	Positive Pairing	0.93	1.92	0.62	2.08
HT-I	Equal Group Size/Equal Covariances	2.58	4.08	1.96	4.37
	Equal Group Size/Unequal Covariances	2.87	4.40	2.08	4.62
	Negative Pairing	6.04	8.93	0.27	9.37
	Positive Pairing	1.06	1.93	0.38	2.05
HT-II	Equal Group Size/Equal Covariances	1.46	2.47	0.21	2.65
	Equal Group Size/Unequal Covariances	1.63	3.15	0.19	3.40
	Negative Pairing	3.82	6.46	0.02	6.96
	Positive Pairing	0.66	1.35	0.00	1.49
SK-I	Equal Group Size/Equal Covariances	2.21	4.04	1.98	4.34
	Equal Group Size/Unequal Covariances	5.24	8.24	3.02	8.68
	Negative Pairing	10.15	14.78	0.43	15.41
	Positive Pairing	1.99	3.56	0.66	3.82
SK-II	Equal Group Size/Equal Covariances	1.83	3.47	0.76	3.70
	Equal Group Size/Unequal Covariances	3.02	5.37	0.77	5.71
	Negative Pairing	6.74	10.60	0.08	11.16
	Positive Pairing	1.11	2.16	0.15	2.35
PK	Equal Group Size/Equal Covariances	0.17	18.00	0.48	1.67
	Equal Group Size/Unequal Covariances	0.46	18.20	0.48	2.60
	Negative Pairing	1.11	17.50	0.15	5.69
	Positive Pairing	0.20	20.12	0.14	1.42

NB: GUP = Gupta's (1968) procedure; RAO = Rao's (1970) procedure; BOT = Bootstrap procedure; BOD = Bodnar-Okhrin's (2011) procedure. Normal distribution has skewness (γ_1) = 0 and kurtosis (γ_2) = 0; HT-I has $\gamma_1 = 0$ and $\gamma_2 = 33$; HT-II has $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-I has $\gamma_1 = 1.8$ and $\gamma_2 = 5.9$; SK-II has $\gamma_1 = 13.2$ and $\gamma_2 = 42.892.9$; PK has $\gamma_1 = 5.9$ and $\gamma_2 = -4.0$.

controlled the FWER to α , whereas the Gupta's asymptotic and bootstrap procedures frequently resulted in conservative FWER, when the data were normally distributed. The average FWERs are summarized in [Table 2](#) by number of outcome variables (p) and correlation structure. When $n = 60$, Gupta's asymptotic and bootstrap procedures frequently resulted in average FWERs that were below the lower bound of Bradley's liberal criterion, whereas both Rao's F -test and Bodnar's procedures controlled the FWER to α , regardless of the number of outcome variables.

The average FWERs for Rao's (1970) F -test, bootstrap, and Bodnar's procedures were slightly higher when the data were sampled from a moderately correlated set of outcomes (i.e., \mathbf{Q}_1 or \mathbf{Q}_3) than when the data were sampled from a strongly correlated set of outcomes (i.e., \mathbf{Q}_2 or \mathbf{Q}_4), regardless of the correlation structure and the number of outcome variables. For example, the average FWERs for Rao's F -test and Bodnar's procedures were 3.90% and 4.52%, respectively, when the data were sampled from a population with \mathbf{Q}_1 correlation structure. The average FWER for both the former and latter procedures were 3.78% and 4.36%, respectively, when the data were sampled from a population with \mathbf{Q}_2 correlation structure.

The pairing of the group covariances and sample sizes and data distribution influenced the procedures' control of the FWER (Table 3). When the data were generated from a multivariate normal distribution and the group sizes were equal, both Rao (1970) F -test and Bodnar's test controlled the FWER to α . Although the FWERs for Gupta (1986) and bootstrap procedures were lower than those for Rao (1970) F -test and Bodnar's test, they were contained within the bounds of Bradley's (1978) liberal criterion. When the data were sampled from a skewed or heavy-tailed distribution, the FWER for Rao's (1970) F -test and Bodnar's procedure were lower, but still contained within the bounds of Bradley's liberal criterion. But when the data were sampled from a multivariate platykurtic distribution, the FWER for Rao (1970) test procedure exceeded the upper bounds of Bradley's liberal criterion while the average FWER for Bodnar's procedure was within the limits of Bradley's liberal criterion. In contrast, Gupta's (1986) and bootstrap test resulted in average FWERs that were below the bounds of Bradley's criterion, regardless of the population distribution.

