DIGITALCOMMONS
—@WAYNESTATE—

December 2017

Robust Measures of Variable Importance for Multivariate Group Designs

Tolulope T. Sajobi
University of Calgary, AB, Canada, tolu.sajobi@ucalgary.ca

Lisa M. Lix
University of Manitoba, Winnipeg, MB

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation

Robust Measures of Variable Importance for Multivariate Group Designs

Robust Measures of Variable Importance for Multivariate Group Designs

Tolulope T. Sajobi
University of Calgary
Calgary, AB

Lisa M. Lix
University of Manitoba
Winnipeg, MB

Variable importance measures based on discriminant analysis and multivariate analysis of variance are useful for identifying variables that discriminate between two groups in multivariate group designs. Variable importance measures are developed based on trimmed and Winsorized estimators for describing group differences in multivariate non-normal populations.

Keywords: Variable importance, discriminant analysis, multivariate analysis of variance, trimmed estimators, assumption violations

Introduction

In educational and behavioral research studies where two groups (e.g., treatment versus control) are compared on a battery of outcome variables, testing for significant differences between the groups and identifying the relative importance of the variables that may discriminate between groups may be of interest. For example, Shih (2012) examined how Taiwanese junior high school students' perfectionistic tendencies and achievement goals were related to their academic burnout versus work engagement. The differences between indicators of burnout versus engagement among students with different subtypes of perfectionism was also investigated. The contributions of perfectionistic tendencies to academic burnout and engagement were found to be far greater than those of achievement goals.

  Bogler (2002) also investigated differences between teachers with a high level of job satisfaction and those with low job satisfaction on occupational, leadership, and demographic characteristics. Teachers' perception of their occupation and their principal's transformational style were the most important characteristics that

discriminated between the two types of teachers. There are several reasons why it may be interesting to identifying the outcomes on which group differences exist.

There may be limited research knowledge about which outcome(s) will be responsive to treatment, or little consensus about which outcome(s) is (are) relevant for behavioral or educational intervention. The intervention might be intended to have a multi-faceted effect, necessitating a research question that focuses on more than one outcome.

Measures to identify the relative importance of outcomes that discriminate between two independent groups were developed based on descriptive discriminant analysis (DDA) and multivariate analysis of variance (MANOVA). They quantify the relative importance of a variable (or outcome) based on its contribution to grouping effects and discriminant function scores (Huberty & Wisenbaker, 1992; Thomas, 1997). DDA measures of variable importance identify one or more linear combinations of the variables that maximize group separation; they are based on functions of the discriminant function coefficients and include standardized discriminant function coefficients, discriminant ratio coefficients, and total discriminant ratio coefficients (Huberty & Wisenbaker, 1992; Thomas, 1992). The *F*-to-remove statistic, a stepwise procedure based on MANOVA, has also been recommended as a variable importance measure; it quantifies the importance of a variable based on its unique contribution to group separation beyond that contributed by the remaining study variables (Huberty, 1984; Huberty & Wisenbaker, 1992). Applications of these measures have appeared in several disciplines including behavioral psychology (Sperling, Schilling, Glosser, Tracy, & Asadi-Pooya, 2008), criminology (Eastman & Bunch, 2009), development of questionnaires (Richardson, 2007), and educational research (Holder, 2007; Curenton, McWey, & Bolen, 2009).

However, DDA and MANOVA rest on the assumption of multivariate normality and covariance homogeneity, two assumptions which may not always be satisfied in practice. Normality of the outcome variables may not be a tenable assumption in behavioral or educational research investigations, which frequently exhibit multi-modal, skewed, or heavy-tailed distributions (Cressie & Whitford, 1986; Micceri, 1989). As well, the treatment group may exhibit greater variability than the control group (Blair & Sawilowsky, 1993; Troendle, Blair, Rumsey, & Moke, 1997). Previous research has shown that measures of variable importance may result in incorrect rank ordering of a set of correlated variables when applied to non-normal data with heterogeneous covariances (McLachlan, 1992). Thus, departures from the assumption of multivariate normality may have serious

consequences for investigators interested in identifying the outcomes that explains the differences between two or more groups.

Linear discriminant analysis procedures that are robust (i.e., insensitive) to departures from the assumption of multivariate normality have been proposed (Todorov, Neykov, & Neytchev, 1994) by replacing the conventional least-squares estimators of means and covariances with robust estimators, such as M-estimators (Croux & Dehon, 2001), minimum covariance determinant (MCD) estimators (Hubert & Van Driessen, 2004; Rousseeuw, 1984), minimum volume ellipsoid (MVE) estimators (Rousseeuw, 1984), and trimmed estimators (Ahmed & Lachenbruch, 1977; Gnanadesikan & Kettenring, 1972; Srivastava & Mudholkar, 2001). However, their emphasis has been primarily on prediction and not on describing the variables that contribute to group separation in non-normal data. There has been limited investigation of robust variable importance measures for evaluating the relative importance of variables in multivariate non-normal data.

Robust measures of variable importance are developed here in which conventional least squares (LS) estimates of the means and covariances are replaced by trimmed means and Winsorized covariance parameters based on coordinatewise trimming (CT) of the multivariate data. Trimmed means and Winsorized covariances possess good theoretical properties for heavy-tailed and skewed distributions, are computationally efficient, and straightforward to implement (Wilcox, 1994; Srivastava & Mudholkar, 2001). The sensitivity of the robust variable importance measures to departures from derivational assumptions and other data-analytic characteristics are investigated using Monte Carlo techniques.

Relative Importance Measures based on Descriptive Discriminant Analysis

Consider the two-group problem, although all of the procedures can be generalized to multi-group designs. Let $\mathbf{y}_{ij}$ be the $p \times 1$ vector of observed measurements for the ith study participant ($i = 1,..., n_j$) in the jth group ($j = 1,\ldots, J$). Initially, assume $\mathbf{y}_{ij} \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ are the population mean and covariance for the jth group and are estimated by $\hat{\boldsymbol{\mu}}_j$ and $\hat{\boldsymbol{\Sigma}}_j$, respectively. For the linear DA procedure, the discriminant function coefficient vector is estimated by

$$\hat{\mathbf{a}} = \hat{\boldsymbol{\Sigma}}^{-1}\left(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2\right) \tag{1}$$

where

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

The number of uncorrelated discriminant functions that separates G groups is equal to $\min(J - 1, p)$.

In DDA, variable importance measures based on discriminant function coefficients can be used to rank variables according to their contributions to group separation (Huberty & Wisenbaker, 1992). The computation and implementation of these measures have been described in detail elsewhere (Sajobi, Lix, Clara, et al., 2012), and are only briefly reviewed here.

