Multivariate methods in rehabilitation

Dimiter Dimitrov\textsuperscript{a,\ast} and Phillip D. Rumrill\textsuperscript{b}
\textsuperscript{a}Graduate School of Education, George Mason University, 4400 University Drive, MS 4B3, Fairfax, VA 22030-4444, USA
Tel.: +1 703 993 3842; E-mail: ddimitro@gmu.edu
\textsuperscript{b}Kent State University, Center for Disability Studies, Kent, OH, USA

Abstract. The complexity of relationships between variables in a variety of rehabilitation settings can be appropriately addressed by the use of multivariate methods of statistical analysis. The purpose of this article is to describe (a) some basic concepts and principles of multivariate methods as well as (b) common misconceptions in the application of such methods. Specifically, the authors focus on the use of multivariate analysis of variance (MANOVA) in rehabilitation research and some pitfalls that rehabilitation researchers face when choosing between MANOVA for a set of outcome variables versus separate analyses of variance (ANOVA) for each outcome variable. Examples of the use of MANOVA in rehabilitation research, as well as references to other multivariate methods, are also provided.

Keywords: Multivariate methods, research design, statistics

1. Introduction

In statistical methods, the word “multivariate” (i.e., many variables) indicates that two or more dependent variables (outcome variables or criterion variables) are simultaneously studied in a single analysis. Multivariate analysis of variance (MANOVA) is employed when two or more groups are compared on a set of two or more dependent (outcome) variables. For example, one can use MANOVA to compare three groups of people categorized by disability type (e.g., blindness, spinal cord injury, mental illness) on a set of three dependent variables (e.g., acceptance of disability, personal control, self-efficacy). With this design, disability type is the independent variable, whereas the 3 variables of psychosocial adjustment are dependent variables. If, however, the disability groups are compared on each dependent variable separately, then 3 separate analyses of variance (univariate ANOVA) are conducted. It is a common, yet not always appropriate, practice in behavioral studies to use MANOVA followed by univariate ANOVAs. The next section addresses this issue and provides an illustrative example in an attempt to help rehabilitation researchers in their choice between MANOVA and separate univariate ANOVA.

2. MANOVA versus univariate ANOVAs

Assume that three groups are compared on a set of five dependent variables, with the null hypothesis stating that the population means in all groups are equal on each dependent variable. One approach to testing this hypothesis is to use five univariate ANOVAs (i.e., to compare the groups separately on each dependent variable using a one-way ANOVA test), say at the 0.05 level of significance for each test. With this approach, however, the overall probability of Type I error (i.e., to falsely reject the null hypothesis) is larger than 0.05 and increases when the number of dependent variables increases. In an attempt to prevent this from happening, researchers usually conduct MANOVA first and then use the statistical significance (if any) of the MANOVA...
omnibus statistic, Wilk’s lambda, as a “shield” against the inflated Type I error rate that is a consequence of univariate ANOVAs. Before discussing in more detail the inappropriateness of such an interpretation, a brief description of the Wilk’s lambda statistic is presented.

