A COMPARISON OF SEVERAL METHODS OF SCORING THE KUDER OCCUPATIONAL INTEREST SURVEY

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

WILLIAM EARL LOADMAN II

1971

This is to certify that the

thesis entitled

A COMPARISON OF SEVERAL METHODS OF SCORING
THE KUDER OCCUPATIONAL INTEREST SURVEY

presented by

William Earl Loadman II

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling, Personnel
Services, and
Educational Psychology

Robert C. Craig
Major professor

Date 10/26/71

O-7639

(-7

ABSTRACT

A COMPARISON OF SEVERAL METHODS OF SCORING THE KUDER OCCUPATIONAL INTEREST SURVEY

By

William Earl Loadman II

The ability of an occupational interest survey to accurately predict occupational classification is related to the method of scoring the instrument. The present research compared four different methods of scoring the Kuder Occupational Interest Survey (Form DD). The purpose was to discover more accurate methods of scoring the Kuder for use in predictive situations. Three scoring procedures used the response pattern of each item. These were: (a) pattern analysis, (b) Chi-square, and (c) Kuder lambda. A fourth method of scoring, discriminant function, was based upon one of the first three, viz. the Kuder lambda.

Method

The basic data for the present study were the same response scores of subjects from nine occupations originally used by Kuder to empirically derive nine lambda scoring keys. In this research, these occupations were divided into two groups of five similar and five dissimilar

occupations. This split provided the researcher with the opportunity to investigate the potential of each scoring method in classifying respondents from (a) similar occupations, and (b) dissimilar occupations. The data from both the similar and dissimilar occupational groups were randomly split in half to allow for cross-validation procedures.

Using a dichotomous (hit or miss) classification score for each subject of the cross-validation sample as the dependent variable, a comparison among the accuracies of the scoring methods was made for each set of data. This comparison was accomplished through the use of a repeated measures analysis of variance model. The five occupations (between subjects) and the four methods of scoring (within subjects) variables were defined as fixed factors in the analysis, while subjects nested within groups was treated as a random factor. Conservative tests of the sources of variation were employed by applying the Greenhouse and Geisser reduction in degrees of freedom. A significant groups by methods disordinal interaction was found in both sets of data. Secondary analyses were employed to determine the significant simple effects of the groups and methods variables. All significant simple effects were subjected to Scheffé post hoc procedures to determine the significant pair-wise contrasts among means.

Results

It was concluded that none of the proposed scoring methods functioned best for all situations. Rather, the post hoc procedures suggested that the scoring procedure which was best in classifying persons within a particular group was dependent upon the set of groups under consideration, the persons within the groups, and in certain cases, the number of persons within the group. The discriminant function, Chi-square and Kuder lambda methods of scoring all performed significantly better than the pattern analytic method. The pattern analytic method classified subjects at about the chance probability level while the remaining three methods each functioned well above the chance level. Of these three methods, no method was consistently superior to the other two methods. For all scoring methods, the classification of subjects was more accurate in the set of dissimilar than in the similar set of occupations.

The Kuder lambda scoring keys (keys now used commercially) were used in this research and the results of the classification based on these keys were compared with those of the three proposed methods of scoring the Kuder. Unfortunately, a true cross-validation sample was not available; this rendered the Kuder lambda keys slightly more predictive of occupational classification than otherwise would be true. The Kuder lambda method was most accurate in classifying five of the ten groups.

The discriminant function method of scoring the Kuder was based on the classification score from each Kuder lambda scoring key. Since there was no true cross-validation for the Kuder lambda technique, it follows that there was no true cross-validation for the discriminant function method. Quasi cross-validation was used and resulted in the most accurate classification by the discriminant function method for two of the ten groups. On the set of similar occupations, the discriminant function method yielded excellent predictions for four of the five groups. For some unknown reason, the correct classification of persons in the fifth group broke down. When considering all of the persons within the set of dissimilar occupations, the discriminant function method accurately classified 82% of the persons on the quasi cross-validation.

A Chi-square method of scoring the Kuder yielded results comparable to those of the Kuder lambda and discriminant function methods for the set of similar occupations. The Chi-square method was slightly less predictive than the Kuder lambda and discriminant function methods on the dissimilar occupational data. Of the four proposed methods of scoring, the Chi-square method was most accurate in classifying persons in three of ten groups.

The pattern analytic scoring method did not function well. This method was found to be significantly worse than the other three methods across all conditions investigated in this research.

A proposed alteration in the procedure for classifying subjects using the pattern analytic method was compared with the original pattern analytic method. Neither method exceeded chance probability in correctly classifying individuals on the cross-validation. The present research was not an adequate test for this comparison.

Conclusions

The results of this research indicated that there was no single best method to be used for scoring the Kuder when either a select set of similar or dissimilar occupations was considered. A significant groups by methods disordinal interaction suggested that the best scoring method for the Kuder was dependent upon the groups under consideration, the computer resources available, and the required degree of accuracy. Because true cross-validation was not used in all aspects of the study, more explicit statements of conclusions are not warranted.

A COMPARISON OF SEVERAL METHODS OF SCORING THE KUDER OCCUPATIONAL INTEREST SURVEY

By

William Earl Loadman II

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirement
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling Personnel Services
and Educational Psychology

1971

ACKNOWLEDGMENTS

A dissertation is not developed and written in a vacuum. There are many persons who have been instrumental in the initial development and subsequent presentation of this thesis. First I wish to thank the members of my committee. Dr. Terrence Allen and Dr. William Mehrens provided encouragement and helpful comments. Dr. Andrew C. Porter generously shared not only his data, but his time and ideas in offering commentary and constructive criticism. He was most helpful in guiding me through several rocky stages in the development of this research. Most of all, I am grateful to the chairman, Dr. Robert C. Craig. His patience, guidance, and encouragement in addition to his valuable suggestions were appreciated in the formative stages and his skillful instruction and insight were highly valued in improving the presentation. I would also like to extend my appreciation to Science Research Associates.

At this point, I wish to say a special thank you to a very good friend, Dr. Donald J. Freeman. It was his initial encouragement that led me to pursue doctoral study. I wish I could find the words to adequately express my gratitude for the positive influence his continued

friendship, interest, and encouragement have been for me.

Finally, I must express my sincere thanks to my wife, Gail. She has doggedly persisted in keeping me on target. I will never be able to repay her many sacrifices in this dissertation effort. It is to her, to our son, William III, and to my parents, who have always been close by with their own special encouragement, that this dissertation is affectionately dedicated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM OF SCORING AN OCCUPATIONAL INTEREST SURVEY

Introduction

It has been repeatedly demonstrated that an individual possesses certain interests which are related to occupational preference. The measurement of these interests has stimulated psychologists for several decades; and, in spite of the considerable time and effort expended in attempting to determine reliable and valid measurement (whether stated, observed, or covertly detected by unobtrusive measures), there still exists marked imprecision in the methodological procedures.

During the past twenty to thirty years, one particular methodological technique has been consistently used; it attempts to quantify the measurement of reported interests and subsequently employ these scores in a prediction model. The quantification of interests through a self-report to a series of items, is called an interest survey. The survey, usually referred to as an instrument, requires the subject to respond to a carefully generated list of items. His responses are compared with responses of one or more reference groups and/or with the responses of a group of men in general. The reference or criterion group generally consists of individuals employed in specific occupations. For example, one reference group may be forest rangers and the responses to the survey items of persons employed as forest rangers serve as the pattern for identifying persons with like interests. Thus, occupations or areas of occupation can be identified and an individual can be "classified" as having reported interests similar to those in the reference group and different from men in the other reference groups or the men in general group. If this is accomplished, discrimination between occupational interests is said to have occurred.

In vocational counseling situations, the derived results of the interest survey are used in conjunction with additional pertinent data (i.e., subject's abilities and current functioning levels) to identify potential occupational courses of action and possible alternatives which are consistent with and palatable to the individual. The ultimate decision regarding any of the suggested actions obviously rests with the individual.

The desired end result of conducting the survey is to provide the individual and/or employer with predictive information regarding prospective occupational satisfaction and job-individual compatability. Clearly, it is imperative that the predictions be as accurate and precise as possible.

Kuder (1961, p. 3) stated:

> More recently, studies of certain jobs have revealed that a worker's preferences among well-known activities are related to his job satisfaction and to the length of time he is likely to stay on a specific job. There is also some evidence that interest is related to achievement.

According to Berdie and Campbell (Whitla, 1968) interest may be a better indicator of satisfaction in a given occupational pursuit than ability. In view of the evidence suggesting that interest plays a primary role in occupational satisfaction, the present study considered only the problem of measuring interests.

The objective of the present study was to build a predictive model based solely on interest. There were several limiting factors which affected the construction and subsequent use of this model. The selection of the sample and the chosen occupations influenced the construction. Actual use of the model is affected by the fact that the interest measure is not the sole criterion upon which a vocational decision is based. A subject may exhibit dramatic interest in a particular job or job field, but be lacking a vital physical or mental capacity to allow successful performance. Consider the occupation of an airline pilot; a man may have significant and compatible interests to pursue this vocation, but have faulty depth perception. Obviously if the occupational prediction were derived solely from the interest criteria, a disastrous situation could result.

An additional and rather serious problem associated with use of a survey is the absence of assurance that the measured interests and the actual interests are the same. The ramifications of this problem are twofold. First, a given subject may report interests of a given nature which may not be reflected in his actual behavior. Thus, the self-report instrument may not yield a valid determination of a subject's interests. Second, there is no assurance that subjects within a particular criterion group espousing particular kinds of interests are in fact satisfied with their present occupation. However, Kuder (1965) provided validation evidence which supported the basic assumption that persons within a given occupation are satisfied and that this satisfaction is positively related to interests.

This study focused on one of the most widely used instruments, the Kuder Occupational Interest Survey, Form DD (1966), which will hereafter be referred to as the Kuder. The Kuder is employed by many counselors and is relied upon in many vocational counseling settings. Therefore, there is an express need for the information supplied by the Kuder to be as accurate as possible when used in predictive situations. To accomplish this, the survey must be able to identify the subject's interests and further, these interests must be applicable in accurately classifying the subject with respect to selected criterion groups.

The purpose of this research was to determine more accurate scoring procedures for the Kuder employing selected criterion groups. As pointed out by Gaier and Lee (1953), the discriminatory power (i.e., the ability of the instrument to accurately differentiate between persons in two or more groups) of a predictive instrument involves the method of scoring as well as the instrument itself. Taking the Kuder instrument as a given, this study considered alternate methods of scoring the Kuder. The proposed methods were compared with the presently employed Kuder system and with a system employed by another author.

Description of the Kuder Instrument

The primary purpose of the Kuder is to classify males into vocational occupations using their interests (as determined by their responses to the items on the survey) as predictors of a specific occupation. The instrument consists of 100 triadic items. Each item is comprised of three sentences or phrases and each of these deals with a common activity which is generally understood and does not directly reflect any particular occupation. The vocabulary is pitched at the sixth grade level, and Kuder claims that the tool is applicable to persons at all levels of educational and vocational sophistication. Responding to each item, the subject indicates the activity which he most prefers and the activity which he least prefers. The directions are easily understood and there is no set time limit for administration. Kuder reports that this time typically ranges from 20 to 30 minutes. The instrument can be applied to an individual or a group with equal ease; the appropriate target population is males from ninth grade through adult. The scoring keys developed by Kuder are empirically based upon the occupation under consideration. There are approximately 77 scoring keys presently available (Kuder, 1967). Further elaboration on the Kuder is unnecessary and not pertinent to the development of the procedures investigated. Science Research Associates has the copyright on the original instrument and additional information can be obtained from that source.

Since Kuder does not provide a "correct" answer for each item, the problem becomes one of weighting each response and obtaining a composite score that will result in a discrimination of subjects along a continuum of interest. Porter (1966, p. 3), stated:

> A solution to the weighting is dependent upon the groups which are to be discriminated among, the equipment available, and the desired limits as to the complexity of the weighting procedure.

Kuder (1957) asserted that there are many potential possibilities for scoring and predicting group membership, and theoretically the very best method would consider all possible combinations of responses, "an astronomical number." Dismissing this alternative as impractical, Kuder proceeded to generate scoring keys based upon the pair-wise comparisons of responses of all occupational groups across each of the 100 items (Kuder, 1966). The keys were derived by determining a lambda coefficient for each response alternative for each occupational group. The occupational category yielding the largest sum of lambda coefficients was considered the classification category for a given subject. This procedure will be discussed in detail in the following section.

Kuder Weighting Technique

When Kuder was originally devising his instrument he encountered a major hurdle in the development of scoring keys. Kuder (1957) devoted considerable time to this problem of scoring and analysis and concluded that it was not possible to specify the one best scoring system; the procedures were dependent upon a host of factors including the number of cases, the composition of the survey, the content and type of item, the range of item validities, the homogenity of the group used, and the distribution of item characteristics. It would appear that the best possible scoring system must be devised for the specific situation. This alternative was thought to be entirely too limiting and Kuder, on the basis of certain empirical evidence, proceeded in another direction.

Reasoning that the research on item weighting was not definitive, a system of weighting was selected which yielded acceptable rather than optimal results; it was computationally simple and easy to interpret. For these reasons a unit weighting system was endorsed for the scale. The unit scoring keys for each occupation were developed and used to differentiate persons in the given occupation from persons in the norm group.

The norm group consisted of 1000 men obtained in a stratified random telephone survey of 138 cities and towns which were assumed to represent sections and segments of the United States population (Kuder, 1961). Kuder stated that this group was not the most representative sample that could be obtained, but that it was adequate for his purposes. The specific characteristics of this norm group had a pervasive and unknown effect upon all the scoring keys constructed by Kuder. The building of a key via the Kuder method* was straightforward and although no specific sample size was insisted upon, a "rule of thumb" was 200 subjects. However, Kuder was quite insistent regarding adequate cross-validation samples to determine the proficiency of the key. With additional research, improved scoring methods were proposed.

---

*Since the Kuder unit weighting system was not used in this study, discussion of the derivation of the original scoring procedure is limited. Derivations of the scoring methods used in this study are presented in the methodology section of the thesis.

Kuder subsequently revised the original procedure and began using a fractional (non-unit) weighting technique. This new system resulted in a more accurate method of prediction (Kuder, 1966), and the Kuder Occupational Interest Survey (Form DD) was originated. The items and the format of the instrument were identical to those employed in Form D; only the method of scoring was altered.

The new scoring system was based upon the work of Findley (1956) and Clemans (1958), and was developed in the following manner. Each item (triad) consists of three statements; a given subject is required to choose one of the statements as the most appropriate for himself and one of the remaining statements as the least appropriate for himself. For a given criterion group, the proportion of subjects that selected each one of the three statements as most liked and the proportion that chose each one of the three statements as least liked were calculated. Six proportions were determined for each item for each criterion group. The following example is helpful in understanding the procedure:

Item XX	Most	Least
Statement 1	$P_{(1)}$ X	$P_{(4)}$ X
Statement 2	$P_{(2)}$ X	$P_{(5)}$ X
Statement 3	$P_{(3)}$ X	$P_{(6)}$ X

$P_{(1)}$ represents the proportion of subjects in group j selecting statement 1 as the most liked alternative; $P_{(2)}$ represents the proportion of subjects in group j selecting statement 2 as the most liked, etc.; and $P_{(4)}$ represents the proportion of subjects in group j selecting statement 1 as the least liked alternative, etc. This procedure was continued until the proportions for each of the 100 items were calculated (a total of 600 proportions for each group). Any given subject within the criterion group would necessarily have chosen a subset of 200 of the 600 possible response alternatives. A single score for each item for each subject was determined by combining the proportions from the two selected responses of the subject in the following fashion, yielding one of six possible response pattern scores for a given subject:

$$1 = P_{(1)} + P_{(5)}$$
$$2 = P_{(1)} + P_{(6)}$$
$$3 = P_{(2)} + P_{(4)}$$
$$4 = P_{(2)} + P_{(6)}$$
$$5 = P_{(3)} + P_{(4)}$$
$$6 = P_{(3)} + P_{(5)}$$

The sum of all proportions for a given triad was 4.00 as the sum of $P_{(1)}$, $P_{(2)}$, and $P_{(3)}$ was equal to 1.00 and the sum of $P_{(4)}$, $P_{(5)}$, and $P_{(6)}$ was equal to 1.00, and each proportion appears twice in the above diagram. For each subject, the sum of the proportions of the 100 endorsed response patterns served as the continuous variable indicating the subject's similarity to the rest of the group. A dichotomous variable was defined by the subject's selection or non-selection of the particular response pattern alternative. By generating the correlation coefficient between the above continuous and dichotomous variables across subjects in an occupation, it was possible to generate 600 correlations; one correlation for each of the six possible alternatives to each of the 100 items. These 600 correlations represent the relationship between the individual's interests and the interests of the group.

Because of differences in the homogeneity of the criterion groups, the range of possible scores varied considerably from group to group. Therefore, it was difficult to make interpretations when comparing the potential classification of an individual based upon the sum of 100 point biserial coefficients. A lambda coefficient was introduced to eliminate the problem of varying ranges of the correlation coefficient; also, the statistic did not make any assumptions concerning the shapes of the distributions of the two variables. In essence, the lambda coefficient was the ratio of the obtained point biserial to the maximum possible point biserial for the data under consideration. The upper limit of this statistic was 1.00 regardless of the homogeneity of the group. Lambda coefficients were used as weights for the Kuder. Thus, there were 600 potential weights for each criterion category and a subset of 100 of these weights was used (according to the response pattern of the subject) to generate the subject's score. A subject's classification score was determined by summing the lambda coefficients for the 100 endorsed response patterns for a particular criterion category and this procedure was repeated for each criterion category under consideration. A subject was classified into the criterion category which yielded the largest sum of lambda coefficients.

Preliminary research comparing the two Kuder scoring techniques (unit and lambda) indicated significantly less overlap in classification (greater accuracy) with the lambda technique (Kuder, 1970).

CHAPTER II

REVIEW OF THE LITERATURE

General History

A review of the literature revealed considerable evidence of efforts to develop weighting systems and scoring keys for interest surveys. Cowdery (1925) basing his work on an earlier study, determined a weighting system for each item when there were two criterion groups and the responses were dichotomous. The weight for each item was $wt = \frac{\Phi}{(1 - \Phi^2)} \sigma$, where Φ was the Phi coefficient and σ was the standard deviation of the (a+b) and (b+d) cells of a two by two matrix. The matrix consisted of groups as one dimension and responses as the other dimension.

