Evaluation of the Use of Standardized Weights for Interpreting Results from a Descriptive Discriminant Analysis

W. Holmes Finch **Teresa Laking**
Ball State University

When conducting descriptive discriminant analysis, many researchers make use of structure coefficients, the correlation between individual predictor variables and a discriminant function. However, previous research has demonstrated that these statistics may lead to an over-identification of variables important for group separation. An alternative to structure coefficients is the standardized discriminant function weights for the individual variables, which can be used to order variables in importance. Relatively little empirical research has been done examining how well they work in this regard. This study examined the utility of standardized weights for interpreting a discriminant function. Results suggest that the standardized weights may be a useful tool for ordering predictor variables and characterizing significant discriminant functions when the assumptions of normality and homogeneity of covariance matrices are met. When these assumptions are violated, the ability of the standardized weights to correctly order predictor variables was somewhat degraded.

Discriminant Analysis (DA) is a commonly used statistical procedure that allows for both a multivariate description of group differences and the prediction of group membership for individual observations based upon a set of predictor variables. In the first context, typically referred to as Descriptive Discriminant Analysis (DDA), the focus of research is on characterizing differences between two or more groups by identifying which variables among a set of predictors most distinguishes among the groups. In contrast, the goal of Predictive Discriminant Analysis (PDA) is to use the predictors as a set for identifying which of the groups an individual is most likely to belong to. While there may be some interest in assessing the relative contribution of the variables to group separation when conducting a PDA, typically the researcher focuses on the accuracy of group prediction and its potential utility for classifying individuals in the future. It should be noted that while the goals of these two types of DA are different, the underlying mathematical model upon which they are built is the same. This model, which is discussed below, is based on the estimation of linear combinations of the predictors that provide maximal group separation for the sample at hand. The specific focus of the current study is on the utility of the standardized weights in DDA for correctly ordering a set of predictor variables in terms of their relative contribution to group separation in the form of statistically significant linear combinations. The organization of the manuscript is as follows: First is a brief description of DDA and the standardized weights used to create the linear combinations. Following this is a discussion of how these standardized weights can be used for ordering variables in terms of their importance in discriminating between groups. Finally, the details of the current simulation study are discussed, followed by the results and discussion of their implications.

As mentioned above, regardless of whether the application involves description or prediction, DA identifies one or more linear combination of the predictor variables that provide maximum group separation. The number of these linear discriminant functions is equal to the smaller of the number of predictor variables or the number of groups – 1. The relative ability of these functions to distinguish between the groups declines from the first through the second and so on. In addition, it should be noted that although it is mathematically possible to have multiple discriminant functions, in practice not all of them need be statistically significant. In other words, some of these functions may not differentiate the groups in a meaningful way. Thus, the first step in interpretation of a DDA analysis is the examination of test statistics (e.g., Wilks' Lambda) indicating which of the functions are statistically significant. Those that are found to be significant can then be interpreted using tools described below. For a more thorough discussion of the various test statistics available for use in such situations, the reader is encouraged to refer to multivariate texts such as Tabachnick and Fidell (2001).

The actual form of the discriminant function appears in equation (1).

$$D_i = d_{i1}z_1 + d_{i2}z_2 + ... + d_{ij}z_j \quad ; \qquad (1)$$

where D_i = the standardized score for an individual on discriminant function i, d_{ij} = the standardized discriminant function coefficient for function i and variable j, and z_j = the value of the standardized predictor variable j.

The discriminant weights, d_{ij}, are determined so as to provide the maximum separation possible on the function value, D_i among the groups in question(Tabachnick & Fidell, 2001). Weights are estimated for each of the *i*discriminant functions separately, and a value of D_i is obtained for each function and each individual in the sample. The means of these D_i are known as group centroids, and their relative proximity can be taken as an indication of the multivariate separation among the groups in question (Huberty & Olejnik, 2006).

The standardized discrimination coefficients take the form:

$$d_{ij} = d_{ij}^{*}\sqrt{s_j^2} \qquad (2)$$

where d_{ij}^{*} = the unstandardized discriminant coefficient and s_j^2 = the variance for variable *j*.

In turn the unstandardizeddiscriminant function coefficients are estimated by solving equation (3) below for $\mathbf{d}^{*}$.

$$(\mathbf{E}^{-1}\mathbf{H} - \lambda_1\mathbf{I})\mathbf{d}^{*} = 0 \qquad (3)$$

where **E** = the Error Sums of Squares and Cross Products matrix, **H** = the Hypothesis Sums of Squares and Cross Products Matrix, λ_1 = the maximum eigenvalue for the product of $\mathbf{E}^{-1}\mathbf{H}$, **I** = an Identity matrix and $\mathbf{d}^{*}$ = a vector of unstandardized discriminant coefficients.

These fundamental equations for DA rely upon three assumptions regarding the data in the population: 1) The predictor variables are normally distributed; 2) The covariance matrices for the groups are homogeneous; and 3) The residuals for individual subjects are independent of one another (Tabachnick & Fidell, 2001). There has been some research published regarding the impact on PDA of violating these assumptions (Finch & Schneider, 2005; Hess, Olejnik & Huberty, 2001; Meshbane & Morris, 1996: McLachlan, 1992). Taken together, results of these prior studies suggest that the accuracy of PDA in terms of correctly placing individuals in the appropriate group was negatively impacted by violations of the assumptions of normality and homogeneity of covariance matrices. Furthermore, the most negative impact was evident when both assumptions were violated simultaneously (Finch & Schneider, 2005). While these studies focused on the performance of PDA, the fact that the underlying model is essentially the same for DDA makes them relevant to the current work. Therefore, one goal of this study is to ascertain the extent to which violations of the normality and homogeneity of covariance matrices assumptions impact the standardized discriminant weights.

