EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT
1979, 39

UNIVARIATE AND MULTIVARIATE ANALYSIS OF CATEGORIAL VARIABLES[1]

MENI KOSLOWSKY
J. C. Penney Co. and Hofstra University

Recent trends in the analysis of categorical or nominal variables were discussed for univariate, multivariate, and psychometric problems. It was shown that several statistical procedures commonly used with these problems have analogues which can be applied to assessing categorical variables. These more general techniques are often required when the data to be analyzed consists of both nominal and interval variables. Even when the data appear to be interval, the investigator can conservatively do an analysis which presumes unordered data. In some cases, the more cautious approach can yield the more meaningful data.

OFTEN in psychological research, the variable of interest is non-continuous and unordered. Common examples in the behavioral sciences include religion, sex, and college attended. Usually the investigator is not willing to assign ordinal values to the categories within each of these variables. When this occurs, statistics which assume a specific shape to the underlying distribution (i.e., normal, Poisson, etc.) cannot be used.[2] Thus, sex and college attended cannot be correlated with each other by using either the Pearson or Spearman correlation coefficients. However, techniques are available which will yield an association measure for this type of data that is similar to the continuous result. Similarly, it is possible to obtain analogues for other bivariate

[1] Request for reprints should be sent to Meni Koslowsky, Department of Counseling, Psychology, and Research, Hofstra University, Hempstead, N. Y. 11550.

[2] Many of the statistical tests such as ANOVA assume data that is at least interval, not just ordinal. In practice, this is often violated. Nunally (1967) defines the different types of data and the operations permissible with each.

and multivariate procedures. The purpose of this paper is to present some recent advances in the analysis of categorical variables.

Discussion of the techniques will be divided into three sections. The first two sections will point out some of the ways that univariate and multivariate procedures can be applied to unordered data. The third section will focus on psychometric issues such as reliability, and how it can be assessed with categorical variables.

First, some clarification of terms is needed. The type of variable discussed here has been called at various times categorical, nominal, qualitative, unordered, and nonmetric. In the literature other terms can also be found but these are probably the most common. In all cases, the variables can be divided into categories which have no necessary mathematical order to them.

Univariate and Bivariate Techniques

Descriptive Measures

With categorical data, frequency counts are the single most common descriptive measures. For calculating central tendency with such data, the mode is the appropriate statistic. McNemar (1969) discusses some of the obvious limitations with this statistic. In the case of dichotomous data it is possible to determine the standard deviation and variance, whereas the usual situation with more than two unordered categories does not readily permit such analysis. Ordinarily categorical variables are not further analyzed descriptively and the researcher is content with presenting the frequency or proportion of cases within each category. However, a possible companion procedure with unordered data, similar to the range and standard deviation with ordinal and interval variables, is to calculate the relative frequencies among categories. By examining the distribution of scores one can get an indication of the dispersion of responses. Thus, a sample of subjects responding to the question of which college they attended might distribute themselves in several ways. Most may have come from two or three colleges. These few categories would then have high relative frequencies and describe most of the distribution. Such a pattern could be called a "dense" distribution. On the other hand, they may have come from many different colleges and combining the two or three most frequently occurring categories would produce only a small proportion of the total responses. Such a pattern could be called a "scattered" distribution. Although the author is not aware of anyone who has developed a measure along these lines, this method appears particularly applicable when different groups are to be compared on the same categories.

Inferential Statistics

With inferential statistics, the possibilities for analysis are much greater. By far, the most common inferential technique with unordered data is chi-square test statistic. Without making any assumptions about the underlying population distribution, one can use the chi-square test to compare the expected to the observed results and make a (probabilistic) decision on whether the present findings could have occurred by chance.

There are two basic assumptions in the chi-square statistic for contingency tables. Observations must be independent and a minimum expected value of five observations per cell is recommended (Hays, 1963). Some statisticians tend to relax the latter assumption. Generally, the rules for minimal cell size are somewhat ambiguous, and, within general guidelines, let the investigator decide. Later, a procedure for handling cells with very low expected frequencies will be discussed.