	Response 1	Response 2
Group 1	a	b
Group 2	c	d

Kelly (1923) demonstrated that the Cowdery procedure was a good estimate of a least squares regression weight (for each item). This procedure was used as the basis for scoring the Strong Vocational Interest Blank

(Stanley, 1964). However, the procedure was limited to dichotomous criteria and dichotomous responses. Tilton (1937), reported a method of weighting which compared the difference in the proportion of subjects selecting particular response alternatives between two criterion groups. This empirical weighting system seemed to function quite satisfactorily, but it was limited to two criterion groups at any one time. Kuder (1963) cited both the advantages and disadvantages of using Tilton's proposed model and provided empirical evidence related to its effectiveness. The results indicated a high degree of accuracy, but marked limitatations in application (dichotomous items and groups). Berdie and Campbell (Whitla, 1968) asserted that there has been a prolonged disagreement with respect to the "best" method of weighting survey items; consideration had been given to unit weighting, fractional weighting, and pattern weighting. At this point in time it is not evident that any of these alternatives is superior to the others.

Pattern Analytic Technique

A large number of studies have employed pattern weighting techniques (rather than item weighting) in an attempt to improve prediction. The following studies represent the major endeavors to improve prediction in a potentially fruitful fashion.

Brigham (1932) considered item patterns or configural scoring as a possible method of response weighting. A response pattern was developed for each individual and a subsequent comparison was made of the individual's responses with responses from specific criterion groups. Zubin (1937) generated a similar technique in conducting a study of likemindedness; and, concluded that although his results were not striking, there appeared to be great promise in the technique. Meehl (1950, 1954) probably made the most significant contributions to the advancement of configural scoring patterns. He strongly supported the contention that one item may not be predictive, but a pattern of two or more items, including the item that was non-predictive when used alone, may be highly predictive. If patterns of items were not taken into account, he argued, important psychological information may be lost.

Others who made contributions to configural scoring techniques were: Cattell (1949), Cronbach (1949), Cronbach and Gleser (1952), and Dumas (1946). Although the results were not conclusive, each of the above mentioned studies indicated the potential value of pattern weighting in addition to item weighting. However, these researchers have not empirically and successfully shown the productivity of the method. More recently, other warriors have undertaken the crusade of configural scoring and have revealed either new techniques or empirical evidence regarding the validity of old techniques.

Ghiselli (1960) verified that items or "tests" as he termed them, can be employed to improve the accuracy of prediction of criterion classification when used in combinations even though no relationship was evident between the individual "test" and the criterion measure. He demonstrated this with a linear and non-linear model. Lee (1961) approached the problem of configural scoring in a somewhat different manner, but clearly demonstrated that the additive model can be improved upon by taking into account certain "interactions" within the data. These "interactions" were in fact pattern configurations. She continued by defining non-additive models and closed with the statement:

> It is at least theoretically possible that these relations, or interaction effects, might in some cases be so extreme that a set of independent variables, each of which independently correlates zero with a given dependent variable, could perfectly predict the dependent variable if considered jointly by appropriate methods (Lee, 1961, p. 804).

Horst (1954) mathematically enhanced the claims of Meehl and presented evidence that the configural scoring methods represented a special case of non-linear item weighting. He concluded that configural scoring should be more fruitful with the introduction of non-linear mathematical models. Fricke (1956) found that the discriminatory power of configural-content intensity items was greater than that of the commonly employed linear methods.

In the late fifties and early sixties, McQuitty (1957, 1958, 1959, 1961a, 1961b, 1963, 1966) produced a significant number of studies and introduced several new pattern analytic, or as he termed them, typal techniques. McQuitty's most prominent contribution was probably in identifying areas of concern and problems associated with patterning techniques. McQuitty developed elementary linkage analysis. Initially this technique appeared to be quite promising, but the related problems quickly negated utilitarian ideas. The foremost drawback was in the method of data reduction. A subject's by items matrix must be mathematically reduced in an attempt to force parsimony on the data. The results were mathematically consistent, but the resulting output presented a distorted picture of psychological reality. Other problems included limitations on the size of the data matrix and the number of criterion groups which can be simultaneously analyzed.

The work on pattern analytic techniques has been continued with a criterion pattern analysis computer program generated by Clark (1968). Through the utilization of data processing equipment, it became feasible to consider a large number of items and a large number of subjects simultaneously. This program allowed for the analysis of nominal level data when both the criterion and predictor variables were dichotomous. A full description of Clark's method of scoring and the functioning of the pattern analysis program will be presented in the methodology section of this paper.

Two sets of data were analyzed in Clark's study and they consisted of: (a) the responses of 99 women to an introversion-extraversion scale (independent variable, dichotomous items) and the hidden figure test of field dependence-independence (criterion variable, dichotomous grouping); and (b) the United Nations roll call voting record of 110 nations on 44 issues (one vote was considered the criterion variable and the remaining 43 votes were considered the independent variables).

Clark compared the results of the pattern analysis program to multiple regression and maximum likelihood estimates on original and cross-validation samples. The pattern analytic, regression and maximum likelihood estimates all predicted well using the original data, but upon cross-validation, the pattern analytic technique proved to be better than the maximum likelihood technique. The most successful technique in terms of cross-validation was the pattern analytic technique in combination with a linear prediction model.

Distinct advantages of Clark's pattern analytic approach over previous configural scoring schemes were: (1) the program could handle large quantities of data; (2) it could consider all combinations of response alternatives; (3) patterns were generated separately for each criterion group; (4) the program could identify configural (several items in combination) as well as non-configural (one item) relationships; (5) the patterns could be used directly in prediction; and (6) the patterns were easily interpretable. His study demonstrated the feasibility and credibility of employing a pattern analytic technique in the analysis of multiple measure data.

Certain inadequacies were also identified in Clark's prediction program; classification of an individual was based upon the highest discrimination index. The length of patterns, number of patterns, and the number of persons having the patterns were not considered in the prediction program.

The present study proposed to explore the generalized use of the Clark technique to include items with more than dichotomous responses and more than two criterion groups. The potential of patterning techniques has consistently been identified but the accompanying methodological procedures have been lacking to some degree. This study planned to make contributions in this area.

Discriminant Function Technique

Another possible way of weighting item responses can be accomplished through the use of the linear discriminant function analysis. The early work in this area reverts back to Sir Ronald Fisher (1937) who initiated the analysis of multiple measures on groups of individuals.

Rao (19
fied as
him to
with a
ables.
among g
their m
mate st
signifi

variabl
used, t
sion an

p. 117)

> dete
> of v
> of s
> squa
> sum

Tatsuoka
history
fifteen
thrust o
analysis
Fisher.
terion g

Rao (1948) worked on the classification of persons identified as representatives of particular groups. This led him to the employment of the discriminant function analysis with a dichotomous criterion and several independent variables. He concluded that it was possible to discriminate among groups on the basis of the likelihood estimates of their multiple variable response distributions. Approximate statistical tests were presented to examine the significance of group discrimination.

Welch (1939) demonstrated that when the criterion variable was dichotomous and the discriminant function was used, the results were equivalent to the multiple regression analysis of these same data.

According to Bock and Haggard (Whitla, 1968, p. 117):

> [The discriminant function] . . . procedures determine the coefficients of the linear combination of variables which best discriminate between groups of subjects, in the sense that between groups sum of squares is a maximum with respect to the within group sum of squares.

The work of Rao was extended by Rulon, Tiedeman, Tatsuoka, and Langmuir (1967). Their book presented a history of personnel classification based on approximately fifteen years of research by the authors. The major thrust of the text concerned the discriminant function analysis and extension of the technique as proposed by Fisher. The method could accommodate more than two criterion groups and its purpose was to generate a G space

such
spec
was a
each
group
multi
genera
determ
signif
where
groups
deviat
weights
linear
techniq
multiva
junctio
The ind
determi
receive
less ove
criminat
was grea
of the s
bility wa
and Morri

such that individuals could be maximally classified on a specific dimension of the G space. The G classification was accomplished through the assignment of a weight to each item which would maximally separate the criterion groups. These weights could be determined through a multivariate analysis. The procedure made it possible to generate the multivariate analysis of variance and then determine the eigenvector associated with the largest significant root of the determinental equation $|W^{-1}A-\lambda I|=0$ where I is an identity matrix and W is the pooled within groups deviation matrix; A represents the between groups deviation matrix. This vector yielded the appropriate weights for the respective items. Thus an item weighting, linear prediction model could be employed when using the technique. For purposes of individual classification, multivariate response contours were generated in conjunction with the significant latent roots of the function. The individual's resultant score was then obtained to determine in which contours his score fell (a subject received a score for each significant latent root). With less overlapping of the contours, there was greater discrimination and predictability of the function. If there was great overlapping of the contours, the classification of the subjects was extremely difficult and the predictability was markedly reduced. According to Anderson (1958) and Morrison (1967), if there were G criterion groups, then

assoc

of sul

the nu

groups

at Pur

analys

(1967)

non-en

functi

the ba

profil

enginee

dents,

in the

cross-

and mul

criter

enginee

and cou

vid
fun
dis
lar
per
par

then it was possible to have G-1 significant latent roots associated with the discriminant function if the number of subjects exceeded the number of variables (items), and the number of variables was greater than the number of groups.

Three recent dissertation studies were completed at Purdue University utilizing the discriminant function analysis of the Strong Vocational Interest Blank. CoBabe (1967) using two criterion groups, student engineers and non-engineers, demonstrated that the multiple discriminant function could be reliably used to separate individuals on the basis of the Strong Vocational Interest Blank (SVIB) profile. The sample was limited to a group of freshmen engineering students and a group of non-engineering students, all at Purdue University. There were 476 students in the sample. The results were verified in subsequent cross-validation procedures.

Chappell (1967) conducted a study using the SVIB and multiple discriminant function analysis. He used as criterion groups, 92 veterinary students, 32 graduate engineering students, and 76 graduate students in guidance and counseling. He asserted:

> The results and conclusions of this study provided suggestive evidence that multiple discriminant function analysis is a powerful statistical tool for discriminating among groups, which, heretofore, have largely resisted sharp discrimination on self report personality measures. Additional research into comparisons of other occupational groups is necessary as

w
h
c
(

cases

this

popula

were r

number

used f

author

parame

present

a high

the thr

apparen

under th

analysis

He attem

engineer

included

the first

31 years

ological

Because o

accuratel

> well as comparisons within occupational domains which have several sub-divisions. The use of multiple discriminant function analysis is hereby recommended (Chappell, 1967, p. 70).

Chappell's study used a relatively small number of cases within each criterion group. With groups that are this small it is difficult to obtain stable estimates of population parameters. Since the three original samples were randomly divided for cross-validation procedures, the number of measures almost equalled the number of subjects used for determining the initial equations. Thus the author assured himself of a relatively good fit of the parameters on the original subjects. However, Chappell presented cross-validation data and the results indicated a high degree of success in correct classification among the three groups. Thus, the estimated parameters were apparently generalizable because the results held up under the cross-validation procedure.

A third study of multiple discriminant function analysis using the SVIB was conducted by Clemens (1969). He attempted to discriminate between 229 practicing engineers and 210 second-year engineering students. He included two SVIB measures of the practicing engineers, the first when they were freshmen in 1935 and the second, 31 years later in 1966. This led to a serious methodological problem due to the non-independence of samples. Because of this problem, the results are difficult to accurately interpret. For example, what would it mean if he f

the

inte

natio

latio

time

into

of th

unawa

porte

did s

seeme

for c

were

analy

into f

functi

This d

classi

conclu

thi
bec
mul
for
sup
func
niqu
wher
(Cle

he found a difference between the practicing engineers on the two measures? Was it due to a change in subject interests, a change in cultural activities, or some combination thereof? A related problem deals with the correlation between the practicing engineers' responses from time one and time two. These correlations were not taken into account and there is no way to determine the effect of this imprecision on the analysis. Clemens was either unaware of the problems or chose to ignore them and reported that in fact the discriminant function technique did separate the three criterion groups. These results seemed to withstand cross-validation although the criteria for correct classification on this aspect of the study were not clearly specified. The author attempted to analyze the data further by breaking the engineers down into fifteen subgroups and then generated the discriminant function using the subgroups as the criterion groups. This discrimination was unsuccessful in accurately classifying individuals in criterion groups. Clemens concluded:

> In summary, based upon the empirical results of this study, meaningful conclusions were made possible because of the statistical discrimination yielded from multiple discriminant function analysis of SVIB data for three overall groups of engineers. These results supported the assertion that multiple discriminant function analysis is an effective statistical technique for discriminating among occupational groups where descriptive-numerical data are available (Clemens, 1969, p. 76).

M

Tiedeman,

discrimin

to milita

continued

of the tw

generaliz

teria and

more subj

case. Th

applicabi

generate

T

function

vard Univ

Vocationa

interest

curriculu

of 32 in

ticipated

function

criterion

which acc

ance. Co

classific

ments wer

Much of the early work conducted by Rulon, Tiedeman, Tatsuoka, and Bryan (1951, 1954) using the discriminant function technique can be directly traced to military personnel classification. Creager (1957) continued in this vein and presented a lucid illustration of the two criterion-two variable case. Creager quickly generalized the arguments to allow for any number of criteria and any number of variables as long as there were more subjects than variables. This was the multivariate case. This discussion demonstrated the feasibility and applicability of the statistical technique being used to generate personnel classification.

Tiedeman and Bryan (1954) applied the discriminant function technique to data obtained from students at Harvard University responding to the Kuder Preference Record, Vocational. Each subject received a score in each of nine interest areas. There were 289 sophomore students in five curriculum areas (the number of students ranged from a low of 32 in biology to a high of 78 in government) who participated in the study. The resulting discriminant function based upon nine predictor variables and five criterion groups yielded two significant latent roots which accounted for approximately 91% of the total variance. Contour scores were generated for each subject for classification purposes and these classification assignments were found to be quite accurate. Since the function

of th
was n

on th
Porte
of hi
impro
occup
group
servec
The si
veteri
the di
social
mechani
weighti
employe
Form D
nature.
only mir
to the n
ated wit
pations n
of Porter
non-unit

of the article was illustrative, cross-validation evidence was not included in the discussion.

Chi-Square Weighting Technique

A multiple weighting scheme for the survey items on the Kuder for selected occupations was generated by Porter (1967) using a Chi-square technique. The purpose of his study was to generate a scoring system which would improve the predictiveness of the Kuder. A group of occupations classified as being similar in nature and a group of occupations classified as dissimilar in nature served as the two major target populations for the study. The similar occupations were optometrists, pediatricians, veterinarians, physical therapists, and X-ray technicians; the dissimilar occupations were clinical psychologists, social case workers, optometrists, foresters, and auto mechanics. Porter provided evidence demonstrating his weighting system superior to the unit weighting system employed by Kuder (Form D) (note the distinction between Form D and Form DD) with occupations that were similar in nature. The results of the cross-validation indicated that only minor shrinkage occurred when the weights were applied to the new sample. Unspecified analytical problems associated with the Kuder scoring system on the dissimilar occupations made comparisons in this area impossible. Results of Porter's study demonstrated the feasibility of using non-unit weighting systems to improve personnel classi

origin

thus r

than w

approa

pation

patter

sion.

frequer

contrib

value.

occupat

given i

100 ite

was ass

sign of

square d

---

by consi
given it
through
each ite
be chose
liked."
cannot be
only two
after the
ble most
the possi
each item
sponse pa
lambda sc

classification. The data analyzed in this study were the original data upon which Kuder based his scoring keys, thus rendering the Kuder weights slightly more predictive than would generally be the case.

The method of obtaining the weights via the Porter approach was accomplished by listing the selected occupations along one margin and listing the six response patterns* available for each item along the second dimension. Porter then obtained the observed and expected frequency counts for each cell and then determined the contribution of each cell toward the total Chi-square value. The resulting cell contribution served as the occupational weight for the response alternative for a given item. This procedure was repeated for each of the 100 items on the inventory. A positive or negative sign was assigned to the weight on the basis of the algebraic sign of the unsquared value of the numerator of the Chi-square definitional formula. Thus a subject's total

---

*The six possible response patterns were derived by considering all possible response alternatives for a given item and then assigning a consistent value of 1 through 6 to each unique alternative. For example, for each item there are three statements; one statement must be chosen as "most liked" and one must be chosen as "least liked." Since the statement that is endorsed as most cannot be the same as the one that is chosen as least, only two alternatives exist for each least selection after the most selection is made. There are three possible most selections for each item. Therefore, there is the possibility of six distinct response patterns for each item. This system exactly coincides with the response patterns presently employed in the Kuder Form DD lambda scoring procedure.

score
patter
patio
occupa
of sco
cedure
in da
the in
consul
for th

Kuder
man and
researc
in term
zations
compone
sisted
versity
they sub
than one
tained e
psycholog
one of th
the occup

score was the sum of all 100 weights of the response patterns that he chose with respect to a specific occupation. A subject was classified by determining the occupational scoring scheme which yielded the largest sum of scores. It was immediately apparent that this procedure was somewhat complex and required several steps in data manipulation. However, as Porter asserted, with the introduction of the computer into educational and consulting situations this was a modest price to pay for the increased accuracy.

Kuder Related Research

Several studies have been completed in which the Kuder has been used as the basis for the research. Tiedeman and Bryan (1954) provided the initial multivariate research on the Kuder. Their study was rather limited in terms of number of subjects and possible generalizations. Shutz and Baker (1962) conducted a principal components analysis of the Kuder. The respondents consisted of 450 freshmen males at a large southwestern university. After they derived the principal components, they submitted all components with eigenvalues greater than one to a varimax rotation. They subsequently obtained eight rotated factors which appeared to make psychological sense. It was interesting to note that one of the factors, health scientist, contained most of the occupations designated by Porter (1967) as being simil

appea

study

midwe

autho

facto

appli

The r

metho

These

posit

keys.

of su

metho

evide

conclu

more

infere

discri

method

Kuder

practi

tive t

of sco

similar. The dissimilar occupations identified by Porter appeared to fall randomly throughout the eight factors.

King and Norrell (1964) generated a factor analytic study on the Kuder using 464 freshmen males from a large midwestern university. It is not clear whether the authors employed a principal components analysis or a factor analysis, but an unnamed rotational procedure was applied to these data resulting in a six-factor solution. The results closely parallel the work of Shutz and Baker.

