Advanced and Multivariate Statistical Methods

Practical Application and Interpretation

Second Edition

Craig A. Mertler
Bowling Green State University

Rachel A. Vannatta
Bowling Green State University

Pyrczak Publishing
P.O. Box 39731 • Los Angeles, CA 90039

Although the author and publisher have made every effort to ensure the accuracy and completeness of information contained in this book, we assume no responsibility for errors, inaccuracies, omissions, or any inconsistency herein. Any slights of people, places, or organizations are unintentional.

Cover design by Robert Kibler and Larry Nichols.

Editorial assistance provided by Sharon Young, Brenda Koplin, Cheryl Alcorn, Monica Lopez, Randall R. Bruce, and Elaine Parks.

Printed in the United States of America by McNaughton and Gunn, Inc.

ISBN 1-884585-41-8

CHAPTER 2

A GUIDE TO MULTIVARIATE TECHNIQUES

One of the most difficult tasks for students conducting quantitative research is identifying the appropriate statistical technique to be utilized for a particular research question. Fortunately, if an accurate and appropriate research question has been generated, the process of determining the statistical technique is really quite simple. The primary factor that determines the statistical test to be used is the variable—more specifically, the type or scale of variables (categorical or quantitative) and the number of independent and dependent variables, both of which influence the nature of the research question being posed. To facilitate this identification process, we provide two decision-making tools so that the reader may select one that is most comfortable. The Table of Statistical Tests begins with the identification of the numbers and scales of independent and dependent variables. In contrast, the Decision-Making Tree is organized around the four different types of research questions: degree of relationship among variables, significance of group differences, prediction of group membership, and structure. This chapter presents these decision-making tools and provides an overview of the statistical techniques addressed in this text as well as basic univariate tests, all of which will be organized by the four types of research questions. Please note that there are additional multivariate techniques that are not addressed in this book.

SECTION 2.1 DEGREE OF RELATIONSHIP AMONG VARIABLES

When investigating the relationship among two or more quantitative variables, correlation and/or regression is the appropriate test. Three statistical tests are presented that address this type of research question.

Bivariate Correlation and Regression

Bivariate correlation and regression evaluate the degree of relationship between two quantitative variables. The Pearson correlation coefficient (r), the most commonly used bivariate correlation technique, measures the association between two quantitative variables without distinction between the independent and dependent variables (e.g., What is the relationship between SAT scores and freshman college GPA?). In contrast, bivariate regression utilizes the relationship between the independent and dependent variables to predict the score of the dependent variable from the independent variable (e.g., To what degree do SAT scores [IV] predict freshman college GPA [DV]?). An overview of bivariate correlation and regression is provided in Chapter 7.

When to use bivariate correlation/regression?

1 IV (quantitative) ⟶ relationship/prediction
1 DV (quantitative)

Multiple Regression

Multiple regression identifies the best combination of predictors (IVs) of the dependent variable. Consequently, it is used when there are several independent quantitative variables and one dependent quantitative variable (e.g., Which combination of risk-taking behaviors [amount of alcohol use, drug use, sexual activity, and violence—IVs] best predicts the amount of suicide behavior [DV] among adolescents?). To produce the best combination of predictors of the dependent variable, a sequential multiple regression selects independent variables, one at a time, by their ability to account for the most variance in the dependent variable. As a variable is selected and entered into the group of predictors, the relationship between the group of predictors and the dependent variables is reassessed. When no more variables are left that explain a significant amount of variance in the dependent variable, then the regression model is complete. Multiple regression is discussed in Chapter 7.

When to use multiple regression?

2+ IV (quantitative) ⟶ relationship/prediction
1 DV (quantitative)

Path Analysis

Path analysis utilizes multiple applications of multiple regression to estimate causal relations, both direct and indirect, among several variables and to test the acceptability of the causal model hypothesized by the researcher (e.g., What are the direct and indirect effects of reading ability, family income, and parents' education [IVs] on students' GPA [DV]?). Before any data analysis is conducted, the researcher must first hypothesize the causal model, which is usually based upon theory and previous research. This model is then graphically represented in a path diagram. Path coefficients are calculated to estimate the strength of the relationships in the hypothesized causal model. A further discussion of path analysis is presented in Chapter 8.