Wilk’s lambda is used in MANOVA to test for the presence (or absence) of an omnibus difference between the compared groups on a set of dependent variables. This statistic takes a value between 0 and 1 indicating what proportion is the error variance of the set of dependent variables (e.g. [1], p. 211). Thus, Wilk’s lambda indicates what percentage of the total variance in the variables (e.g. [1], p. 211). Thus, Wilk’s lambda indicates what proportion is the error variance of the set of dependent variables. This statistic takes a value between 0 and 1 indicating what proportion is the error variance of the set of dependent variables (e.g. [1], p. 211). Thus, Wilk’s lambda indicates what percentage of the total variance in the set of dependent variables is NOT explained by differences between the groups being compared. Evidently, the closer Wilk’s lambda is to zero, the larger the effect size for the difference between the groups on the set of dependent variables (Wilk’s lambda is tested for statistical significance through a multivariate F statistic [1, p. 212]). It is important to emphasize that Wilk’s lambda takes into account the correlations among the dependent variables, whereas separate univariate ANOVAs ignore such correlations. Also, a statistically significant Wilk’s lambda does not necessarily mean that the groups differ on separate dependent variables – it rather indicates that the groups differ on some combination of dependent variables that have a common substantive meaning. Therefore, Wilk’s lambda does not provide information about the overall Type I error rate produced by using separate ANOVAs [1–3]. It may even happen that Wilk’s lambda is statistically significant, yet none of the univariate ANOVA test statistics is statistically significant. Conversely, a statistically nonsignificant Wilk’s lambda may be followed by a statistical significant result for some univariate ANOVAs. Thus, conducting MANOVA as a preliminary step to using univariate ANOVAs is unnecessary. In fact, MANOVA and separate univariate ANOVAs address different research questions and provide different information about the data and their interpretation. The nature of the research question(s) should guide researchers in their choice of MANOVA versus univariate ANOVAs. The results from the initial analysis may suggest the investigation of additional (refining) research questions that involve appropriate statistical methods (e.g., t tests, factor analysis, discriminant analysis).

2.1. Univariate research questions

Using univariate ANOVAs is appropriate when the researcher wants to know the individual dependent variables on which the compared groups differ. One scenario in which this is the right question to ask is when the dependent variables are conceptually independent of one another. In this case, it is not expected that some linear composite of dependent variables will reveal an underlying construct, thereby necessitating the use of MANOVA to compare groups. Another situation in which univariate ANOVAs are appropriate occurs when the researcher wants to examine bivariate relationships between the independent (i.e., treatment) variable and individual dependent variables (e.g., for exploratory purposes or to compare the results with previous studies on such relationships). In a rehabilitation study [4], for example, the purpose was to determine whether individuals with work-related injuries receiving worksite analysis would have less lost workdays than individuals not receiving worksite analysis. To explore for differences in functional status and pain levels among the groups, two separate ANOVAs were conducted on each of the two dependent variables (functional status and pain).

As noted earlier, it is a common misconception that if the omnibus MANOVA statistic, Wilk’s lambda, is statistically significant, this will maintain the overall probability of Type I error with univariate ANOVAs at the adopted level of significance, say \( \alpha = 0.05 \). Instead of following a significant MANOVA, one can maintain the overall (familywise) level of significance for univariate ANOVAs by conducting each of them at a level lower than \( \alpha \). Specifically, using the Bonferroni adjustment, one should use for each ANOVA a level of significance equal to \( \alpha / k \), where \( k \) is the number of univariate ANOVAs. For example, for the comparison of groups on four dependent variables, each separate univariate ANOVA should be tested at the level of significance 0.0125 (= 0.05/4) to maintain the familywise level of significance at 0.05. It should be kept in mind, however, that univariate ANOVAs ignore interrelationships among the dependent variables, thus opening the door for analysis and interpretation of redundant information.

2.2. MANOVA research questions

Comparing groups on separate dependent variables through the use of separate univariate ANOVAs does not allow researchers to address broader and more insightful questions that involve interrelationships between the dependent variables. Such questions can be answered by using MANOVA, thus allowing researchers to view substantive relationships from new
perspectives. For example, the researcher may want to know (a) which subsets of dependent variables account for group separation, (b) what are the underlying constructs for such subsets, and (c) what is the relative contribution of individual dependent variables to group separation. In a rehabilitation study [5], for example, the purpose was to determine if standing dynamic balance was affected by carrying a backpack. MANOVA was used to compare two groups (backpack vs. no backpack) on five dependent variables: reaction time, movement velocity, end point excursion, maximum excursion, and directional control.