Kuder (1957) did a comparative study on various methods of developing scoring keys for his instrument. These included configural scoring, unit scoring, answer position scoring, and unit scoring on pattern analysis keys. The study was limited in both the characteristics of subjects (very homogeneous) and the relatively weak methodologies employed. The results did not provide evidence that one method was superior to another. He concluded this study with the assertion that considerably more research was necessary in this area before definitive inferences could be drawn; and to be certain of maximum discrimination, the exploration of the widest variety of methods should be attempted. In a later publication, Kuder (1961) argued that fractional weights were not practical due to the high cost of determining them relative to the limited payoff. Cost was explained in terms of scoring complexity and payoff in terms of increased validity.

limitatic

as compu

ramifica

fication

stated,

by hand

thereby

in scorir

used, why

differenc

of mechan

adopted a

availabil

man and B

weighting

to the un

system is

T

concernin

items. T

weighting

fied in a

on the Ku

pattern a

validity, reliability, or accuracy. The mechanical limitations espoused by Kuder represented a weak argument as computer utilization became more accessible and the ramifications of lack of precision in occupational classification were considered. In addition (Kuder, 1961, p. 4), stated, "the occupational inventory can be scored either by hand or by machine." The reference to machine scoring thereby weakens his claim for the necessity of simplicity in scoring. If a machine scoring procedure is going to be used, why not use the most accurate procedure if the only difference between procedures stems from the complexity of mechanical computations. Kuder (1966), subsequently adopted a non-unit weighting system. Considering the availability of computers, and the works of Porter, Tiedeman and Bryan, and Clark, it is apparent that fractional weighting systems are realistic and feasible alternatives to the unit weighting system. The question is, which system is most accurate?

Summary

Through the years, there has been much controversy concerning the most accurate method of weighting inventory items. The argument persists. Three distinct methods of weighting the items on an interest survey have been identified in addition to the lambda technique currently used on the Kuder. They were: (a) item unit weighting; (b) pattern analytic; and (c) Chi-square multiple weighting.

A discriminant function multiple weighting technique was also identified as a candidate to be used with at least ordinal level data. All the weighting schemes were empirically based scoring systems. The generation of scores from any scoring system that is empirically based will fluctuate when the participant subjects or criterion groups are altered. The question most frequently posed was: can the utilization of a multiple weighting scoring scheme significantly improve the predictability of the Kuder, whether using a linear or non-linear model? Porter demonstrated that predictability could be improved using the Chi-square technique. Kuder and Strong have adopted scoring procedures that function "adequately" even though there may be more efficient methods than those currently practiced. Therefore, is it possible to identify which of the pattern analytic, Chi-square, discriminant function, or lambda predictive systems most accurately classifies individuals with respect to selected occupations?

vestig

scorin

weight

nique,

is def

fied i

valida

the res

H_1:

is descr

CHAPTER III

DATA AND METHOD

Statement of Problem

The research reported here was an empirical investigation of the relative accuracy of a pattern analytic scoring technique, a discriminant function sum of items weighting technique, the Chi-square item weighting technique, and the Kuder lambda weighting technique. Accuracy is defined as the proportion of persons correctly classified into their respective occupations using the cross-validation samples as the criterion groups.

The following hypothesis specifies the nature of the research more precisely:

> H_1: When considering more than two similar or more than two dissimilar occupational groups the discriminant function technique will most accurately classify the subjects into occupations according to the sum of item responses to the Kuder (Form DD), the pattern analytic approach will be next best, the Chi-square approach will be next best, and the Kuder lambda system will be least accurate.

Overview of Procedure

The procedure followed in testing the hypothesis is described under the seven main divisions of the remainde

pational

discrimi

scoring

validati

among va

score th

it was n

discrimi

tem, and

on the f

analysis

correct c

was used

aspects c

Th

responses

from nine

data used

for the n

could not

G. Fredri

the data

the data

remainder of this chapter. These are: (1) Kuder occupational interest data; (2) Kuder scoring system; (3) discriminant function scoring system; (4) pattern analytic scoring system; (5) Chi-square scoring system; (6) cross-validation techniques; and (7) classification comparisons among various systems.

To evaluate the hypothesis it was necessary to score the data using the Kuder lambda weights; likewise, it was necessary to score the data using the multiple discriminant function weights, the pattern analytic system, and the Chi-square weights. Scoring procedures based on the four systems were compared using a repeated measures analysis of variance on the cross-validation data. The correct classification for each cross-validation sample was used as an indicator of accuracy. Each of the seven aspects of the data and method will now be described.

Occupational Data on the Kuder

The data for the present study consisted of the responses on the Kuder of a total of 3906 male subjects from nine different occupations. These were the original data used by Kuder to construct the lambda scoring keys for the nine particular occupations. The present research could not have been conducted without the cooperation of G. Fredric Kuder (originator of the Kuder), who collected the data initially, and Andrew C. Porter, who had obtained the data from Kuder.

sub

fac

ger

het

gro

406

386

het

452

and

appe

occu

and -

purpo

occup

halve

ident

for ea

an occ

patter

informa

weightin

Porter (1967) divided the nine groups into two subgroups of similar and dissimilar occupations. Later factor analytic work gave evidence supporting the homogeneity of the occupations in his similar group and the heterogeneity among the occupations in his dissimilar group.

The homogeneous occupational set was composed of 406 optometrists, 436 pediatricians, 400 veterinarians, 386 physical therapists, and 274 X-ray technicians. The heterogeneous set included 500 clinical psychologists, 452 social case workers, 406 optometrists, 348 foresters, and 298 auto mechanics. Note that the optometrist data appeared in both sets. Thus, there were two sets of five occupations. The similar occupations were labeled Set I and the dissimilar occupations were labeled Set II. For purposes of cross-validation, the data cards for each occupation were divided into random halves; these random halves were labeled A and B respectively for purposes of identification within each occupation. The information for each subject consisted of an identification number, an occupation or criterion group number, and a response pattern number for each of the 100 Kuder items; this information was contained on two Hollerith cards.

Kuder Scoring System

The scoring keys derived by Kuder via the lambda weighting were used to predict occupational classification

on the
through
for use
Kuder w
halves
accurac
was lat
systems

Analysis
that thi
ations w
were dic
of handl
porate mo
dence of
exist. I
use with
his techn

Th
the resea
the numbe
pations w

on the sets of data. These lambda weights were obtained through the cooperation of Science Research Associates for use in this study. Since there was only one set of Kuder weights, these weights were applied to both random halves of the data in a quasi cross-validation. The accuracy of classification with the Kuder scoring system was later compared with that of the three other scoring systems.

Pattern Analytic Scoring System

Clark (1968) developed the Criterion Pattern Analysis. There was empirical evidence which suggested that this approach might be useful in predictive situations when both the criterion and predictor variables were dichotomous. Clark's computer program was capable of handling criterion and predictor variables that incorporate more than dichotomous alternatives. However, evidence of the feasibility of these generalizations did not exist. If Clark's procedures could be generalized for use with the occupational groups of the present study, his technique might be useful in scoring the Kuder.

The computer program for this approach required the researcher to specify the following parameters: (1) the number of criterion groups; (2) the number of occupations within each group; (3) the total number of observations; (4) the discrimination level* for a pattern to be stored; (5) the minimum number of observations necessary for a pattern to be stored; and (6) a significance level for determining additional pattern searches (typically set at .05). A flow diagram, Figure 1 (Clark, 1968, p. 32), is presented to facilitate the understanding of the workings of the program.

The parameters for this study were set by the researcher to allow for as many patterns as possible to be generated and still remain within the capacity of the Control Data Corporation 3600 computer. Since there were five occupational groups within each data set, the criterion group parameter was set at five. The number of observations within each criterion group corresponded with the number of subjects within each occupation (this number varied from a low of 138 to a high of 250 subjects within any given group). The total number of observations was determined by summing the number of observations within each of the occupations for each set of data. The maximum number of observations the computer program could accommodate was 999. For these data, these parameters were set at 965 and 999 for Set I and Set II respectively. The discrimination level, the number of observations necessary for the pattern to be stored, and the significance level

---

*The discrimination level is defined as the ratio of subjects within a criterion group having a specific pattern to the total number of subjects in all criterion groups that have the pattern.

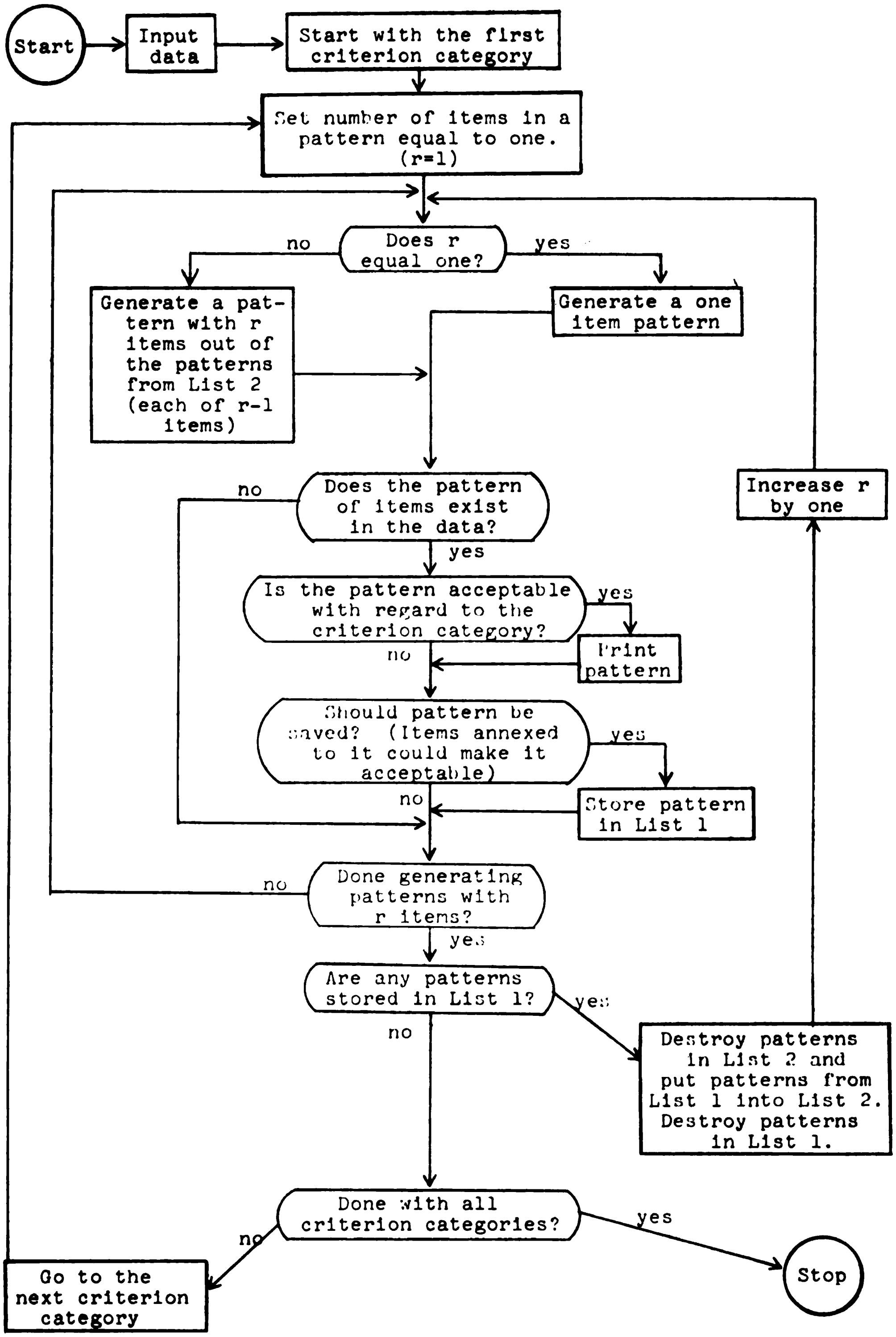

Figure 1

A Flow Chart for Performing Criterion Pattern Analysis

for de

lated

most a

.35, a

manipu

that a

the pa

of cer

contro

and is

good p.

upon t

After

rived

group.

items

said t

of dis

given

all th

pattern

Suppose

groups,

respons

for determining additional pattern searches were manipulated freely by the researcher in an effort to obtain the most accurate predictions; the final resolution was .70, .35, and .05 for the three parameters respectively. This manipulation is discussed in a later section of the thesis.

The program, when executed, generates item patterns that are predictive of criterion groups with respect to the parameters specified by the researcher. The setting of certain of these parameters is completely under the control of the researcher (e.g., discrimination level), and is problem specific. Therefore the ascertainment of good predictive patterns was an interative procedure based upon the subjective manipulation of the program parameters. After these parameters had been specified, the program derived predictive patterns separately for each criterion group. This was accomplished by identifying patterns of items which occurred with some frequency. The pattern is said to be predictive if a certain high proportion (level of discrimination set by the researcher) of subjects in a given criterion group have the pattern with respect to all the subjects in the study. Typical lengths of the patterns ranged from two through five items.

A simplified example will illustrate the procedure. Suppose we just have data from the Kuder on two criterion groups, and as already stated, there are six possible response alternatives for each item. Also suppose the

rese
minin
patt
patt
of th
have
plus
(60%)
or th
items
consi
crite
the a
natio

in the
subjec
prints
ing to
criter
dictive
dictive
group A
this is
paired
indicate

researcher set the discrimination level at .60 and the minimum number of observations necessary to store the pattern was set at 25. This would indicate that before a pattern would be stored in computer memory, at least 60% of the persons in one of the two criterion groups must have the pattern and a minimum of 25 persons (group A plus group B) must have the pattern. Thus 15 persons (60%) in one criterion group must have the given pattern or the pattern and all other possible combinations of the items included in the pattern will be rejected from further consideration. A pattern is defined as predictive of the criterion group when the discrimination index generated on the actual data is equal to or greater than the discrimination index specified by the program parameter card.

The program continues to consider scoring patterns in the data until all possible response patterns for all subjects have been exhausted. At this point, the program prints all acceptable patterns found in the data (according to the original parameter specifications) from each criterion group. Criterion group A might have one predictive pattern identified and group B might have two predictive patterns identified. A possible pattern for group A might consist of the following: 1 (2), 3 (1) .84; this is indicative of item one and response number two paired with item three response number one. The .84 indicates that 84% of the persons who had the pattern were in gro

of per

patter

Interp

analog

are pl

by Cla

persons

generat

terion

of the

that pat

patterns

pattern

criterio

subjects

100 and

gram is

ject one

group A.

Subject

dictive

classifi

subject

cates cl

in group A. The program also indicates the total number of persons who had the pattern. Group B might have the patterns: 1 (1), 2 (1), 3 (2) .71; and 1 (2), 3 (2) .68. Interpretations of the patterns from group B would be analogous to those in group A.

To predict or classify individuals, the patterns are placed directly into the prediction program generated by Clark (1968). This program systematically lists all persons involved in the study from one through N. It generates a symbolic representation of each of the criterion groups beside each subject. If a subject has any of the patterns predictive of any of the criterion groups, that pattern is momentarily stored until all predictive patterns of each criterion group are identified. The pattern with the highest discrimination index for each criterion group is printed out for each subject. Suppose subjects one and two were from group A while subjects 100 and 101 were from group B. The output from the program is illustrated in Figure 2. Note in Figure 2, subject one has only one pattern and it is predictive of group A. This would be a "hit" or correct classification. Subject number two also has one pattern, but it is predictive of group B. This would be a miss or incorrect classification. Subject number 100 is a "hit" while subject number 101 is in a state of quandry. Clark advocates classification to be made in this situation with

Subject
1
2
⋮
100
101

An

respect t

would be

than .71

to the or

patterns

index for

something

classific

manner fo

A

subjects

might le

Subject	Criterion Group	Number of Patterns	Pattern	Discrimination Index
1	A	1	1 (2), 3 (1)	.84
	B	0		
2	A	0		
	B	1	1 (2), 3 (2)	.68
⋮				
100	A	0		
	B	1	1 (2), 3 (2)	.68
101	A	1	1 (2), 3 (1)	.84
	B	2	1 (1), 2 (1), 3 (2)	.71

Figure 2

An Illustration of the Output from the Prediction Program Associated with Computer Pattern Analysis

respect to the highest discrimination index. Subject 101 would be classified into group A because .84 is greater than .71 and therefore a miss would be recorded according to the original procedure. Note that subject 101 has two patterns predictive of group B but the discrimination index for the two patterns is .71 (listed pattern) and something between .60 and .71 (unlisted pattern). This classification procedure was modified in the following manner for the present study.

A large number of patterns was generated for subjects in this study. As a result, any given individual might legitimately have had several patterns. It was possible

that the

"relaxir

Thus, th

from whi

cedures

possible

nation (

patterns

a subjec

A positi

take bot

criminat

weighted

the disc

the numb

values o

indices

deviated

1, 2, and

subject

printing

along wi

criterio

squared

criterio

possible to increase or decrease the number of patterns that the original program generated by systematically "relaxing" or "tightening" the controlling parameters. Thus, there appeared to have been a number of continuums from which possible pattern analytic classification procedures could have been selected. At one extreme, one possible continuum could be the highest level of discrimination (Clark technique) regardless of the number of patterns and at the other extreme, the number of patterns a subject had regardless of the levels of discrimination. A position somewhere between the two extremes which would take both the number of patterns and the level of discrimination into account was adopted. This new procedure weighted each pattern according to the squared value of the discrimination index thus taking into account both the number and the magnitude of the patterns. The squared values of the index was chosen to give more weight to indices close to 1.0 and less value to indices as they deviated from 1.0. In the example of Figure 2, subjects 1, 2, and 100 would not be affected by the change, but subject 101 would be affected. This change required the printing of all predictive patterns for a given subject along with the corresponding discrimination indices. The criterion group which had the highest summation of the squared discrimination indices was designated as the criterion for that subject. Using the new standard, subject 101 [$(.84)^2$ for group A compared with $(.71)^2$ + $(.60)^2$ for group B] would now be classified as a member of group B. As the discrimination index was relaxed, more patterns emerged from the program (other parameters held constant) and thus the distinction between the two methods became more crucial.

A minor aspect of the present study compared the relative accuracy of the proposed pattern analytic classification procedure with Clark's original procedure.

The item scores for this aspect of the study were the same six response patterns for each item as those used by Porter (1967) in the Chi-square study and those currently used in the Kuder lambda technique. The patterns generated by the computer for these subjects and these items were placed directly into the prediction program for purposes of cross-validation.