Another technique for testing the significance of relationship in a contingency table is the likelihood ratio test for categorical data (Mood, 1950; Sokal and Rohlf, 1969). Essentially, the test involves the sum of a linear combination of natural logarithms for several observed values (cell, row, column, and total frequencies) in the contingency table. This statistic often called G, has several advantages over the conventional sample chi-square. The arithmetic is somewhat easier and the distribution of "G" more closely approximates theoretical chi-square, especially for small sample sizes. One disadvantage, particularly, when working by hand is the need to consult logarithm tables for doing the calculations. As with sample chi-square, the assumption of independence of observations still holds. Minimum sample size requirements may be relaxed somewhat (Hays, 1963).

However, even a significant chi-square leaves many questions unanswered. It provides little information on the strength of association. Particularly with a large sample, significant relationships are commonly found but tend to be meaningless in practical applications. Similarly in the case of large contingency tables, overall significance is usually not sufficient for the researcher. He may want to identify the significant, as well as the nonsignificant, relationships within the table. Therefore, procedures for further analyzing contingency tables are needed.

Statisticians have proposed several measures to handle most of these problems. These include the phi statistic,[3] contingency coeffi-

[3] When specific sources are not cited any standard statistical test will do. Particularly useful ones are Hays (1963), Dixon and Massey (1969), and McNemar (1969).

cient, and the Cramer index (Goodman and Kruskal, 1954; Hays, 1963). Each provides some measure of association derived from the distribution of scores in the contingency table. These indices are usually difficult to interpret. None of them have an analogue to r^2, proportion of explained variance.

A better index for understanding the results in a contingency table is lambda, λ. Developed by Goodman and Kruskal (1954), λ is equal to the proportion of error reduction in predicting one variable by specifying a category of the other variable. This index is asymmetric and will usually be different for each predictor. It is a close analogue of explained variance with continuous data and appears to be an appropriate value to report in the analysis of contingency tables.

An example applying λ might be a 3×3 contingency table with rows A_1, A_2, A_3 and columns B_1, B_2, B_3. If no other information existed about the B variable, the best guess would be the modal category in the marginal distribution of B. The probability of an error would be one minus the proportion of scores in the modal category. However, if the two variables are associated, usually (see Hays, 1963) knowledge of a category in A will reduce the error in predicting B. This proportional reduction in error is called lambda.[4]

A large chi-square value indicates that the row and column variables are associated, but does not tell us where the relationship is located. A general procedure for handling this is called "partitioning." An $n \times n$ table with either rows or columns greater than two can be divided into several combinations of tables (Bresnahan and Shapiro, 1966; Maxwell, 1961a). The chi-square values of the individual components, which sum to equal the chi-square value of the entire table, can be tested individually for significance.

One of the major advantages of partitioning is the ability to handle missing data or cells with expected values less than five. Usually such tables are analyzed by combining cells or discarding them. However, it is possible to consider the table with missing data as a component of the larger table. The equation used for calculating chi-square in a partitioned table would then be used. Although the degrees of freedom are equal to what they would be if the cells were eliminated or pooled, the chi-square value is different.

Using another approach for post-hoc analysis, Marascuilo (1966) showed that multiple comparisons analogous to those in the analysis

[4] A measure similar in some respects to Lambda and "G" is derived from information theory. Basically, it supplies an index of reduction in uncertainty in one variable given information on the other variable. For further discussion of this measure, see Hays (1963) and Attneave (1959).

of variance were possible with K independent contingency tables. For example, suppose we have five contingency tables each measuring the opinions of students on the issue of legalizing marijuana at five universities and each with identical categories for rows and columns. One can then test the hypothesis of equality of interactions i.e., H_0: $I_1 = I_2 = I_3 = I_4 = I_5$, where the subscripted numbers represent the contingency tables. If the null hypothesis is rejected at the α level, then contrasts can be set up at the $(1 - \alpha)\%$ confidence level to identify the interaction(s) which are significantly different from the others.

Multivariate Techniques

The analysis of contingency tables, particularly those of higher order than a two way table, is also a legitimate component of multivariate analysis (Goodman, 1970). For example, one can have three variables each with several categories and test for association among the variables two at a time, or even three at a time.