When to use path analysis?

2+ IV (quantitative) ⟶ relationship/causal
1+ DV (quantitative)

SECTION 2.2 SIGNIFICANCE OF GROUP DIFFERENCES

A primary purpose of testing for group differences is to determine a causal relationship between the independent and dependent variables. Comparison groups are created by the categories identified in the IV(s). The number of categories in the IV, the number of IVs, and the number of DVs determine the appropriate test.

t Test

The most basic statistical test that measures group differences is the *t* test, which analyzes significant differences between two group means. Consequently, a *t* test is appropriate when the IV is defined as having two categories and the DV is quantitative (e.g., Do males and females [IV] have significantly different SAT scores [DV]?). Further explanation of *t* tests is provided in most introductory level statistical texts and therefore is not included in this text.

When to use a t test?

1 IV (2 categories) ⟶ Group differences
1 DV (quantitative)

One-Way Analysis of Variance

One-way analysis of variance (ANOVA) tests the significance of group differences between two or more means as it analyzes variation between and within each group. ANOVA is appropriate when the IV is defined as having two or more categories and the DV is quantitative, (e.g., Do adolescents from low, middle, and high socioeconomic status families [IV] have different scores on an AIDS knowledge test [DV]?). Since ANOVA only determines the significance of group differences and does not identify which groups are significantly different, post hoc tests are usually conducted in conjunction with ANOVA. An overview of ANOVA is provided in Chapter 4.

When to use a one-way ANOVA?

1	IV (2+ categories)	⟶	Group differences
1	DV (quantitative)		

One-Way Analysis of Covariance

One-way analysis of covariance (ANCOVA) is similar to ANOVA in that two or more groups are being compared on the mean of some DV, but ANCOVA additionally controls for a variable (covariate) that may influence the DV (e.g., Do preschoolers of low, middle, and high socioeconomic status [IV] have different literacy test scores [DV] after adjusting for family type [covariate]?) Many times the covariate may be pretreatment differences in which groups are equated in terms of the covariate(s). In general, ANCOVA is appropriate when the IV is defined as having two or more categories, the DV is quantitative, and the effects of one or more covariates need to be removed. Further discussion of one-way ANCOVA is provided in Chapter 5.

When to use a one-way ANCOVA?

1	IV (2+ categories)	⟶	Group differences
1	DV (quantitative)		
1+	covariate		

One-Way Multivariate Analysis of Variance

Similar to ANOVA in that both techniques test for differences among two or more groups as defined by a single IV, one-way multivariate analysis of variance (MANOVA) is utilized to simultaneously study two or more related DVs while controlling for the correlations among the DVs (Vogt, 1993). If DVs are not correlated, then it is appropriate to conduct separate ANOVAs. Since groups are being compared on several DVs, a new DV is created from the set of DVs that maximizes group differences. After this linear combination of the original DVs is created, an ANOVA is then conducted to compare groups based on the new DV. A MANOVA example follows: Does ethnicity [IV] significantly affect reading achievement, math achievement, and overall achievement [DVs] among 6th grade students? Chapter 6 discusses one-way and factorial models of MANOVA and MANCOVA.

When to use a one-way MANOVA?

1	IV (2+ categories)	⟶	Group differences
2+	DVs (quantitative)		

One-Way Multivariate Analysis of Covariance

An extension of ANCOVA, multivariate analysis of covariance (MANCOVA) investigates group differences among several DVs while also controlling for covariate(s) that may influence the DVs (e.g., Does ethnicity [IV] significantly affect reading achievement, math achievement, and overall achievement [DVs] among 6th grade students after adjusting for family income [covariate]?).

When to use a one-way MANCOVA?

1 IV (2+ categories) ⟶ Group differences
2+ DVs (quantitative)
1+ covariate

Factorial Multivariate Analysis of Variance

Factorial multivariate analysis of variance (factorial MANOVA) extends MANOVA to research scenarios with two or more IVs that are categorical (e.g., Does ethnicity and learning preference [IVs] significantly affect reading achievement, math achievement, and overall achievement [DVs] among 6th grade students?). Since several independent variables are used, different combinations of DVs are created for each main effect and interaction of the IVs.