Discriminant Function Scoring System

It has been successfully demonstrated that the multiple linear discriminant function analysis is extremely proficient in classifying individuals with respect to criterion groups. This differentiation is achieved by a multivariate statistical procedure discussed by T. W. Anderson (1958). Conceptually the process is quite complex as it generates multivariate density functions in a number of dimensions (four for this study). The mathematical manipulations are somewhat less complex and can be

Es

or

cc

cr

su

la

nc

vi

su

se

as

ly

in

Ti

to

a

st

of

the

(19

alt

cri

lam

pred

be mastered through an application of matrix algebra. Essentially, the problem reduced to the determination of one or more vectors of weights which when applied to the corresponding response data maximally separated the five criterion groups with respect to the ratio of among groups sum of squares to the within groups sum of squares. Violation of the assumptions of the statistical model are not crucial in the development of a scoring system, providing the initial results were meaningful and these results hold up under cross-validation. There have been several studies which clearly did not satisfy the assumptions of the multivariate normal populations underlying the sample data, yet the results have been encouraging, e.g., Chappell (1967), CoBabe (1967), Clemens (1969), Tiedeman and Bryan (1954). These findings lend support to the position that positive results may be obtained in a study of this nature even if the assumptions of the statistical model are not met. The level of measurement of the data has also not seemed to have seriously hampered the functioning of the technique in the studies of Chappell (1967), Clemens (1969), and Tiedman and Bryan (1954), although the data of the present study used in the discriminant function are ordinal.

Essentially, the present research used the total lambda score of each subject for each criterion group as predictor scores. Thus there was a criterion variable (group membership) and five predictor variables (one predictor from each group summated lambda score). For example, a forester would have the classification score of the forester group plus a total lambda score for each of the five occupations in Set II. The total lambda score based upon each occupational key would be obtained by scoring the 100 response patterns of the subject with the lambda scoring key for the occupation under consideration. The lambda scores were analyzed and the resulting vectors of weights were used in the prediction model to differentiate among the five occupational groups. The prediction of group membership was determined through the use of Mahalanobis' D^2 statistic. Essentially, this procedure determines the centroid of the multivariate response contour for each of the groups under consideration. Then each subject's score vector is compared with the centroid score of each group and the deviations of the vector scores (squared and summed) around the centroid are used as an indicator of the subject's similarity to the rest of the group. The centroid yielding the lowest D^2 results in the subject being classified into that group.

Because the derivations of the lambda scores were based upon the data used in this study, a true cross-validation procedure was not possible for this aspect of the study. Therefore a quasi cross-validation procedure was employed.* This was accomplished by deriving the weights for the discriminant function on half A of the data and then applying these weights to half B of the data. The effect of this non-independence of data samples probably resulted in a slight "over fitting" of the model, which ultimately translated into increased accuracy in prediction.

Chi-Square Scoring System

Porter (1967) developed an empirical weighting system for the Kuder based upon a Chi-square approach to group differentiation. The response pattern for each item was analyzed by creating a two-dimensional matrix of six response patterns by five occupational groups. The frequency of selection of each response alternative for subjects within each occupational group was determined and placed in the data matrix. Using the marginals and total frequencies, and applying the Chi-square statistic, the expected values for each of the 30 cells in the matrix were determined. The individual cell contributions to the total Chi-square value were subsequently obtained and used as weights in the Porter study. A signed value (+ or -) was given to each weight on the basis of the sign of the unsquared numerator value of the Chi-square

---

*From this point, use of the term cross-validation will refer to cross-validation or quasi cross-validation, depending upon appropriateness.

defini

under

item a

occupa

a give

600 we

the su

have 1

weight

respec

pation

result

the re

sulted

subjec

classi

Chi-sq

from h

from h

cross-v

of othe

predict

to insu

definitional formula $[\Sigma\frac{(O-E)^2}{E}]$ for the particular cell under consideration. This procedure was followed for each item and resulted in a total of 600 weights for each occupation. The scoring for a given occupation for a given subject was based upon 100 of the possible 600 weights according to the response pattern of the subject. For example, subject one would necessarily have 100 response patterns (one for each item). The weights corresponding to these patterns for each of the respective 100 items would be summed (within an occupational key) and this sum would be compared to the sums resulting from application of the other scoring keys to the response patterns. The occupational key which resulted in the highest sum of weights based upon the subject's response patterns was used to designate the classification group for that subject.

Cross-validation procedures were initiated. After Chi-square weights were determined by analyzing the data from half A, these weights were used to score the data from half B. The accuracy of prediction based upon the cross-validated Chi-square scores was compared with that of other scoring systems.

Cross-Validation Technique

Cross-validation is used in studies dealing with prediction or classification of individuals in an attempt to insure the general applicability of the derived scoring syst

the

unde

scor

one-

half

tinu

to d

model

or pr

paren

sampl

Cross

dicti

ships

paren

prese

cross

scori

--dic

the ac

miss c

repeat

four m

discuss

system. As applied in this study, the technique involved the random separation of a large sample of individuals under study into two independent groups. A predictive scoring procedure was then derived (independently) for one-half of the data. The derived scoring weights for half A were then applied to half B to determine the continued accuracy of the scoring procedure. The purpose is to determine the amount of "shrinkage" in the prediction model. Shrinkage is defined as the reduction in accuracy or predictiveness due to chance relationships in the parent sample when comparing the results of the parent sample to the results of the cross-validation sample. Cross-validation procedures are designed to identify prediction equations which are the result of chance relationships or relationships which are idiosyncratic to the parent sample rather than "true" relationships. The present study employed cross-validation and/or quasi cross-validation procedures with each of the proposed scoring systems and then used these results (hit or miss --dichotomous variable) as the criterion for determining the accuracy of the four predictive systems. The hit or miss criterion was used as the dependent variable in a repeated measures analysis of variance based upon the four methods of scoring and five occupational groups.

Herzberg (1969) in a Psychometric Supplement discussed several properties of cross-validation studies and the a

Carlo stu

size tenc

measures

tended t

measures

the squa

interest

cross-va

estimate

cross-va

cross-va

relation

ables.

the pres

number o

sample s

age of t

results

of predi

case the

for an "

cross-va

variable

variable

and the associated statistics and parameters in a Monte Carlo study. Herzberg found that an increase in sample size tended to decrease bias associated with predictive measures, but that an increase in the number of measures tended to increase the bias associated with the predictive measures. This was true for the population parameter ρ^2, the squared population multiple correlation. Of most interest was the finding that the sample and population cross-validities were approximately equal and both underestimated ρ^2. These findings indicated that the sample cross-validation was a good estimate of the population cross-validation, but probably underestimated the true relationship between the predictor and criterion variables. This finding had significant implications for the present study, particularly because of the very large number of subjects in each of the criterion groups. With sample sizes this large, there should be minimal shrinkage of the cross-validation sample data results from the results based upon the parent sample. Also, the number of predictor variables in two cases was 100 and in a third case there were five predictor variables; this may account for an "overfitting" of the model. Greater shrinkage on cross-validation should occur in the methods using 100 variables compared to the method using five predictor variables.

Classification Comparisons Among Various Systems

The comparison of scoring key accuracy was ascertained by placing the "hits" in the cross-validation samples for the four selected techniques (Kuder, Pattern Analysis, Discriminant Function, and Chi-square) in a repeated measures analysis of variance. Recent Monte Carlo studies on the analysis of variance model using nominal level data for the dependent variable have indicated that the parameter estimates and the expected values were remarkably close to the population parameters (Hsu and Feldt, 1969; Lunney, 1970; Draper, 1971). Draper (1971) also determined the expected values for a repeated measures design when the dependent variable was a Bernouli variable. These results indicated that a conservative test of the main effect (Greenhouse and Giesser, 1959) would yield valid results, i.e., the actual and theoretical values for Type I errors are extremely close. The Draper study was also a Monte Carlo procedure.

The independent variables in the design were: (a) occupation (five); (b) repeated measures (four scoring systems); and (c) subjects (nested within occupation). The dependent measure was a hit or miss classification score for each of the four measures. Since this was a repeated measures design, there were proportional subclass frequencies, thus the unequal number of subjects among occupations was not a problem. Two separate analyses were conducted; one analysis was conducted on the similar occupations and one analysis on the dissimilar occupations.

In addition, a comparison was made between Clark's original pattern analytic scoring technique and the modified procedure used in this study. This comparison was based upon the proportion of correct classifications in the cross-validations.

Summary

The major emphases of this research were: (a) to develop scoring systems for selected occupations on the Kuder (one derived by a pattern analytic and the other derived by a discriminant function technique); (b) to demonstrate the feasibility of these procedures; and (c) to compare the accuracy of classification of these new systems to each other and to the Kuder and Chi-square scoring systems. Cross-validation or quasi cross-validation procedures were employed in each of the four scoring systems. The data for the study were the original response patterns upon which Kuder built nine lambda occupational keys.

ar
si
oc
tw
wi
bo
re
Th
is
se
me
Ku
th
an
se
sc
dis

CHAPTER IV

RESULTS AND DISCUSSION

General Overview

The accuracy of the methods of scoring the Kuder are described in detail in this chapter. Data for five similar occupations (Set I) and data for five dissimilar occupations (Set II) were independently analyzed; thus, two distinct analyses were done. The first section deals with the proportions of subjects classified correctly in both analyses. The second section is concerned with the repeated measures analyses for both Set I and Set II. The third section reports the simple effects analyses and is followed by the Scheffé post hoc analyses. The fifth section is a report on the relationships among the scoring methods. The sixth section reports on the results of the Kuder lambda technique. The pattern analytic method and the comparison between Clark's original scoring procedure and the proposed scoring procedure are illustrated in the seventh section. A report on the discriminant function scoring system results follows. Section nine is a brief discussion of the Chi-square scoring method. Section ten present

based u

results

sample

the sco

and dis

weights

occupati

Kuder la

Associat

should b

scoring

obtained

classifi

analysis

there we

found in

were mos

duction

that all

able con

manual

presents a brief comparison among the scoring systems based upon the parent sample results (uncross-validated results).

Proportions of Subjects Correctly Classified

For Set I and Set II data, one-half of the total sample for each occupation (half A) was used to generate the scoring weights for the pattern analysis, Chi-square and discriminant function respectively. These scoring weights were then applied to the remaining data from each occupation (half B) for cross-validation purposes. The Kuder lambda weights were obtained from Science Research Associates and applied directly to half B data. It should be noted that the original data and the Chi-square scoring weights from half A for Set I and Set II were obtained from Porter (1967) and used in these analyses.

A comparison among Porter's percentage of correctly classified males and the proportions obtained in this re-analysis revealed a rather close correspondence. However, there were minor discrepancies, and these data can be found in Appendix B. The reasons for the differences were most likely due to either: (a) a small random reduction in the number of cases within a set of data so that all methods of analysis could be handled by available computer capabilities; (b) mechanical errors in the manual classifying procedure; (c) errors in the recording or key punching of the classification results; or (d) some combination of the above. These discrepancies did not appear to be of sufficient magnitude to cause concern.

Science Research Associates were very cooperative in the research effort. They provided the Kuder lambda weights for eight of the nine occupational groups . . . twice. After the data had been scored and summarized with the original set of lambda weights, there appeared to be a lack of accuracy associated with the scoring system. A check with SRA confirmed the suspicion of the researcher. The initial set of weights were not the correct occupational keys. The situation was remedied and progress on the research continued.

Another interesting facet of the study unfolded when the research firm could not locate the scoring key for the social case worker occupation. The original raw data were given to the researcher along with a formula for the lambda scoring system. From verbal reports at the testing firm, certain occupations just do not lend themselves readily to differentiation from other occupations, and apparently social case workers fall into that category as there appeared to be minimal usage of the scoring key. After applying the scoring formula to the raw data, a set of weights was obtained and used in the scoring procedure. Unfortunately, when 0% of the 226 social case workers were classified as social case workers, the researcher again suspected the scoring key. After obtaining the derivation of the scoring formula from the research firm, two errors were found in the derivation. After the errors were corrected, the formula was applied to the original raw data for the social case workers and a new scoring key was obtained. Using half B of the data, the newly derived lambda weights correctly classified approximately 56% of the social case workers. Also, the discriminant function statistic based on half A data weights classified 76% of the social case workers in half B of the data in the quasi cross-validation.

Table 1 presents proportions of males correctly classified by each of the four methods of scoring for each of the five occupations in the cross-validation using the weights obtained from half A, Set I data to score the data from half B, Set I.

Table 2 presents the proportions of males correctly classified by each of the four methods of scoring for each of the five occupations in the cross-validation using the weights obtained from half A, Set II data to score the data from half B, Set II.

Repeated Measures Analyses

The results of the dichotomous classification cross-validation procedures were submitted to a mixed design analysis of variance with one between and one within subjects variable. In the repeated measures

TABLE 1.--The proportions of males correctly classified on the cross-validation sample for Set I data.

		Pattern Analysis	Chi-square	Kuder Lambda	Discriminant Function	Group Mean
1.	Optometrists	.1232	.5567	.6946	.6847	$\bar{x}_{1.}$ = .5148
2.	X-ray Technicians	.0000	.5725	.6739	.6667	$\bar{x}_{2.}$ = .4783
3.	Pediatricians	.1972	.6835	.6972	.6422	$\bar{x}_{3.}$ = .5550
4.	Physical Therapists	.0146	.4806	.6117	.2087	$\bar{x}_{4.}$ = .3289
5.	Veterinarians	.5450	.9000	.6100	.8200	$\bar{x}_{5.}$ = .7188
	Methods Mean	$\bar{x}_{.1}$=.1865	$\bar{x}_{.2}$=.6425	$\bar{x}_{.3}$=.6570	$\bar{x}_{.4}$=.5990	

$N_1 = 203$; $N_2 = 138$; $N_3 = 218$; $N_4 = 206$; $N_5 = 200$

TABLE 2.--The proportions of males correctly classified on the cross-validation sample for Set II data.

	Pattern Analysis	Chi-square	Kuder Lambda	Discriminant Function	Group Mean
1. Clinical Psychologist	.3880	.8840	.8240	.8360	$\bar{x}_{1.} = .7330$
2. Auto Mechanic	.2267	.9733	.7933	.9067	$\bar{x}_{2.} = .7250$
3. Forester	.0231	.5202	.8092	.8497	$\bar{x}_{3.} = .5560$
4. Social Case Worker	.2788	.6150	.5575	.7611	$\bar{x}_{4.} = .5531$
5. Optometrist	.0150	.4050	.8050	.7550	$\bar{x}_{5.} = .4950$
Methods Mean	$\bar{x}_{.1}=.2012$	$\bar{x}_{.2}=.6777$	$\bar{x}_{.3}=.7528$	$\bar{x}_{.4}=.8158$	

$N_1 = 250$; $N_2 = 150$; $N_3 = 173$; $N_4 = 226$; $N_5 = 200$

design, groups (between subjects) and methods (within subjects) were placed in the same design with subjects nested within groups but crossed with methods. There was an unequal number of subjects within each group, but because of the repeated measures on each subject the sub-class frequencies were proportional. The research of Hsu and Feldt (1969), Lunney (1970), and Draper (1971) indicated that the actual F distribution obtained when using a dichotomous dependent variable closely approximates the theoretical F distribution when the null hypothesis is true and the number of subjects is large. Repeated measures designs using dichotomous criterion scores were found by Draper (1971) to yield interpretable results if conservative tests of the sources of variation are applied. Greenhouse and Geisser (1959) suggested appropriate conservative tests which were applied in the present research (see Table 3).

The model used to analyze the dichotomous data was $Y_{ijk} = \mu + \alpha_j + \eta_{i/j} + \beta_k + (\alpha\beta)_{jk} + (\eta\beta)_{i/jk} + \varepsilon_{ijk}$ where α_j represents the groups effects, β_k represents the methods effects, $\eta_{i/j}$ represents the subjects within groups effects; $(\alpha\beta)_{jk}$ represents the interaction of groups by methods effects; and, $(\eta\beta)_{i/jk}$ represents the methods by subjects nested within groups effects. The groups and methods variables were considered fixed factors and subjects were designated as random factors.

TABLE 3.--Repeated measures analyses of variance on the dichotomous classification data for Set I and Set II.

Source	df	MS	F
Set I Analysis			
Groups	4	15.93	46.57
Subjects/Groups	960	.34	
Methods	3	48.64	363.85
Groups x Methods	12	3.34	25.03
Methods x Subjects/Groups	2880	.13	
Set II Analysis			
Groups	4	9.75	41.29
Subjects/Groups	994	.23	
Methods	3	78.06	582.71
Groups x Methods	12	3.47	25.87
Methods x Subjects/Groups	2982	.13	

The procedures appropriate for the classical one-way analysis of variance model were used to test for the groups source of variation; and, the results were found to be statistically significant with F values of 46.57 and 41.29 for Set I and II respectively ($.999F_{4,960} = 4.67$). The null hypothesis of equality of groups with respect to the proportion of correct classification within each group was rejected at the given level of confidence.

The methods source of variation was tested with the resulting F ratios of 363.85 and 582.71 for data Sets I and II respectively. According to Greenhouse and Geisser (1957) and Draper (1971), the methods source can be tested when the data are dichotomous and there is an arbitrary variance covariance matrix. This can be accomplished by a reduction in the degrees of freedom associated with the mean squares used to generate the appropriate F ratio. For Set I data, the degrees of freedom used to test the F ratio associated with the methods source were reduced from 3 and 2880 to 1 and 960 respectively. For Set II data, the degrees of freedom were reduced from 3 and 2982 down to 1 and 994 respectively for the numerator and denominator of the appropriate F ratio. The results were consistent when either the liberal or conservative tests were applied to the data ($.999F_{1,960} = 10.88$). The null hypothesis of the equality of methods, with regard to the proportion of persons correctly classified by the various procedures, was rejected.

The test of the groups by methods interaction, a test of group profiles having the same shape, was generated. F ratio values of 25.00 and 25.87 were obtained from Set I and Set II data; the liberal and conservative tests were applied to this interaction and both results were consistent ($.999F_{4,960} = 4.67$). The degrees of freedom for the conservative test were reduced from 12 and 2880 for Set I and 12 and 2982 for Set II down to 4 and 960 for Set I and 4 and 994 for Set II. The significant groups by methods interaction presented the possibility of potentially misleading results when the main effects of this design were considered. The groups by methods interaction was graphed and is illustrated in Figures 3, 4, 5, and 6. It was apparent that one method functioned very poorly, but none of the other methods were consistent in most accurately classifying subjects within each set of data. The three remaining methods were substantially better than the one weak method.

