CRITERION PATTERN ANALYSIS: A METHOD FOR IDENTIFYING PREDICTIVE ITEM CONFIGURATIONS

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

JAMES ARTHUR CLARK

1968

This is to certify that the

thesis entitled

CRITERION PATTERN ANALYSIS:
A METHOD FOR IDENTIFYING PREDICTIVE
ITEM CONFIGURATIONS

presented by

JAMES ARTHUR CLARK

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Louis L. McQuitty
Major professor

Date 22 January 1968

O-169

ABSTRACT

CRITERION PATTERN ANALYSIS: A METHOD FOR IDENTIFYING PREDICTIVE ITEM CONFIGURATIONS

by James Arthur Clark

Meehl, in 1950, demonstrated the potential usefulness of patterns of items for predicting. The various approaches subsequently developed to capitalize on this predictive power of patterns are classified and evaluated. In the light of their marked lack of success, nine principles for an effective pattern prediction method are proposed. The method should (1) find all major patterns which predict the criteria; (2) find patterns separately for each criterion category; (3) isolate non-configural as well as configural relationships; (4) be capable of predicting directly from patterns; (5) be capable of predicting better than linear methods on the analysis sample; (6) predict better than linear methods upon cross validation; (7) be applicable to small samples; (8) yield readily interpretable results; (9) provide readily obtainable results. Criterion Pattern Analysis (CPA), a method conforming to the above principles, has been developed.

CPA operates on discrete data, typically a matrix of the responses to a set of items made by people who have

been previously classified into two or more criterion categories. For each of the criteria, patterns of responses are sought which are highly related only to that category. These patterns may be in one item, two items, three items, etc. A pattern is accepted as relating to the criterion category if it is significantly more predictive of the criterion than are any of its subpatterns; the hypergeometric distribution is utilized in making the significance test. In checking all possible patterns, all one-item patterns are scrutinized first, then two-item patterns, then three-item patterns, etc. To overcome the impossible task of checking all possible patterns one by one, a technique of rejecting many patterns at one stroke is employed. Thus a pattern of r items, whether or not acceptable itself, is also tested as to whether it can possibly be improved through the addition of more items. Only if it can be improved significantly will it be used in the formation of patterns of r+1 items. This procedure for CPA was carefully programmed to make efficient use of the capabilities of an electronic computer.

A method for predicting directly from patterns was developed. A person for whom prediction is desired is checked for the patterns previously extracted. That pattern which he has which is most highly related to its criterion determines the highest prediction. In this way a hierarchy of prediction can be obtained. In an alternative prediction scheme, each person in the original response matrix is given scores of 1's and 0's according to whether or not he has each pattern. This set of scores can then be employed in one of the linear prediction methods. In this guise, CPA functions as an extension of item analysis.

Two sets of data were used to compare CPA with two linear methods, multiple regression and a multivariate normal maximum likelihood procedure. The first set involved prediction of field dependence and independence from items of the I-E scale; the second involved predicting voting behavior on a selected issue from votes on other issues in the UN General Assembly. On the analysis samples, of 50 subjects and 55 nations, the maximum likelihood procedure predicted better than did CPA; multiple regression did better than CPA on the UN data, but not as well on the Crego data. On the cross validation samples of 49 subjects and 55 nations, CPA consistently predicted better. The combination prediction scheme yielded better results than did predicting directly from patterns. On both sets of data patterns from CPA offered greater opportunity for substantive interpretation than did the results of the linear methods.

Various ways of applying CPA are indicated. Areas of improvement of the present method are pointed out, such as establishment of an over-all significance level for patterns. CPA is compared with several other methods purported to utilize configural properties found in data. It is suggested that types as determined by patterns from CPA might be capable of helping revise typal theory in general.

CPA is measured against the nine principles initially formulated, and is found to meet all of them with the exception of number 5; here the necessity of sacrificing maximum prediction to the analysis sample in order to obtain best cross validation prediction is asserted. Importantly, the seemingly impossible task of examining all possible patterns in search of the highly predictive ones has been achieved, and with the aid of a high speed computer the application of CPA is made a practical procedure.