Those discriminant functions that have been identified as statistically significant, can be viewed as effectively differentiating the groups in question. Given such a significant outcome, a researcher would very likely want to gain an understanding as to the nature of such differences; i.e. what do each of the predictor variables contribute to the overall discriminant function that has been shown to differentiate the groups (Rencher, 1995)? There are multiple approaches that have been suggested for use in characterizing the functions based upon the contribution of the individual predictor variables, with the two most common being interpretation of the standardized discriminant coefficients and interpretation of structure coefficients (SC's) (Huberty&Olejnik, 2006).

SC's can be interpreted as the correlation between a discriminant function and the individual predictor variables upon which it is based (Huberty & Olejnik, 2006). Researchers have argued that the SC's are appropriate for interpreting DDA because they provide direct information regarding the relationship between a discriminant function that significantly differentiates among the groups and the individual predictors (Stevens, 2000). Thus, if a variable has a high SC, it can be concluded that it is highly associated with group separation. It should be noted that SC's in DDA are similar in concept to factor loadings, which are used routinely in characterizing the nature of latent factors (Huberty&Olejnik, 2006). Therefore, it may be reasonable to use them for characterizing discriminant functions in much the same fashion.

Because they are similar (though not identical) to factor loadings, some authors have suggested applying arbitrary cut-off values for identifying "important" variables, as is commonly done in exploratory factor analysis. For example, Tabachnick and Fidell (2001) have recommended that SC's larger than 0.32 be considered "important" in terms of understanding the nature of the discriminant function. They selected this value because 0.322 is roughly 0.1, indicating that 10% of the variance in the predictor variable is accounted for by the discriminant function. Pedhazur (1997) recommended using a cut value of 0.3, while other authors (e.g., Huberty & Olejnik, 2006; Stevens, 2000) suggest that researchers not use a single value, but rather focus on the relative magnitude of the SC's, placing greater emphasis on interpreting those variables with larger values. Dalgleish (1994) introduced a bootstrap confidence interval for use with SC's in DDA. He hoped that this approach would obviate the need for applying arbitrary cut off values by providing information regarding whether, in the population, a given SC differs from 0. If this were the case, Dalgleish argued that a practitioner could then know, with some level of confidence, that a given predictor variable was associated with a significant discriminant function.

Researchers have studied the effectiveness of SC's for interpreting significant discriminant functions in DDA. For example, Dalgleish (1994) found that the bootstrap confidence intervals that he developed for SC's had somewhat conservative Type I error rates, but generally did a better job at maintaining Type I error near the nominal 0.05 level than did arbitrary cut values, including 0.3, 0.4 and 0.5. Finch (2007) conducted a simulation study examining both the Type I error rates (incorrectly identifying a predictor variable as "important" in group separation) and power (correctly identifying a predictor variable as "important" in group separation) of various methods for interpreting SC's, including cut values (0.3, 0.4 and 0.5), relative ordering of importance and the bootstrap confidence interval. Results of this study indicated that in general, the use of SC's led to an over identification of variables associated with group separation. In other words, a researcher using any of these approaches for interpreting SC's could expect to conclude that one or more variables are related to the significant discriminant function when in fact they are not. In addition, the Finch study reported that when the assumptions of normally distributed predictors with equal covariance matrices across groups were violated, the Type I error inflation was particularly severe.

Some researchers have long advocated against using SC's for interpreting significant discirminant functions, and in favor of the standardized weights described above (e.g., Rencher, 1992). The argument in favor of this approach, set forth by Rencher (1995), is that the standardized weight for a particular variable reflects its contribution to the discriminant function in the presence of the other predictors. On the other hand, Rencher argued that the SC relating this variable to the discriminant function demonstrates only the univariate contribution of the individual predictor in question, totally ignoring the presence of the others. For this reason, he asserted that "…these correlations are useless in gauging the importance of a given variable in the context of others because they provide no information about how the variables contribute jointly to separation of the groups. Consequently, they become misleading if used for interpretation of discriminant functions" (Rencher, 1995, p. 317). Instead, he argued on behalf of referring to the discriminant weights when interpreting DDA, because they do account for all of the variables in the model and are therefore more appropriate when one is interested in characterizing significant discriminant function results.

This opinion that standardized weights are more appropriate than SC'sfor use in interpreting discriminant functions is not universally shared. Huberty and Wisenbaker (1992) objected to using the weights because, they stated, simply ordering variables in importance does not communicate anything regarding the different degrees of variable importance, only that one is more important than another. Huberty and Olejnik (2006) go on to argue against the notion of ordering variables in terms of relative importance as a generally useful exercise, and instead focus on characterizing the discriminant function by ascertaining which of the predictor variables were most highly correlated with it, based on the SC's.

Clearly, given the discussion above, the disagreement between methodologists regarding the appropriate approach for interpreting significant discriminant functions has not been resolved to date. In addition to the studies described above that focused on SC's, Huberty (1975) also conducted a simulation study in which he compared the ability to identify predictor variables relevant to group separation of standardized weights and SC's. The outcome variable in this study was the consistency of variable ranking in terms of relative contribution to a significant discriminant function. The data were generated from a normal distribution with equal covariance matrices across 3 and 5 groups for 10 predictor variables. Sample sizes were set at 90, 150, 300 and 450. Huberty concluded that in the 5 groups case, the SC's were slightly more effective at ordering the predictors, while in the 3 groups case the standardized weights performed slightly better in this regard. As he stated, these results are limited to the case where groups are of equal size and the assumptions of equality of covariances and normality are met.