In a three way table, at least five different null hypotheses of non-independence can be tested. Birch (1963), who wrote one of the original theoretical articles in the field, showed how to calculate the expected values under each hypothesis. Although the equations can become somewhat involved, they are similar in concept to the analysis of variance. In general the expected values are composed of a linear combination of the logarithm of observed frequencies. The expected values obtained from each hypothesis can be compared to the observed results and a decision made on the most appropriate hypothesis. Fienberg (1970) runs through the entire computation algorithm and shows how to check for the goodness of fit for a particular model (or hypothesis).

Birch (1963) also proved that these maximum likelihood estimates of the expected values are unique and are equivalent across several sampling procedures. Furthermore, Goodman (1969) has shown that the three way table can be partitioned into additive components which can be tested for three way interactions and two way interactions. For further applications with multidimensional tables, see Fienberg (1972), and Bishop (1971). In his recent book, Bock (1975) has discussed some of these techniques in detail.

The more traditional multivariate procedures such as regression or discriminant analysis can also handle unordered data. The typical regression problem is concerned with explaining the variance of a dependent variable from some combination of independent variables. With categorical data, the usual procedure, though not the only one, assigns a "1" to indicate the presence of the independent variable (or category) and a "0" to indicate its absence. Cohen (1968b) discussed at length the use of these so-called "dummy" variables as part of a general exposition on multiple regression. This technique which is commonly employed by economists allows for testing main effects as well as interaction with either categorical or continuous data. The coefficients associated with the independent variables can also represent effect coding or orthogonal contrasts. See Kerlinger and Pedhazur (1973) for a good discussion of the different coding techniques.

Discriminant analysis and canonical analysis which are commonly used for purposes of classification within groups and discriminating between groups are two of the more widely employed multivariate procedures. Although the usual canonical variate analysis seeks to identify the relationships between two sets of continuous variables, it can also be shown to yield the identical results as the multiple discriminant analysis. In this formulation, "dummy" variables are used to represent the groups or categories. Thus, a person who is a member of the *K*th group is assigned a "1" on G_K; for all other G's he is assigned a "0". Tatsuoka (1971) showed that the discriminant criterion value obtained from the usual procedures of maximizing the discriminant equation is a simple function of the canonical correlation value.

This dummy variable scheme is similar to the one used in regression analysis. Typically, however, in discriminant analysis the categorical data occur in the dependent variable whereas in multiple regression it occurs in the independent variable.

However, Claringbold (1958) and Maxwell (1961b) have also applied the procedures of discriminant analysis to categorical independent variables. In the typical discriminant analysis, the objective is to obtain weights on the independent variables which maximize the ratio of variance between groups to the variance within groups. With categorical data, the same procedures are followed. The only difference, involves the form of the data in the original matrix. As before, a "1" is assigned if a particular attribute (the independent variable) is applicable to a group and a "o" if it is not applicable. Different answer patterns result which are summed to obtain frequency of occurrence. Once these sums are obtained, the calculations proceed as usual (Maxwell, 1961b).

An example of the previous procedure would be the following: Assume three different universities *A*, *B*, and *C* as the dependent variables. Assume also that demographic information such as sex, religion (Jewish, Catholic, Protestant) and place of residence (in state or out of state) had been collected on *N* subjects from the three schools. An investigator would like three prediction equations for classifying potential students into one of the schools. A table is set up with 12 columns, each column representing one of the categories in the demographic data. A "1" or "0" indicating presence or absence is assigned to each subject for each category. Pattern of responses are then obtained for the three colleges. The between sums of squares matrix and within sums of squares matrix, which are necessary for determining the coefficients for each function, are calculated from the answer pattern.