The standardized discriminant function coefficient (SDFC) is one commonly adopted variable importance measure. It quantifies the importance of a variable by taking into account the presence of other variables in the study. The SDFC for the kth variable, denoted by $\hat{a}_k^*$, is

$$\hat{a}_k^* = \hat{a}_k s_k \tag{3}$$

where $\hat{a}_k$ and s_k are the corresponding estimated k^{th} discriminant function coefficient and standard deviation, respectively. By placing a constraint on the discriminant function coefficients such that $\hat{\mathbf{a}}^{\text{T}}\mathbf{S}_{\mathbf{E}}\hat{\mathbf{a}} = 1$, where $^{\text{T}}$ is the transpose operator, $\mathbf{S}_{\mathbf{E}} = \hat{\boldsymbol{\Sigma}}$, and the coefficients range in value from -1 to +1. SDFCs can be positive or negative and the absolute magnitude determines relative importance. Although there have been arguments in favor of using SDFCs to measure variable importance (Rencher, 2002), they are known to be sensitive to variable correlations (Rencher, 2002).

Discriminant ratio coefficients (DRCs) are sometimes recommended instead of SDFCs (Thomas, 1992; Thomas & Zumbo, 1996). DRCs measure the importance of a variable as a proportion of the group differences explained by the variable. The kth DRC is given by

$$q_k = a_k^* f_k \tag{4}$$

where f_k is the kth structure coefficient, the correlation between the kth outcome variable and the discriminant function. DRCs generally range in value from zero to one, with larger values indicating greater importance. However, they can have

negative values; a large negative value may be indicative of collinearity or suppression (Thomas & Zumbo, 1996). In MANOVA, suppression occurs when a variable makes little or no direct contribution to group separation on its own but contributes indirectly through another variable.

The F-to-remove statistic (Huberty & Wisenbaker, 1992) is obtained by conducting p MANOVA tests, each time removing one variable from the analysis. For the kth domain, the statistic is

$$F_{(k)} = \frac{k_2 \left(\hat{a}_k / s_{(kk)} \right)^2}{\left(\overline{z}_1 - \overline{z}_2 \right) + k_3 - \left(\hat{a}_k / s_{(kk)} \right)^2} \tag{5}$$

where $k_2 = (n_1 + n_2 - 2 - q)$, $k_3 = (n_1 + n_2)(n_1 + n_2) / n_1 n_2$, $\hat{a}_k$ is the discriminant function coefficient for the kth domain, $\overline{z}_1$ and $\overline{z}_2$ are the group means for the discriminant function score corresponding to $\hat{\mathbf{a}}$, and $s_{(kk)}$ is the positive square root of the kth diagonal element of the inverse of $\mathbf{E}$, the error sums of square and cross product matrix. F-to-remove statistics have a lower bound of zero, but no upper bound. Variable importance is assessed by the magnitude of F-to-remove-statistic, with the most important domain yielding the largest statistic.

The total discriminant ratio coefficient (TDRC) for the kth variable is

$$t_k = \frac{| \hat{a}_k | \mathbf{S}_{\mathbf{T}kk}}{\sqrt{\left(\hat{\mathbf{a}} \mathbf{S}_{\mathbf{T}} \hat{\mathbf{a}} \right)}} \tag{6}$$

where

$$\mathbf{E} = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left(\mathbf{y}_{ij} - \overline{\mathbf{y}}_j \right) \left(\mathbf{y}_{ij} - \overline{\mathbf{y}}_j \right)^{\mathrm{T}} \tag{7}$$

is the error sum of squares and cross product matrix, $\mathbf{S}_{\mathbf{T}kk}$ is the $(k, k)^{\text{th}}$ element of $\mathbf{S}_{\mathbf{T}}$, $\mathbf{S}_{\mathbf{T}} = \mathbf{T} / (N - 1)$, $\mathbf{T} = \mathbf{H} + \mathbf{E}$,

$$\mathbf{H} = \sum_{j=1}^{2} n_j \left(y_j - \overline{\mathbf{y}} \right) \left(\mathbf{y}_j - \overline{\mathbf{y}} \right)^{\mathrm{T}} \tag{8}$$

is the hypothesis sum of squares and cross product matrix, and ^{T} is the transpose operator. Like DRCs, the TDRCs can range in value from zero to one, with a larger value indicating greater relative importance.

Robust Estimation

In CT of multivariate data, each outcome variable is independently trimmed by removing a pre-specified proportion of the data at both of tails of the variable's distribution. Let $y_{(1)jm} \leq y_{(2)jm} \leq \ldots \leq y_{(n_j)jm}$ denote the order statistics of the jth group for the mth ($m = 1, \ldots, p$) outcome variable (Srivastava & Mudholkar, 2001). Define $b_j = [\delta n_j]$, where δ represents the proportion of the observations to be trimmed, or censored, from each tail of the distribution and $[x]$ denotes the integer less than or equal to x. When symmetric trimming is adopted, so that the same number of observations are removed from each tail of the distribution, the effective sample size for the jth group is $f_j = n_j - 2b_j$. The trimmed mean for the jth group on the kth outcome variable is

$$\bar{y}_{tjk} = \frac{1}{f_j} \sum_{i=b_j+1}^{n_j-b_j} y_{(i)jk} \tag{9}$$

Define z_{ijk} as

$$z_{ijk} = \begin{cases} y_{(b_j+1)jk} & \text{if } y_{ijk} \leq y_{(b_j+1)jk} \\ y_{ijk} & \text{if } y_{(b_j+1)jk} < y_{ijk} < y_{(n_j-b_j)jk} \\ y_{(n_j-b_j)jk} & \text{if } y_{ijk} \geq y_{(n_j-b_j)jk} \end{cases}$$

Then

$$\bar{y}_{wjk} = \frac{1}{n_j} \sum_{i=1}^{n_j} z_{ijk} \tag{10}$$

is the jth group Winsorized mean for the kth outcome variable. The Winsorized sum of squared deviations for the kth and l^{th} outcome variable in the jth group is

$$ss_j = \sum_{i=1}^{n_j} \left(z_{ijk} - \overline{z}_{jk} \right) \left(z_{ijl} - \overline{z}_{jl} \right) \tag{11}$$

and $\mathbf{S}_{jw} = (ss_{jk})$ is the estimated Winsorized sum of squares and cross product matrix. Given that CT estimators are derived by independently trimming each outcome variable (Wilcox, 1994; Maronna, Martin, & Yohai, 2006), they share similar robustness properties to univariate trimmed estimators. Robust variable importance measures are derived by replacing the least squares means and covariances by trimmed means and Winsorized covariances of the data.

Simulation Study

A Monte Carlo study was used to evaluate the performance of measures of relative importance in rank ordering a set of correlated variables under a variety of data analytic conditions. Measures of relative importance investigated include (a) SDFC, (b) DRC, (c) FTR, and (d) TDRC. These were developed based on least squares and CT means and Winsorized covariances. All variable importance measures were investigated for the case of two independent groups.