As noted earlier, a significant omnibus MANOVA (i.e., statistically significant Wilk’s lambda) means that the groups being compared differ on some linear combination(s) of the dependent variables. In multivariate analysis, such linear combinations are referred to as linear discriminant functions (LDFs). Thus, LDF is a latent construct that discriminates (separates) the groups being compared. With m groups, the number of all possible LDFs is m − 1. With two groups, for example, there is only one LDF that may contribute to the difference between the two groups. The main questions are then (a) which dependent variables define (i.e., highly correlate with) the latent construct, (b) what is the substantive interpretation of the construct, and (c) what is the relative contribution of individual dependent variables in separating the groups. With more than two groups, however, these questions may be asked for differences (pairwise or more complex) between specific groups (for details, the reader may refer, e.g., to [1, p. 217] and [2]).

### 2.3. Assumptions of MANOVA

The statistical assumptions of MANOVA are closely parallel to those of ANOVA [1,3]:

1. The subjects are randomly sampled from the target population.
2. The observations (e.g., persons’ scores) are statistically independent of one another.
3. The dependent variables have a multivariate normal distribution within each group.
4. All groups have the same variance on each dependent variable (i.e., the homogeneity of variance assumption in ANOVA must be met for each dependent variable).
5. The correlation between any two dependent variables must be the same in all groups.

One cannot expect, of course, all of the above assumptions to be met precisely in practical MANOVA applications. Violation of any of the first two assumptions may invalidate the results with MANOVA. The violation of the second assumption, also referred to as the “independence assumption,” may occur, for example, when the participants in an experiment are working in small groups and interacting with each other. Fortunately, MANOVA is robust to a large extent to violation of the last three assumptions. Some notes of caution, however, are necessary. The third assumption is difficult to test, but for practical purposes one can simply test the normality of each dependent variable in each group separately (e.g., using computer programs such as SPSS or SAS). It should also be noted that violation of normality has little effect on the Type I error rates, but serious departures from normality may reduce the statistical power of MANOVA test statistics (e.g., Wilk’s lambda).

Taken together, the last two assumptions are equivalent to the assumption that all groups have the same within-group population covariance matrix. A test for homogeneity of covariance matrices, referred to as Box’M, is widely available with statistical programs (e.g., SPSS, SAS), but it is extremely sensitive to violations of normality (i.e., Box’M is not dependable when the assumption of normality is not met). The assumption for homogeneity of covariance matrices (assumptions 4 and 5 together) is taken care of relatively well when the groups have equal sample size (i.e., with a balanced MANOVA design). This, however, is not the case with sharply unequal group sizes; with this assumption violated, the MANOVA test becomes very liberal (i.e., makes it unduly easy to reject the null hypothesis) when larger sample sizes are associated with smaller variances. Conversely, the test becomes very conservative (i.e., makes it unduly difficult to reject the null hypothesis) when larger variances are associated with groups with the larger sample size. An approach to dealing with this problem is to transform the original scores on the dependent variable (e.g. [1], p. 274). For example, if the scores on a given dependent variable are proportions, one can stabilize the group variances on this variable by using the “square root” transformation (i.e., replacing each score by its square root value). The reader may also refer to ( [1, p. 256]) for transformations related to violations of the normality assumption and to [6–8] for additional discussions on analyzing and interpreting MANOVA results.
3. Example

The purpose of this example is to illustrate some basic MANOVA concepts, research questions, and interpretations in the context of real data for a population of persons with multiple sclerosis. The data were taken from an existing data pool produced with the Employment Preparation Survey Project funded by the National Multiple Sclerosis Society (see, also [9]).

The illustrative research goal here is to investigate whether job satisfaction affects quality of life. Specifically, the main research question is what aspects of the general factor “quality of life” separate employed persons with multiple sclerosis who have different levels of satisfaction with their present jobs: Satisfied (S), Undecided (U), and Not satisfied (NS). The relative importance of individual dependent variables to the separation among the groups is also of interest in this example. With the Employment Preparation Survey, “quality of life” is measured by the following seven variables on a 7-point Likert-type scale (from 1 = not satisfying to 7 = completely satisfying):

- Social life and experiences ($Y_1 = SLE$);
- Family life and experiences ($Y_2 = FLE$);
- Hobbies and recreational experiences ($Y_3 = HRE$);
- Educational and intellectual development ($Y_4 = EID$);
- Experience of daily living (e.g., work) ($Y_5 = EDL$);
- Romantic experiences ($Y_6 = RE$);
- Expectations and hopes for the future ($Y_7 = EHF$).