In contrast to the Huberty study, Rencher (1992) described analytically why there may be problems with using the SC's to interpret discriminant functions, and in turn why the standardized weights might be preferable. As noted above, he showed that in the 2-groups case, the SC'sare mathematically proportional to the univariate t-test comparing the means on the predictor variable between the two groups. Thus, he argued, a researcher making use of the SC's has simply taken what is inherently a multivariate problem and reduced it to a series of univariate ones (Rencher, 1992). Rencher concluded his paper by stating that standardized weights, rather than SC's, are most appropriate for interpreting significant discriminant functions because they allow for a direct ordering of individual predictors in terms of importance while accounting for the presence of all of the other predictors.

Based upon prior research examining the performance of SC's (Finch, 2007, Dalgleish, 1994) there remain some doubts regarding their effectiveness in helping researchers interpret significant discriminant functions. Specifically, regardless of the rule used, SC's appear to over-identify the importance of individual variables in terms of their contribution to group separation. In addition, based upon Rencher's (1992) arguments, these SC values may not be addressing the appropriate multivariate question, namely which variables contribute the most to group separation, in the presence of the other variables in the analysis? Given these potential problems with SC's described by Rencher and highlighted in prior simulation studies, and the relative lack of Monte Carlo research examining the performance of standardized weights in characterizing group differences in DDA, the primary goal of the current study was to use simulations to ascertain how well the standardized weights could order variables in terms of relative importance in group separation under a variety of conditions, which are outlined below. It is hoped that this effort will add to the literature regarding interpretation of DDA and provide some additional guidance to researchers in the field. The performance of the standardized weights was measured in terms of how well they ordered predictors with varying degrees of between group difference, and what aspects of the data might impact this ordering.

Methods

This Monte Carlo simulation study involved the manipulation of a number of data conditions in order to identify factors influencing the utility of standardized weights for correctly ordering variables based on their relative importance in defining the discriminant function. All analyses were conducted with 2 groups and 6 predictor variables using the SAS software system, version 9.1 (SAS, 2005) PROC DISCRIM. Initially, standardized weights based on both the total and within groups covariance matrices were estimated and retained for further investigation. However, subsequent analysis of the results demonstrated that across all conditions manipulated in this study, the performance of the two types in terms of variable ordering was virtually identical. For this reason, outcomes are reported only for the weights based on the total sample covariance matrices. The manipulated conditions described below were completely crossed with another.

Distribution of the Predictor Variables

The predictor variables were simulated to be normal or non-normal with skewness of 1.75 and kurtosis of 3.75. In order to maintain the desired levels of correlation (described below) among these predictors, the approach for simulating data described by Headrick and Sawilowsky (1999) were employed. These values of skewness and kurtosis were selected because they have been shown to impact the performance of discriminant analysis (Hess, Olejnik, & Huberty, 2001).

Homogeneity of groups' covariance matrices

In addition to the normality of the predictors, a second major assumption underlying DA is the homogeneity of group covariance matrices. Therefore, in order to evaluate the performance of the standardized weights under a range of conditions, the covariance matrices were manipulated to be either equal or unequal. In this study, inequality of covariance matrices was simulated with one group having variances for the predictors that were 5 times larger than that of the other group.

Sample Size

Total sample sizes took four different values across the simulations: 30, 60, 100 and 150. These values correspond to values seen in the applied DA literature (e.g., Glaser, Calhoun, & Petrocelli, 2002; Russell & Cox, 2000; Matters & Burnett, 2003). They represent conditions from small to moderately large samples.

Sample Size Ratio

Three conditions for relative group size were used. In the first condition, the two groups were simulatedwith equal numbers of subjects. In conditions two and three, sample sizes were different such that the larger group had twice the number of subjects as the smaller. In condition two, group 1 had the larger sample size, while in condition three group 2 was the larger. Sample size ratio was completely crossed with covariance matrix equality/inequality. Therefore, in one set of conditions, the larger group had the larger variance while in another the smaller group had the larger variance. In the third combination, the covariances were equal, even as group size ratios were unequal. It was believed that examining the combination of sample size ratio and covariance matrix equality was important to examine because of previous evidence that the interaction of unequal sample sizes and unequal group covariance matrices has an impact on the performance of PDA (Finch & Schneider, 2005).

Group Separation

Separation between the two groups was simulated using Cohen's d, univariate effect size (Cohen, 1988). Table 1 contains the pattern of mean differences for the various combinations of effect sizes. The data were simulated so that group 2 had a mean of 0 and standard deviation of 1 for all of the predictors, while the predictor values for group 1 were generated using the means displayed in Table 1, for each condition respectively. For example, in the 8/0 condition, group 1 had a mean of 0.8 on the first predictor, and means of 0 on the other five, while data for group 2 were generated with means of 0 on all six predictors.

Table 1. Differences (in Cohen's *d*) in Predictor Means between Group 1 and Group 2.

Predictor Variable						Condition Label
X_1	X_2	X_3	X_4	X_5	X_6	
0.5	0	0	0	0	0	5/0
0.8	0	0	0	0	0	8/0
0	0.5	0.5	0.5	0.5	0.5	0/5
0	0.8	0.8	0.8	0.8	0.8	08
3.0	2.5	2.0	1.5	1.0	0.5	5/5
4.8	4.0	3.2	2.4	1.6	0.8	8/8
0.8	0.5	0.5	0.5	0.5	0.5	8/5
0.5	0.8	0.8	0.8	0.8	0.8	5/8

Correlation Between the Predictor Variables

The correlations among the predictors were manipulated at three levels: 0.3, 0.5 and 0.8. In order to maintain these correlations even as the skewed distribution was simulated, the methodology outlined by Headrick and Sawilowsky (1999) was used.