When to use a factorial MANOVA?

2+ IVs (categorical) ⟶ Group differences
2+ DVs (quantitative)

Factorial Multivariate Analysis of Covariance

Factorial multivariate analysis of covariance (factorial MANCOVA) extends factorial MANOVA to research scenarios that require the adjustment of one or more covariates on the DVs (e.g., Does ethnicity and learning preference [IVs] significantly affect reading achievement, math achievement, and overall achievement [DVs] among 6th grade students after adjusting for family income [covariate]?).

When to use a factorial MANCOVA?

2+ IVs (categorical) ⟶ Group differences
2+ DVs (quantitative)
1+ covariate

Section 2.3 Prediction of Group Membership

The primary purpose of predicting group membership is to identify specific IVs that best predict group membership as defined by the DV. Consequently, the following statistical techniques are appropriate when the DV is categorical.

Discriminant Analysis

Discriminant analysis is often seen as the reverse of MANOVA in that it seeks to identify which combination of quantitative IVs best predict group membership as defined by a single DV that has two or more categories (e.g., Which risk-taking behaviors [amount of alcohol use, drug use, sexual activity, violence—IVs] distinguish suicide attempters from nonattempters [DV]?). In contrast, MANOVA identifies group differences on a combination of quantitative DVs. Discriminant analysis seeks to interpret the pattern of differences among the predictors (IVs); consequently, the analysis will often produce several sets or combinations of IVs that predict group membership. Each IV set, referred to as a function, represents a mathematical attempt to maximize a linear combination of the IVs to discriminate among groups. Discriminant analysis is best used when groups are formed naturally based on some characteristic and not randomly. Chapter 10 discusses discriminant analysis in further detail.

When to use discriminant analysis?

2+ IVs (quantitative)
1 DV (2+ categories) ⟶ Group prediction

Logistic Regression

Logistic regression is similar to discriminant analysis in that both identify a set of IVs that best predict group membership. Although SPSS provides both binary and multinomial logistic regression, our discussion will address only the binary logistic regression in which the DV is a dichotomous (having only two categories) variable. The IVs may be categorical and/or quantitative. Since the DV consists of only two categories, logistic regression estimates the odds probability of the DV occurring as the values of the IVs change. For example, a research question that would utilize logistic regression is: To what extent do certain risk-taking behaviors (amount of alcohol use, drug use, sexual activity, and the presence of violent behavior—IVs) increase the odds of a suicide attempt (DV) occurring? Logistic regression is discussed in Chapter 11.

When to use logistic regression?

2+ IVs (categorical/quantitative)
1 DV (2 categories) ⟶ Group prediction

SECTION 2.4 STRUCTURE

When the researcher questions the underlying structure of an instrument or is interested in reducing the number of IVs, factor analysis and/or principal components are appropriate methods. Although factor analysis and principal components are different techniques, they are very similar and will be presented together under the heading of factor analysis. Both of these techniques will be discussed in Chapter 9.

Factor Analysis and Principal Components Analysis

Factor analysis allows the researcher to explore underlying structures of an instrument or data set and is often used to develop and test theory. Principal components is generally used to reduce the number of IVs, which is advantageous when conducting multivariate techniques in which the IVs are highly correlated. For example, principal components can reduce a 100-item instrument to ten factors that will then be utilized as IVs in subsequent data analysis. This IV reduction can also aid the researcher in exploring, developing, and testing theories based upon how the items are grouped. Consequently, factor analysis/principal components combine several related IVs into fewer, more basic underlying factors. Independent variables that share common variance are grouped together. Once factors are created, they are often adjusted (rotated) so that these factors are not highly related to one another and more accurately represent the combined IVs. Since research questions that utilize factor analysis/principal components typically only address IVs, this statistical technique is not included in the Table of Statistical Tests, which relies upon the identification of both IVs and DVs. An example research question that would utilize factor analysis/principal components is as follows: What underlying structure exists among the variables of male life expectancy, female life expectancy, birth rate, infant mortality rate, fertility rate among women, number of doctors, number of radios, number of telephones, number of hospital beds, and gross domestic product?