Testing for Simple Effects

A simple effect is defined as the comparison of the levels of one independent variable while considering only a single level of another independent variable. According to Kirk (1968), when an interaction is significant, there is little interest in testing the main effects of the model because one independent variable

M_1 = Pattern Analysis
M_2 = Chi-square
M_3 = Kuder Lambda
M_4 = Discriminant Function

G_1 = Optometrists
G_2 = X-ray Technicians
G_3 = Pediatricians
G_4 = Physical Therapists
G_5 = Veterinarians

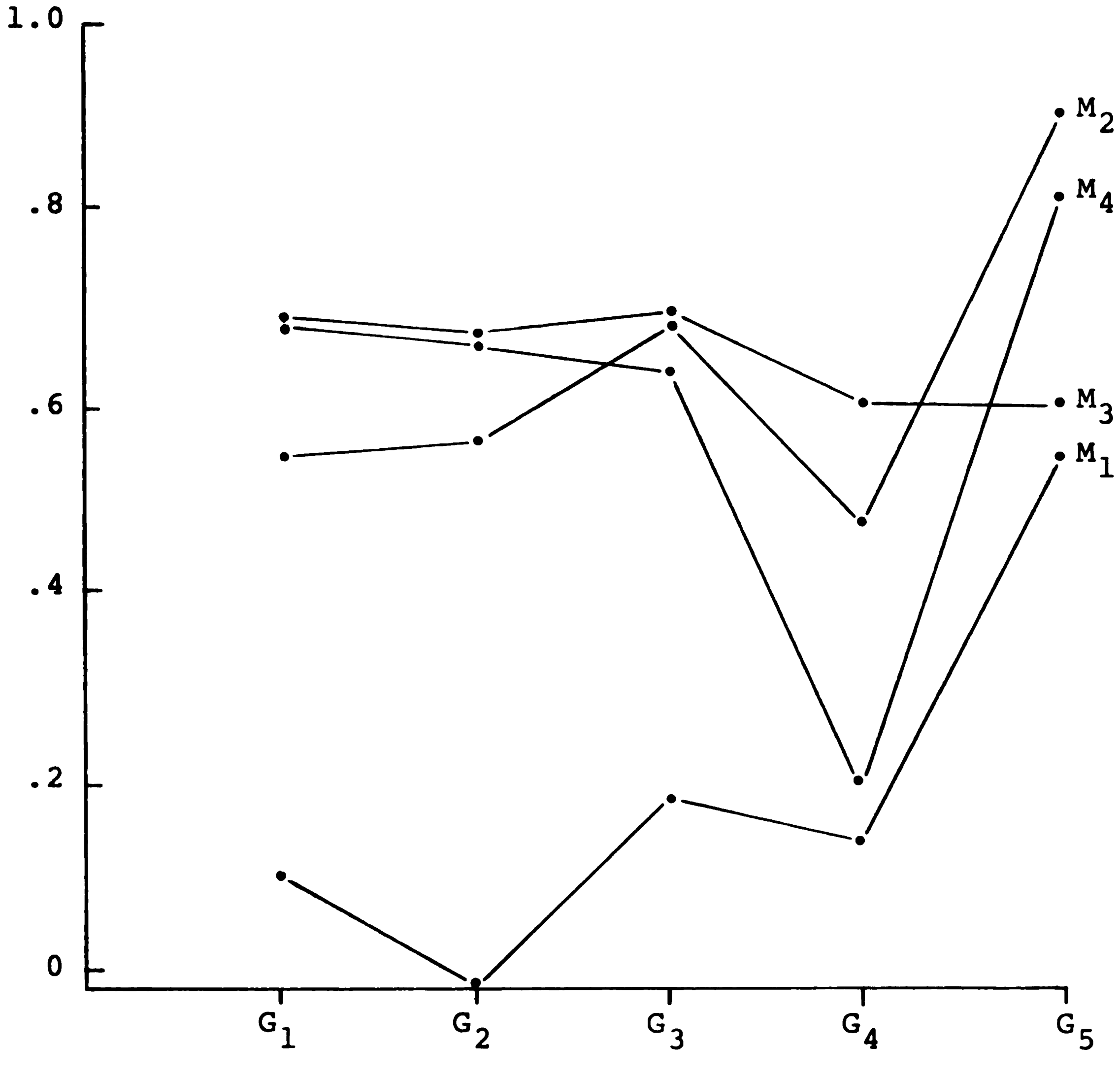

Figure 3

The Graphic Relationship Among Groups, Methods, and the Proportion of Correctly Classified Individuals for Set I Data; Groups Variable Placed on the Abscissa

G_1 = Optometrists

G_2 = X-ray Technicians

G_3 = Pediatricians

G_4 = Physical Therapists

G_5 = Veterinarians

M_1 = Pattern Analysis

M_2 = Chi-square

M_3 = Kuder Lambda

M_4 = Discriminant Function

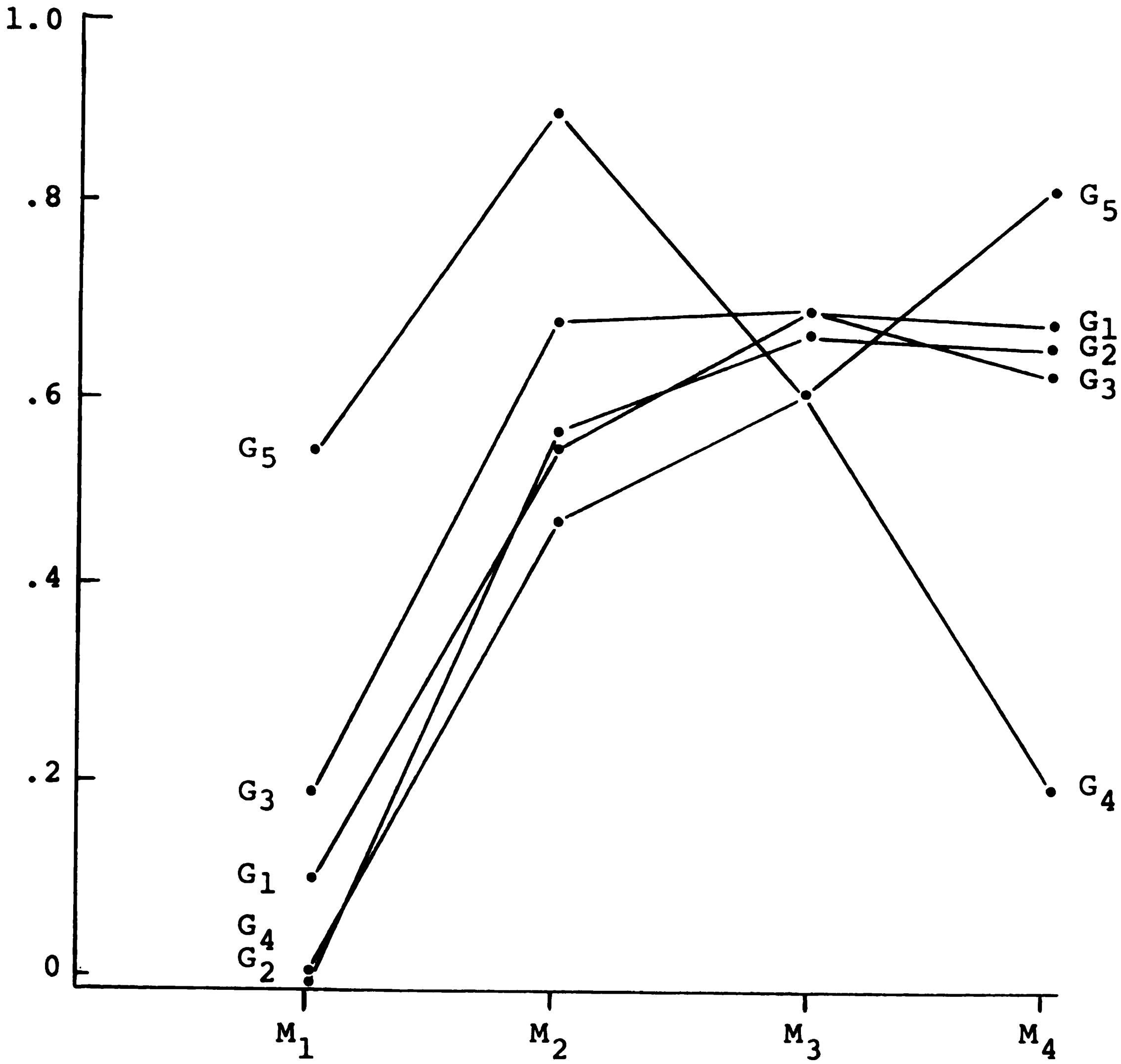

Figure 4

The Graphic Relationship Among Groups, Methods, and the Proportion of Correctly Classified Individuals for Set I Data; Methods Variable Placed on the Abscissa

M_1 = Pattern Analysis
M_2 = Chi-square
M_3 = Kuder Lambda
M_4 = Discriminant Function

G_1 = Clinical Psychologists
G_2 = Auto Mechanics
G_3 = Foresters
G_4 = Social Case Workers
G_5 = Optometrists

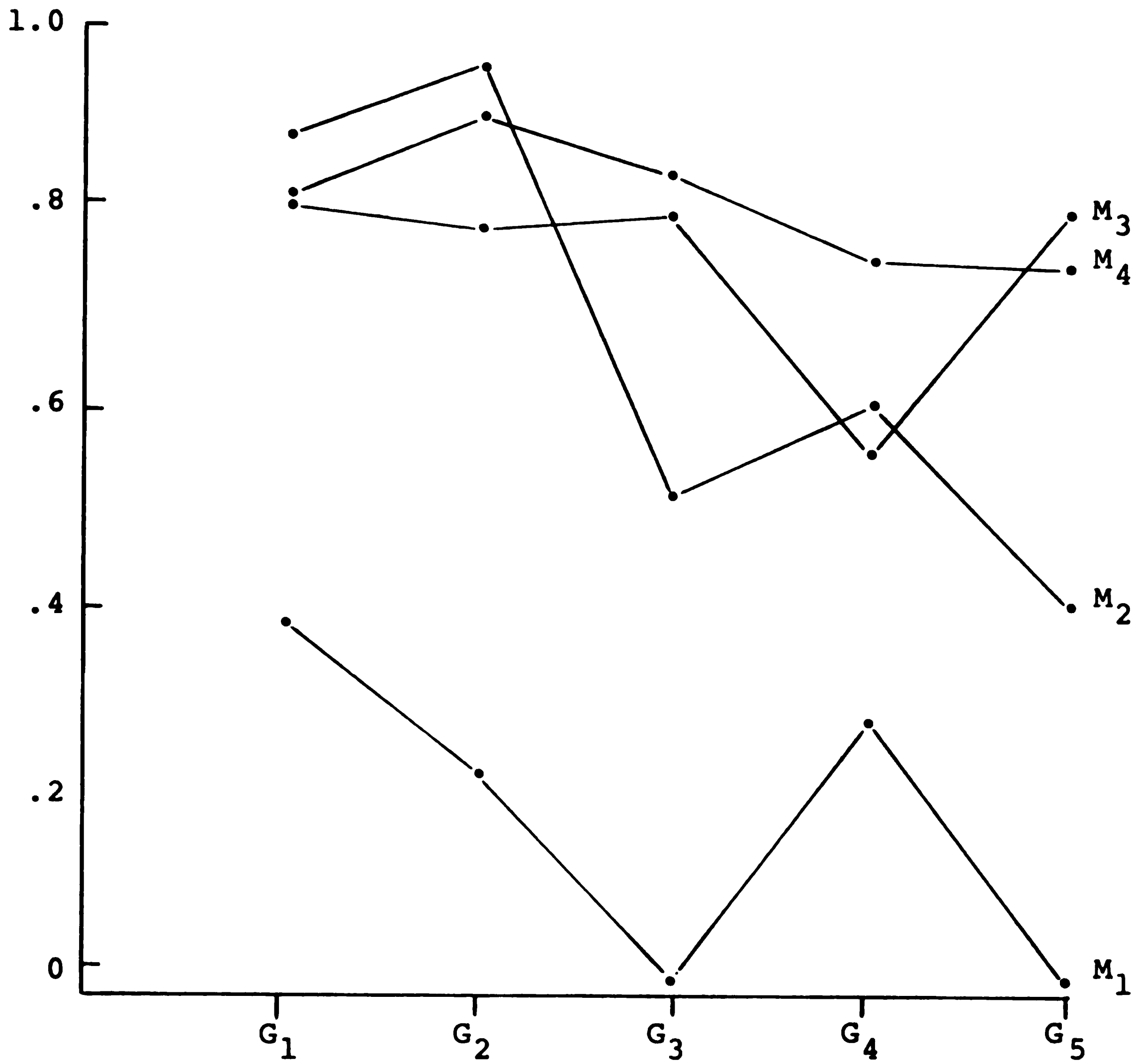

Figure 5

The Graphic Relationship Among Groups, Methods, and the Proportion of Correctly Classified Individuals for Set II Data; Groups Variable Placed on the Abscissa

G_1 = Clinical Psychologists
G_2 = Auto Mechanics
G_3 = Foresters
G_4 = Social Case Workers
G_5 = Optometrists

M_1 = Pattern Analysis
M_2 = Chi-square
M_3 = Kuder Lambda
M_4 = Discriminant Function

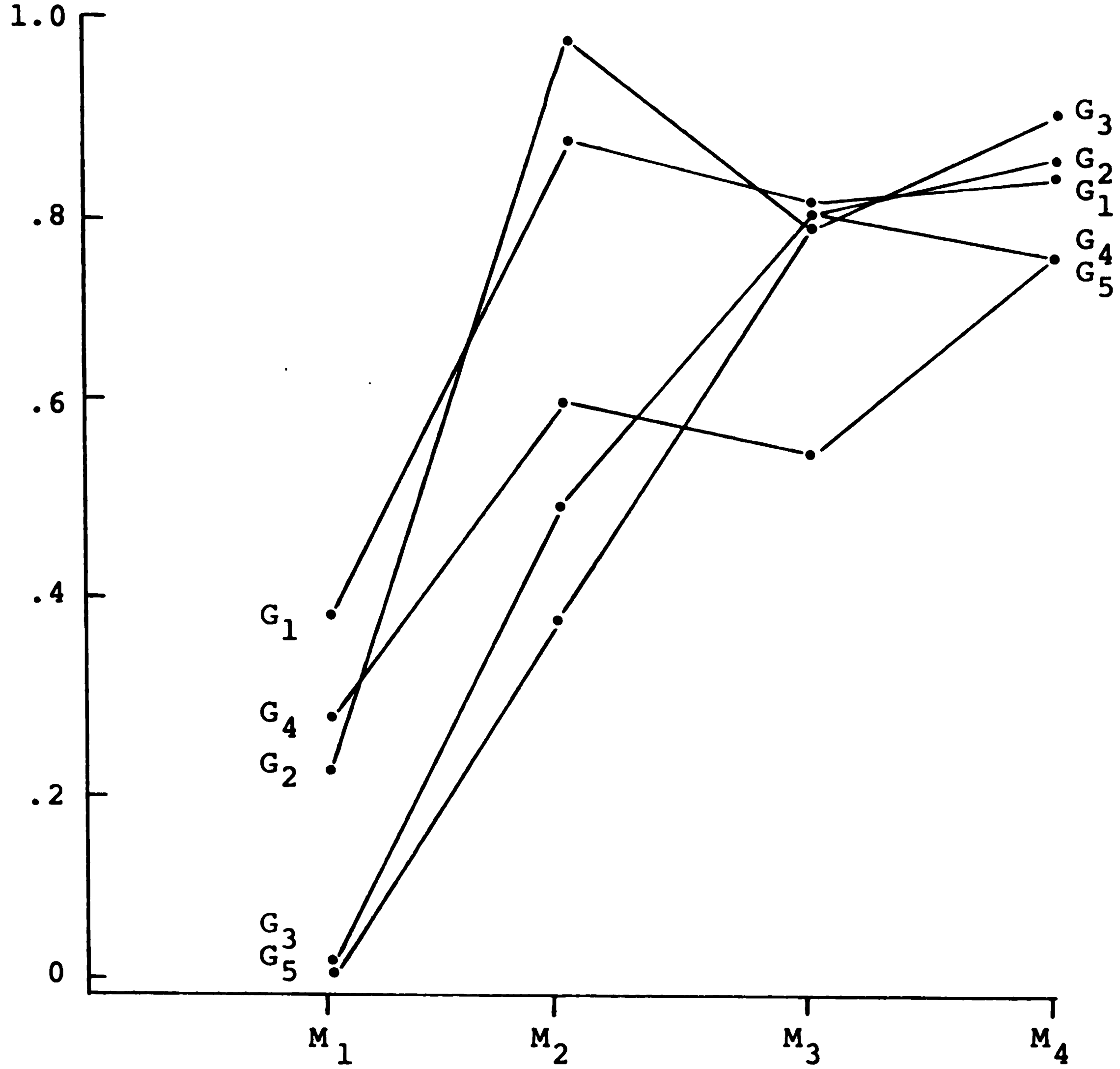

Figure 6

The Graphic Relationship Among Groups, Methods, and the Proportion of Correctly Classified Individuals for Set II Data; Methods Variable Placed on the Abscissa behaves differentially within the different levels of the second independent variable. For this study, the simple effects were tested by four one-way analyses of variance across the five groups of subjects (one analysis on each scoring method) and, five simple repeated measures analyses (one analysis on each group with four repeated measures). After the one-way and simple repeated measures analyses were generated, the error variances among groups and methods were checked for homogeneity. The variance estimates for each of the one-way analysis of variance problems within a data set were found to be very close (maximum range of .10 for Set I and maximum range of .04 for Set II data). The same results held for both data sets on the simple repeated measures analyses (maximum range of .05 for Set I and maximum range of .10 for Set II). Therefore, the respective variances were pooled according to the method presented by Kirk (1968) (see Tables 4 and 5). The results of the analyses using the pooled variance estimates were consistent with the results of the unpooled separate analyses.

The pooled variance component (error term) for the repeated measures simple effects (between methods at G_i) and for the groups x methods interaction was the same component that was used in the repeated measures analyses, i.e., the methods x subjects within groups error term (see Tables 3, 4, and 5). However, the pooled variance

TABLE 4.--Analysis of variance: Tests of simple effects on the classification scores of Set I data using pooled estimates of the variance.