Recently, Goodman (1972) has taken a matrix similar to the one required for discriminant analysis of categorical data, and developed a multiple regression approach for a given dichotomous dependent variable. His procedure enables the investigator to determine the weight for main effects and interactions using categorical independent and dependent variables. In Goodman's model the dependent variable can represent "odds" (i.e., probability) of the occurrence of either of two events. The computations, which are relatively straightforward, yield results similar to a multiple regression or an analysis of variance while relaxing some of their crucial assumptions (such as normality or homoscedasticity).[5]

Although traditional factor analysis is not possible with qualitative data, procedures recently developed by Guttman (1968) and Lingoes and Guttman (1967) allow for the reduction of all types of data into a space of lowest dimension, according to a specified set of criteria. The technique called smallest space analysis (SSA) is adaptable to several different kinds of measures including correlations, frequencies, likelihoods, etc. A solution in a particular *n*-space tries to maintain the rank order of distances or similarity of the input data.

The calculation can get quite involved and must be handled on a computer. Programs are presently available which will print out the coefficients of each variable on each dimension (or factor), plot the dimensions two at a time, and translate the input data (continuous as well as qualitative) into distance values. Finally, the goodness of fit for a given number of dimensions is calculated and printed out as the *coefficient of alienation.* Bloombaum (1969) discusses on a very elementary level some of the advantages of SSA even where the data is continuous. Schlesinger and Guttman (1969) have shown that SSA is more parsimonious i.e., yields a smaller space than a factor analysis of the same data.

Smallest space analysis is particularly useful for determining those variables that are associated with particular categories. For example, an investigator may be interested in identifying correlates of three major psychotic types: paranoia, schizophrenia, and mania. Data which

[5] For a lucid description of the equivalence of analysis of variance and regression analysis, see Jennings, 1967.

are both ordered and unordered are collected on individuals from a psychotic population. The ordinal data, in this type of analysis, would be divided into levels such as high income, middle income and low income. The results of the smallest space analysis can then be examined for constellations of variables around each psychotic type. Each constellation could be considered a description of individuals within that category. For an informative illustration of smallest space analysis with all types of data, see Guttman, Guttman, and Rosenzweig (1967).

Other techniques which try to group people from categorical data are also available. One of these is a pattern analytic method or more specifically the linkage analysis developed by McQuitty (1955, 1967). Basically, McQuitty's method allows for the classification of categories into types. The types are simply a function of the pattern of responses. As applied to individuals, this procedure makes it possible to form groups composed of similar or likeminded individuals. Although inter-individual correlations across several variables are commonly used to define similarity, it is just as reasonable to use frequency counts or agreement scores. The latter, as developed by McQuitty (1957), involves computing an agreement score for each subject paired with every other subject. For example, in the case of 5 categorical variables, subjects A and B can have agreement scores of .80 (i.e., they gave similar responses to 4 of the 5 items) whereas subjects A and C can have an agreement score of .20 (i.e., they gave similar responses to only 1 of the 5 items). Individuals with high agreement scores are combined to form groups which in turn define specific types.

In his comprehensive review of the field (often referred to as numerical taxonomy), Baker (1972) describes a general procedure for analyzing such data. After similarity coefficients or agreement scores are obtained among all combinations of subjects, the scores can be discarded from future calculations. It is this feature which makes the categorical variable as amenable to numerical taxonomy as the continuous variable.[6] As Baker points out, sophisticated algorithms and computer programs in this field are approximately 15 years old and applications of these techniques to the behavioral sciences await the innovative researcher.

Some Psychometric Applications

The biserial and point biserial correlation are commonly employed measures for estimating the correlation between a dichotomous vari-

[6] It is important to note that several of the procedures reviewed by Baker (1972) require continuous data in order to determine the similarity index.

able and a continuous variable (Walker and Lev, 1952; Mcnemar, 1969). The biserial r is recommended when one of the two continuous variables has been artificially dichotomized (this is the reason it has been called a short cut version of the product moment r) whereas the point biserial r is to be used when one of the variables is inherently dichotomous. In such cases, the dichotomized variable may be unordered. Therefore, when analyzing categorical data such as the sex variable, it would seem appropriate to use point biserial r for calculating the correlation coefficient.