The outcome of interest in this study was the degree to which standardized weights provided correct information regarding the order of importance of predictor variables in terms of group separation. It would be expected that the absolute value of these weights should be larger for those variables associated with greater group separation (Rencher, 1995). Thus, in the context of this study, the weights associated with larger values of Cohen's d should themselves be larger than those weights associated with smaller effect sizes. The specific outcome in this study then, is the proportion of cases across simulation replications in which, for adjacent pairs of variables, the predictor associated with the larger effect size (greater group separation) had the larger standardized weight. In the cases where predictor effect sizes were the same, we anticipate the standardized weights for one of the variables in a pair will be higher than the other roughly half of the time. The appendix contains an S-PLUS program that generated the simulated data for the study. The *rnorm* function in S-PLUS generated 100 random normal data points and output nine variables listed in the data command [data <-c(y,x,z,gain,ypost,xc,xz,xsq,xsqz)]. The post test scores (*Y*) and pre test scores (*X*) were created by adding residual error (*ey* or *ex*) to this random normal variable (*true*). Group assignment (*Z*) was determined based on subtracting a cut score of 20 from the pre test score (1–treatment, 0–comparison). This 10 point treatment gain was added to the post test score (*Y*). Optional *print* and *write* statements are included to either view or save the data in a file.

Results

As mentioned above, results for the total and within groups weights were essentially identical across conditions, therefore the discussion henceforth will focus only on the performance of the total values. In addition, an examination of results revealed virtually identical outcomes regardless of the sample size ratios simulated. Therefore, in order to limit the length of the manuscript unnecessarily, this variable will also not be included in the following discussion of results. The results are organized by the assumptions of normality and homogeneity of covariance matrices. As stated above, the outcome of interest in this case was the proportion of cases in which the standardized weights correctly reflected the variables' order of importance in terms of group separation. For example, referring to Table 1, in the 5/0 condition, the weight for the first variable should be larger than the weights for the other variables, while the weights for variables 2-6 should be equal (within sampling error) so that no one of these should consistently be larger than the others.

Normal Distribution, Homogeneous Covariance Matrices

Table 2 reflects the results for the case when both assumptions of normality and homogeneous covariance matrices were met, by effect size and correlation among the predictors. Across correlations, when variable 1 had the null effect size and variable 2 did not (05, 08 conditions), the proportion of cases in which variable 2 correctly had the larger standardized weight was greater than 0.90, and increased concomitantly with the correlation value. In contrast, when only the first variable was associated with group separation (80, 50 conditions), the first weight was correctly larger than that of variable 2 at much lower rates. Indeed, for correlations of 0.5 and 0.8, the first weight was correctly larger in less than 30% of the simulation replications.

The proportion of cases in which the standardized weight of the first variable was correctly larger than that of the second in the 88 and 55 conditions, where all of the variables were involved in group separation though to a different degree, was much higher than when only variable 1 differed between groups. Furthermore, as with the 05 and 08 cases, the proportion of replications where the first standardized weight was larger than the second increased along with the correlation among predictors, with the exception of the 55 case for a correlation of 0.8. Finally, when considering the 85 and 58 conditions, the standardized weights were better able to order the variables in the latter case versus the former. In other words, the proportion of cases displaying correct ordering was greater when the second variable had the larger effect size, as opposed to when the first variable had the larger effect size.Note that this outcome follows a very similar pattern to the 05/08 versus 50/80, where variable ordering was correct more frequently in the former than the latter. The proportion of correct ordering outcomes increased with increasing correlation, except for the 85 condition with $r = 0.8$.

When considering the comparisons of the standardized weights for the adjacent pairs in variables 2 through 6, it is important to remember that these variables were all simulated with the same effect size values separating the groups, except for the 88 and 55 conditions. Thus, we would expect them to have very similar standardized weight values across simulation replications. In fact, results for the 80, 50, 08, 05, 85, and 58 conditions revealed that the proportion of times the weights for one of these variables was larger than that of the adjacent one was very close to 0.5 in all cases, indicating that they were comparable in size across replications. Given the similarity of these results in the expected way, the data presented in the tables for variables 2 through 6 only includes rates for the 88 and 55 conditions, where effect size values were not uniform. It is hoped that the tables will more clearly display relevant outcomes that are not obscured by a large number of redundant results.

In general, for both the 88 and 55 conditions it appears that the proportion of cases exhibiting a correct ordering of standardized weights declined somewhat for variable pairs further down the list (e.g., X_3 vs. X_4, X_4 vs. X_5, etc.).For example, in the 88 condition the weight for variable 2 was correctly larger than that of variable 3 at rates comparable to those for the variable 1 versus 2 comparison. In contrast, for the final adjacent pair in the set, variable 5 correctly had a larger standardized weight than variable 6 at lower rates, generally differing by between 0.06 and 0.10 for different values of r. The rate of correct ordering by the standardized weights was higher for larger correlation values, with the exception of the 55 condition with $r = 0.8$.

Table 2. Proportion of cases in which variable ordering is correct based on ldf weights, by effect size and correlation among predictors: ***Normal distribution and homogeneous covariance matrices***.

r	Effect Size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
0.2	05	0.912				
	08	0.966				
	50	0.578				
	80	0.650				
	58	0.789				
	85	0.809				
	88	0.853	0.860	0.844	0.797	0.761
	55	0.824	0.829	0.797	0.784	0.769
0.5	05	0.945				
	08	0.984				
	50	0.293				
	80	0.227				
	58	0.849				
	85	0.831				
	88	0.885	0.894	0.862	0.827	0.777
	55	0.863	0.866	0.839	0.814	0.803
0.8	05	0.985				
	08	0.999				
	50	0.053				
	80	0.021				
	58	0.940				
	85	0.626				
	88	0.895	0.914	0.883	0.838	0.801
	55	0.603	0.611	0.609	0.588	0.582

Table 3 displays the proportion of correctly ordered variables by effect size and sample size when the assumptions of normality and homogeneity of covariances were met. In general, the pattern of results across effect sizes was very similar to those described above. The proportion of cases correctly ordered for the first 2 variables increased concomitantly with sample size, except for the 50 and 80 conditions. In other words, when only the first variable was simulated to be different between the groups, the proportion of times that the standardized weight for variable 1 was larger than that of variable 2 declined as sample size increased. With respect to the comparisons among the adjacent pairs for variables 2 through 6, the proportion of correctly ordered pairs declined for variables further down the list. In addition, the rate of correct ordering improved with larger sample sizes. Indeed, for a total sample size of 150, the standardized weights were ordered correctly in more than 80% of cases for all adjacent pairs. Even for a sample size of 100, the lowest proportion of accurately ordered pairs was 0.774 for variables 5 and 6 in the 55 condition.