The conditions investigated were (a) number of outcome variables, (b) total sample size and equality/in equality of group sizes, (c) magnitude and pattern of variable correlation, (d) mean configuration, and (e) shape of the population distribution. The number of outcome variables was set at $p = 4$, 6, and 8. Similarly investigated numbers of outcome variables ranging from 4 to 10 were previously considered (LeBreton, Polyhart, & Ladd, 2004). Total sample sizes of $N = 60$, 90, 140, and 200 were investigated. Although previous simulation studies for relative importance measures have primarily focused on equal group sizes (Finch & Laking, 2008), unequal group sizes have also been shown to influence the size of discriminant function coefficients (Barön, 1991; He & Fung, 2000), and may influence the consistency of the measures in accurately rank ordering the variables. Therefore, both equal and unequal group size conditions were investigated. For $n = 60$, the group sizes were $(n_1, n_2) = (30, 30)$ and $(n_1, n_2) = (24, 36)$. For $n = 90$, they were $(n_1, n_2) = (45, 45)$ and $(n_1, n_2) = (36, 52)$. For $n = 140$, the group sizes were $(n_1, n_2) = (70, 70)$ and $(n_1, n_2) = (56, 84)$. For $n = 2000$, the group sizes were $(n_1, n_2) = (100, 100)$ and $(n_1, n_2) = (80, 120)$. These group sizes were chosen based on previous research to represent small to large degrees of group size imbalance (Barön, 1991; Sajobi, Lix, Laverty, & Li, 2011).

Previous investigations have shown that variable importance measures are sensitive to the magnitude of correlation among the outcome variables (Huberty & Wisenbaker, 1992; Thomas & Zumbo, 1996). Therefore, investigated here are different correlation structures and sizes of correlations: (a) $\mathbf{Q}_1$: an independent correlation structure with $\rho = 0.0$, (b) $\mathbf{Q}_2$: compound symmetric structure with $\rho = 0.3$, (c) $\mathbf{Q}_3$: compound symmetric structure with $\rho = 0.6$, (d) $\mathbf{Q}_4$: unstructured with average correlation among the off-diagonal elements of 0.3, and (e) $\mathbf{Q}_5$: unstructured with average correlation among the off-diagonal elements of 0.6. Previous research studies have investigated variable correlations ranging between 0.1 and 0.8 (Tonidandel, LeBreton, & Johnson, 2009).

The measures of relative importance were investigated when the data were both multivariate normal and non-normal. For the former, in which skewness (γ_1) and kurtosis (γ_2) values were equal to $\gamma_1 = 0$ and $\gamma_2 = 0$, respectively, pseudo-random observation vectors $\mathbf{X}_{ij}$ from a multivariate normal distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$ were generated using the RANNOR function in SAS (SAS Institute, 2008). A vector of p standard normal deviates, $\mathbf{R}_{ij}$, was transformed to a vector of multivariate observations via $\mathbf{X}_{ij} = \boldsymbol{\mu}_j + \mathbf{LR}_{ij}$. The Cholesky decomposition was used to obtain $\mathbf{L}$, an upper triangular matrix of dimension m satisfying the equality $\mathbf{L}^T\mathbf{L} = \boldsymbol{\Sigma}_j$. The measures were also considered when the data were sampled from populations with multivariate skewed and heavy-tailed distributions.

For non-normal distributions, two skewed distributions and two heavy-tailed distributions were investigated. In the former, a moderately skewed non-normal distribution with $\gamma_1 = 1.8$ and $\gamma_2 = 5.9$ (SK-I) and a largely skewed non-normal distribution with $\gamma_1 = 13.2$ and $\gamma_2 = 42892.9$ (SK-II) were investigated. Two heavy-tailed non-normal distributions were also considered; the first is a moderately heavy-tailed distribution with $\gamma_1 = 0$ and $\gamma_2 = 33$ (HT-I), while the second distribution is a heavy-tailed non-normal distribution similar to a heavy-tailed Cauchy distribution for which γ_1 and γ_2 are undefined (HT-II). 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 R_{ijk} via

$$Y_{ijk} = \left(\frac{\exp\left(gR_{ijk}\right)-1}{g} \right)\left(\exp\frac{h}{2}R_{ijk}^2 \right) \tag{12}$$

Table 1. Values of μ_1 selected for the Monte Carlo study

# Variables	Mean Pattern	Mean Vector	D^2
4	I	(2.5,2,1.5,1)	13.50
	II	(1,0.75,0.5,0.25)	1.88
	III	(1.5,1,0.5,2)	7.50
	IV	(0.75,0.5,0.25,1)	1.88
6	I	(3.5,3,2.5,2,1.5,1)	34.75
	II	(1.25,1,0.75,0.5,0.5,0.25)	3.69
	III	(2,1.5,1,0.5,2.5,3)	22.75
	IV	(1,0.75,0.5,0.25,1.25,1.5)	5.69
8	I	(4.5,4,3.5,3,2.5,2,1.5,1)	71.00
	II	(2.5,2.25,2,1.75,1.5,1.25,1,0.75,0.5)	17.50
	III	(2.5,2,1.5,1,0.5,3,3.5,4)	51.00
	IV	(1.25,1,0.75,0.5,0.25,1.5,1.75, 2)	12.75

when $g = 0$, this equation reduces to

$$Y_{ijk} = \left(\exp \frac{h}{2} R_{ijk}^2 \right) R_{ijk} \qquad (13)$$

The g and h parameters provide a convenient approach to generate data from a wide range of multivariate non-normal distributions. When $g = h = 0$, the distribution is multivariate normal. The parameter h determines the heaviness of the tails of a distribution, while the parameter g controls the magnitude of skewness of the distribution.

Four configurations of the variable means were selected for group 1 (Table 1). Configurations I and II had a linearly decreasing mean trend across the variables for group 1, while configuration III and IV had a quadratic mean trend across the variables for group 1. In all cases, the mean vector for group 2 was the null vector. For robust variable importance measures, a 20% symmetric trimming rule was adopted as recommended by Wilcox (1994). All combinations of conditions were investigated for each procedure and each method of estimation, resulting in a total of 1200 combinations of simulation conditions with 10,000 replications for each combination. The Monte Carlo study was conducted using SAS/IML version 9.2 (SAS Institute, 2008).

To evaluate the ranking accuracy of each variable importance measure, the estimated rank ordering of variables in the data was compared to the population variable rankings, which were derived from the rank ordering the variables based on the magnitude differences in the population group means. Other methods for

generating the population variable ranking are described in the Appendix. Three indices were used to compare the ranking accuracy of these measures: (a) any-variable correct ranking (ACR) percentage, (b) average all-variable correct ranking (ALLCR) percentage, and (c) average per-variable correct ranking (PCR) percentage. The average ACR rate is the percent of simulations in which at least one variable in the sample was ranked the same as the corresponding variable in the population. The average ALLCR is the percentage of simulations in which all variable in the sample was ranked the same as the population rank ordering of the variables. The average PCR percentage is the overall average percentage of simulations in which a variable in the sample had the same rank as the variable in the population.

Results

Described in Table 2 are the average ACR percentages for variable importance measures based on least squares and CT estimators by the type of distribution and mean configuration. For normally distributed data, the ACR percentage values for all importance measures increased as the magnitude of group separation increased. However, under large departures from a multivariate normal distribution, the magnitude of change in ACR percentage values were substantial for variable importance measures based on least squares estimation as the magnitude of group separation increased.