When the group sizes and covariances were positively paired, the average FWERs for Gupta (1986) and bootstrap procedures were consistently below the lower bound of Bradley's (1978) liberal criterion, regardless of the population distribution. The average FWERs for Bodnar and Rao (1970) F -test procedures were consistently below the lower bound of Bradley's (1978) liberal criterion when the data were sampled from a population with multivariate heavy-tailed distribution or severely skewed distribution. However, for these same conditions, the FWER values for the latter procedure, exceeded the upper bound of Bradley's (1978) liberal criterion, whereas the FWER values for the former procedure was smaller than the lower bound of Bradley's (1978) liberal criterion when the data were sampled from a platykurtic (PK) distribution.

However, when the group sizes and covariance were negatively paired, the average FWER values for Rao (1970) F -test and Bodnar's procedures mostly exceeded the upper bound of Bradley's (1978) liberal criterion for normal, skewed, heavy-tailed, and platykurtic distributions. The FWER values for the bootstrap procedure were

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

consistently below the lower bound of Bradley's liberal criterion, regardless of the population distribution. The FWERs for Gupta (1986) procedure were slightly higher than α , but were within the bounds of Bradley's liberal criterion for the normal and heavy-tailed distributions.

Power Rates

The average any-variable power rates (percentages) for the procedures are reported in Tables 4 through 6. Table 4 and 5 describes the average any-variable power values of the test procedures for a variety of population distributions when group covariances were equal and unequal. For equal group covariance conditions (Table 4), the average any-variable power rates for all procedures were higher under the multivariate normal distribution and lowest under extremely skewed or heavy-tailed distributions, regardless of the configuration of the population means. When the data were sampled from a multivariate normal distribution with equal group covariances, the bootstrap procedure resulted in higher average any-variable power rates than any other procedure. However, this procedure was less powerful when the data were sampled from a multivariate non-normal distribution than when the data were sampled from a multivariate normal distribution.

Bodnar's procedure was most powerful among the investigated procedures, when the data were sampled from a heavy-tailed or extremely skewed distribution, regardless of the configuration of the population means. For these same conditions, there were negligible differences in the average any-variable power values of Bodnar and Rao (1970) *F*-test procedures. However, when the data were sampled from a platykurtic distribution, the latter procedure exhibited higher any-variable power than the former procedure, regardless of the population mean configuration. For example, the average any-variable power rates for Bodnar and Rao (1970) procedures were 89.8% and 89.3%, respectively, when the data were sampled from a multivariate normal population with mean configuration IV. The average power for both procedures were 53.8% and 52.2%, respectively, when the data were sampled from an extremely heavy-tailed distribution, and 71.2% and 99.9%, when the data were sampled from a platykurtic distribution for the same mean configuration.

When the data were sampled from a multivariate skewed or heavy-tailed population distributions and heterogeneous group covariances, Bodnar's procedure achieved the highest any-variable power. (Table 5). Nevertheless, there were negligible differences in the power rates for Bodnar's and Rao (1970) procedures in almost all the conditions investigated. But when the data were sampled from a platykurtic distribution, Rao (1970) *F*-test was most powerful among the investigated procedures.

Table 4. Mean Any-Variable Power (%) of Discriminant Function Coefficient Test Procedures by Population Distribution and Mean Configuration when Group Covariances are Equal

Distribution	Mean Configuration	GUP	RAO	BOT	BOD
Normal	II	62.27	70.10	73.76	74.96
	III	30.75	40.56	48.99	41.46
	IV	79.65	89.34	95.95	89.77
	V	48.89	60.93	86.32	61.72
HT-I	II	58.69	67.35	60.17	68.15
	III	30.36	38.62	36.83	39.49
	IV	88.66	93.69	92.88	94.03
	V	60.45	70.00	78.01	70.86
HT-II	II	14.69	22.20	6.42	22.98
	III	5.86	12.48	2.65	13.29
	IV	33.85	52.29	21.56	53.78
	V	20.08	37.76	12.65	39.19
SK-I	II	56.52	67.68	62.76	68.57
	III	27.20	36.66	36.69	37.53
	IV	78.57	88.41	91.41	88.86
	V	49.50	61.09	77.24	61.86
SK-II	II	38.15	48.86	32.29	49.84
	III	19.16	26.88	16.78	27.75
	IV	72.51	82.96	73.50	83.61
	V	47.15	58.23	54.04	59.05
PK	II	35.42	99.00	54.00	56.77
	III	36.95	91.67	41.42	54.64
	IV	69.56	99.96	87.23	81.22
	V	56.44	99.89	73.79	71.23