Normal Distribution, Heterogeneous Covariance Matrices

Results for the case where the predictors were simulated to be normally distributed and the covariance matrices between the groups were heterogeneous appear in Tables 4 and 5. Across correlation conditions (Table 4), the proportion of correctly ordered weights was lower than when both assumptions were met (Table 2). The lone exception to this result was for correlations of 0.5 and 0.8 in conjunction with the 50 and 80 effect size conditions, where the proportion correctly ordered was somewhat higher when the covariance matrices were heterogeneous. It should be noted, however, that in general, for the 50 and 80 cases the proportion of correctly ordered weights remained low. The most dramatic reduction in the proportion of correct ordering for the normally distributed heterogeneous covariance case occurred in

Table 3. Proportion of cases in which variable ordering is correct based on ldf weights, by effect size and sample size: ***Normal Distribution and Homogeneous Covariance Matrices***.

N	Effect Size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
30	05	0.860				
	08	0.946				
	50	0.365				
	80	0.332				
	58	0.753				
	85	0.636				
	88	0.758	0.779	0.750	0.704	0.682
	55	0.688	0.702	0.681	0.655	0.652
60	05	0.951				
	08	0.990				
	50	0.314				
	80	0.288				
	58	0.837				
	85	0.754				
	88	0.869	0.879	0.842	0.801	0.756
	55	0.772	0.781	0.748	0.731	0.717
100	05	0.981				
	08	0.998				
	50	0.279				
	80	0.263				
	58	0.900				
	85	0.833				
	88	0.927	0.931	0.912	0.872	0.813
	55	0.829	0.826	0.808	0.783	0.774
150	05	0.994				
	08	0.999				
	50	0.258				
	80	0.241				
	58	0.936				
	85	0.874				
	88	0.965	0.971	0.947	0.913	0.864
	55	0.863	0.863	0.847	0.830	0.813

the 88 and 55 effect size conditions with $r = 0.8$. When the data were normally distributed with heterogeneous covariance matrices, the proportion of correctly ordered cases dropped by approximately 0.35 to 0.45 for all adjacent pairs of variables, as compared to the normal homogeneous case.

With respect to the impact of sample size for the normal distribution and heterogeneous covariance condition, results in Table 5 suggest that larger samples did ameliorate the negative impact of heterogeneous covariance matrices for some effect size combinations, but not others. For example, when groups differed on all but the first variable (05, 08), the proportions of correctly ordered standardized weights in Table 5 become very similar to those in Table 3 for samples of 100 and particularly 150. On the other hand, when group separation was isolated in the first variable only (50, 80), the proportion of correctly ordered ldf weights declined with increasing sample size, a pattern also apparent in Table 3. In the other effect size conditions simulated in this study, a larger sample size was associated with improved accuracy in ordering the variables, though the rates did not match those found when both assumptions of normality and homogeneity of variance were satisfied.

Table 4. Proportion of cases in which variable ordering is correct based on ldf weights, by effect size and correlation among predictors: ***Normal Distribution and Heterogeneous Covariance Matrices***.

r	Effect Size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
0.2	05	0.798				
	08	0.904				
	50	0.522				
	80	0.574				
	58	0.706				
	85	0.708				
	88	0.804	0.803	0.773	0.752	0.736
	55	0.762	0.749	0.741	0.722	0.720
0.5	05	0.843				
	08	0.940				
	50	0.353				
	80	0.304				
	58	0.758				
	85	0.688				
	88	0.838	0.838	0.816	0.785	0.779
	55	0.778	0.772	0.759	0.746	0.741
0.8	05	0.934				
	08	0.979				
	50	0.131				
	80	0.055				
	58	0.854				
	85	0.519				
	88	0.429	0.443	0.453	0.434	0.430
	55	0.203	0.225	0.229	0.236	0.244

Non-Normal Distribution, Homogeneous Covariance Matrices

The third combination of conditions to be examined in this study was the non-normal, homogeneous covariance case. One pattern of results apparent across values of the correlation was that the proportion of correctly ordered weights in the X_1 versus X_2 comparison was higher in the non-normal homogeneous covariance condition than for the normal heterogeneous covariance condition when the first variable was associated with a larger group difference (50, 80, 88, 55). The lone exception to this pattern was the 85 condition, in which the first variable was associated with a large effect while the other variables were associated with a medium effect. Conversely, when the first variable was associated with a null effect size (05, 08) as well as in the 58, 85 cases, the proportion of correct ordering was lower in the non-normal, homogeneous covariance situation. In general, the proportion of correctly ordered standardized weights was lower than when both assumptions were met.

With respect to the adjacent variable comparisons other than X_1 versus X_2, the proportion of correctly ordered weights was somewhat higher earlier in the sequence for the non-normal homogeneous case as compared to the normal heterogeneous data, and somewhat lower for X_4 versus X_5 and X_5 versus X_6. In addition, the sharp decline in accuracy that occurred in the normal heterogeneous case for $r = 0.8$ was not in evidence in the non-normal, homogeneous case. With the exception of the 50 and 80 conditions, the proportion of correctly ordered standardized weights increased with increasing sample sizes in Table 7. In addition, for the 88 and 55 effect size cases, the proportion of correctly ordered weights was comparable or slightly higher in this condition than when both assumptions were met.