CRITERION PATTERN ANALYSIS: A METHOD FOR IDENTIFYING PREDICTIVE ITEM CONFIGURATIONS

By

James Arthur Clark

A THESIS

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

DOCTOR OF PHILOSOPHY

Department of Psychology

1968

ACKNOWLEDGMENTS

No dissertation is developed and written in isolation. There are many people who have been influential in the various stages of conceptualization and presentation of this thesis. First, I wish to acknowledge the members of my committee: Dr. William Mueller and Dr. Bertram Karon provided encouragement and many helpful comments. Dr. Terrence Allen generously extended thorough word for word criticism in the final stages of preparation, as well as valuable suggestions in the development. Most of all, I owe a debt of gratitude to the chairman, Dean L. L. McQuitty. Not only were his patience, encouragement, and suggestions appreciated in the formative stages, but also his skillful instruction led to a better presentation. Certainly without Dean McQuitty's pioneer work in pattern analytic methods, this thesis would not have been possible.

Sincere thanks are owed to John Hafterson who developed many ideas similar to my own and was always ready to listen and react to my attempts to solve the problems at hand. Many people in and around the Michigan State University Computer Laboratory were very

helpful in the development and running of the computer programs connected with the thesis.

Finally, I must express my gratitude to my wife, Jan, who contributed in so many ways I could not begin to recount. It is to her and our son, Paul, that this dissertation is affectionately dedicated.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table		Page

Table Page

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

THE PROBLEM OF PATTERN PREDICTION METHODS

Introduction

In 1950, Meehl, in an influential paper, asserted the importance of patterns for predicting. While single items alone may not be predictive, a pattern of two items together may be perfectly predictive. Meehl argued that how a person responds to a pair of items can uniquely reflect psychological characteristics, particularly in the clinical area (see also Meehl, 1954). Unless patterns are utilized, important psychological information can be lost.

Meehl's paper helped stimulate many workers in psychology to develop quantitative techniques incorporating patterns (Gaier and Lee, 1953; Sells *et al*., 1955). Some techniques were analytical in nature: patterns were used to classify people (e.g., McQuitty, 1957a, 1963, 1966). Other methods were predictive: patterns were used for predicting *a priori* categories of people. Many of these methods are reviewed and classified below.

In this paper, an effective new technique for isolating predictive patterns is presented. The identification of these patterns is viewed here as an extension of item analysis, which only identified predictive single items. The patterns extracted by the new method may be used for predicting either directly or in combination with linear procedures.

Certainly the ideal approach would be to examine all possible patterns and pick out only the most reliably predictive ones; however, the large number of patterns to be examined in most cases renders this approach too laborious even with the aid of a high speed computer. The method proposed here, to be called Criterion Pattern Analysis, identifies the same patterns as does the ideal approach, but without having to examine all possible patterns one by one.

It is the thesis here that Criterion Pattern Analysis is a practical method, more informative than linear methods, and at least as predictive as linear methods--more predictive when there are configural properties in the data.

Methods of Predicting from Patterns

The various methods of predicting from patterns fall into six classes: (1) cumulative, (2) reductive, (3) classification, (4) total-pattern, (5) small-pattern, and (6) pattern-search. The first two classes were proposed by McQuitty. The third was first hinted at by McQuitty under the term dual-pattern, but is now only one of several classification methods (McQuitty, 1957b). The last three classes of methods are unique to the present paper. They include some of the more often proposed methods. Criterion Pattern Analysis is best classified under pattern-search methods.

The Cumulative Method

An illustration of this rare method is contained in a study by Lubin (1954). The single item which best predicts the criterion is selected. Next, the item is selected which, when paired with the first, most improves prediction. A third item is added which most improves the item pair, and so on. Lubin applied this technique to 1474 subjects responding to 20 items from the MMPI. He compared the results of this method with multiple regression and found that while the pattern method predicted significantly better on the analysis sample, multiple regression predicted significantly better upon cross validation.