For example when the data for group 1 were sampled from a population with a linear decreasing mean configuration (configuration I), the average ACR percentage values for DRCs based on least squares estimation under mean configuration I and II were 99.1% and 96.1%, respectively. When the data were sampled from a skewed distribution (SK-II), the average ACR percentage values for DRCs based on least squares estimation under mean configuration I and II were 88.5% and 77.9%, respectively. The magnitude of change in ACR percentage values for the robust variable importance measures decreased as the amount of group separation increased. However, it was not substantially different irrespective of the population distribution. For example, the average ACR percentage values for the robust FTR under mean configurations III and IV were 97.15% and 95.22%, respectively when the data were sampled from normal distribution.

Table 2. Average percentage of any-variable correct ranking (ACR) for variable importance measures by population distribution and mean configuration

Distribution	Mean Configuration	Least Squares Estimators				Robust Estimators			
		SDFC	DRC	FTR	TDRC	SDFC	DRC	FTR	TDRC
Normal	I	89.81	99.08	95.62	97.18	93.98	97.80	93.22	98.52
	II	85.45	96.10	91.11	89.32	89.99	95.18	89.44	94.00
	III	87.75	98.55	91.44	95.29	88.58	97.15	90.78	95.71
	IV	84.47	96.70	87.49	88.61	85.31	95.22	87.26	89.99
HT-I	I	89.15	97.29	91.68	92.88	94.69	97.70	94.88	98.37
	II	83.01	91.35	86.20	84.95	90.16	94.06	90.38	92.34
	III	83.14	96.62	86.72	89.85	80.14	95.76	82.67	91.07
	IV	81.22	92.52	83.27	83.79	78.97	92.08	80.32	85.18
SK-I	I	91.07	97.94	93.40	95.49	94.29	97.53	94.26	98.19
	II	84.99	93.79	88.68	87.72	90.02	94.50	90.17	93.00
	III	84.34	97.81	88.23	91.20	83.70	96.94	86.49	93.02
	IV	82.22	94.38	84.84	85.68	81.47	93.49	83.05	87.14
HT-II	I	86.28	94.19	88.58	90.99	95.04	97.58	95.46	98.12
	II	79.45	87.26	82.43	80.24	89.84	93.57	90.54	91.36
	III	81.34	94.32	83.80	86.34	76.00	94.74	77.09	88.14
	IV	78.11	87.67	79.46	79.38	74.80	88.87	75.57	81.89
SK-II	I	80.99	88.50	83.56	83.00	95.51	97.59	96.05	97.88
	II	72.92	77.91	75.32	70.43	89.57	92.45	90.42	89.31
	III	78.50	87.95	79.44	79.02	70.04	91.05	68.96	82.92
	IV	74.17	79.54	74.82	72.85	71.63	85.52	71.36	78.88

Note: See Table 1 for a description of mean configurations. SDFC = standardized discriminant function coefficient; DRC = discriminant ratio coefficient; FTR = F-to-remove statistic; TDRC = total discriminant ratio coefficient; HT-I = heavy-tailed distribution with $\gamma_1 = 0$ and $\gamma_2 = 33$; HT-II = a heavy-tailed distribution with $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-I = skewed distribution with $\gamma_1 = 1.8$ and $\gamma_2 = 5.9$; SK-II = skewed distribution with $\gamma_1 = 13.2$ and $\gamma_2 = 42892.9$

Table 3. Average percentage of all-variable correct ranking (ALLCR) by population distribution and correlation structure

Distribution	Mean Configuration	Least Squares Estimators				Robust Estimators			
		SDFC	DRC	FTR	TDRC	SDFC	DRC	FTR	TDRC
Normal	Independent	26.69	34.93	26.85	37.92	24.46	29.75	24.60	37.87
	CS(0.3)	14.14	35.81	13.94	17.27	15.50	30.94	15.29	21.21
	CS(0.6)	1.59	38.24	1.52	2.44	2.66	32.77	2.57	5.21
	UN(0.3)	9.69	24.84	9.27	12.80	11.02	24.46	10.76	16.53
	UN(0.6)	2.26	7.03	3.84	5.61	4.98	8.80	6.05	14.01
HT-II	Independent	14.74	16.38	14.76	20.65	23.78	27.53	23.86	33.41
	CS(0.3)	10.02	18.24	9.99	12.04	11.61	28.05	11.56	15.62
	CS(0.6)	2.91	23.09	2.86	3.34	3.23	27.38	3.18	4.91
	UN(0.3)	9.40	16.52	9.20	11.01	11.27	19.77	11.13	14.99
	UN(0.6)	3.78	6.46	4.91	5.71	5.10	6.18	5.45	10.38
SK-II	Independent	8.28	8.81	8.33	10.14	22.17	25.48	22.26	30.07
	CS(0.3)	6.04	9.59	6.11	6.61	9.13	23.58	9.10	11.75
	CS(0.6)	3.06	12.31	3.25	3.13	3.37	21.20	3.38	4.51
	UN(0.3)	6.38	9.48	6.39	6.72	10.75	15.50	10.80	13.12
	UN(0.6)	3.30	4.63	3.93	3.61	4.56	4.75	4.54	8.40

Note: See Table 1 for a description of mean configurations. SDFC = standardized discriminant function coefficient; DRC = discriminant ratio coefficient; FTR = F-to-remove statistic; TDRC = total discriminant ratio coefficient; HT-II = a heavy-tailed distribution with $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-II = skewed distribution with $\gamma_1 = 13.2$ and $\gamma_2 = 42892.9$; CS(ρ) = compound symmetric correlation structure with $\rho = 0.3$; UN(ρ) = unstructured correlation with average off-diagonal correlation of ρ

Compiled in Table 3 are the results for average ALLCR percentage values when the data were sampled from multivariate normal and non-normal distributions for independent, compound symmetric, and unstructured correlations. The average ALLCR percentage values for SDFC, TDRC, and FTR decreased as the magnitude of correlation among the variables increased, regardless of the population distributions, number of variables, or method of estimation. In contrast, the change in ALLCR percentage values for the DRCs based on least squares and robust estimators as the magnitude of correlation increased varied across correlation structures. For example, when the data were sampled from a population with compound symmetric correlation structure, the average ACR values for DRCs based on least squares estimators increased but there was negligible change in the average values of DRC based on robust estimators, as the magnitude of correlation increased from 0.0 to 0.6, when the data were sampled from a normal distribution.

However, when the data were sampled from a population with an unstructured variable correlation, the decrease in average ALLCR values were more than 10%. Moreover, the magnitude of change in average ALLCR percentage values for the variable importance measures varied across population distribution and correlation

structure. The average ALLCR percentage values for each variable importance measure were largest when the data were sampled from a multivariate normal distribution but smallest under a multivariate skewed distribution, as the magnitude of correlation increased. For example, when the data were sampled from a normally distributed population, the average ALLCR percentage values for the FTR measure based on the least-squares method under independent and unstructured variable correlations (i.e. UN[0.6]) were 26.9% and 3.8%, respectively. However, under a SK-II distribution, average ALLCR percentage values were 8.3% and 3.9% when the data were sample from a population with independent and unstructured variable correlations, respectively.