Table 5. Proportion of cases in which variable ordering is correct based on ldf weights, by effect size and sample size: ***Normal Distribution and Heterogeneous Covariance Matrices***

N	Effect Size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
30	05	0.719				
	08	0.844				
	50	0.390				
	80	0.375				
	58	0.670				
	85	0.545				
	88	0.650	0.665	0.648	0.621	0.618
	55	0.570	0.570	0.576	0.568	0.567
60	05	0.837				
	08	0.945				
	50	0.363				
	80	0.317				
	58	0.744				
	85	0.611				
	88	0.710	0.718	0.703	0.675	0.665
	55	0.772	0.621	0.607	0.598	0.595
100	05	0.912				
	08	0.980				
	50	0.312				
	80	0.281				
	58	0.790				
	85	0.688				
	88	0.762	0.757	0.742	0.716	0.709
	55	0.656	0.651	0.635	0.630	0.628
150	05	0.951				
	08	0.994				
	50	0.294				
	80	0.264				
	58	0.862				
	85	0.751				
	88	0.788	0.782	0.764	0.743	0.732
	55	0.680	0.676	0.669	0.655	0.655

Non-normal Distribution, Heterogeneous Covariance Matrices

This combination of conditions represents the situation where neither of the foundational assumptions underlying DA were met. Table 8 reveals that across nearly all conditions the ordering of the standardized weights was correct at markedly lower rates than when both assumptions were met (Table 2). The only exceptions to this pattern were for the 50 and 80 cases, when all group difference was isolated in the first variable only. The pattern of declining accuracy for variables entered later in the equation that was evident in the other distribution and covariance conditions was also apparent when neither assumption was met. In fact, the relative decline in accuracy rates for adjacent pairs further down the sequence was greater in this condition than when both assumptions were met. Larger correlations among the predictors were associated with greater accuracy rates for the 88 and 55 conditions particularly, for the X_1 versus X_2, X_2 versus X_3 and X_3 versus X_4 adjacent pairs. However, for the X_4 versus X_5 and X_5 versus X_6 variable pairs, the proportion of correctly ordered standardized weights actually declined with increasing correlation values.

Table 6: Proportion of cases in which variable ordering is correct based on ldf weights, by effect size and correlation among predictors: ***Non-Normal Distribution and Homogeneous Covariance Matrices***

r	Effect Size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
0.2	05	0.579				
	08	0.734				
	50	0.527				
	80	0.617				
	58	0.682				
	85	0.683				
	88	0.869	0.836	0.800	0.754	0.637
	55	0.852	0.817	0.775	0.726	0.611
0.5	05	0.571				
	08	0.703				
	50	0.506				
	80	0.575				
	58	0.697				
	85	0.629				
	88	0.907	0.871	0.830	0.733	0.490
	55	0.897	0.853	0.809	0.707	0.472
0.8	05	0.577				
	08	0.693				
	50	0.470				
	80	0.472				
	58	0.763				
	85	0.490				
	88	0.905	0.881	0.812	0.616	0.220
	55	0.665	0.658	0.609	0.514	0.366

The total sample size appears to have been associated with standardized weight ordering accuracy for only some of the effect size combinations when the data were not normally distributed and covariance matrices were not equal between groups. Specifically, from Table 9 when the first variable accounted for more of the group separation (50, 80, 85, 88 and 55 effect size combinations) the proportion of correctly ordered weights for the X_1 versus X_2 comparison increased concomitantly with sample size. This increase in accuracy was most notable in the 88 and 55 cases. For adjacent pairs other than X_1 and X_2, there was a clear positive relationship between sample size and weight ordering accuracy for the X_2 versus X_3 and X_3 versus X_4 comparisons. On the other hand, for the last two pairs in the sequence, there appears not to have been this positive relationship between sample size and the accuracy rate.

The goal of this Monte Carlo study was to examine the potential utility of standardized weights for ordering predictor variables in terms of their relative importance in defining a significant discriminant function. Prior simulation research has found that other methods for characterizing group separation in DDA, such as the use of SC's, may be less than optimal in many situations. Thus, the current research was designed to ascertain how effective an alternative the standardized weights might be for this purpose. The study conditions were selected so as to replicate those in earlier studies that focused on SC's, and the outcome of interest was the proportion of cases in which the weights correctly ordered the variables in terms of their relative importance in separating two groups.

Table 7: Proportion of cases in which variable ordering is correct based on ldf weights, by effect size combination and sample size: ***Non-Normal Distribution and Homogeneous Covariance Matrices***

N	Effect size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
30	05	0.511				
	08	0.570				
	50	0.502				
	80	0.534				
	58	0.609				
	85	0.532				
	88	0.780	0.744	0.694	0.617	0.468
	55	0.711	0.681	0.644	0.588	0.488
60	05	0.555				
	08	0.673				
	50	0.501				
	80	0.538				
	58	0.687				
	85	0.578				
	88	0.879	0.836	0.782	0.681	0.447
	55	0.792	0.758	0.707	0.635	0.483
100	05	0.599				
	08	0.772				
	50	0.502				
	80	0.572				
	58	0.762				
	85	0.620				
	88	0.943	0.916	0.855	0.736	0.440
	55	0.840	0.814	0.763	0.672	0.483
150	05	0.638				
	08	0.825				
	50	0.500				
	80	0.576				
	58	0.798				
	85	0.671				
	88	0.974	0.954	0.915	0.772	0.441
	55	0.876	0.851	0.810	0.699	0.484