Despite the relatively small number of variables (items), the Cronbach alpha coefficient for the internal consistency of the QOL scale used in this example was high (0.90). MANOVA is appropriate because the research question relates to differences among three job satisfaction groups (S, U, and NS) on aspects of quality of life that may emerge as linear composites of substantive related dependent variables ($Y_1, \ldots, Y_7$). The Pearson correlations for all pairs of these seven variables were statistically significant and varied in the range from 0.41 to 0.65.

The Wilk’s lambda (= 0.747) for the omnibus MANOVA test was found to be statistically significant, $F(14, 822) = 1.78$, $p = 0.038$. At this point, we can conclude that there are statistically significant differences between the S, U, and NS groups on some linear combinations of dependent variables that represent aspects of quality of life. The normality assumption, which is important primarily for the correctness of the Box’M statistic, was supported to a large extent by the normality (P-P plot) test provided in SPSS for each dependent variable within each group. The Box’M value (67.35) was not statistically significant, $F(56, 101541) = 1.07$, $p = 0.19$, thus indicating that the assumption of homogeneity of covariance matrices is also met.

As a next step, the discriminant analysis option in SPSS was used with MANOVA to determine which specific dependent variables define linear composites (linear discriminant functions, LDFs, discussed in the previous section). The reader may refer to [1, p. 219], for the SPSS syntax of this procedure. As three groups are being compared, there are two LDFs (LDF1 and LDF2) that may separate the three groups of job satisfaction (S, U, and NS). The chi-square test was statistically significant for both LDF1, $\chi^2(14) = 152.94$, $p < 0.001$, and LDF2, $\chi^2(6) = 12.62$, $p < 0.05$, thus indicating that there are two linear composites (constructs) of the dependent variables that separate persons with multiple sclerosis at the three levels of satisfaction: Satisfied, Undecided, and Not satisfied.

Table 1 provides the correlations and the standardized coefficients for each dependent variable with LDF1 and LDF2. Examination of the correlation coefficients clearly indicates that $Y_1$, $Y_2$, $Y_5$, $Y_6$, and $Y_7$ correlate with LDF1, whereas $Y_3$ and $Y_4$ correlate with LDF2. To determine possible redundancy in these correlations, we examine the standardized coefficients. The comparison of the standardized coefficients for $Y_1$, $Y_2$, $Y_5$, $Y_6$, and $Y_7$ on LDF1 shows that $Y_1$ and $Y_6$ provide redundant information about LDF1 because their standardized coefficients (–0.068 and 0.069, respectively) are much smaller than those for $Y_5$, $Y_6$, and $Y_7$. Therefore, the substantive meaning of LDF1 is determined by using three variables: $Y_5$ (“the experience of daily living, e.g., work”), $Y_6$ (“romantic experience”), and $Y_7$ (“expectations and hopes for the future”), with a leading role of $Y_5$ as indicated by its highly strong correlation (0.919) with LDF1. Given this, we use the working label “Daily experience and hopes for the future” for the first latent construct, LDF1.

On the other side, the examination of the standardized coefficients of the variables that correlate with LDF2 ($Y_3$ and $Y_4$) shows that $Y_3$ provides redundant information as its standardized coefficient (0.107) is much smaller than that for $Y_4$ (1.051). Evidently, LDF2 is represented by $Y_4$ and, therefore, labeled here as “intellectual and educational development.”
...function, LDFs at group means (Centroids).