SECTION 2.5 THE TABLE OF STATISTICAL TESTS

The Table of Statistical Tests is presented in Figure 2.1. This tool organizes statistical methods by the number and type (categorical versus quantitative) of IVs and DVs. Steps for using this table are listed as follows:

1. Identify the variables in the research question.
2. Indicate which variables are the independent and dependent variables and covariates.
3. Determine the type (categorical or quantitative) of all variables. If a variable is categorical, determine the number of categories.
4. Use the table to identify:
    - the appropriate row for IVs;
    - the appropriate column for the DVs;
    - the row and column intersection indicates the statistical test to be used.

These steps are applied to the following research question: Does ethnicity significantly affect reading achievement, math achievement, and overall achievement among 6th grade students after adjusting for family income?

Step 1: Underline the variables in the research question.

> Does ethnicity significantly affect reading achievement, math achievement, and overall achievement among 6th grade students after adjusting for family income.

Step 2: Identify which variables are the independent and dependent variables and the covariates. It is helpful to examine the sentence order of variables since the first variables are usually the IVs. The verb of the research question can also help in the identification process.

Indicates preceding variable is IV

IV — Does ethnicity significiantly (affect) reading achievement (DV), math achievement (DV), and

overall achievement (DV) among 6th grade students (after adjusting) for family income (Covariate)?

Indicates following variable is covariate

Step 3: Determine the type (categorical or quantitative) of all variables. This is dependent on how you decide to operationalize your variables.

> Does ethnicity (IV) 3+ categories significantly affect reading achievement (DV) quantitative, math achievement (DV) quantitative, and overall achievement (DV) quantitative among 6th grade students after adjusting for family income (covariate) quantitative?

Consequently, this research question includes the following: one IV (3+ categories), 3 DVs (all quantitative), and one covariate (quantitative).

Step 4: Use the table to:

- identify the appropriate row for IVs.
  Example: IV→categorical→one IV→2+ categories with one covariate
- identify the appropriate column for the DVs.
  Example: DV→quantitative→several DVs
- the row and column intersection indicates the statistical test to be used.
  Example: The intersection of the preceding row and column indicates that one-way MANCOVA should be conducted.

Section 2.6 The Decision-Making Tree for Statistical Tests

The Decision-Making Tree for Statistical Tests is presented in Figure 2.2. This tool organizes statistical methods by the purpose of the research question. Once the purpose has been identified, the process is then guided by the number and type of variables. Although the Decision-Making Tree begins with the purpose of the research question, we recommend first identifying the number and types of variables as this will guide the process of determining the purpose. The steps for using the Decision-Making Tree are listed as follows.

1. Identify the variables in the research question.
2. Indicate which variables are the independent and dependent variables and covariates.
3. Determine the type (categorical or quantitative) of all variables. If a variable is categorical, determine the number of categories.
4. Determine the purpose of the research question: degree of relationship, group differences, prediction of group membership, and structure. Here are a few helpful hints in using the variable information to determine the research question purpose.
   - When the IVs and DVs are all quantitative, the purpose is degree of relationship.
   - When the IVs are categorical and the DVs are quantitative, the purpose is group differences.
   - When the DVs are categorical, the purpose is predicting group membership.
5. Apply the information from the preceding steps to the Decision-Making Tree following the process of decisions—research question, number and type of DV, number and type of IV, and covariates—to the appropriate test.

These steps are applied to the following research question: Which combination of risk-taking behaviors [amount of alcohol use, drug use, sexual activity, and violence] best predict the amount of suicide behavior among adolescents?

Step 1: Identify the variables in the research question.
Which combination of risk-taking behaviors [amount of alcohol use, drug use, sexual activity, and violence] best predict the amount of suicide behavior among adolescents?

Step 2: Indicate which variables are the independent and dependent variables and covariates.

IV IV
Which combination of risk-taking behaviors [amount of alcohol use, drug use,

IV IV DV
sexual activity, and violence] best predict the amount of suicide behavior among

adolescents?

Indicates that previous variables are IVs

Step 3: Determine the type (categorical or quantitative) of all variables.

(IV) quantitative (IV) quantitative
Which combination of risk-taking behaviors [amount of alcohol use, drug use,

(IV) quantitative (IV) quantitative (DV) quantitative
sexual activity, and violence] best predict the amount of suicide behavior among

adolescents?