NB: GUP = Gupta's (1968) procedure; RAO = Rao's (1970) procedure; BOT = Bootstrap procedure; BOD = Bodnar-Okhrin's (2011) procedure. Normal distribution has skewness (γ_1) = 0 and kurtosis (γ_2) = 0; HT-I has $\gamma_1 = 0$ and $\gamma_2 = 33$; HT-II has $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-I has $\gamma_1 = 1.8$ and $\gamma_2 = 5.9$; SK-II has $\gamma_1 = 13.2$ and $\gamma_2 = 42.892.9$; PK has $\gamma_1 = 5.9$ and $\gamma_2 = -4.0$. See Table 1 for description of the mean configurations

For example, when the data were sampled from a multivariate normal distribution and unequal group covariances, the average power rates for Rao (1970) *F*-test and Bodnar procedures were 75.2% and 76.1%, respectively, when the data were sampled from mean configuration II. However, the average power rates for the procedures were 99.3 and 57.3%, respectively, when the data were sampled from a platykurtic distribution for the same mean configuration.

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

Table 5. Mean Any-variable Power (%) of Discriminant Function Coefficient Test Procedures by Population Distribution and Mean Configuration when Group Covariances are Unequal

Distribution	Mean Configuration	GUP	RAO	BOT	BOD
Normal	II	63.56	75.24	61.73	76.05
	III	34.07	44.47	36.53	45.37
	IV	81.86	90.85	92.70	91.20
	V	54.17	66.68	80.16	67.45
HT-I	II	58.94	67.48	41.32	68.33
	III	33.34	41.35	22.37	42.26
	IV	88.55	93.56	86.02	93.96
	V	63.80	72.87	67.84	73.65
HT-II	II	16.83	25.25	3.13	26.17
	III	7.54	14.82	1.13	26.17
	IV	36.74	55.65	13.34	57.04
	V	23.63	41.91	8.02	43.32
SK-I	II	59.59	70.93	46.91	71.81
	III	34.53	44.80	27.95	45.69
	IV	79.28	89.11	85.72	89.56
	V	53.31	65.31	68.85	66.05
SK-II	II	43.13	54.44	17.74	55.46
	III	24.74	33.78	9.50	34.65
	IV	74.05	84.50	63.35	85.08
	V	49.65	61.12	42.20	61.95
PK	III	35.87	99.30	44.21	57.32
	II	37.52	93.69	32.13	55.34
	IV	69.60	100.00	82.48	81.23
	V	56.10	99.98	68.92	71.35

NB: GUP = Gupta's (1968) procedure; RAO = Rao's (1970) procedure; BOT = Bootstrap procedure; BOD = Bodnar-Okhrin's (2011) procedure; GUP = Gupta's (1968) procedure; RAO = Rao's (1970) procedure; BOT = Bootstrap procedure; BOD = Bodnar-Okhrin's (2011) procedure. Normal distribution has skewness (γ_1) = 0 and kurtosis (γ_2) = 0; HT-I has $\gamma_1 = 0$ and $\gamma_2 = 33$; HT-II has $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-I has $\gamma_1 = 1.8$ and $\gamma_2 = 5.9$; SK-II has $\gamma_1 = 13.2$ and $\gamma_2 = 42.892.9$; PK has $\gamma_1 = 5.9$ and $\gamma_2 = -4.0$. See Table 1 for a description of mean configurations.

All the test procedures were not equally sensitive to the group mean configurations that were investigated. The average any-variable power for the procedures increased as the Euclidean distance between the two group mean vectors increased. The average any-variable power for each procedure was significantly higher for constant population means (II or IV) than for non-constant mean configurations (III or V), regardless of the population distribution. However, for each procedure, the

Table 6. Mean Any-Variable Power (%) for Discriminant Function Coefficient Test Procedures by Population Distribution and Correlation Structure