Moreover, when the data were sampled from a multivariate normal distribution, differences between the average ALLCR percentage values for variable importance measures based on least squares estimators and those based on robust estimators were not more than 6%, except for the TDRC measure. However, when the data were sampled from non-normal distributions, the robust variable importance measures resulted in substantially higher average ALLCR percentage values than the variable importance measures based on least squares estimators. For instance, when the data were sampled from a normal distribution with an independent correlation structure, the average ALLCR values for DRC based on least squares and robust estimators were 34.9% and 289.8%, respectively. But when the data were sampled from a population with a multivariate heavy-tailed non-normal distribution (HT-II), the corresponding average PCR values were 16.4% and 27.5%, respectively.

Depicted in Figure 1 are the average PCR for all the investigated measures by method of estimation, and total sample size. The analyses reveal the average PCR values for each variable importance measure were smallest when $N = 50$ and largest when $N = 200$, regardless of the population distribution and the method of estimation. The average PCR value for each variable importance measure increased as N increased, regardless of the method of estimation. Specifically, the DRC was most sensitive to changes in sample size when the data were sampled from a multivariate normal distribution. For example, for DRC based on least squares estimators and robust estimators, the change in average ACR percentage values were 20.6% and 22.70%, respectively, as N increased from 50 to 200. For other variable importance measures, the increase in average PCR percentage values were not more than 15% as N increased from 50 to 200. When the data were sampled from a highly skewed non-normal distribution, the average PCR values increased as N increased from 50 to 200, but the magnitude of change were not as large as under multivariate normal distribution. For example, as N increased from 50 to 200,

the change in average PCR values for the TDRC based on least squares estimators were 9.7% and 4.9% when the data were sampled from a multivariate normal and highly skewed distributions, respectively. For example, the change in average ACR values for the FTR measure based on robust and least squares estimators as n increased from 50 to 200 were 8.9% and 7.2%, respectively, when the data were sampled from a heavy-tailed distribution (Figure 1).

The average PCR values for all the variable importance measures increased as p increased from 4 to 8 (Figure 2). There were negligible differences among the average PCR values of SDFC, DRC, and FTR procedures based on least squares and robust estimators when the data were sampled from a normal distribution. But the robust variable importance measures resulted in higher PCR values than measures based on least-squares estimators as p increased from 4 to 8, when the data were sampled from a highly skewed or heavy-tailed non-normal distribution.
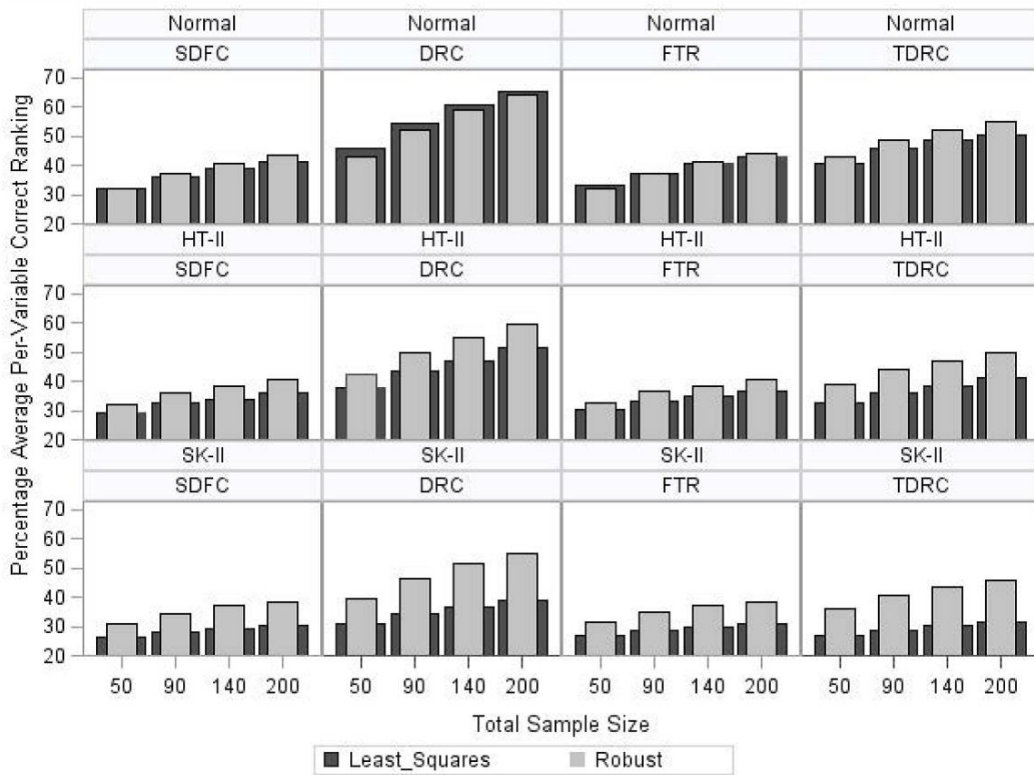

Figure 1. Average percentage of per-variable correct ranking (ACR) for variable importance measures by population distribution and total sample size
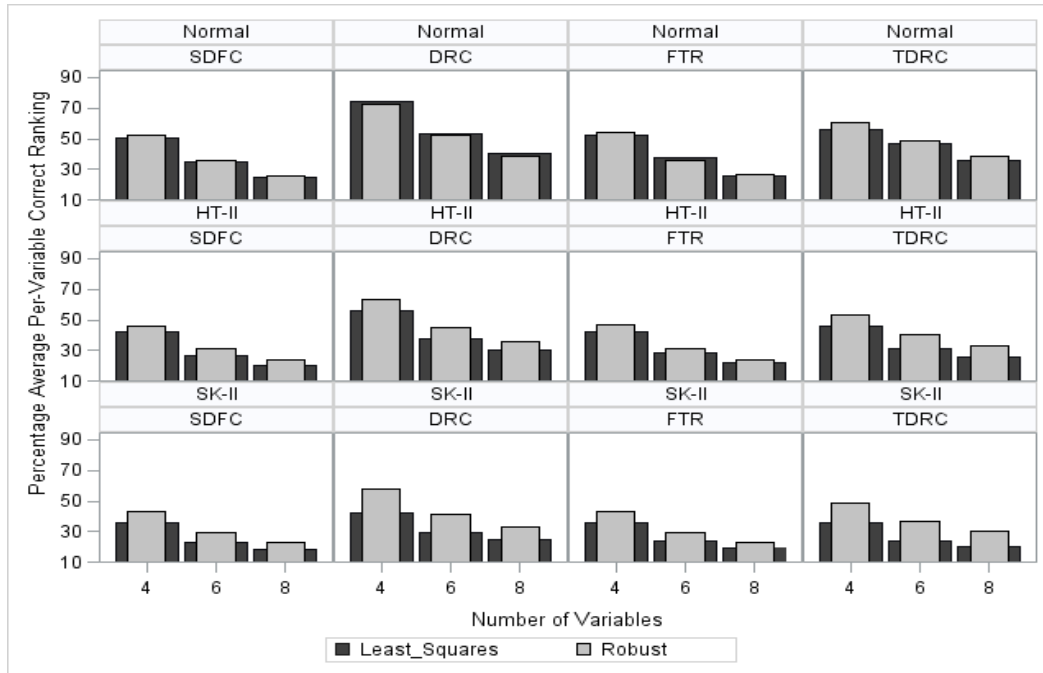
Figure 2. Average percentage of per-variable correct ranking (PCR) for variable importance measures by population distribution and number of variables