With more than two categories most psychometric indices require ordered data. However, in the area of reliability, particularly the inter-rater variety, several innovative ideas have been proposed. Goodman and Kruskal (1954) suggested a measure, λ_r which estimates the degree to which two raters agree on the assignment of individuals to different categories. A contingency table is used for calculating λ_r. The categories are placed in the same order for both rows and columns. The rows represent the responses for one judge and the columns the responses for the other judge. The range of λ_r is -1 to 1. The lowest value occurs when there is complete disagreement and the sum of the two marginal proportions corresponding to the modal class is 1. The highest value occurs when the two raters always agree. The practical feature about λ_r is that it is easily interpretable. The authors define the statistic "as the relative decrease in error probability as we go from the no information situation to the other method (rater) situation." (p. 758).

More recently Cohen (1960, 1968a) has developed two indices, kappa (K) and weighted kappa, which are similar to lambda. The indices are calculated by subtracting the mean observed degree of disagreement from the mean degree of disagreement by chance and dividing this result by the latter value. Both kappa and weighted kappa vary from -1 to 1 with a negative value indicating poorer than chance agreement, 0, chance agreement, and a positive value, better than chance agreement. The only difference between the two measures is that weighted kappa allows the investigator to specify the relative seriousness of each kind of disagreement. Interestingly when the categories for the variables can be ordered, the weighted kappa and the intra class correlation coefficient are equivalent (Fleiss and Cohen, 1973).

Finn (1970) proposed another technique for determining the reliability of categorical data. However, the procedure he used implied an order to the categories and would not be applicable here. In general, one must be careful not to infer order from numerals whose only purpose is discrimination between categories.

All of the preceding measures, kappa, weighted kappa, and lambda involve the classification of many individuals into one of several categories by two raters. Suppose the information on these categories can be obtained from several sources or items. We then have something quite similar to the reliability measure for an entire test. Koslowsky and Bailit (in press) have recently derived a measure anologous to λ_r but based on combining the information from several items. This index describes the reduction in error caused by knowledge of the judges' ratings on a series of items.

Conclusion

This paper has focused on recent techniques for analyzing categorical variables. Most of the commonly used statistical procedures such as r^2, factor analysis and reliability determination have their counterparts in Kruskal and Goodman's lambda, Guttman's smallest space analysis, and Cohen's Kappa, respectively.

Although these pairs yield somewhat similar measures, one must be cautious with the interpretation of the results. A reliability coefficient obtained by correlating the ratings of two judges on several individuals is not equal to kappa which compares agreement to disagreement frequencies for the two judges. A decision on what index to use must be based on two criteria: (a) the type of data the researcher is required to analyze (nominal, ordinal, or interval) and (b) the kind of inference the researcher would like to make.

In general, the problem of determining the best way for analyzing data is an old one in psychology. The researcher should, whenever feasible, extract all useful information from the study. If the data is ordered, statistics are available that are more efficient than those mentioned in this paper (Siegal, 1959). With legitimate interval data, analysis of variance and *t* tests are still more powerful than the non-parametric ordinal procedures, provided other parametric assumptions of these techniques can be justified. However, the increase in power is paid for by an increase in the stringency of the assumptions that must be met by the data. Often the researcher is unwilling or unable to make these assumptions.

The investigator who has collected unordered data obviously has little choice but to use the techniques described here. However, even the individual who has data that appear to be interval can conservatively do an analysis which presumes unordered data. In some cases, such as smallest space analysis, there are some advantages for performing categorical analyses rather than the interval counterparts. Similarly, an investigator might have difficulty in obtaining a quantitative measure but may be confident of some qualitative classification for each subject. Instead of considering his independent variables as continuous, dummy variable analysis might be an appropriate alternative. Depending on the objectives of the study, multiple regression or discriminant analysis with dummy variable coding are two possible procedures for analyzing the data. In any case, one might well consider some of the procedures discussed here when analyzing interval as well as categorical variables.

REFERENCES

Attneave, F. *Application of information theory to psychology.* New York: Holt, Rinehart & Winston, 1959.

Baker, F. B. Numerical taxonomy for educational researchers. *Review of Educational Research,* 1972, 42, 345–358.

Birch, M. W. Maximum likelihood in three-way contingency tables. *Journal of Royal Statistical Society,* 1963, 25, 220–233.

Bishop, Y. M. Effects of collapsing multidimensional contingency tables. *Biometrics,* 1971, 27, 545–562.

Bloombaum, M. Doing smallest space analysis. *Journal of Conflict Resolution,* 1969, 14, 409–416.