Distribution	Correlation Structure	GUP	RAO	BOT	BOD
Normal	Q_1	74.98	85.60	83.02	86.22
	Q_2	49.93	60.79	62.16	61.52
	Q_3	70.04	82.50	80.88	83.22
	Q_4	52.80	62.77	63.34	63.44
	Q_5	40.79	51.55	55.49	52.38
SK-II	Q_1	60.65	72.47	46.78	73.33
	Q_2	42.27	52.07	30.10	52.84
	Q_3	60.24	72.00	46.19	72.88
	Q_4	41.22	50.78	29.46	51.56
	Q_5	32.46	41.95	21.29	42.80
HT-II	Q_1	25.47	39.03	9.21	40.25
	Q_2	19.46	31.94	5.73	32.97
	Q_3	25.01	39.15	9.05	40.41
	Q_4	18.41	31.19	5.43	32.22
	Q_5	10.19	18.94	2.50	19.81
PK	Q_1	43.62	73.33	49.95	58.31
	Q_2	31.59	35.82	33.73	46.02
	Q_3	34.46	71.65	40.88	47.81
	Q_4	34.00	74.21	38.78	47.42
	Q_5	33.07	73.10	37.43	42.11

GUP = Gupta's (1968) procedure; RAO = Rao's (1970) procedure; BOT = Bootstrap procedure; BOD = Bodnar-Okhrin's (2011) procedure. Normal distribution has skewness (γ_1) = 0 and kurtosis (γ_2) = 0; HT-II has $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-II has $\gamma_1 = 13.2$ and $\gamma_2 = 42\ 892.9$; PK has $\gamma_1 = 5.9$ and $\gamma_2 = -4.0$. ρ = variable correlation; Q_1 : compound symmetric (CS) structure with $\rho = 0.3$; (b) Q_2 : CS structure with $\rho = 0.6$, (c) Q_3 : unstructured correlation matrix with average off diagonal elements of 0.3, (e) Q_4 : unstructured correlation matrix with average off-diagonal elements of 0.7, and (f) Q_5 simplex correlation matrix with off-diagonal elements of 0.3 and 0.7.

differences in any-variable power rates across the mean configurations were highest under the multivariate normal distribution and smallest under the extremely heavy-tailed distribution.

For example, the average any-variable power for Gupta (1986) and Bodnar procedures were 76.1% and 63.6%, respectively, when the data were sampled from a multivariate normal distribution with mean configuration II. The average power rates for these two procedures were 34.1% and 45.4%, respectively, when the data were sampled from a multivariate normal distribution with mean configuration II. On the other hand, when the data were sampled from an extremely heavy-tailed distribution

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

(i.e., HT-II), the average any-variable power rates for Gupta (1986) and Bodnar's procedures were 16.8% and 26.2%, respectively, when the data were sampled from a HT-II distribution with mean configuration I. The average power rates for the procedures were 7.5% and 16.2%, respectively, when the data were sampled from a population with mean configuration II under HT-II distribution.

The any-variable power rates for all the procedures were largest when the data were sampled from a population with the \mathbf{Q}_1 structure (i.e., all variables are independent), but smallest when the data were sampled from a population with the \mathbf{Q}_6 (i.e., simplex) correlation structure, regardless of the population distribution (Table 6). The any-variable power rates for Gupta and bootstrap procedures were more sensitive to smaller total sample size than Bodnar and Rao (1970) procedures, regardless of the population distribution. For example, the average any-variable power for Rao's (1970) and bootstrap procedures were 24.9% and 14.6%, respectively, when $n = 60$. Whereas, the average power rates for the former and latter procedures were 59.0% and 56.7%, respectively, when $n = 200$.

Discussion

This aim of this study was to investigate four test procedures to evaluate statistical significance of discriminant function coefficients for multiple correlated outcomes in grouped data. Based on the simulation study, Gupta's asymptotic test and bootstrap test procedures had values that were often below the lower bound of Bradley's (1978) liberal criterion, which suggests that they are conservative in controlling the FWER, except when sample size is very large for multivariate normal distributions. Rao (1970) F and Bodnar's test procedures were effective in controlling the overall FWER when the data were sampled from a multivariate normal distribution or were slightly non-normal.

However, all procedures were conservative in controlling the FWER when the data were sampled from multivariate non-normal data distributions. Furthermore, the pairing of group covariances and group sample sizes influenced the FWER control for all investigated procedures. The average FWERs for Gupta (1986) asymptotic test, Rao (1970) F -test, and Bodnar test exceeded the upper bound of Bradley's (1978) liberal criterion when the data were sampled from a population where group sample and covariances are negatively paired. But these procedures resulted in conservative FWERs when the data were sampled from a population where group size and covariances were positively paired.