Conclusion

Robust measures of relative importance were developed based on trimmed means and Winsorized covariances for non-normal two-group multivariate data. The performance of these measures were investigated in a variety of data analytic conditions including multivariate non-normality, number of variables, sample size, magnitude of correlation, and mean configuration. Our findings suggest that variable importance measures based on coordinatewise trimmed means and Winsorized covariance estimators result in higher percentages of correct variable ranking than least-squares measures under moderate to strong departures from multivariate normality. Robust DRCs, which achieved the highest proportion of correctly ranked variables, were the least sensitive to group separation, variable correlation, and sample size, regardless of the distribution of the data.

One criticism of trimmed estimators is that they may not perform well when sample size is small and the data are normally distributed (Wilcox, 1994). This study results suggest robust measures of relative importance performed equally well as measures based on least squares estimators under small sample size

conditions and when data are normally distributed. Hence, they can be adopted for rank ordering the variables regardless of the shape of the population distribution.

The ability of a variable importance measure to correctly rank order a set of correlated variables in a simulation study may depend on the method for generating the population rank order of the variables. Although the variable importance measures described in this study may not be equally sensitive to choice of methods for generating population variable rankings (See Tables A1 and A2), one finding that is consistent across these methods is that the robust variable importance measures developed in this study resulted in higher correct ranking percentage values than variable importance measures based on least squares estimators regardless of the population ranking method.

Although previous research investigations have shown that variable importance measures based on DDA and MANOVA are sensitive to non-normal data (Sajobi, Dansu, & Lix, 2013), adopting these measures for evaluating the relative importance of variables in non-normal data is likely to lead to incorrect conclusions about the study variables. Based on the findings, it is recommended robust variable importance measures should be adopted for rank ordering variables in studies that are likely to be characterized by non-normal data distributions.

The limitations of this study should be noted. The simulation study focused on conditions in which group covariances were homogeneous. This may not be a reasonable assumption in all clinical studies; the treatment group may exhibit greater variability than the control group (Blair & Sawilowsky, 1993). The DRC is the only measure among the investigated measures that is designed to detect the presence of suppression effects among the variables (Thomas & Zumbo, 1996). Although, the presence of suppression effects among the variables can reduce the percentage of correctly rank variables, previous studies have suggested the exclusion of the suppressor variable from the relative importance analysis since it makes little direct contribution to group discrimination (Thomas, Hughes, & Zumbo, 1998; Ochieng & Zumbo, 2001). However, the exclusion of collinear or suppressor effects may also reduce statistical power and can affect the rank ordering of the study variables. In our simulation study, we ignored the presence of the suppression among the variables by ranking all the variables according to the signed DRC values. We must also reiterate that other importance measures are also may not always be robust to suppression effects among the variables. Moreover, the robust procedures described in this study are sensitive to strong variable correlations. This might be attributed to the coordinate-wise approach for trimming multivariate data. An alternative trimmed estimator is the maximum trimmed likelihood estimators (MTLE), which trims the log-likelihood function of the

multivariate data (Bednarski & Clarke, 1993). Future investigations might examine the development of variable importance measures based on MTLE for describing group separation in multivariate data.

Another criticism of variable importance measures is that the absolute importance of a set of correlated outcome variables is derived based on estimates of variable importance indices obtained from sample data and do not account for sampling variations around these estimates. There are currently limited recommendations to guide researchers about contextual relevance/meaning of the rank ordering of outcome variables from. Rules of thumb and descriptive analyses of relative importance are being used. For example, Dalgleish (1994) suggested that variables with structure coefficients above 0.4 should be considered as important. But this approach fails to account for sampling error in the observed data. More formal parametric methods that assess the statistical significance of the estimated variable importance index have been proposed (Rao, 1970; Huberty & Wisenbaker, 1992).

Although the variable importance measures developed in this study are robust to non-normal data, other variable importance measures that do not assume multivariate normality have been developed based on logistic regression model. This includes relative weights (Azen & Traxel, 2009; Tonidandel & LeBreton, 2010), dominance analysis (Azen & Budescu, 2003), standardized logistic regression coefficients, and Pratt's index for logistic regression (Thomas, Zumbo, Zhu, & Dutta, 2008). The relative weights are derived from an orthogonal transformation of the predictor variables, whereas the dominance analysis method relies on regression model's pseudo coefficient of determination (i.e., model R^2) to evaluate variable importance. Although some of these measures (e.g., rescaled relative weights) may perform better under some conditions (e.g., strong variable correlations), there has not been a comparison of the performance of these logistic regression-based variable importance measures under a broad range of data analytic conditions.

There are additional considerations for conducting relative importance analysis. The conclusion that one outcome variable is more important than another outcome variable depends on the set of variables under investigation. Hence, changing the mix of variables included in a study could change a researcher's conclusions about variable importance. The choice of study variables and the grouping variables are assumed to have been determined *a priori*. However, conclusions about a variable's importance as estimated from a sample may not be generalizable to other populations because it does not account for sampling variability in the data. An internal or external validation of the ranks should also be

considered, in order to assess the generalizability of the results. For example, a split-sample validation or resampling based methods such as bootstrap may be adopted to examine statistical significance of a variable's importance.

In conclusion, robust measures for describing the relative importance of investigated set of correlated variables for group discrimination in normal and non-normal data were developed. Measures of relative importance have a number of applications (Baek et al., 2008; Johnson & LeBreton, 2004). They can be used to build a parsimonious statistical model for use in further research or to choose a subset of variables on which to focus in clinical investigations.