Source	SS	df	MS	F
Between subjects				
Groups at M_1	37.4340	4	9.3585	50.3686
Groups at M_2	21.2012	4	5.3003	28.5269
Groups at M_3	1.5448	4	.3862	2.0818
Groups at M_4	43.6736	4	10.9184	58.7642
Within cell (pooled)	713.4720	3840	.1858	
Within subjects				
Between methods				
at G_1	43.9164	3	14.6388	109.4899
at G_2	42.9711	3	14.3237	107.1331
at G_3	37.5687	3	12.5229	93.6641
at G_4	44.5377	3	14.8459	111.0388
at G_5	17.0238	3	5.6746	42.4427
Groups x methods	40.1064	12	3.3422	24.997
Methods x subjects:				
groups	385.0560	2880	.1337	
Total	963.2064	3859		

$_{.99}F_{4,\ 3840} = 3.32$

$_{.99}F_{3,\ 2880} = 3.78$

$_{.99}F_{12,\ 2880} = 2.18$

where M_1 = Pattern Analysis; M_2 = Chi-square; M_3 = Kuder lambda; M_4 = Discriminant Function; G_1 = Optometrists; G_2 = X-ray Technicians; G_3 = Pediatricians; G_4 = Physical Therapists; G_5 = Veterinarians

TABLE 5.--Analysis of variance: Tests of simple effects on the classification scores of Set II data using pooled estimates of the variance.

Source	SS	df	MS	F
Between subjects				
Groups at M_1	22.6008	4	5.6502	35.4022
Groups at M_2	43.9284	4	10.9821	68.8101
Groups at M_3	10.8412	4	2.7103	16.9818
Groups at M_4	2.9560	4	.7390	4.6303
Within cell (pooled)	634.5696	3976	.1596	
Within subjects				
Between methods				
at G_1	40.1790	3	13.3930	99.9477
at G_2	52.1517	3	17.3839	129.7305
at G_3	75.3453	3	25.1151	187.4261
at G_4	27.5487	3	9.1829	68.5291
at G_5	80.4399	3	26.8133	200.0092
Groups x methods	41.5836	12	3.4653	25.8604
Methods x subjects:				
groups	399.5880	2982	.1340	
Total	948.8125	3995		

$_{.99}F_{4,\ 3976} = 3.32$

$_{.99}F_{3,2982} = 3.78$

$_{.99}F_{12,\ 2982} = 2.18$

where M_1 = Pattern Analysis; M_2 = Chi-square; M_3 = Kuder lambda; M_4 = Discriminant Function; G_1 = Clinical Psychologists; G_2 = Auto Mechanics; G_3 = Foresters; G_4 = Social Case Workers; G_5 = Optometrists component used to test the between subjects simple effects was obtained by combining the pooled estimate of the variances of the one-way analyses with the methods by subjects within groups variance.

When different error terms are used to test the main effect and the interaction respectively, the appropriate error term for the simple effect is the pooled variance estimates of the two components (Kirk, 1968). The pooling of the variance for the simple effects of groups is done because a simple effect not only contains the estimate of the effect, but also contains the particular cell interaction effect.

After the appropriate error terms were generated for this study, they were used to construct the F ratios for each of the ten sources within each set of data. The level of significance was partitioned into equal pieces ($\alpha' = .01$) so that the overall level of significance for a main effect was $\leq .05$. Nineteen of twenty of these sources were found to be statistically significant at $p < .01$. The only non-significant ratio was the simple effect of Set I groups for the Kuder lambda method.

Scheffé Post Hoc Analyses

Post hoc procedures were initiated to determine which pair-wise contrasts of means were significantly different (see Tables 6, 7, and 8). The Scheffé post hoc

TABLE 6.--Scheffé post hoc analyses of Set I groups data at each level of the methods variable.

Mean Differences	Pattern Analysis M_1	Chi-square M_2	Discriminant Function M_4	Scheffé Critical Value
$G_1 - G_2$	.1232	-.0158	.0180	.173
$G_1 - G_3$	-.0540	-.1268	.0425	.152
$G_1 - G_4$	.1086	.0761	.4760*	.155
$G_1 - G_5$	-.4218*	-.3433*	-.1353	.156
$G_2 - G_3$	-.1972*	-.1110	.0245	.169
$G_2 - G_4$	-.0145	.0919	.4580*	.172
$G_2 - G_5$	-.5450*	-.3275*	-.1533	.174
$G_3 - G_4$	.1826*	.2029*	.4335*	.151
$G_3 - G_5$	-.3478*	-.2165*	-.1778*	.153
$G_4 - G_5$	-.5304*	-.4194*	-.6113*	.155

α = .01 level of significance * = Significant difference

G_1 = Optometrists; G_2 = X-ray Technicians; G_3 = Pediatricians; G_4 = Physical Therapists; G_5 = Veterinarians

TABLE 7.--Scheffé post hoc analyses of Set II groups data at each level of the methods variable.

Mean Differences	Pattern Analysis M_1	Chi-square M_2	Kuder Lambda M_3	Discriminant Function M_4	Scheffé Critical Value
$G_1 - G_2$	.1613	-.0893	.0307	-.0707	.162
$G_1 - G_3$	.3649*	.3638*	.0148	-.0137	.154
$G_1 - G_4$	.1092	.2690*	.2665*	.0749	.144
$G_1 - G_5$	.3730*	.4790*	.0190	.0860	.149
$G_2 - G_3$	.2036*	.4483*	-.0159	.0570	.174
$G_2 - G_4$	-.0521	.3583*	.2358*	.1456	.165
$G_2 - G_5$	.2117*	.5683*	-.0217	.1517	.169
$G_3 - G_4$	-.2557*	-.0948	.2517*	.0886	.158
$G_3 - G_5$	.0081	.1152	.0042	.0947	.163
$G_4 - G_5$	.2638*	.2100*	-.2475*	.0061	.152

α = .01 level of significance   * = Significant difference

G_1 = Clinical Psychologists; G_2 = Auto Mechanics; G_3 = Foresters; G_4 = Social Case Workers; G_5 = Optometrists

TABLE 8.--Scheffé post hoc analyses of Set I and Set II data at each level of the groups variable.

Methods at each level of G_i

Mean Differences	G_1	G_2	G_3	G_4	G_5
			Set I Data		
M_1-M_2	-.5714*	-.5725*	-.4863*	-.4660*	-.3550*
M_1-M_3	-.4335*	-.6739*	-.5000*	-.5971*	-.0650NS
M_1-M_4	-.5615*	-.6667*	-.4450*	-.1941*	-.2750*
M_2-M_3	-.5615*	-.1014NS	-.0137NS	-.1311NS	.2900*
M_2-M_4	-.1280NS	-.0942NS	.0413NS	.2719*	.0800NS
M_3-M_4	.0099NS	.0070NS	.0050NS	.4030*	.2100*
Scheffé Critical Value	.135	.163	.129	.133	.136
			Set II Data		
M_1-M_2	-.4960*	-.7466*	-.4971*	-.3362*	-.3900*
M_1-M_3	-.4360*	-.5666*	-.7861*	-.2787*	-.7900*
M_1-M_4	-.4480*	-.5891*	-.8266*	-.4823*	-.7400*
M_2-M_3	.0600NS	.1800*	-.2890*	.0575NS	-.4000*
M_2-M_4	.0480NS	.0666NS	-.3295*	-.1461*	-.3500*
M_3-M_4	-.0120NS	-.1134NS	-.0405NS	-.2036*	-.0500NS
Scheffé Critical Value	.122	.150	.145	.128	.136

* = Significant at α=.01 NS = Non-significant

M_1 = Pattern Analysis; M_2 = Chi-Square; M_3 = Kuder Lambda; M_4 = Discriminant Function; For Set I: G_1 = Optometrist; G_2 = X-ray Technician; G_3 = Pediatrician; G_4 = Phys. Ther.; G_5 = Veterinarians; For Set II: G_1 = Clin. Psych.; G_2 = Auto Mech.; G_3 = Foresters; G_4 = Soc. Case Wkr.; G_5 = Optometrists.

procedure was used in this study because it was least sensitive to violations of assumptions; and, there were unequal cell sizes at each level of the groups variable.

Table 6 illustrates the pair-wise contrasts of group means for each significant analysis for each level of the methods variable for Set I data. Table 7 contains similar information for Set II; both tables contain the critical values for the Scheffé test. An interval was created by the addition and subtraction of the critical value to the difference between means. If the interval crossed zero, the contrast was non-significant. If the interval did not cross zero, the contrast was significant. These analyses were generated using $\alpha = .01$. The Scheffé critical values ranged from a high of .174 to a low of .152 for Set I data, and from .174 down to .144 for Set II data. Unless the differences between means exceeded these values, the differences were non-significant.

Optometrists (Set I) were significantly less accurately classified for the pattern analytic and Chi-square methods than the veterinarians, but more accurately classified than the physical therapists on the discriminant function method. The X-ray technicians were more accurately classified than the physical therapists using the discriminant function method, but less accurately classified than the veterinarians using the pattern analytic and Chi-square techniques; and, they were less accurately classified than the pediatricians using the pattern analytic method. The pediatricians were classified significantly more accurately than the physical therapists, but significantly less accurately than the veterinarians on all techniques except the Kuder lambda. The veterinarians were classified significantly more accurately than the physical therapists on all but the Kuder lambda technique. All other pair-wise contrasts on Set I data at each level of methods variable were non-significant.

Since the simple effects analysis for groups on Set I data at the Kuder lambda level was non-significant, this post hoc analysis was not generated. All pair-wise contrasts would have yielded non-significant results. An interesting phenomenon resulted when the simple effects analysis of Set II data yielded a significant F ratio for the groups at the discriminant function level and the Scheffé post hoc analysis failed to detect a significant pair-wise contrast. Two possible reasons for this were: (a) with the pooling of the variance estimates, any among groups heterogeneity of variance could have obscured the difference; and (b) the conservative nature of the Scheffé test and its lack of sensitivity to detect relatively small differences. Kirk (1968) argued that with sample sizes over 30 there is minimal bias associated with the pooling procedure. And, with a less conservative test, or upon using a larger level of significance, significant contrasts could be obtained.

In the Set II data, the clinical psychologists were classified significantly more accurately than the foresters and optometrists when the pattern analytic method was used; and, they were classified significantly more accurately than the foresters, social case workers, and optometrists when the Chi-square method was used. The clinical psychologists were classified significantly more accurately than the social case workers when the Kuder lambda technique was used.

The auto mechanics were classified significantly more accurately than the foresters and optometrists when the pattern analytic method was used; and, they were classified more accurately than foresters, social case workers, and optometrists when the Chi-square method was used. When the Kuder lambda method was used, the auto mechanics were classified more accurately than the social case workers.

With the pattern analytic method, the social case workers were more accurately classified than the foresters and the optometrists. The social case workers were also more accurately classified than optometrists when the Chi-square method was used. The results were reversed for the foresters and social case workers on the Kuder lambda method from that of the pattern analytic method, i.e., foresters were more accurately classified than social case workers. All other pair-wise contrasts on the Set II data at each level of the methods variable were non-significant.

Comparisons of Methods Means

It was evident that method one, the pattern analytic, was significantly inferior to the other three methods of scoring across all levels of groups in both sets of data, except the Kuder lambda procedure on the veterinarian group (see Table 8). The pattern analytic method simply failed to function.

No single method was always superior to the other methods. In Set I, the Chi-square method was significantly worse than the Kuder lambda method on the optometrist group. The discriminant function analysis was significantly better than the Kuder lambda on the veterinarian group, but significantly worse than the Kuder lambda on the physical therapist group. All other pair-wise mean differences within Set I at each level of the groups variable were non-significant. The Scheffé critical values ranged from a low of .129 to a high of .163.

On Set II data, the Chi-square was significantly better than the Kuder lambda on two groups and significantly worse on two groups; there was no difference in the fifth group. The discriminant function was significantly better than the Chi-square for the forester, social case worker, and optometrist (Set II) groups. The discriminant function was also significantly better than the Kuder lambda for the social case worker group. All other pair-wise mean differences within Set II at each level of the groups variable were non-significant. The Scheffé critical values ranged from a low of .122 to a high of .150.

Excluding the contrasts involving the pattern analytic method, the discriminant function had five significant contrasts, the Kuder lambda had four and the Chi-square had three significant contrasts. Overall, the discriminant function method classified subjects more accurately than the two remaining methods on the Set II data. When the occupations under consideration were dissimilar in nature, the discriminant function method classified individuals into their appropriate groups better than the Kuder lambda and Chi-square methods. For Set I data (occupations similar in nature), there was no one method which demonstrated consistent superiority over the other methods.

Ignoring the pattern analytic technique, it appeared that the determination of the best method for classifying individuals appropriately for both Set I and Set II data was dependent upon both the method of scoring and the group under consideration. For example, the

Chi-square weights were most accurate in predicting the cross-validation classification of auto mechanics (97%), but could accurately classify only 40% of the optometrists from Set II. Whereas the discriminant function weights were most accurate for classifying social case workers and foresters; and, the Kuder lambda weights were most accurate for classifying the optometrists in both sets of data.

Given the significant interaction, the major hypothesis of this study was difficult to evaluate. Clearly it was apparent that there was no one best method to use to score the data from all groups to consistently obtain the greatest accuracy. It was obvious that one method was inappropriate. The pattern analytic method failed to function and was eliminated from consideration as a method of weighting the Kuder occupational interest survey. This finding was in opposition to the research hypothesis which predicted that the pattern analytic method would be the second most accurate method of scoring. The interaction of groups with scoring methods was not anticipated and also negated the research hypothesis. The Chi-square method proved to classify individuals better than anticipated as did the Kuder lambda method which was predicted to be the least accurate method but was the most accurate for five of ten groups. The Chi-square method was most accurate on three groups and the discriminant function classified most accurately on the remaining two groups. Again, the resulting scores for comparisons were not all based upon true cross-validation thus putting the Chi-square method at a slight disadvantage.

Relationships Among Scoring Methods

The variation and covariation of the dichotomous scores of correct or incorrect classifications defined by each of the four methods of scoring were considered. Phi coefficients for all pairs of variables are presented in Table 9. The pattern analytic method correlated minimally with the other three methods. The remaining three methods moderately correlated with each other (.45, .48, and .54 for Set I; and .31, .24, and .39 for Set II), based upon 965 and 999 observations respectively. With the exception of the pattern analytic method, there was sufficient evidence to suggest that the methods of scoring had considerable overlap in the subjects who were accurately and inaccurately classified. Apparently (for the three functional classification methods) certain subjects lend themselves to classification (or non-classification) consistently across measures as noted by the moderate correlations among the measures.

Appendix B contains the variance-covariance matrices of data Sets I and II based on the four methods of scoring. With the exception of the pattern analytic

TABLE 9.--The Φ coefficient: The relationships among the four repeated measures.

Scoring Method	M_1	M_2	M_3	M_4
		Set I B, N = 965		
M_1	1.000	.0963	.0042	.1530
M_2		1.000	.4494	.4837
M_3			1.000	.5401
M_4				1.000
		Set II B, N = 999		
M_1	1.000	.1378	-.0596	.0259
M_2		1.000	.3147	.2359
M_3			1.000	.3861
M_4				1.000

M_1 = Pattern Analysis; M_2 = Chi-square; M_3 = Kuder Lambda; M_4 = Discriminant Function method, both matrices appeared reasonably close to the second data model discussed by Greenhouse and Geisser (1959), i.e., equal variance elements and covariance elements in the population parameters. Perhaps the conservative tests were not necessary for the repeated measures analyses. However, this is a moot point since the same results were achieved via both the liberal and conservative tests.

Kuder Lambda Scoring Results

Scoring based on the Kuder lambda weights predicted the classification of the subjects on the cross-validation procedure slightly better than Chi-square and the discriminant function weights for Set I data. The discriminant function predicted most accurately for Set II data. The Kuder lambda, the Chi-square and the pattern analytic scoring procedures followed in that order. The Kuder lambda proportions of correct classification for Set I data on cross-validation ranged from .61 to .69, while the range on Set II data was from .56 to .82. The Kuder lambda method was at least competitive with both the Chi-square and discriminant function methods for these data.

Pattern Analytic Scoring Results

The results of the pattern analytic scoring technique were extremely disappointing (see Table 10).

TABLE 10.--Comparison of the proposed and original pattern analytic scoring procedures for Set I and Set II data.

Occupation	N Correctly Classified		N of Generated Patterns Used for Scoring	N Expected Correct by Chance	N
	Modified	Original			
Set I--Similar Occupations					
1. Optometrists	25	15	97	40.6	203
2. X-ray Technicians	0	0	10	27.6	138
3. Pediatricians	44	6	165	43.6	218
4. Physical Therapists	3	2	40	41.2	206
5. Veterinarians	108	169	345	40.0	200
Total	180	192	757	193.0	965
Set II--Dissimilar Occupations					
1. Clinical Psychologists	96	81	329	50	250
2. Auto Mechanics	34	32	177	30	150
3. Foresters	4	4	35	36.8	173
4. Social Case Workers	63	73	144	45.2	226
5. Optometrists	3	3	18	40	200
Total	200	193	703	199.8	999

An enormous amount of computer time was expended. A reasonable and meaningful number of predictor patterns were derived for each set of data. Through a manual iterative process, it was determined that the discrimination index must be set at .70 or higher, the level of significance set at .05 or lower, and the minimum number of observations set at 25 or higher in order to generate a solution to the problem. If any of the three basic limits was exceeded, the program would exceed the capacity of the CDC 3600 computer and subsequently abort the run, but only after consuming approximately one hour of central processing time. Even within the limitations of the program and the computer, "successful" runs consumed up to 90 minutes of central processing time. In all fairness to the pattern analytic technique, it should be noted that the research problems presented a severe test for the program. The program is capable of handling up to 999 observations with up to ten groups of subjects and a potential of over 100 variables with up to 10 responses per variable. The problems in this research involved up to 999 observations, 100 variables, five groups and six responses per variable for each set of data.

There were problems with the determination of the predictor patterns. If the program controlling parameters were "relaxed," e.g., minimum number of observations necessary for a pattern to be stored was dropped to 10, the program typically exceeded the computer capacity. But, if the parameters were "tightened," e.g., minimum number of observations necessary for a pattern to be stored was increased to 75, the program would not exceed the computer capacity, but a very small number of predictor patterns would be generated. Following many iterations, an acceptable number of patterns was obtained by setting the discrimination index at .75, the level of significance at .05 and the minimum number of observations necessary for the pattern to be stored at 35. This was true for both sets of data. Other combinations of these parameters failed to yield acceptable results. Typical occurrences for other combinations were: (a) zero patterns for less homogeneous groups, i.e., limited within group similarity; (b) the possible number of patterns for a particular group exceeded computer capabilities; (c) the job card time limit was exceeded; (d) minimal patterns were generated for all groups; and (e) few patterns were generated for the less homogeneous groups while the more homogeneous groups yielded an overflow of patterns.