The the bootstrap procedure was conservative, which is consistent with previous research. For example, simulation studies that investigated the use of bootstrap

methods for statistical inference about variable importance in regression literature revealed that the bootstrap procedure can result in overly conservative Type I error rates (Dagleish, 1994; Tonidandel & Lebreton, 2009). In terms of statistical power to detect non-zero discriminant function coefficients, our results suggest that bootstrap procedure can be more powerful than other procedures when the data are sampled from a multivariate normal distribution and when group sizes and covariances are equal. However, Rao (1970) *F*-test and Bodnar's procedures were more powerful than Gupta (1986) asymptotic test and bootstrap test procedure when the data were characterized by unequal group sizes, heterogeneous group covariances, or multivariate non-normal distributions.

Consequently, Rao (1970) *F*-test or Bodnar's test procedure should be adopted for testing hypotheses about the discriminant function coefficients when the data were sampled from multivariate normal distributions because they provide good control of the FWER and yield maximum any-variable power, regardless of the sample size and number of outcome variables. When the data were sampled from multivariate skewed or heavy-tailed distributions, we recommend that these two procedures should be adopted provided that the group covariances are equal. But none of the investigated procedures provided a good control of FWER and maximum statistical power when the data were sampled from multivariate non-normal distribution with heterogeneous group covariances. This suggest the need for new robust test procedures that provide good control of FWER and high statistical power when the data are sampled from multivariate non-normal distributions and heterogeneous group covariances.

The limitations of this study should be noted. Discriminant analysis procedures that are insensitive to departures from the assumptions of multivariate normality and covariance homogeneity have been developed, including robust discriminant analysis based on minimum covariance determinant estimators (Hubert & van Driessen, 2004), and *S* estimators (He & Fung, 2000; Croux & Dehon, 2001). Further research is needed to investigate the performance of these test procedures when the assumptions of multivariate normality and covariance homogeneity are violated in the study populations.

The Bonferroni correction was adopted, which is a commonly used method to adjust for multiple tests of statistical significance. However, it assumes independence among the variables and may result in conservative control of the FWER and low statistical power for correlated outcomes (Sankoh, Huque, & Dubey, 1997). Multiple testing procedures such as Hochberg's step-up procedure (Hochberg, 1988) and Holm's step down procedure (Holm, 1979) that account for correlation among the outcomes, are alternative procedures that can be adopted to control the FWER. Further research could compare these multiple testing procedures for discriminant

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

function coefficients. However, the simulation study compared the FWER and statistical power of the investigated procedures for testing if the discriminant function coefficient associated with each outcome variable is significantly different from zero. These test procedures can also be extended to test whether or not the discriminant function coefficient associated with a variable is significantly larger than the coefficient associated with another study variable.

Repeated measures discriminant analysis procedures based on parsimonious means and/or covariance structures have been proposed for predicting group membership and/or describing group separation in longitudinal studies (Tomasko, Helms, & Snapin, 1999; Roy & Khattree, 2005a, 2005b). But there has been no research on statistical inference about the relative importance of variables in multivariate longitudinal studies. Future research will investigate the extension of the proposed test procedures for evaluating statistical inference for coefficients of repeated measures discriminant analysis procedures that assume parsimonious means and/or covariance structures in multivariate longitudinal designs.

Acknowledgment

This research was supported by the National Science and Engineering Research Council Discovery Grant to the first author.

References

- Bargmann, R. E., Bose, R. C., Chakravarti, I. M., Mahalanobis, P. C., Rao, C. R., & Smith, K. J. (1970). Interpretation and use of a generalized discriminant function. *Essays in Probability and Statistics*, Chapel Hill: University of North Carolina Press.
- Baron, A. E. (1991). Misclassification among methods used for multiple group discrimination: the effects of distributional properties. *Statistics in Medicine*, *10*(5), 757–766. doi: [10.1002/sim.4780100511](https://doi.org/10.1002/sim.4780100511)
- Beaumont, J. L., Lix, L. M., Yost, K. J., & Hahn, E. A. (2006). Application of robust statistical methods for analysis of health-related quality of life outcomes. *Quality of Life Research*, *15*(3), 349–356. doi: [10.1007/s11136-005-2293-1](https://doi.org/10.1007/s11136-005-2293-1)
- Bodnar, T., & Okhrin, Y. (2011). On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory. *Scandinavian Journal of Statistics*, *38*(2), 311–331. doi: [10.1111/j.1467-9469.2011.00729.x](https://doi.org/10.1111/j.1467-9469.2011.00729.x)
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. doi: [10.1111/j.2044-8317.1978.tb00581.x](https://doi.org/10.1111/j.2044-8317.1978.tb00581.x)
- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate Analysis of Variance*. Newbury Park, CA: Sage. doi: [10.4135/9781412985222](https://doi.org/10.4135/9781412985222)