References

Ahmed, S. W., & Lachenbruch, P. A. (1977). Discriminant analysis when scale contamination is present in the initial sample. In: J. Van Ryzin (Ed.), *Classification and clustering* (pp. 351-353). New York, NY: Academic Press. doi: 10.1016/b978-0-12-714250-0.50016-4

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods, 8*(2), 129-148. doi: 10.1037/1082-989x.8.2.129

Azen, R., & Traxel, N. M. (2009). Using dominance analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics, 34*(3), 319-347. doi: 10.3102/1076998609332754

Baek, S., Moon, H., Ahn, H., Kodell, R. L., Lin, C., & Chen, J. J. (2008). Identifying high-dimensional biomarkers for personalized medicine via variable importance ranking. *Journal of Biopharmaceutical Statistics, 18*(5), 853-868. doi: 10.1080/10543400802278023

Barön, 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

Bednarski, T., & Clarke, B. R. (1993). Trimmed likelihood estimation of location and scale of the normal distribution. *Australian & New Zealand Journal of Statistics, 35*(2), 141-153. doi: 10.1111/j.1467-842x.1993.tb01321.x

Blair, R. C., & Sawilowsky, S. (1993). Comparison of two tests useful in situations where a treatment is expected to increase variability relative to controls. *Statistics in Medicine, 12*(23), 2233-2243. doi: 10.1002/sim.4780122308

Bogler, R. (2002). Two profiles of school teachers: A discriminant analysis of job satisfaction. *Teaching and Teacher Education, 18*(6), 665-673. doi: 10.1016/s0742-051x(02)00026-4

Cressie, N. A. C., & Whitford, H. J. (1986). How to use the two sample *t*-test. *Biometrical Journal, 28*(2), 131-148. doi: 10.1002/bimj.4710280202

Croux, C., & Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *The Canadian Journal of Statistics, 29*(3), 473-493. doi: 10.2307/3316042

Curenton, S. M., McWey, L. M., & Bolen, M. G. (2009). Distinguishing maltreating versus non-maltreating at-risk families: Implications for foster care and early childhood education interventions. *Families in Society: The Journal of Contemporary Social Services, 90*(2), 176-182. doi: 10.1606/1044-3894.3871

Dalgleish, L. I. (1994). Discriminant analysis: Statistical inference using the jackknife and bootstrap procedures. Psychological Bulletin, 116(3), 498-508. doi: 10.1037/0033-2909.116.3.498

Eastman, B. J., & Brunch, S. G. (2009). Variables associated with treatment outcome in a sample of female offenders. *The Open Criminology Journal, 2*, 18-23. doi: 10.2174/1874917800902010018

Field, C., & Genton, M. G. (2006). The multivariate **g**-and-**h** distribution. *Technometrics, 48*(1), 104-111. doi: 10.1198/004017005000000562

Finch, W. H., & Laking, T. (2008). Evaluation of the use of standardized weights for interpreting results from a descriptive discriminant analysis. *General Linear Model Journal, 34*(1), 19-34.

Gnanadesikan, R., & Kettenring, J. R. (1972). Robust estimates, residuals, and other outlier detection with multiresponse data. *Biometrics, 28*(1), 81-124. doi: 10.2307/2528963

He, X., & Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis, 72*(2), 151-162. doi: 10.1006/jmva.1999.1857

Holder, B. (2007). An investigation of hope, academics, environment, and motivation as predictors of persistence in higher education online programs. *The Internet and Higher Education, 10*(4), 245-260. doi: 10.1016/j.iheduc.2007.08.002

Hubert, M., & Van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics & Data Analysis, 45*(2), 301-320. doi: 10.1016/s0167-9473(02)00299-2

117

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

Huberty, C. J., & Wisenbaker, J. M. (1992). Variable importance in multivariate group comparisons. *Journal of Educational and Behavioral Statistics, 17*(1), 75-91. doi: 10.3102/10769986017001075

Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods, 7*(3), 238-257. doi: 10.1177/1094428104266510

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

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust statistics: Theory and methods*. Hoboken, NJ: Wiley. doi: 10.1002/0470010940

McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. New York, NY: Wiley. doi: 10.1002/0471725293

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166. doi: 10.1037/0033-2909.105.1.156

Ochieng, C. O., & Zumbo, B. D. (2001). *Examination of a variable ordering index in linear regression models: An assessment of the relative Pratt index in Likert data*. Paper presented at the Bob Conry Conference on Measurement Evaluation Research and Methodology, Vancouver, Canada.

Rao, C. R. (1970). Inference on discriminant coefficients. In R. C. Bose (Ed.), *Essays in probability and statistics* (pp. 587-602). Chapel Hill, NC: University of North Carolina Press.

Rencher, A. C. (2002). *Methods of multivariate analysis* (2nd ed.). Hoboken, NJ: Wiley. doi: 10.1002/0471271357

Richardson, J. T. E. (2007). Mental models of learning in distance education. *British Journal of Educational Psychology, 77*(2), 253-270. doi: 10.1348/000709906x110557

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*(388), 851-857. doi: 10.1080/01621459.1984.10477105

Sajobi, T. T., Dansu, B. M., & Lix, L. M. (2013). Variable importance measures for non-normal data: An application to patient-reported outcomes.

Chronic Diseases and Injuries in Canada, 33(2). Retrieved from https://www.canada.ca/en/public-health/services/reports-publications/health-promotion-chronic-disease-prevention-canada-research-policy-practice/vol-33-no-2-2013/cseb-student-conference-2012-abstract-winners.html

Sajobi, T. T., Lix, L. M., Clara, I., Walker, J., Graff, L. L., Rawsthorne, P.,… Bernstein, C. N. (2012). Measures of relative importance for health-related quality of life. *Quality of Life Research, 21*(1), 1-11. doi: 10.1007/s11136-011-9914-7

Sajobi, T. T., Lix, L. M., Laverty, W. H., & Li, L. (2011). Discriminant analysis for repeated measures data: Effects of 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

SAS Institute. (2008). *SAS/IML user's guide* (Version 9.2). Cary, NC: SAS Institute.

Shih, S. (2012). An examination of academic burnout versus work engagement among Taiwanese adolescents. *The Journal of Educational Research, 105*(4), 286-298. doi: 10.1080/00220671.2011.629695

Sperling, M. R., Schilling, C. A., Glosser, D., Tracy. J. I., & Asadi-Pooya, A. A. (2008). Self-perception of seizure precipitants and their relation to anxiety level, depression, and health locus control in epilepsy. *Seizure, 17*(4), 302-307. doi: 10.1016/j.seizure.2007.09.003

Srivastava, D. K., & Mudholkar, G. S. (2001). Trimmed $\tilde{T}^2$: A robust analog of Hotelling's T^2. *Journal of Statistical Planning and Inference, 97*(2), 343-358. doi: 10.1016/s0378-3758(00)00239-1

Thomas, D. R. (1992). Interpreting discriminant functions: a data analytic approach. Multivariate Behavioral Research, 27(3), 335-362. doi: 10.1207/s15327906mbr2703_3

Thomas, D. R. (1997). A note on Huberty and Wisenbaker's "views of variable importance". *Journal of Educational and Behavioral Statistics, 22*(3), 309-322. doi: 10.3102/10769986022003309

Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research, 45*(1-3), 253-275. doi: 10.1023/A:10069540

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

Thomas, D. R., Zumbo, B. D., Zhu, P, & Dutta, S. (2008). On measuring the relative importance of explanatory variables in a logistic regression. *Journal of Modern Applied Statistical Methods, 7*(1), 21-38. doi: 10.22237/jmasm/1209614580

Todorov, V., Neykov, N., & Neytchev, P. (1994). Robust two-group discrimination by bounded influence regression. A Monte Carlo simulation. *Computational Statistics & Data Analysis, 17*(3), 289-302. doi: 10.1016/0167-9473(94)90122-8