Bock, D. *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill, 1975.

Breshahan, J. L. and Shapiro, M. N. A general equation and technique for the exact partitioning of chi-square contingency tables. *Psychological Bulletin,* 1966, 66, 252, 262.

Cohen, J. A coefficient of agreement for nominal scales. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1960, 20, 37–46.

Cohen, J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin,* 1968, 70, 213–220. (a)

Cohen, J. Multiple regression as a general data analytic system. *Psychological Bulletin,* 1968, 70, 426–443. (b)

Claringbold, P. H. Multivariate quantial analysis. *Journal of the Royal Statistical Society,* 1958, 20, 398–405.

Dixon, W. J. and Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1969.

Fienberg, S. E. The analysis of multidimensional contingency tables. *Ecology,* 1970, 51, 419–433.

Fienberg, S. E. The analysis of incomplete multi-way contingency tables. *Biometrics,* 1972, 28, 177–202.

Finn, R. H. A note on estimating the reliability of ctegorical data. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1970, 30, 71–76.

Fleiss, J. L. and Cohen, J. The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1973, 33, 613–619.

Goodman, L. A. On partitioning X^2 and detecting partial association in three-way contingency tables. *Journal of Royal Statistical Society,* 1969, 31, 486–498.

Goodman, L. A. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association,* 1970, 45, 226–256.

Goodman, L. A. A modified multiple regression approach to the analysis of dichotomous variables. *American Sociological Review,* 1972, 37, 28–46.

Goodman, L. A. and Kruskal, W. H. Measures of association for cross-classifications *Journal of American Statistical Association,* 1954, 49, 732–764.

Guttman, L. A. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika,* 1968, 33, 469–506.

Guttman, R., Guttman, L. and Rosenzweig, K. A. Cross-ethnic variation in dental, sensory, & perceptual traits: A nometric multibivariate derivation of distance for ethnic groups and traits. *American Journal of Physical Anthropology,* 1967, 27, 259–276.

Hays, W. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Jennings, E. Fixed effects analysis of variance by regression analysis. *Multivariate Behavioral Research.* 1967, 2(1), 95–108.

Kerlinger, F. and Pedhazur, E. *Multiple regression in behavioral research.* New York: Holt, Rinehart, & Winston, 1973.

Koslowsky, M. and Bailit, H. A measure of reliability using qualitative data. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, in press.

Lingoes, J. C. and Guttman, L. Nonmetric factor analysis. A rank reducing alternative to linear factor analysis. *Multivariate Behavioral Research,* 1967, 2, 485–505.

Marascuilo, L. Large sample multiple-comparisons. *Psychological Bulletin,* 1966, 65, 280–290.

Maxwell, A. E. *Analyzing qualitative data.* London: Methuen, 1961. (a)

Maxwell, A. Canonical variate analysis when the variables are dichotomous. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1961, 21, 259–271. (b)

McNemar, Q. *Psychological Statistics.* New York: Wiley, 1969.

Mcquitty, L. L. Pattern analysis illustrated in classifying patients and normals. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1955, 89, 598–604.

Mcquitty, L. L. Elementary Linkage Analysis for isolating orthogonal and oblique types and typal relevancies. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1957, 17, 207–229.

Mcquitty, L. L. A mutual development of some typological theories and pattern-analytic methods. EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 1967, 27, 21–46.

Mood, A. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

Nunnally, J. *Psychometric theory.* New York: McGraw-Hill, 1967.

Schlesinger, I. M. and Guttman, L. Smallest space analysis of intelligence and achievement tests. *Psychological Bulletin,* 1969, 71, 95–100.

Siegel, S. Nonparametric Statistics for the behavioral sciences. New York: McGraw Hill, 1956.

Sokal, R. R. and Rohlf, F. H. *Biometry.* San Francisco: W. H. Freeman & Company, 1969.

Tatsuoka, M. M. *Multivariate analysis: Techniques for educational and psychological research.* New York: Wiley, 1971.

Walker, H. M. and Lev, J. *Statistical inference.* New York: Holt, Rinehart, & Winston, 1952.