- Dalgleish, L. I. (1994). Discriminant analysis: statistical inference using the jackknife and bootstrap procedures. *Psychological Bulletin*, *16*(3), 498–508. doi: [10.1037/0033-2909.116.3.498](https://doi.org/10.1037/0033-2909.116.3.498)
- Croux, C., & Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *Canadian Journal of Statistics*, *29*(3), 473–493. doi: [10.2307/3316042](https://doi.org/10.2307/3316042)
- Das Gupta, S. (1968). Some aspects of discriminant function coefficients. *Sankhya Series A*, *30*, 387–400.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*(293), 52–64. doi: [10.1080/01621459.1961.10482090](https://doi.org/10.1080/01621459.1961.10482090)
- Field, C., & Genton, M. G. (2006). The multivariate g- and h- distribution. *Technometrics*, *48*(1), 104–111. doi: [10.1198/004017005000000562](https://doi.org/10.1198/004017005000000562)
- Finch, W. H., & Laking, T. (2008). Evaluation of the use of standardized weights for interpreting results from a descriptive discriminant analysis. *Multiple Linear Regression Viewpoints*, *34*, 19–34.
- Harwell, M. R., Rubenstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational & Behavioral Statistics*, *17*(4), 315–339. doi: [10.3102/10769986017004315](https://doi.org/10.3102/10769986017004315)
- He, X., & Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis*, *72*(2), 15–162. doi: [10.1006/jmva.1999.1857](https://doi.org/10.1006/jmva.1999.1857)
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800–802. doi: [10.1093/biomet/75.4.800](https://doi.org/10.1093/biomet/75.4.800)
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.
- Hubert, M., & van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, *45*(2), 301–320. doi: [10.1016/s0167-9473\(02\)00299-2](https://doi.org/10.1016/s0167-9473(02)00299-2)
- Huberty, C. J., & Smith, J. D. (1982). The study of effect in MANOVA. *Multivariate Behavioral Research*, *17*(3), 417–432. doi: [10.1207/s15327906mbr1703_7](https://doi.org/10.1207/s15327906mbr1703_7)
- Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin*, *105*(2), 302–308. doi: [10.1037/0033-2909.105.2.302](https://doi.org/10.1037/0033-2909.105.2.302)
- Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and Discriminant Analysis*. New York: Wiley. doi: [10.1002/047178947x](https://doi.org/10.1002/047178947x)
- Huberty, C. J., & Wisenbaker, J. M. (1992). Variable importance in multivariate group comparisons. *Journal of Educational Statistics*, *17*(1), 75–91. doi: [10.3102/10769986017001075](https://doi.org/10.3102/10769986017001075)
- Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. *Psychological Bulletin*, *95*(1), 156–171. doi: [10.1037/0033-2909.95.1.156](https://doi.org/10.1037/0033-2909.95.1.156)
- Lebreton, J. M., Ployhart, R. E., & Ladd, R. T. (2004). A Monte Carlo comparison of relative importance methodologies. *Organizational Research Methods*, *7*(3), 258–282. doi: [10.1177/1094428104266017](https://doi.org/10.1177/1094428104266017)
- Lix, L. M., Deering, K. N., Fouladi, R. T., & Manivong, P. (2008). Comparing treatment and control groups on multiple outcomes: robust procedures for testing a directional alternative hypothesis. *Educational and Psychological Measurement*, *69*(2), 198–215. doi: [10.1177/0013164408322027](https://doi.org/10.1177/0013164408322027)