The information given in Table 10 indicates that the number of patterns generated for specific groups varied from a low of 10 to a high of 345. Efforts to increase the minimum and decrease the maximum number of patterns within a data set were unsuccessful. The range of the number of item patterns was almost directly proportional to the number of persons correctly classified in the occupation, i.e., veterinarians had 345 patterns and more than 100 persons correctly classified on the cross-validation procedure, whereas X-ray technicians had only 10 generated patterns and zero persons were correctly classified. Less than 5% of the total sample received no classification prediction on the cross-validation procedure.

If, given no prior information, one considers chance to be the probability of being correctly classified (one out of five for this study), correct classification exceeded the chance level in only five of the 10 groups. Correct classification exceeded the .50 level for but one of the five groups using the modified scoring procedure. Similar results were obtained when the original pattern analytic scoring procedure was employed.

At this point, the researcher suspected a data error. After the Chi-square and lambda scoring procedures were completed, a reanalysis of the data was conducted using the pattern analytic method. The prediction results were unchanged. The lack of precision can probably be attributed to limitations imposed on the program and the computer by the massive amount of data and the restrictions on the program parameters due to the "taxing" of the model. A third plausible explanation is that the model will provide accurate classification for nominal level data only if the data are dichotomous (both for the predictor and the criterion).

It is hereby recommended that the pattern analytic technique not be included as a research tool for purposes of occupational classification with respect to the Kuder. When the predictor and criterion variable are on a nominal scale of measurement and the data are not dichotomous, the pattern analytic technique should not be used. Additional research may demonstrate the utility of the technique under modified conditions, of course.

Comparison of the Original and Modified Pattern Analytic Scoring Procedures

The comparison of the pattern analytic scoring system proposed by Clark with the modified version of the pattern analytic scoring procedure was a minor aspect of this research. Clark used only the highest discrimination index as the criteria for classification; the proposed method used the sum of the squared discrimination indices. Any argument about the relative merits of the two methods was academic because over both sets of cross-validation data, neither method performed beyond the chance level (see Table 10).

The only groups that were classified with any degree of accuracy within both sets of data were the veterinarians and possibly the clinical psychologists. This can probably be explained by the large number of predictor patterns obtained for those specific groups in the prediction program. The veterinarians and the clinical psychologists had 44% and 45% of the patterns associated with Set I and Set II data respectively. With a great number of patterns one would expect a large number of hits. The number of predictor patterns a group had, compared with the total number of patterns within the set, almost dictated the number of hits that were likely to occur for a given occupation.

Based upon the number of hits in the cross-validation data, the modified pattern analytic scoring method was equal to or better than the original method in eight of ten groups. When summed over groups in Set I, the original method performed slightly better than the modified method (but both were worse than chance). The modified procedure exceeded the original procedure for Set II data with both procedures hovering close to the chance level. It was concluded that neither method should be employed in attempts to classify persons if the problem considerations are similar to those of this research.

Discriminant Function Scoring Results

The weights for the discriminant function procedure were derived by scoring half A of the data with the Kuder lambda weights. These resulting scores were used (one for each occupation for each subject) in the discriminant function analysis. Both Set I and Set II data yielded four significant latent roots (see Appendix A). The eigenvectors associated with the four latent roots were used to score a subject's response vector. The response vector (for each subject) was obtained from Kuder lambda scores in half B of the data. Mahalanobis' D_2 statistic was employed to determine a subject's predicted occupational classification. The sum of the squared distance of the observed scores from the group contour scores was the criterion for the classification. The classification total for each cross-validation sample is included in Appendix A.

On the discriminant function, by definition, 100% of the trace was accounted for with the four latent roots of the determinental equation $|W^{-1}A-\lambda I| = 0$. For Set I, the four latent roots accounted for 41, 30, 20, and 8% of the trace respectively. For Set II, the four roots accounted for 48, 23, 19, and 11% of the trace respectively. The Chi-square test of significance for each of the latent roots was significant ($p < .0001$). This was interpreted to mean that the first latent root of Set I accounted for 41% of the explained variance, the second latent root of Set I accounted for 30% of the explained variance, etc. Similar results and explanations hold for Set II.

Even though the Chi-square approximations for significance testing were significant well beyond the chance level, it was entirely possible that this resulted because the statistic had a great deal of statistical power due to the large sample sizes; i.e., high statistical significance does not necessarily insure a high degree of discriminatory power for a predictive instrument. One measure of the discriminatory power is the proportion of correct classifications on the cross-validation sample. Unfortunately, there is no single index of total discriminatory power or "percent of variance accounted for" with respect to the multiple discriminant function; no such multivariate statistic appears in the literature. However, Tatsuoka (1970) argued for the generalization of the univariate statistic ω^2. The argument reduced to the comparison of the variance attributed to group differences compared with the variance attributed to the total sum of squares and cross-products. The multivariate analog to ω^2 becomes

$$\omega^2_{\text{mult}} = \frac{|T| - |W| - \frac{k-1}{N-k}|W|}{|T| - \frac{1}{N-k}|W|},$$

where T is the sum of the squares and cross-products matrix and W is the within groups sum of squares and cross-products matrix; k is the number of groups, and N is the total sample size.

When applied to the research data, Tatsuoka's index yielded ω^2_{mult} values of .8019 and .9378 respectively for Set I and Set II. This was interpreted to mean that approximately 80% and 94% of the total variability of the four discriminant functions was attributable to group differences for Set I and Set II respectively. These results indicated a high degree of discrimination for these data. The data for dissimilar occupations (Set II) had slightly greater discriminatory power (based on the generated functions) than did the data for similar occupations (Set I). One would expect persons within similar occupations to respond similarly on the criterion measure and thus render differentiation more difficult. However, the group similarity did not restrict the "discriminability" of the Tatsuoka statistic.

With the present research data, Wilk's lambda statistic provided a test of the equality of the five groups population centroids. The Λ values were computed as .197 and .062 respectively for data Set I and Set II. These results were both statistically significant (see Appendix A). The smaller the values of the Λ the higher the probability the group centroids are significantly different. Lambda is defined as the ratio of the pooled within groups deviation score cross-products matrix to the total deviation score cross products matrix. As the proportion of total variance attributable to among groups differences increases (within groups variability remaining constant), the Λ score approaches zero. For this study, a very high proportion of the total variability was attributable to the among groups centroid differences.

The use of discriminant function analysis on these data was successful in generating equations that were ultimately employed in a prediction model. Subsequently, the equations were used to classify individuals into one of five occupational groups. As illustrated in Table 1, for nine of the ten occupational groups the proportion of correct classification on the cross-validation procedure ranged from a high of .91 to a low of .64.

The discriminant function technique worked very well for all but one group, physical therapists. Upon cross-validation, only 21% of the physical therapist group was accurately classified. This result puzzled the researcher in light of the total data. A data error was suspected, but none was found. The contour score for this group was in close proximity to at least one other contour when considered with the other similar occupations. Also, the physical therapist group did not seem to be as homogeneous as the other four groups. The predictability for this group with respect to the other scoring methods was lower than all but one of the groups using the Kuder lambda and lower than all groups using the Chi-square method.

Set II proportions of persons correctly classified on the cross-validation procedure were generally higher than Set I proportions for all scoring methods. This was expected because Set II data contained the dissimilar occupations which typically lend themselves to more accurate classification than the similar groups.

Chi-square Scoring Results

The results for Chi-square scoring procedure were comparable in accuracy to the Kuder lambda and the discriminant function procedures. The Chi-square method over both sets of data was slightly lower in the proportion of correct classifications on the cross-validation than the Kuder lambda. The Chi-square technique was slightly better than the discriminant function on Set I data and slightly worse than discriminant function on Set II data. For certain groups, e.g., veterinarians and auto mechanics, the Chi-square procedure was the best method of scoring to obtain the most accurate results. The proportion of correctly classified persons on the cross-validation procedure based on the Chi-square scoring system ranged from a low of .40 to a high of .97 across ten occupations. The Chi-square method was at a slight disadvantage compared with the Kuder lambda and discriminant function techniques because it used true cross-validation whereas the other two methods employed quasi cross-validation.

Comparison of Methods in Quasi Cross-Validation

In order to accurately compare the Chi-square procedure with the discriminant function and Kuder lambda procedures, the results of each method must be based on the same point of origin. Since the Chi-square method used true cross-validation and the other two methods used quasi cross-validation, the points of origin were different. Therefore an attempt was made to place the methods on equal footing.

The results Porter (1967) presented for Set I, half A data weights used to score Set I, half A data via the Chi-square procedure (non cross-validation) were compared with the discriminant function and Kuder lambda procedures when applied to Set I data based upon the weights generated from Set I data in this study (non cross-validation) (see Appendix B). The results indicated over the five groups that the Chi-square procedure correctly classified approximately 10% more of the persons than did the discriminant function and Kuder lambda procedures. The Chi-square was more accurate than the discriminant function procedure in all but the optometrist group for Set I data. The proportions of correct classifications on the parent samples ranged from .61 to .91 for the Chi-square method, from .21 to .78 for the discriminant function method, and from .53 to .71 for the Kuder lambda procedure.

A similar comparison was made for Set II data. The results were almost completely opposite from those for Set I data, i.e., the proportion of persons correctly classified on the discriminant function and Kuder lambda procedures exceeded the Chi-square results by approximately 8% and 5% respectively over the five groups. The discriminant function results for three of the five groups for Set II data were more accurate than the Chi-square results. The proportions ranged from .73 to .89 and from .47 to .97 respectively. The discriminant function method was slightly more accurate than the Kuder lambda method, but there was no consistent pattern from group to group. These results supported the conclusion that no one method is best for all groups.

With the discriminant function and Kuder lambda analyses based upon the rather large sample sizes, there was minimal "shrinkage" on the cross-validation sample. In several groups there was an increase in the proportion of correct classification on the cross-validation sample. When a similar comparison was made by Porter (1967) on the Chi-square scoring system, minimum to moderate shrinkage occurred on the cross-validation sample when compared with results of scoring based on weights which were derived from and used with the parent sample. In all groups there was a decrease from the parent to the cross-validation proportions based on the Chi-square technique.

These results may be indicative of a highly generalizable set of weights based upon the Kuder lambda and discriminant function procedures or possibly it is indicative of "overfitting" the prediction model through the use of quasi cross-validation. Also, it must be remembered that the Kuder lambda scoring keys were derived on samples twice as large as the samples used to derive the Chi-square scoring keys. Therefore, the Kuder lambda keys were probably more precise and yielded better estimates of the population parameters than did the Chi-square keys. This would account for the large shrinkage on the cross-validation sample when the Chi-square method was applied to these data.

Herzberg (1969) presented evidence indicating that the population cross-validity is very accurately estimated by the sample cross-validity when the number of predictors is small and the sample size is large. The discriminant function results tended to support that finding. However, one must remember that these results were not derived through true cross-validation. His findings might explain the difference in the amount of shrinkage on the Chi-square and discriminant function scoring methods. With the sample size constant, the number of variables used in the predictor model for the Chi-square was 100 whereas the discriminant function used only five variables.

CHAPTER V

SUMMARY AND CONCLUSIONS

Four methods of scoring the Kuder Occupational Interest Survey (Form DD) were used in this research in an attempt to determine the most accurate method of classifying individuals into a select number of occupations. A set of five similar and a set of five dissimilar occupations were used in conjunction with the pattern analytic, Chi-square, Kuder lambda, and discriminant function scoring procedures. The hit or miss classification of individuals within an occupation on a cross-validation sample was the basis for the comparison of methods. The data were analyzed by a repeated measures analysis of variance, using conservative tests for the sources of variation, and subsequent tests of all simple effects. The results indicated that the most accurate method of classification was dependent upon the homogeneity of the group under consideration, the composition of the remaining groups in the data set being inspected, and, the method of scoring, i.e., no one method performed consistently better than the other methods.

The pattern analytic method was unacceptable for use as a method for scoring the Kuder. The method consistently failed to classify individuals correctly at more than the chance level for the majority of the occupations. Additional research is indicated to determine the usefulness of the pattern analytic program when the predictor and criterion variables are nominal level data but not dichotomous and when the sample size is smaller than that which was considered in this research. Limited extensions of the technique as proposed by Clark may be helpful in trying to determine the program's limitations.

A minor aspect of this study considered the accuracy of a modified version of the prediction program for the pattern analytic technique compared with the original prediction program. Neither method functioned very well; both resulted in chance level predictions.

On Set I data, the Chi-square method was competitive (over all groups) with the Kuder lambda and the discriminant function methods. However, the accuracy of the Chi-square method in classifying persons among groups varied considerably from occupation to occupation as did the accuracy of the discriminant function technique. This variability was a contributing factor to the interaction between groups and methods.

On Set II data, the accuracy of the Chi-square method again displayed considerable variability from group to group, but across groups it did not perform as well as the Kuder lambda and discriminant function methods. Again, the variability contributed to the interaction between groups and methods. For certain occupational groups the Chi-square method performed extremely well, while in other groups it was not as accurate as the other methods.

The Kuder lambda method performed the best of the four methods on five of the ten occupational groups; four of the groups appeared in Set I of the data. The proportion of correct classification on the quasi cross-validation ranged from .56 to .81 for all ten groups. Thus, the Kuder lambda was the most consistent method of scoring.

The Kuder lambda scoring method had a definite advantage over the remaining three scoring methods. There was no true cross-validation for the Kuder scoring system, and the derivations of the Kuder weights were based upon sample sizes approximately twice the size of those used with the remaining three methods. With the larger sample sizes, parameter estimates are usually more accurate and this ultimately translates into higher proportions of correct classifications. It was possible that these two factors accounted for the consistent accuracy of the Kuder lambda scoring procedure.

When this study was originally planned, there was a strong possibility that only the Kuder scoring keys and not the formula for deriving the scoring keys would be made available to the researcher. Thus, a decision was made to incorporate quasi cross-validation into the design of the study. The scoring formula was subsequently released to the researcher but use of the formula was not integrated into this study.

The discriminant function method performed very similarly to the Kuder lambda except on Set I, one group had only 21% accurately classified. On each of the other nine groups, the discriminant function method was either within chance error range or more accurate than the Kuder lambda results. On Set II data, the discriminant function method accurately classified 82% of the subjects across the five groups. As with all other studied scoring procedures, the discriminant function procedure was more accurate in classifying the dissimilar groups than the similar groups.

The four scoring methods were empirically based and changes in the composition of the five occupations comprising a set of data would probably alter the weights for all methods except the Kuder lambda. Replacing the current subjects with a new set of subjects within a particular group may or may not have an effect upon the results. There would probably be minimal changes associated with the Chi-square results. The Kuder lambda weights had the advantage of being derived and tested with the same data. Thus, the Kuder lambda classification proportions were probably slightly inflated due to the idiosyncrasies inherent in the data which contributed to the prediction model. These results probably would change if a new set of subjects were used in further research. The discriminant function procedure was based upon the Kuder lambda scoring procedure and therefore, the derived classification proportions are probably also slightly inflated. These results would probably change if a new set of subjects were used in the research.

The results of the study were clear. There was no one best method to score the Kuder Interest Survey to consistently obtain the most accurate occupational predictions. If one were forced to choose one particular method and exclude the others, the discriminant function method would probably be chosen for dissimilar occupations. No method could be identified as best for similar occupations. However, if one could select the scoring method to fit the occupations under consideration for a specific situation, a variety of procedures would probably be recommended. These results must be viewed as tentative as true cross-validation comparisons were not available for all scoring methods.

An issue which has not been considered in this study is the expenditure of resources necessary to ascertain the discriminant function weights. Parameters which would be involved in the choice of a method would include:

(a) the desired or required accuracy of the scoring procedure; (b) the purpose of the instrument; (c) the resources and facilities at hand; (d) the sophistication of the researcher; and (e) the composition of the groups to be discriminated among.

If the method of scoring can improve the accuracy with which classification decisions are made, as occurred in this study, then the ramifications for education are obvious. With the increased availability of computer hardware and software, gaining additional accuracy will benefit those for whom the instrument was intended. The cost of this increased accuracy seems minimal when compared with an error in classification.

Suggestions for additional research include a weighted composite of the three (or more) scoring methods in an attempt to "tease out" the maximum possible prediction. The inclusion of additional groups in the research design to determine the effects of increasing the number of variables entering the multiple discriminant function equation might prove fruitful. Using the Chi-square scores as a starting point for the discriminant function analysis as was done with the Kuder lambda scores in this research, is a third possibility.

Probably the most obvious suggestion for additional research would be to redesign the present study allowing for true cross-validation of all methods of scoring. This would provide more accurate comparisons among all scoring methods as well as stronger evidence regarding the potential of any single technique or combination of techniques for scoring the Kuder.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Barnard, M. M. The secular variations of skull characteristics on four series of Egyptian skulls. Annals of Eugenics, 1935, 6 (Part IV), 352-371.

Bartlett, M. S. Multivariate analysis. J. Roy. Statist. Soc. Suppl., 1947, 9, 176-197.

Bartlett, M. S. The statistical significance of cononical correlations. Biometricka, 1941, 32, 29-38.

Beech, H. R., & Maxwell, A. E. Differentiation of clinical groups using cononical variates. J. Consult. Psychol., 1958, 22, 113-121.

Bock, R. D. Multivariate statistical methods in behavioral research. Statistical Laboratory, University of Chicago, 1969. (Mimeo, chapters 1, 2).

Box, G. E. P. A general distribution theory for a class of likelihood criteria. Biometricka, 1949, 36, 317.

Box, G. E. P. The exploration and exploitation of response surfaces. Biometrics, 1954, 10, 16-60.

Brogden, H. E. An approach to the problem of differential prediction. Psychometrika, 1946, 11, 139-154.

Brown, G. W. Discriminant functions. Ann. of Math. Stat., 1947, 18, 514-528.

Bryan, J. G. The generalized discriminant function: Mathematical foundation and computational routine. Harvard Educ. Rev., 1951, 21, 90-95.

Burt, C. On the discrimination between members of two groups. In A. Lubin, Linear and non-linear discriminating functions. Brit. J. Psychol. (Stat. Section), 1950, 3, 104.