McQuitty (1957b and 1959) has reviewed the properties of the cumulative method and found much to be desired. The cumulative method may grossly miss important predictive patterns. Perhaps the appropriate starting point is not with the best single item, but with a poor single item. The item which pairs best with a poor item may prove to have a better predictive value. On the other hand, the best single item, being highly correlated with the criterion, is limited in its configural relations with the criterion through some other item or items. While a cumulative approach might be feasible, there is, as yet, no practical procedural scheme proposed which will get at the highly predictive item combinations.

The Reductive Method

Whereas the cumulative method starts with one item and builds it to two, three, four, etc., items; the reductive method starts with the total pattern over all items used and reduces it to a subset of fewer items. Both methods, by their respective processes, hope to improve the validity and the practicality of prediction.

McQuitty (1958) provides an example of this procedure. McQuitty first selects the person who has more item responses in common with more members of his criterion group than does any other member. The person most like him is then used to select items which they both answer in common. A third person can be added in a similar way. The resulting responses to a subset of items is treated as a scoring key on which each person in the analysis sample can be scored. A cutting score can then be determined in an effort to separate people belonging to one category from those belonging to another.

This technique was tried on a small sample with four criterion groups. One large group contained 64 people, the others 14, 12, and 15 respectively. When compared with a linear method, this pattern method did better even on the cross validation sample, but on the large group only. On the small groups the linear method did better. This difference was explained by pointing out that only in the large group were there enough people for different patterns to emerge corresponding to different types of people within the group.

This promising method could be exploited further. McQuitty's use of it was not strictly in accordance with what Meehl (1950) called configural scoring, which is treating each combination of items differently. When McQuitty used the patterns as a scoring key to obtain a total score for each person, some "configural" information was lost. This is because two people who have the same score can have two different patterns of responses.

The Classification Methods

In this class of methods, the people within each criterion category are first classified (by some appropriate method) into two or more groups. Predictions are then based on patterns associated with each of the groups.

McQuitty used this method in several different forms. In one form, called the dual-pattern method, (McQuitty, 1957b) people were classified into groups on the basis of external criterion scores. A subset of item responses held in common by all members was found for each criterion group. This pattern was then used as a predictor of the criterion category on other samples of people.

In another form, the classification of people within each category is based on the test items themselves (McQuitty, 1959, 1961a). Selection of the level of classification for identifying predictor patterns can be at the highest levels (major patterns) or at lower levels (minor patterns).

Finally, a modification of this procedure was offered (McQuitty, 1961b) in which each classification in one criterion group is paired with each classification in the other criterion groups. Items are then found for each pair which distinguish one category from another. In this fashion, a scoring key for each pair is developed. Such keys can reflect configural properties if types are present in the data.

The results of applying these methods have been inconclusive or disappointing as compared with linear methods (McQuitty, 1956, 1957b). It seems that many configural properties can be missed while concentrating on a few. Classifying within each criterion group may result in a pattern which has little to do with discriminating between criterion groups. And even if the pattern does discriminate, it was not selected on the basis of being one of the best discriminating patterns. This method does have its merits, however. When there is a large number of items, this method offers a way of getting at predictive patterns which otherwise might never be sought. Even with the advent of high speed computers, looking at all patterns seems a formidable task. These methods offer a compromise between what is possible and what is theoretically best.

The Total-Pattern Methods

When the number of items is very small, say less than ten, and the number of people is very large, say more than a thousand, then the total-pattern methods can be employed. In these procedures the pattern of responses of each person to all the items is used. Since there are relatively few items, all empirical patterns are likely to occur with at least moderate frequencies. Cochran and Hopkins (1961) developed one such predictive model which they used for predicting election outcomes. In such a situation there are very few items and many people who are divided into two criterion categories. The frequency of occurrence of each pattern in each criterion category can be readily counted and used as the basis for probability statements. The prediction for a new person is made by first ascertaining his pattern of responses to the set of items and then predicting to that criterion group which has the largest probability for that pattern.

A similar method for medical diagnosis was proposed by Ledley and Lusted (1959).

Another method which is quite common (Fricke, 1957; Lykken, 1956) is for situations which have quantitative criteria. For each pattern, the mean criterion score is computed over all people who have the pattern. A prediction for a new person is made by first ascertaining his pattern and then assigning to him the mean score associated with his pattern.