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

- Lix, L. M., & Fouladi, R. T. (2007). Robust step-down tests for multivariate independent group designs. *British Journal of Mathematical and Statistical Psychology*, 60(2), 245–265. doi: [10.1348/000711006x117853](https://doi.org/10.1348/000711006x117853)
- Mackinnon, J. G. (2006). Bootstrap methods in Econometrics. *Economic Record*, 82(s1), S2–S18. doi: [10.1111/j.1475-4932.2006.00328.x](https://doi.org/10.1111/j.1475-4932.2006.00328.x)
- McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. New York: Wiley. doi: [10.1002/0471725293](https://doi.org/10.1002/0471725293)
- Onur, E., Alkin, T., & Tural, U. (2007). Panic disorder subtypes: further clinical differences. *Depression and Anxiety*, 24(7), 479–486. doi: [10.1002/da.20253](https://doi.org/10.1002/da.20253)
- Pratt, J.W. (1987). Dividing the indivisible: using simple symmetry to partition variance explained. In T. Pukkila and S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics* (pp. 245-260). Tampere, Finland: University of Tampere.
- Rao, C. R. (1970). Inference on discriminant coefficients. In *Essays in probability and statistics*, Bose et al. (Eds), 587 – 602. North Carolina: University of North Carolina.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. New Jersey: Wiley. doi: [10.1002/0471271357](https://doi.org/10.1002/0471271357)
- Roy, A., & Khattree, R. (2005a). Discrimination and classification with repeated measures data under different covariance structures. *Communications in Statistics – Simulation and Computation*, 34(1), 167–178. doi: [10.1081/sac-200047072](https://doi.org/10.1081/sac-200047072)
- Roy, A., & Khattree, R. (2005b). On discrimination and classification with multivariate repeated measures data. *Journal of Statistical Planning and Inference*, 134(2), 462–485. doi: [10.1016/j.jspi.2004.04.012](https://doi.org/10.1016/j.jspi.2004.04.012)
- Sajobi, T. T., Lix, L. M., Laverty, W. H., & Li, L. (2011) Discriminant analysis for repeated measures data: effects of mean and covariance misspecification on bias and error in discriminant function coefficients. *Journal of Modern Applied Statistical Methods*, 10(2), 571–582. doi: [10.22237/jmasm/1320120840](https://doi.org/10.22237/jmasm/1320120840)
- Sankoh, A. J., Huque, M. F., & Dubey, S. D. (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine*, 16(22), 2529–2542. doi: [10.1002/\(sici\)1097-0258\(19971130\)16:22<2529::aid-sim692>3.0.co;2-j](https://doi.org/10.1002/(sici)1097-0258(19971130)16:22<2529::aid-sim692>3.0.co;2-j)
- SAS Institute Inc. (2011). SAS/IML user’s guide, version 9.3. Cary NC: SAS Institute, Inc.
- Sherry, A. (2006). Discriminant analysis in counselling psychology research. *The Counseling Psychologist*, 34(5), 661–683. doi: [10.1177/0011000006287103](https://doi.org/10.1177/0011000006287103)
- Thomas, D. R. (1992). Interpreting discriminant functions: a data analytic approach. *Multivariate Behavioral Research*, 27(3), 335–362. doi: [10.1207/s15327906mbr2703_3](https://doi.org/10.1207/s15327906mbr2703_3)
- Thomas, D. R. (1997). A note Huberty and Wisenbaker’s “Views of variable importance”. *Journal of Educational and Behavioural Statistics*, 22(3), 309–322. doi: [10.3102/10769986022003309](https://doi.org/10.3102/10769986022003309)
- Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational and Behavioral Statistics*, 21(2), 110–130. doi: [10.3102/10769986021002110](https://doi.org/10.3102/10769986021002110)
- Timm, N. H. (2002). *Applied Multivariate Analysis*. Springer: New Jersey.
- Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods*, 14(4), 387–399. doi: [10.1037/a0017735](https://doi.org/10.1037/a0017735)

- Tomasko, L., Helms, R. W., & Snappin, S. M. (1999). A discriminant analysis extension to mixed models. *Statistics in Medicine*, 18(10), 1249–1260. doi: [10.1002/\(sici\)1097-0258\(19990530\)18:10<1249::aid-sim125>3.3.co;2-r](https://doi.org/10.1002/(sici)1097-0258(19990530)18:10<1249::aid-sim125>3.3.co;2-r)
- Urbakh, V. Y. (1971). Linear discriminant analysis: Loss of discriminating power when a variate is omitted. *Biometrics*, 27(3), 531–534 doi: [10.2307/2528594](https://doi.org/10.2307/2528594).