Note: LDF1 = "Daily experience and hopes for the future" (LDF1), with the strongest separation between S (satisfied) and NS (not satisfied). Somewhat weaker separation among the groups is provided by the second latent construct, "intellectual and educational development" (LDF2), with the larger distinction being observed between groups S (satisfied) and NS (not satisfied). The location of the group centroids also deserves attention. For example, the mean of group S ("satisfied") is positive, whereas the mean of group U ("undecided") is negative, on both LDF1 and LDF2. It should be emphasized, however, that MANOVA takes into account not only the group means, but also the correlations between dependent variables.

The relative importance of a given dependent variable (if deleted) on the decrease of the group separation. As one can see from Table 1, the F-to-remove statistic for $Y_5$ (44.351) is much higher than the values of this statistic for all other dependent variables.

Thus, $Y_5$ ("the experience of daily living, e.g., work") is the most important contributor to the separation of persons with multiple sclerosis by the level of their job satisfaction. Conversely, the very small F-to-remove values for $Y_1$ (0.686) and $Y_6$ (0.191) indicate that these two variables have a negligible contribution to the multivariate group separation. In fact, as shown earlier, $Y_1$ and $Y_6$ correlate with the latent trait LDF1, but they were not used in its interpretation because of their redundancy with other dependent variables that also correlate with LDF1.

It should be noted that the SPSS printout for MANOVA also provides the univariate F tests for pairwise group comparisons on each dependent variable separately. In this example, there was a statistically significant difference between any two groups (S, U, and NS) on each of the seven dependent variables. One may wish to investigate other group contrasts – for example, to compare group S (satisfied) versus the other two groups together (U and NS). For information on multivariate contrast between groups, one can refer to [1,2, pp. 217–243], and [6].

4. Notes on other multivariate methods

4.1. Discriminant analysis

As illustrated in the above example, discriminant analysis can be used with MANOVA to provide information about the separation of groups on linear com-
posites of dependent variables. Each linear composite is interpreted as a latent construct underlying a specific set of dependent variables (e.g., \textit{LDF1} and \textit{LDF2} in our example.) Thus, discriminant analysis can be used for the description of MANOVA results. Another purpose for using discriminant analysis is to classify subjects into groups, i.e., to predict group membership. For example, Stevens [1, p. 303], illustrates the use of discriminant analysis to classify (predict) kindergarten children into two levels (low risk or high risk) of reading problems based on the children’s scores on three predictor variables: word identification, word comprehension, and passage comprehension. As Kerlinger [10, p. 562] noted “when dealing with two groups, the discriminant analysis is nothing more than a multiple regression equation with the dependent variable a nominal variable (coded 0, 1) representing group membership. With three or more groups, however, discriminant analysis goes beyond multiple regression methods.” The reader may also refer to [11–14] for readable treatments and practical applications of discriminant analysis.

4.2. Canonical correlation

Whereas MANOVA and discriminant analysis deal with linear composites within a set of variables, the analysis of canonical correlations deals with correlations between linear composites in two sets of variables; the linear composites of variables are referred to as \textit{canonical variates}. As Stevens [1] noted, “canonical correlation is appropriate if the wish is to parsimoniously describe the number and the nature of mutually independent relationships existing between two sets” (p. 471). He describes, for example, a study by Tetenbaum [15] that addresses the issue of the validity of student ratings of teachers [1, p. 479]. Specifically, the research question is whether in the process of rating a teacher, the students focus on the need-related aspects of the perceptual situation and base their judgment on those areas of the teacher’s performance most relevant to their own needs [1, p. 418]. The canonical analysis applied in this study revealed that there were three statistically significant canonical correlations between linear composites of 12 need variables and linear composites of the corresponding 12 rating variables (in general, the two sets do not necessarily have to contain equal numbers of variables). As a result, it was found that there were three statistically significant canonical correlations that account for most of the between association for the two sets of variables (needs and ratings). For example, the first pair of canonical variates (i.e., with the largest canonical correlation) showed a strong relationship between needs and ratings on the same latent continuum (construct) – the “positive” direction of this latent continuum is interpreted as \textit{ascendancy} and the opposite, “negative,” as \textit{intellectual striving needs} of the students (for more details, see [15]).