Discussion

The results described above indicated that under some conditions, the standardized weights did indeed provide an accurate ordering of the predictor variables, particularly when both the assumptions of normality and homogeneity of covariance matrices were met. These accuracy rates were frequently over 90% for samples of 100 and 150 subjects. Furthermore, the ordering accuracy rates for all adjacent pairs improved when the correlations among the predictors increased in several of the conditions simulated here. The major exception to these positive results when both assumptions were met occurred when group separation was only present for the first predictor variable. In this case, the accuracy rates were much lower than for the other conditions, and they declined with increasing correlations among the variables. In other words, when the group difference was truly univariate in nature and centered in the first variable, the standardized weight for the second variable was frequently (incorrectly) larger than that of the first. Finally, the accuracy of the standardized weight ordering approach was somewhat higher for variable pairs earlier in the sequence, even though the relative difference in group separation later in the

Table 8. Proportion of cases in which variable ordering is correct based on ldf weights, by effect size and correlation among predictors: Non-Normal distribution and heterogeneous covariance matrices

r	Effect Size	X_1 vs X_2	X_2 vs X_3	X_3 vs X_4	X_4 vs X_5	X_5 vs X_6
0.2	05	0.503				
	08	0.484				
	50	0.547				
	80	0.587				
	58	0.510				
	85	0.518				
	88	0.759	0.714	0.648	0.565	0.439
	55	0.648	0.596	0.557	0.525	0.444
0.5	05	0.487				
	08	0.445				
	50	0.569				
	80	0.634				
	58	0.509				
	85	0.527				
	88	0.821	0.757	0.679	0.527	0.303
	55	0.708	0.640	0.576	0.473	0.371
0.8	05	0.488				
	08	0.440				
	50	0.597				
	80	0.695				
	58	0.479				
	85	0.569				
	88	0.888	0.830	0.698	0.430	0.167
	55	0.793	0.726	0.596	0.433	0.253

sequence was identical. For example, Table 1 shows that the difference between group means for variable 2 was simulated to be 4.0 in the 88 effect size case, while the difference for variable 3 was simulated to be 3.2. Thus the difference in conditions was 0.8 (4.0-3.2). The difference between group means for variable 4 was simulated to be 2.4, which was 0.8 units different from the group separation for variable 3. However, the proportion of correctly ordered weights for variable 2 versus variable 3 was greater than that for variable 3 versus variable 4 across correlation conditions. A similar pattern was evident for the other adjacent variable pairs further down the sequence.

In general, the results of this study demonstrated that when the assumptions of normality and/or homogeneity of covariance matrices were not met, the standardized weights were less accurate in ordering predictor variables based on their relative importance in group separation. The performance of these weights was generally most degraded when neither assumption was met. The lone exception to this last pattern occurred when the predictors were normally distributed but the covariance matrices were unequal and the correlation among the predictors was 0.8. In this case, the ordering accuracy rates were well below 50% for both the 88 and 55 effect size conditions. Under most conditions where one or both of these assumptions were unmet, larger sample sizes served to mitigate problems with ordering accuracy to some extent, though rarely did accuracy match that when both assumptions were met. The positive impact of increased sample sizes was particularly evident when the data were non-normal. Indeed, in the 88 and 55 effect size conditions, the accuracy rates were comparable (or nearly so) to the normal, homogeneous covariance case for both of the non-normal situations when the sample size was 150. It should also be noted that when the first variable was not associated with group separation (08, 05) the accuracy rates in the normal distribution, heterogeneous covariance condition were higher than when the data were not normally distributed, and for samples of 100 and 150 were above 0.9.

Implications for Practice

Some authors (e.g., Rencher, 1995) have recommended that researchers using DDA to differentiate two or more groups in the multivariate case consider relying on these standardized weights to characterize the nature of the significant discriminant functions. Rencher (1992) argued that they are superior to other tools, such as SC's, because they incorporate information about all of the variables in the analysis, rather than simply reproducing univariate analyses. The results of this study appear to support the potential utility of these standardized weights for characterizing multivariate group differences in some situations, but not others. Following are some potential implications for practice based on results discussed above. It should be noted that guidelines for what would be acceptable performance are not available. Ideally, of course, the rates of correct variable ordering would be 100%, though such a perfect outcome would be unlikely for any statistical procedure. Rather than select an arbitrary cut off for what is acceptable performance, we have elected in this manuscript to discuss the rates in relative terms and allow readers to make their own judgments regarding the acceptability (or not) of the standardized weights' performance.

First of all, it does appear that when the assumptions of normality and group homogeneity of covariance matrices are both satisfied, variables are accurately ordered in terms of relative contribution to group separation at rates above 80% when the sample size is 100 or greater and the group differences are multivariate in nature (all effect size conditions except for 80 and 50). Indeed, when the sample size was as least 60 and all the variables were associated with group separation, the standardized weights would accurately order variables 1 and 2 in importance more than 80% of the time, except when the second variable was associated with a moderate effect and the first was associated with a moderate or large effect (85, 55 conditions).

While performance of the weights in variable ordering was often relatively goodwhen the groups were separated on multiple predictors (and the foundational assumptions were met), in cases where the groups only differed on one variable (the first in the sequence in this study), theydid not accurately reflect this fact very well, regardless of sample size. This problem was more acute when the predictor variables were more highly correlated. Therefore, researchers using DDA should carefully consider the variables that they have selected as predictors so that any significant group differences not beunivariate in nature. Furthermore, if results of the analysis appear to indicate that the groups differ on only one variable, the researcher should be very careful when interpreting variable ordering with these standardized weights.

When the predictor variables do not conform to the assumptions of normality and homogeneity of covariance matrices, researchers should also exercise caution when using standardized weights to interpret discriminant functions. Results of this study suggest that when the predictor variables are not normally distributed and/or the group covariance matrices are not equal, the weights may frequently order the variables incorrectly in terms of their relative importance, particularly when both assumptions are violated simultaneously. Therefore, researchers considering the use of these weights for characterizing the nature of significant group separation should be very careful to check these assumptions. If they do not hold, the weights may not be appropriate for ordering the variables. It is important to note that larger overall sample sizes do not fully ameliorate this problem.