Consequently, this research question includes the following: four IVs (all quantitative) and one DV (quantitative).

Step 4: Determine the purpose of the research question: degree of relationship, group differences, prediction of group membership, and structure. Since all our variables are quantitative, the purpose of the research question is degree of relationship.

Step 5: Apply the information from the preceding steps to the Decision-Making Tree: research question, number and type of DV, number and type of IV, and covariates. Continuing with the example, the decisions would be as follows:
degree of relationship→1 DV (quant.)→2+ IVs (quant.)→ multiple regression

SUMMARY

Determining the appropriate statistical technique relies upon the identification of the type of variables (categorical or quantitative) and the number of IVs and DVs, all of which influence the nature of the research questions being posed. This chapter introduces the statistical tests to be presented in the upcoming chapters. The statistical methods are organized under four purposes of research questions: degree of relationship, significance of group differences, prediction of group membership, and structure. Statistical tests that analyze the degree of relationship include bivariate correlation and regression, multiple regression and path analysis. Research questions addressing degree of relationship have all quantitative variables. Methods that examine the significance of group differences are *t* test, one-way and factorial ANOVA, one-way and factorial ANCOVA, one-way and factorial MANOVA, and one-way and factorial MANCOVA. Research questions that address group differences have categorical IV(s). Statistical tests that predict group membership are logistic regression and discriminant analysis. Research questions that address prediction of group membership have a categorical DV. Statistical tests that address the purpose of structure are factor analysis and principal components; questions that address structure usually do not distinguish between independent and dependent variables.

Two decision-making tools are provided to assist in identifying which statistical method to utilize—the Table of Statistical Tests and the Decision-Making Tree. The Table of Statistical Tests is organized by the type and number of IVs and DVs, while the Decision-Making Tree is organized by the purpose of the research question.

Figure 2.1 Table of Statistical Tests

				DEPENDENT VARIABLE(s)			
				Categorical		*Quantitative*	
				2 categories	**2+ categories**	**One DV**	**Several DVs**
INDEPENDENT VARIABLE(s)	*Categorical*	**One IV**	**2 categories**			*t* Test	One-way MANOVA
			2+ categories			One-way ANOVA	
			With covariate			One-way ANCOVA	One-way MANCOVA
		Several IVs	**No covariate**	Logistic Regression		Factorial ANOVA	Factorial MANOVA
			With covariate			One-way ANCOVA	Factorial MANCOVA
	Quantitative	**One IV**				Bivariate Correlation Bivariate Regression	
		Several IVs		Discriminant Analysis Logistic Regression	Discriminant Analysis	Multiple Regression Path Analysis	Path Analysis

Figure 2.2 Decision-Making Tree

Research Question	*Number & Type of DV*	*Number & Type of IV*	*Covariates*	*Test*	*Goal of Analysis*
Degree of Relationship	1 quantitative	1 quantitative		Bivariate Correlation and/or Regression	Determine relationship and prediction
		2+quantitative		Multiple Regression	Create linear combination that best predicts DV
	1+quantitative	2+quantitative		Path Analysis	Estimate causal relations among variables in a hypothesized model
Group Differences	1 quantitative	1 categorical (2 categories)		*t* Test	Determine significance of mean group differences
		1 categorical (2+ categories)	None	One-way ANOVA	
			Some	One-way ANCOVA	
		2+ categorical	None	Factorial ANOVA	
			Some	Factorial ANCOVA	
	2+ quantitative	1 categorical	None	One-way MANOVA	Create linear combo of DVs to maximize mean group differences
			Some	One-way MANCOVA	
		2+ categorical	None	Factorial MANOVA	
			Some	Factorial MANCOVA	
Prediction of Group Membership	1 categorical (2 categories)	2+ mixed		Logistic Regression	Create linear combo of IVs of the log of odds of being in one group
	1 categorical (2+ categories)	2+ quantitative		Discriminant Analysis	Create best linear combo of predict group membership
Structure	3+ quantitative			Factor Analysis (theoretical)	Create linear combinations of observed variables to represent latent variable
				Principal Components (empirical)	