Appendix: Correlation Structures

$$p = 4$$

Compound Symmetric Correlation

$$Q_k = \begin{pmatrix} 1 & \rho_k & \rho_k & \rho_k \\ \rho_k & 1 & \rho_k & \rho_k \\ \rho_k & \rho_k & 1 & \rho_k \\ \rho_k & \rho_k & \rho_k & 1 \end{pmatrix}$$

$$k = 1, 2; \rho_1 = 0.3, \rho_2 = 0.7$$

Unstructured Correlation

$$Q_3 = \begin{pmatrix} 1 & 0.25 & 0.5 & 0.1 \\ 0.25 & 1 & 0.1 & 0.4 \\ 0.5 & 0.1 & 1 & 0.5 \\ 0.1 & 0.4 & 0.5 & 1 \end{pmatrix}$$

$$Q_4 = \begin{pmatrix} 1 & 0.5 & 0.7 & 0.3 \\ 0.5 & 1 & 0.7 & 0.5 \\ 0.7 & 0.7 & 1 & 0.4 \\ 0.3 & 0.5 & 0.4 & 1 \end{pmatrix}$$

Simplex Correlation

$$Q_5 = \begin{pmatrix} 1 & 0.3 & 0.7 & 0.3 \\ 0.3 & 1 & 0.3 & 0.7 \\ 0.7 & 0.3 & 1 & 0.3 \\ 0.3 & 0.7 & 0.3 & 1 \end{pmatrix}$$

$$p = 7$$

INFERENCE FOR DISCRIMINANT FUNCTION COEFFICIENTS

Compound Symmetric Structure

$$Q_k = \begin{pmatrix} 1 & \rho_k & \rho_k & \rho_k & \rho_k & \rho_k & \rho_k \\ \rho_k & 1 & \rho_k & \rho_k & \rho_k & \rho_k & \rho_k \\ \rho_k & \rho_k & 1 & \rho_k & \rho_k & \rho_k & \rho_k \\ \rho_k & \rho_k & \rho_k & 1 & \rho_k & \rho_k & \rho_k \\ \rho_k & \rho_k & \rho_k & \rho_k & 1 & \rho_k & \rho_k \\ \rho_k & \rho_k & \rho_k & \rho_k & \rho_k & 1 & \rho_k \\ \rho_k & \rho_k & \rho_k & \rho_k & \rho_k & \rho_k & 1 \end{pmatrix}$$

$$k = 1, 2; \rho_1 = 0.3, \rho_2 = 0.7$$

Unstructured Correlation

$$Q_3 = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.45 & 0.6 & 0.3 & 0.5 \\ 0.5 & 1 & 0.3 & 0.55 & 0.1 & 0.4 & 0.25 \\ 0.5 & 0.3 & 1 & 0.4 & 0.3 & 0.2 & 0.4 \\ 0.45 & 0.55 & 0.4 & 1 & 0.15 & 0.3 & 0.45 \\ 0.6 & 0.1 & 0.3 & 0.15 & 1 & 0.35 & 0.6 \\ 0.3 & 0.4 & 0.2 & 0.3 & 0.35 & 1 & 0.25 \\ 0.5 & 0.25 & 0.4 & 0.45 & 0.6 & 0.25 & 1 \end{pmatrix}$$

$$Q_4 = \begin{pmatrix} 1 & 0.65 & 0.66 & 0.7 & 0.72 & 0.65 & 0.75 \\ 0.65 & 1 & 0.7 & 0.7 & 0.67 & 0.72 & 0.74 \\ 0.66 & 0.7 & 1 & 0.71 & 0.75 & 0.63 & 0.74 \\ 0.7 & 0.7 & 0.71 & 1 & 0.7 & 0.65 & 0.73 \\ 0.72 & 0.67 & 0.75 & 0.7 & 1 & 0.68 & 0.74 \\ 0.65 & 0.72 & 0.63 & 0.65 & 0.68 & 1 & 0.75 \\ 0.75 & 0.74 & 0.74 & 0.73 & 0.74 & 0.75 & 1 \end{pmatrix}$$

Simplex Correlation

$$Q_5 = \begin{pmatrix} 1 & 0.3 & 0.7 & 0.3 & 0.7 & 0.3 & 0.7 \\ 0.3 & 1 & 0.3 & 0.7 & 0.3 & 0.7 & 0.3 \\ 0.7 & 0.3 & 1 & 0.3 & 0.7 & 0.3 & 0.7 \\ 0.3 & 0.7 & 0.3 & 1 & 0.3 & 0.7 & 0.3 \\ 0.7 & 0.3 & 0.7 & 0.3 & 1 & 0.3 & 0.7 \\ 0.3 & 0.7 & 0.3 & 0.7 & 0.3 & 1 & 0.3 \\ 0.7 & 0.3 & 0.7 & 0.3 & 0.7 & 0.3 & 1 \end{pmatrix}$$