A fourth implication of these results is that the correlations among the predictor variables have an impact on the performance of standardized weights when the assumptions of normality and homogeneity of covariances are met. In general, higher correlations among the predictors were typically associated with more accurate ordering based on the standardized weights. The lone exception to this outcome occurred when only the first variable was associated with group difference, in which case higher correlations resulted in the weight of the second variable (not different between groups) being larger (incorrectly) than that of the first, at very high rates. Researchers considering the use of standardized weights for interpreting DDA thus need to be cognizant of these correlations. If they select a number of variables that have relatively low correlations, they may have more difficulty in correctly identifying which of these is most associated with the significant discriminant function, and the associated group differences. It is also interesting to consider this result in light of Rencher's (1992) argument in favor of using standardized weights: namely that they account for the presence of the other predictors in the model. The fact that performance generally improved with higher correlations appears to validate this earlier observation.

Finally, when compared with results of earlier simulation research examining the SC's as a tool for interpreting discriminant functions, the standardized weights appear to perform favorably. Finch (2007) reported very high rates (often in excess of 0.5) of incorrect identification of "important" variables using these SC's. In addition, under several data conditions similar to those included in this study, rates of correct identification of such "important" variables were not higher than those reported here for the standardized weights. Therefore, given the high Type I error rates for the SC's, along with the comparable power, it would appear that the standardized weights may prove to be a worthwhile alternative for interpreting significant discriminant functions.

Limitations and Directions for Future Research

Future studies should be designed to improve on the current research. For example, results described in this manuscript are limited to the two groups case. Thus, one logical next step in this area is to examine the utility of standardized weights for differentiating among more than two groups. By including multiple groups, interpretation of more than one significant discriminant function would also be possible.

A second area for future research is the examination of the performance of standardized weights for a different set of effect size combinations. In the current study, most of the differences among the predictors with respect to group separation were between variable 1 and the others. With the exception of the 88 and 55 conditions, variables 2 through 6 were associated with the same effect size difference between the groups. Future studies should use a different variety of such group differences in order to provide a more complete understanding of the effectiveness of the weights for ordering the predictor variables.

Future studies in this area should also examine a different set of non-normal distributions for the predictors. While this is the first study in this area to use non-normal data, generalizations of the results herein are limited to those non-normal cases where the predictors have skewness of 1.75 and kurtosis of 3.75. For example, some research has shown that a related statistical analysis, Multivariate Analysis of Variance (MANOVA), is impacted by variables with truncated tails (e.g., Finch, 2005). Thus, it seems reasonable that DDA, which is based upon the same multivariate linear model, might also experience problems with such a distribution.

References

Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Dalgleish, L.I. (1994). Discriminant Analysis: Statistical inference using the Jackknife and Bootstrap procedures. *Psychological Bulletin, 116,* 498-508.

Finch, W.H. (2007). *Identification of variables associated with group separation in Descriptive Discriminant Analysis: Comparison of methods for interpreting Structure Coefficients*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Finch, W.H. (2005). Comparison of the performance of the nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology, 1,* 27-38.

Finch, W.H. & Schneider, M.K. (2005). Misclassification rates for four methods of group classification: Impact of predictor distribution, covariance inequality, effect size, sample size and group size ratio. *Educational and Psychological Measurement, 66,* 240-257.

Glaser, B.A., Calhoun, G.B., & Petrocelli, J.V. (2002). Personality characteristics of male juvenile offenders by adjudicated offenses as indicated by the MMPI-A. *Criminal Justice Behavior, 29,* 183-201.

Headrick, T.C. & Sawliowsky, S.S. (1999). Simulating correlated multivariate nonnormal distributions: Extending the Fleishman power method. *Psychometrika, 64,* 25-35.

Hess, B., Olejnik, S., &Huberty, C.J. (2001). The efficacy of two improvement-over-chance effect sizes for two-group univariate comparisons under variance heterogeneity and nonnormality. *Educational and Psychological Measurement, 61*, 909-936.

Huberty, C.J. (1975). The stability of three indices of relative variable contribution in discriminant analaysis. *Journal of Experimental Education, 2,* 59-64.

Huberty, C. J. & Olejnik, S. (2006). *Applied Manova and Discriminant Analysis.* New York: Wiley.

Huberty, C.J. & Wisenbaker, J.M. (1992). Variable importance in multivariate group comparisons. *Journal of Educational Statistics, 17,* 75-91.

McClachlan, G.J. (1992). *Discriminant Analysis and Pattern Recognition.* New York: Wiley.

Matters, G. & Burnett, P.C. (2003). Psychological predictors of the propensity to omit short-response items on a high-stakes achievement test. *Educational and Psychological Measurement, 63,* 239-256.

Meshbane, A. & Morris, J.D. (1996). *Predictive Discriminant Analysis versus Logistic Regression in two-group classification problems*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Fort Worth, TX: Harcourt Brace College Publishers.

Rencher, A.C. (1995). *Methods of Multivariate Analysis*. New York: Wiley.

Rencher, A.C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components. *The American Statistician, 46,* 217-225.

Russell, W.D. & Cox, R.H. (2000). Construct validity of the Anxiety Rating Scale-2 with individual sport athletes. *Journal of Sports Behavior*, *23(4)*, 379-388.

SAS Version 9.1, (2005). Cary, NC: The SAS Institute.

Stevens, J. (2000). *Applied Multivariate Statistics for the Social Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, publishers.

Tabachnick, B.G. & Fidell, L.S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn and Bacon.

Send correspondence to: W. Holmes Finch
Ball State University
Email: whfinch@bsu.edu