Tonidandel, S., & LeBreton, J. M. (2010). Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Methods, 13*(4), 767-781. doi: 10.1177/1094428109341993

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

Troendle, J. F., Blair, R. C., Rumsey, D., & Moke, P. (1997). Parametric and non-parametric tests for the overall comparison of several treatments to a control when treatment is expected to increase variability. *Statistics in Medicine, 16*(23), 2729-2739. doi: 10.1002/(sici)1097-0258(19971215)16:23<2729::aid-sim694>3.3.co;2-y

Wilcox, R. R. (1994). Some results on the Tukey-McLaughlin and Yuen methods for trimmed means when distributions are skewed. *Biometrical Journal, 36*(3), 259-273. doi: 10.1002/bimj.4710360302

Appendix

Although this simulation study defined the rank ordering of variables based on the differences amongst population means, there are other methods for defining the population rank ordering, including the population discriminant function coefficients and the population Mahalanobis distance (M-distance). In the former, the rank ordering of variables is based on the population discriminant function coefficients derived from the population covariance matrix and group means. Consider two population groups (A and B) measured on p outcome variables, where μ_1 and μ_2 denote the $p \times 1$ vector of means for the population groups A and group B, respectively.

Population Discriminant Function Coefficients

The population vector of discriminant function coefficients is

$$\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)$$

where $\boldsymbol{\Sigma}$ is the $p \times p$ pooled covariance matrix for both population groups. The population variable ranking is derived by rank ordering the variables based on the magnitude of the elements of $\boldsymbol{\alpha}$.

Population M-Distance

Given the population means and pooled covariance matrix, $\boldsymbol{\Sigma}$, the M-distance for the two populations is

$$\boldsymbol{\Delta} = \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^{\mathrm{T}}$$

where $^{\mathrm{T}}$ is the transpose operator. The contribution of the kth ($k = 1,\ldots, m$) variable to the group differences can be estimated as

$$\Delta_k = \boldsymbol{\Delta} - \boldsymbol{\Delta}_{-k}$$

Where Δ_{-k} is value of Δ when the kth variable is excluded from the analyses. A variable's importance in the population can be derived by ranking according to the

magnitude of each variable's Δ_k. We present some summary results from both approaches (See Table A1 and A2).

Table A1. Average percentage per-variable correct ranking (PCR) of variable importance measures by population distribution and correlation structure

Distribution	Mean Configuration	Least Squares Estimators				Robust Estimators			
		SDFC	DRC	FTR	TDRC	SDFC	DRC	FTR	TDRC
Normal	Independent	55.36	63.76	55.59	66.34	53.36	59.28	53.56	66.48
	CS(0.3)	51.80	43.52	51.70	54.61	47.75	40.65	47.61	49.01
	CS(0.6)	55.10	25.06	56.19	46.28	48.10	24.94	48.68	39.04
	UN(0.3)	54.27	39.14	51.13	52.79	47.08	36.16	44.94	46.54
	UN(0.6)	73.72	21.39	40.57	47.97	33.88	16.56	26.02	30.49
HT-II	Independent	42.59	45.31	42.60	49.54	52.10	56.46	52.23	61.59
	CS(0.3)	39.33	37.74	39.57	42.34	46.24	41.68	46.44	50.48
	CS(0.6)	42.61	28.01	44.07	43.12	45.01	28.53	45.32	43.48
	UN(0.3)	42.84	36.38	41.62	44.23	48.36	39.35	47.79	49.15
	UN(0.6)	56.08	22.02	41.24	50.35	37.11	20.38	32.83	34.66
SK-II	Independent	33.70	34.80	33.78	36.85	51.06	54.98	51.17	59.12
	CS(0.3)	31.14	32.14	31.51	31.72	42.70	40.48	42.97	47.31
	CS(0.6)	32.98	27.47	34.55	32.97	41.14	29.09	41.43	41.64
	UN(0.3)	33.43	32.19	33.11	33.44	45.43	39.61	45.62	47.24
	UN(0.6)	44.59	22.84	35.79	39.63	36.27	22.36	33.61	34.54

Note: See Table 1 for a description of mean configurations. SDFC = standardized discriminant function coefficient; DRC = discriminant ratio coefficient; FTR = F-to-remove statistic; TDRC = total discriminant ratio coefficient; HT-II = A heavy-tailed distribution with $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-II = Skewed distribution with $\gamma_1 = 13.2$ and $\gamma_2 = 42892.9$; CS(ρ) = Compound Symmetric correlation structure with $\rho = 0.3$; UN(ρ) = Unstructured correlation with average off-diagonal correlation of ρ

Appendix II: Results Based on Population Variable Ranking Derived from Population Mahalanobis Distance

Table A2. Average percentage any-variable correct ranking (PCR) of variable importance measures by population distribution and correlation structure

Distribution	Mean Configuration	Least Squares Estimators				Robust Estimators			
		SDFC	DRC	FTR	TDRC	SDFC	DRC	FTR	TDRC
Normal	Independent	95.81	97.63	95.90	97.77	95.14	96.78	95.22	97.69
	CS(0.3)	95.01	93.65	94.89	96.35	92.97	91.18	92.85	94.52
	CS(0.6)	96.73	92.73	97.14	97.58	94.67	88.85	95.15	95.34
	UN(0.3)	95.95	90.56	95.86	97.14	93.51	89.15	92.60	95.10
	UN(0.6)	99.16	74.43	99.09	98.92	93.21	88.96	90.74	96.50
HT-II	Independent	87.58	89.02	87.59	90.45	92.08	93.60	92.17	94.54
	CS(0.3)	86.08	84.14	86.39	87.93	90.44	86.97	90.55	92.47
	CS(0.6)	89.65	82.57	90.95	90.96	92.71	86.53	93.62	94.41
	UN(0.3)	88.07	81.54	87.78	89.74	91.87	80.03	91.75	93.40
	UN(0.6)	94.66	73.33	94.34	93.67	87.86	83.87	87.97	93.82
SK-II	Independent	81.14	81.93	81.18	83.15	92.19	93.65	92.28	94.55
	CS(0.3)	78.58	78.79	79.23	78.50	88.63	84.30	88.83	90.96
	CS(0.6)	80.57	76.97	82.72	80.63	90.46	84.34	91.56	92.46
	UN(0.3)	81.24	77.14	81.24	81.22	90.55	76.69	90.54	92.21
	UN(0.6)	88.13	72.38	88.15	85.09	84.01	81.82	85.17	90.45

Note: See Table 1 for a description of mean configurations. SDFC = standardized discriminant function coefficient; DRC = discriminant ratio coefficient; FTR = F-to-remove statistic; TDRC = total discriminant ratio coefficient; HT-II = A heavy-tailed distribution with $\gamma_1 = \infty$ and $\gamma_2 = \infty$; SK-II = Skewed distribution with $\gamma_1 = 13.2$ and $\gamma_2 = 42892.9$; CS(ρ) = Compound Symmetric correlation structure with $\rho = 0.3$; UN(ρ) = Unstructured correlation with average off-diagonal correlation of ρ