It should be noted that the number of possible canonical correlations equals the smaller number for the variables in the two sets. For example, if the first set has 10 variables and the second set 7 variables, then the
number of possible pairs of canonical variates is 7. An important feature of canonical correlation is that the pairs of canonical variates are mutually independent. This means that (a) the canonical variates within each set are uncorrelated and (b) any two canonical variates that belong to different pairs and different sets are uncorrelated. The reader may also refer, for example, to [7,16] for readable treatments of canonical correlation and to [17,18], for practical studies using canonical correlation analysis.

4.3. Structural equation modeling

Structural Equation Modeling (SEM) also provides methods for testing differences between groups on a set of variables. Which method (SEM or MANOVA) is more appropriate in a particular study depends on the type of relationships between observed variables and constructs (latent variables, factors). MANOVA is more appropriate with an emergent variable system in which the construct “emerges” as a linear composite of observed variables that represent causal agents of the construct. For example, Hancock [19] describes “stress” as a latent construct that emerges as a linear composite of observed dependent variables such as relationship with parents, relationship with spouse, and demands of the work place. Conversely, SEM is more appropriate with a latent variable system in which the construct has a causal influence on the observed variables. For example, the self-esteem is a construct that has a causal influence on the observed variables. Thus, using MANOVA with a latent variable system (i.e., when a construct has causal effects on observed variables) may provide inaccurate results about group differences on the construct of interest. For readable presentations of SEM methods in comparing groups on constructs, the reader may refer, for example, to [19–21].

5. Conclusion

Multivariate methods are used to address the complexity of phenomena and research questions in behavioral studies. Most phenomena in the lives of people with disabilities are complex, interdependent, and rooted in multiple explanations for multiple outcomes, so simple statistical analyses like ANOVA and multiple regression often fail to provide sufficient explanatory power in adding new knowledge to the field of rehabilitation. Despite a general awareness of this message, there are misconceptions at large that are still practiced in some applications of multivariate methods, thus preventing researchers from taking appropriate advantage of MANOVA when analyzing multiple variables.

Particular attention is warranted for misconceptions related to the use of separate ANOVAs as a follow up on a statistically significant omnibus MANOVA. Researchers should be aware that (a) the decision about using MANOVA or separate ANOVAs should be governed by the nature of the research question (e.g., related to substantive relationships between the dependent variables) and (b) statistically significant omnibus MANOVA does not necessarily protect against inflated Type I error rate in follow-ups with univariate ANOVAs.

MANOVA is appropriate with research questions about group differences on linear composites of dependent variables. A linear composite (linear discriminant function, LDF) represents a weighted sum of the original dependent variables. An LDF is interpreted as a latent variable (construct) defined by the meaning of those original dependent variables that have high weights (loadings) on the construct, excluding the variables that provide redundant information.

Discriminant analysis is used with two main purposes: (a) description of MANOVA results and (b) prediction of group membership. In the first case, the example provided earlier used discriminant analysis with MANOVA to determine and interpret linear composites that separate the groups, as well as to determine the relative importance of individual dependant variables in the group separation.

Canonical correlation is another multivariate method used to determine relationships between linear composites (constructs) for two sets of variables. The correlation between a linear composite of variables in the first set and a linear composite of variables in the second set is referred to as canonical correlation and the linear composites are referred to as canonical variates.

Structural Equation modeling (SEM) also provides methods for testing differences between groups in a latent variable system (i.e., when the construct of interest is believed to have a causal influence on the observed outcome variables). It is important to emphasize that constructs that are derived within SEM are error-free,
whereas constructs that emerge as linear composites within MANOVA are NOT free from measurement errors associated with the observed variables.

In conclusion, knowledge of the use of MANOVA, discriminant analysis, canonical correlation, and structural equation modeling enabled rehabilitation researchers to address more properly interesting research questions about group differences and relationships on a set of variables. By understanding the assumptions and procedures inherent in multivariate analyses, researchers can continue to build a knowledge base that reflects the complexity of the lived experience of disability.

References