A more elaborate version was proposed by Horst (1954) and refined by Lubin and Osburn (1957). It was shown that a pattern of responses could be translated into a polynomial function involving all possible interaction terms. With this mathematical representation, the usual regression analysis could be performed to predict to a quantitative criterion variable. Lubin and Osburn assert that this polynomial technique will produce a minimum number of misclassifications when the criterion score is normally distributed for each pattern.

Alf (1957) and Lee (1957) tried the polynomial technique and found that the usual linear methods were better upon cross validation, although not significantly so. Lee's explanation was that configural methods tend to capitalize on chance patterns which then throw off prediction upon cross validation. Osburn and Lubin (1957) agree that while all information is considered, all information is also conserved whether reliable or not. In other words, there are often too many degrees of freedom, and the parameters in the regression equation will not be accurately estimated. All researchers with the total-pattern methods reiterate the need for very large samples of people.

The Small-Pattern Method

In an attempt to get away from the need for large samples of people in order to make accurate estimates, patterns of only two or three items are utilized instead of patterns over all items. In other respects the small-pattern methods are the same as the total-pattern methods. Lee (1957), in addition to using total patterns of eight items, also used small patterns of five and three items. Hoffman (1960) proposed the polynomial technique using all item pairs. Saunders (1955), using such a technique, found it no better than using linear predictors.

While enormous samples of people are not required, the small-pattern methods still tend to capitalize on random patterns, again leading to erroneous predicting on cross validation samples.

The Pattern-Search Methods

Instead of indiscriminately using all patterns or some subset of patterns, pattern-search methods are selective. The various methods employ different criteria for selecting. Forehand and McQuitty (1959) selected patterns according to their departure from chance occurrence. They found that using significant patterns predicted better upon cross validation than did using all patterns. However, multiple regression still predicted better. They comment that part of the trouble is again too many patterns with too few subjects in each. It might be added that their selection criteria of significance of occurrence did not pick patterns which are necessarily significantly related to the criteria.

Horst (1957) suggested seeking patterns which are highly related to the criteria, after first checking to see whether they could be expressed as a linear function of smaller patterns. This was followed up by Wainwright (1966) who defined the configural phenomenon as a non-linear combination of items. Thus he selected patterns which could not be expressed in terms of single items. However, Wainwright was not interested in predicting to a criterion. His conclusion was that a linear combination of items does not account for all information, which is what Meehl initially asserted.

In general, there seems to be a paucity of pattern-search techniques, even though this might be a fruitful approach. The widespread availability of high speed computing facilities might change the balance in the years to come.

Resumé

By the end of the Fifties, most of the proposed methods to predict with patterns had been tested. The majority of the results were inconclusive or disappointing in comparison with linear methods. Loevinger (1959), in reviewing some of these studies, believed the case for configural predicting was closed. While able to predict better in the analysis sample, pattern methods failed to hold up on the cross validation sample. The linear methods did at least as well or better. Apparently the potential of McQuitty's reductive method and the success of Cochran and Hopkins' probability model had been forgotten and not fully explored. And, of course, the pattern-search methods had hardly gotten off the ground.

While the problem of predicting from patterns had apparently been solved in theory, only unusually large samples of people responding to a few items could be handled reliably. Seldom does the worker in psychology have these kinds of data. Not fully solved was how to use patterns to predict reliably on smaller samples of the kind of qualitative data with which psychologists often work. Hence, the need for a satisfactory pattern prediction method still exists.

Requirements for a Pattern Prediction Method

In the previous section various predictive approaches which attempted to capitalize on configural properties in the data were discussed. In this section a critique for an effective pattern prediction method is outlined.

1. All major patterns which predict the criteria should be found.

This requirement is fundamental, and is implicit in other requirements. What is wanted is a set of patterns which are very reliable and which predict the criteria. No reliable predictive pattern should be excluded. If any such patterns are missing after analysis, predicting in cross validation can be jeopardized.