Cattell, R. B. r_p and other coefficients of pattern similarity. Psychometrika, 1949, 14, 279-298.

Cattell, R. B. (Ed.) Handbook of multivariate experimental psychology. Chicago: Rand McNally and Co., 1966.

Chappell, J. S. Multivariate discrimination among selected occupational groups utilizing self report data. Unpublished doctoral dissertation, Purdue University, 1967.

Claringbold, P. J. Multivariate quantal analysis. J. Roy. Statis. Soc., 1958, 20, 398-405.

Clark, J. A. Criterion Pattern Analysis: A method for identifying predictive item configurations. Unpublished doctoral dissertation, Michigan State University, 1968.

Clark, J. A. Program Predict. Undocumented Fortran program. Michigan State University, 1969.

Clemans, W. V. An analytical and empirical examination of some properties of ipsative measures. Psychometric Monographs, 1968, 14.

Clemans, W. V. An index of new criterion relationships. Educ. and Psychol. Measmt., 1958, 18, 1.

Clemens, B. T. Discriminant function analysis of inventoried interests among selected engineering groups. Unpublished doctoral dissertation, Purdue University, 1969.

CoBabe, T. A. The Strong Vocational Interest Blank as a predictor of success in engineering. Unpublished doctoral dissertation, Purdue University, 1967.

Cochran, W. G., & Hopkins, C. E. Some classification problems with multivariate qualitative data. Biometrics, 1961, 17, 10-32.

Cooley, W. W., & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: John Wiley and Sons, 1962.

Creager, John A. Discriminant analysis and its role in the classification of airmen. Research report: Air Research and Development Command, Lackland Air Force Base, Texas, 1957.

Crego, C. A. A pattern analytic approach to the measure of modes of expression of psychological differentiation. Unpublished doctoral dissertation, Michigan State University, 1966.

Cronbach, L. J., & Gleser, G. C. Assessing similarity between profiles. Psychol. Bull., 1953, 50, 456-473.

Draper, J. Personal communication on properties of repeated measure designs with dichotomous dependent variables. Unpublished doctoral dissertation, Michigan State University, 1971.

DuMas, F. M. A quick method of analyzing similarity of profiles. J. Clin. Psychol., 1946, 2, 80-83.

Dwyer, P. S. Solution of the personnel classification problem with the method of optimal regions. Psychometrika, 1954, 19, 11-26.

Dwyer, P. S. The detailed method of optimal regions. Psychometrika, 1957, 22, 43-52.

Findley, Warren. A rationale for evaluation of item discrimination. Educ. and Psychol. Measmt., 1956, 16, 175-180.

Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 1936, 7, 179-188.

Fisher, R. A. The statistical utilization of multiple measurements. Annals of Eugenics, 1938, 13, 376-386.

Fisher, R. A. The sampling distribution of some statistics obtained from non-linear equations. Annals of Eugenics, 1939, 9, 238.

Fricke, B. G. A configural-content-intensity item for personality measurement. J. Educ. Psychol. Measmt., 1956, 16, 54-62.

Fricke, B. G. A coded profile method for predicting achievement. J. Educ. Psychol Measmt., 1957, 17, 98-104.

Gaier, E. L., & Lee, Marilyn C. Pattern analysis: The configural approach to predictive measurement. Psychol. Bull., 1953, 50, 140-148.

Ghiselli, C. E. Differentiation of tests in tests of the accuracy with which they predict for a given individual. Educ. Psychol. Measmt., 1960, 20, 675-684.

Graybill, F. A. An introduction to linear statistical models. Vol. I. New York: McGraw-Hill, 1963.

Greenhouse & Geisser. On the analysis of profile data. Psychometrika, 1959, 24, 95-111.

Herzburg, P. A. The parameters of cross-validation. Psychometric Monograph Supplement, 1969, 16.

Horst, P. Pattern analysis and configural scoring. J. Clin. Psychol., 1954, 13, 107-114.

Hsu, Tse-Chi, & Feldt, L. S. The effect of limitations on the number of criterion score values on the significance level of the F-test. Educational Research Journal, 6(4), 1959, 515-527.

Jones, L. V. Problems of devising and selecting appropriate measurement tools. Am. J. Ment. Def., 1959, 64, 384-394.

Jones, L. V. Some illustrations of psychological experiments designed for multivariate statistical analysis. The Psychometric Laboratory, University of North Carolina, December, 1960. (Mimeograph)

Jones, L. V., & Bock, R. D. Multiple discriminant analysis applied to "ways to live" ratings from six cultural groups. Sociometry, 1960, 23, 162-176.

Kelley, T. L. Statistical method. New York: Macmillan, 1923.

Kirk, Roger. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks/Cole, 1968.

King, P., & Norrell, Gwen. A factorial study of the Kuder Preference Record Occupational, Form D. Educ. Psychol. Measmt., 1964, 24, 57-63.

Kuder, G. F. Kuder Preference Record Occupational, Form D Test Booklet. Chicago: SRA, 1956.

Kuder, G. F. A comparative study of some methods of developing occupational keys. Educ. Psychol. Measmt., 1957, 17, 105-114.

Kuder, G. F. Kuder Preference Record Occupational, Form D Manual. Chicago: SRA, 1961.

Kuder, G. F. A rationale for evaluating interests. Educ. Psychol. Measmt., 1963, 23, 3-10.

Kuder, G. F. Kuder Occupational Interest Survey, Form DD, General Manual. Chicago: SRA, 1970.

Kuder, G. F. Kuder Occupational Interest Survey, Form DD, Test Booklet. Chicago: SRA, 1966.

Lee, M. C. Configural vs. linear prediction of collegiate academic performance. Dissertation Abstracts, 1957, 17, 297-398.

Lee, M. C. Interactions, configurations and non-additive models. Educ. Psychol. Measmt., 1961, 21, 797-805.

Lohnes, Paul R. Test space and discriminant space classification models and related significance tests. Educ. Psychol. Measmt., 1961, 21, 559-574.

Lubin, A. Linear and non-linear discriminating functions. Brit. J. Psychol. (Stat. Sect.), 1950, 3, 90-104.

Lubin, A. A methodological study of configural scoring. U.S. Army Personnel Research Bureau, 1954, Note No. 42, 13pp.

Lunney, Gerald H. Using analysis of variance with a dichotomous dependent variable: An empirical study. J. of Educ. Measmt., 1970, 7(4), 263-269.

Marascuilo, L., & Levin, J. Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of type IV errors. A.E.R.J., 1970, 11, 397-421.

Maxwell, A. E. Cononical variate analysis when the variables are dichotomous. Educ. Psychol. Measmt., 1961, 21, 259-271.

McQuitty, L. L. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. Educ. Psychol. Measmt., 1957a, 17, 207-229.

McQuitty, L. L. Isolating prediction patterns associated with major criterion patterns. Educ. Psychol. Measmt., 1957b, 17, 3-42.

McQuitty, L. L. Job knowledge scoring keys by item versus configural analysis for assessing levels of mechanical experience. Educ. Psychol. Measmt., 1958, 18, 661-680.

McQuitty, L. L. A method for selecting patterns to differentiate categories of people. Educ. Psychol. Measmt., 1961a, 21, 85-94.

McQuitty, L. L. Item selection for configural scoring. Educ. Psychol. Measmt., 1961b, 21, 925-928.

McQuitty, L. L. Rank order typal analysis. Educ. Psychol. Measmt., 1963, 23, 55-61.

McQuitty, L. L. Single and multiple hierarchial classification by reciprocal pairs and rank order types. Educ. Psychol. Measmt., 1966, 26, 253-265.

Mood, A. M., & Graybill, F. A. Introduction to the theory of statistics. New York: McGraw-Hill, 1963.

Morrison, D. F. Multivariate statistical methods. New York: McGraw-Hill, 1967.

Mosier, C. I. Problems and designs of cross-validation. Educ. Psychol. Measmt., 1951, 11, 5-11.

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

Pillai, K. C. S. On the distribution of the largest or smallest root of a matrix in multivariate analysis. Biometrika, 1956, 43, 122.

Pillai, K. C. S. On the distribution of the largest seven roots of a matrix in multivariate analysis. Biometrika, 1964, 51, 270-275.

Pillai, K. C. S. On the distribution of the largest characteristic root of a matrix in multivariate analysis. Biometrika, 1965, 52, 405-414.

Pillai, K. C. S., & Samson, P., Jr. On Hotelling's generalization of T squared. Biometrika, 1959, 46, 160-168.

Porter, A. C. A chi-square approach to discrimination among occupations using an interest inventory. Technical report No. 24. University of Wisconsin, Center for Cognitive Learning, Madison, Wisconsin, 1967.

Rao, C. R. Utilization of multiple measurements in problems of biological classifications. _J. Roy. Statist. Soc._, 1948, _10_, 159-193.

Rao, C. R. _Advanced statistical methods in biometric research_. New York: Wiley, 1952.

Roy, S. N. P-statistics or some generalizations in analysis of variance appropriate to multivariate problems. _Sankhya_, 1939, _4_, 381.

Roy, S. N. On a heuristic method of test construction and its use in multivariate analysis. _Ann. Math. Stat._, 1953, _24_, 220-238.

Roy, S. N. _Some aspects of multivariate analysis_. New York: Wiley, 1957.

Roy, J., & Borgmann, R. E. Tests of multiple independence and the associated confidence bounds. _Ann. Math. Stat._, 1958, _29_, 491-503.

Roy, J. Step-down procedure in multivariate analysis. _Ann. Math. Stat._, 1958, _29_, 1177-1187.

Rozeboom, W. W. The fallacy of the null-hypothesis significance test. _Psychol. Bull._, 1960, _57_, 416-428.

Rulon, P. J. Distinctions between discriminant and regression analysis and geometric interpretation of the discriminant function. _Harvard Educ. Rev._, 1951, _21_, 80-90.

Rulon, P., Tiedeman, D., Tatsuoka, M., & Longmuir, C. _Multivariate statistics for personnel classification_. New York: John Wiley and Sons, 1967.

Rulon, P., Tiedeman, D., Langmuir, C., & Tatsuoka, M. _The profile problem: A methodological study of the interpretation of multiple test scores_. Cambridge, Mass., Educ. Research Corporation, February 1954.

Schotzoff, M. Exact distributions of Wilks Likelihood Ratio Criterion. _Biometrika_, 1966, _53_, 347-358.

Schutz, R. E., & Baker, R. L. A factor analysis of the Kuder Preference Record Occupational, Form D. _Educ. Psychol. Measmt._, 1962, _22_, 97-104.

Stellwagon, W. R. An examination of the use of linear and non-linear discriminant functions in the classification of college students into academic groups. _Dissertation Abstr._, 1960, _20_, 4435.

Stephenson, Richard. A new pattern analysis technique for the SVIB. J. of Counsel. Psych., 1961, 8, 355-362.

Stormes, Lowell, H. Discrepancies between factor analysis and multivariate discrimination among groups as applied to personality theory. J. Ment. Sci., 1958, 104, 713-721.

Tatsuoka, M. M. The relationship between cononical correlation and discriminant analysis; and a proposal for utilizing qualitative data in discriminant analysis. Cambridge, Mass.: Educational Research Corp., 1955.

Tatsuoka, M. M., & Tiedeman, D. V. Discriminant analysis. Rev. Educ. Res., 1954, 24, 402-420.

Thorndike, R. L. The problem of classification of personnel. Psychometrika, 1950, 15, 215-235.

Tiedeman, D. V. The utility of the discriminant function in psychological and guidance investigations. Harvard Educ. Rev., 1951, 21, 71-80.

Tiedeman, D. V., & Bryan, J. G. Prediction of college field of concentration. Harvard Educ. Rev., 1954, 24, 122-139.

Tiedeman, D. V., Bryan, J. G., & Rulon, P. J. The utility of the Airman Classification Battery for assignment of airmen to eight Air Force specialties. Cambridge, Mass.: Educ. Research Corp., June, 1951.

Tiedeman, D. V., Rulon, P. J., & Bryan, J. G. The multiple discriminant function: A symposium. Harvard Educ. Rev., 1951, 21, 71-95.

Tilton, J. W. The measurement of overlapping. J. Educ. Psychol., 1937, 28, 656-662.

Tintnes, G. A note on rank, multi-colinearity, and multiple regression. Ann. Math. Stat., 1945, 16, 304-308.

Travers, R. M. W. The use of a discriminant function in the treatment of psychological group differences. Psychometrika, 1939, 4, 25-32.

Votaw, D. F., Jr. Methods of solving some personnel-classification problems. Psychometrika, 1952, 17, 255-266.

Weiss, D. J. A technique for curvilinear multivariate prediction. Dissertation Abstracts, 1964, 24(11), 4788.

Welch, B. L. Note on discriminant functions. Biometrika, 1939, 31, 218-220.

Wilks, S. S. Certain generalizations in the analysis of variance. Biometrika, 1932, 24, 471-474.

Zubin, J. The determination of response patterns in personality adjustment inventories. J. Educ. Psychol., 1937, 28, 401-413.

Zubin, J. Nomographs for determining the significance of the differences between the frequencies of events in two controlled series or groups. J. Amer. Statis. Assoc., 1939, 34, 539-548.

APPENDICES

APPENDIX A

DISCRIMINANT FUNCTION ANALYSES OF SET I AND SET II DATA

Set I.--Discriminant Function for Set I Half A Used to Score Set I Half B Data.

5 Variables		Number of Total Classifications on Cross-Validation				
		5 Groups				
		AOpts	AX-ray	APeds	APT	AVets
N=203B	Optometrists	140	12	17	19	15
N=138B	X-ray Technicians	10	93	12	15	8
N=218B	Pediatricians	12	13	140	32	21
N=206B	Physical Therapists	51	55	42	43	15
N=200B	Veterinarians	10	6	12	9	163

Trace = 2.0903	Wilk's lambda = .197	$F_{20,3079} = 97.23$
Root 1 = 41.41% of trace	$\chi^2 = 581.133$ Df = 8	P = .0001
Root 2 = 30.00% of trace	$\chi^2 = 453.653$ Df = 6	P = .0001
Root 3 = 20.36% of trace	$\chi^2 = 330.431$ Df = 4	P = .0001
Root 4 = 8.24% of trace	$\chi^2 = 148.130$ Df = 2	P = .0001

Discriminant Function Weights Based on Set I Half A Data

	1	2	3	4
Optometrists	-.0775	-.8030	.9372	-.0898
X-ray Technicians	-.4921	-.0368	-.1792	-.5970
Pediatricians	-.0223	.4777	.6063	-.1691
Physical Therapists	-.2314	.3541	-.6598	.7692
Veterinarians	.8353	.0174	-.0849	.1236

Set II.--Discriminant Function for Set II Half A Used to Score Set II Half B Data.

5 Variables		Number of Total Classifications on Cross-Validation: 5 Groups				
		AClin. Psy	AAuto	AForst.	ASCW	AOpts
N=250B	Clin. Psychologists	209	0	5	22	14
N=150B	Auto Mechanics	1	136	6	2	5
N=173B	Foresters	6	7	147	6	7
N=226B	Social Case Workers	29	4	6	172	15
N=200B	Optometrists	9	11	9	17	154

Trace = 4.3258	Wilk's lambda = .062	$F_{20,3284} = 215.37$
Root 1 = 47.48% of trace	χ^2 = 1116.708 Df = 8	P = .0001
Root 2 = 22.67% of trace	χ^2 = 679.368 Df = 6	P = .0001
Root 3 = 18.64% of trace	χ^2 = 587.808 Df = 4	P = .0001
Root 4 = 10.71% of trace	χ^2 = 378.386 Df = 2	P = .0001

Discriminant Function Weights Based on Set II Half A Data

	1	2	3	4
Clin. Psychologists	.7134	-.3789	.0541	.5130
Auto Mechanics	.0815	.5195	-.3488	.1444
Foresters	-.3945	.2287	.7299	-.1471
Social Case Workers	.3488	-.2003	.2107	-.8331
Optometrists	-.4551	.7030	-.5461	.0165

APPENDIX B

A PRESENTATION OF THE NON CROSS-VALIDATED CLASSIFICATION PREDICTION FOR THE CHI-SQUARE, KUDER LAMBDA AND DISCRIMINANT FUNCTION TECHNIQUES

A presentation of the proportions of correctly classified males in Set I and Set II data based upon the weights derived from half A and used to score half A and half B data for the Chi-square, Kuder lambda, and discriminant function techniques.

	Chi-square			Kuder Lambda		Discriminant Function	
	Half A*	Half B*	Half B**	Half A**	Half B**	Half A**	Half B**
Set I							
Optometrists	.6108	.5419	.5567	.6896	.6946	.6798	.6847
X-ray Technicians	.6423	.4964	.5725	.6617	.6739	.5735	.6667
Pediatricians	.7478	.6422	.6835	.7064	.6972	.6972	.6422
Physical Therapists	.6062	.3679	.4806	.5278	.6117	.2111	.2087
Veterinarians	.9050	.8950	.9000	.5850	.6100	.7750	.8200
Total	.7040	.5940	.6425	.6429	.6570	.6051	.5990
Set II							
Clinical Psychol.	.9080	.8840	.8840	.8440	.8240	.7880	.8360
Auto Mechanics	.9732	.9597	.9733	.8263	.7933	.8867	.9067
Foresters	.5517	.4483	.5202	.8093	.8092	.8208	.8497
Social Case Workers	.6903	.5973	.6150	.5575	.5575	.7301	.7611
Optometrists	.4680	.4236	.4050	.8200	.8050	.8150	.7550
Total	.7180	.6629	.6788	.7650	.7528	.8008	.8158

*Provided from Porter's 1967 study.
**Provided from the present study.

APPENDIX C

VARIANCE COVARIANCE MATRICES OF THE DICHOTOMOUS CROSS-VALIDATION SCORES FOR SET I AND SET II DATA

APPENDIX C

VARIANCE COVARIANCE MATRICES OF THE DICHOTOMOUS CROSS-VALIDATION SCORES FOR SET I AND SET II DATA

Set I	M_1	M_2	M_3	M_4
M_1	.1517	.0180	.0007	.0292
M_2		.2297	.1022	.1136
M_3			.2253	.1257
M_4				.2402

Set II	M_1	M_2	M_3	M_4
M_1	.1607	.0258	-.0103	.0040
M_2		.2185	.0635	.0427
M_3			.1861	.0646
M_4				.1502

M_1 = Pattern Analysis
M_2 = Chi-square
M_3 = Kuder lambda
M_4 = Discriminant Function