2. The patterns should be found separately for each criterion category.

The patterns which predict to one criterion group may be ineffective in predicting to another. Fricke (1956) was the first to point this out in a modification of Meehl's example of configural scoring. The requirement does not mean that classifying within each criterion group will be acceptable; as pointed out earlier, the procedure may miss predictive patterns. The analogous assertion for linear methods was made by Stormes (1958).

3. The method should isolate non-configural as well as configural relationships.

For example if single items or linear combinations of them are highly predictive, they should be extracted as such. Some configural methods already allow for single-item "patterns."

4. Patterns extracted by the method should be capable of being used directly for prediction.

This implies that a method may be developed for predicting with patterns themselves rather than with derivatives or functions of the patterns. The prediction method should work separately from the method used to extract patterns and yet be tied to it logically. By using patterns directly, interpretation should be simplified (see requirement 8).

5. The method should be capable of predicting better than linear methods on the analysis sample.

In those cases where the analysis sample is the only one to which prediction is desired, the method certainly should be capable of predicting better than linear methods.

Although satisfaction of this requirement is desirable, a method which predicts better than linear methods on the analysis sample has no guarantee of predicting better on cross validation samples. Prediction to a cross validation sample is based on information common to both samples which is gleaned from the analysis sample. A high prediction on an analysis sample is likely based on much idiosyncratic information which cannot predict to a cross validation sample.

6. The method should predict better than linear methods on a cross validation sample.

Previously this requirement has been the toughest to meet; and yet, if reliable patterns are isolated, as in requirement 1, pattern methods will begin to do much better.

7. The method should be applicable to small as well as to large samples of people.

The repeated assertion about needing larger samples to show that configural properties are present can only weaken the appeal of pattern methods. The aim should be to develop methods effective on the small sample.

8. The results of a configural prediction method should be readily interpretable.

This is a plea for simplicity. With many methods, both linear and configural, it is difficult to understand the relationships between the predictors and the criterion. Configural methods have an opportunity to present a clear picture.

9. The results of a configural prediction method should be readily obtainable.

The method should not exist in theory only, but should be translatable into a practical tool. This translation can be one of the more difficult tasks in the development of a method, and will almost certainly have to be implemented on an electronic computer. Indeed, a method which meets in full the previous eight requirements might be fully translated only with difficulty, even when a computer is available.

Resumé

Suggested requirements for a configural prediction method have been listed. The method proposed in this thesis will be measured against these standards in the final section.

CHAPTER II

THE METHOD OF CRITERION PATTERN ANALYSIS

Introduction

The method of Criterion Pattern Analysis is directed toward solving the problem of predicting with nominal data: if a person responds to a set of items, what prediction can be made from these responses about the criterion category to which this person belongs, judging from a similar set of people who have responded to the same set of items and for whom the criterion categories are known? According to the critique in the last section, the "major" patterns associated with each criterion category should be ascertained first. After this is accomplished, these major patterns are used for predicting.

The details of the method are developed in three sections: (1) the definition of an acceptable predictive pattern, (2) the steps in finding patterns in the data which meet the definition, and (3) the use of patterns in predicting.

Definition of an Acceptable Predictive Pattern

Zubin (1938) was one of the first to point out the usefulness of patterns in a set of items. Meehl, however, emphasized a unique role for patterns in predicting behavior. Meehl used the term configural to indicate a combination of items which predict to a criterion when the single items treated separately do not predict (Meehl, 1950). A contrived example of configural prediction is given in Table 1.

TABLE 1.--An example of configural prediction.

Observation	Item A	Item B	Criterion Category
1	1	1	1
2	2	2	1
3	1	2	2
4	2	1	2

Item A answered "1" is equally associated with both categories of the criterion; item A answered "2" is equally associated to both categories of the criterion. Likewise both responses to item B are equally associated to both criterion caregories. These are the linear relationships; item A alone and item B alone are obviously not helpful in predicting the criterion. The configural relationships, however, are helpful in predicting the criterion. Item A answered "1" together with item B answered "1" perfectly predict category 1 of the criterion. Also items A and B both answered "2" predict 2 of the criterion. Criterion category 2 is perfectly predicted by either item A "1" and item B "2", or A "2" and B "1".

Configural, then refers to a greater predictability by a pattern than by its unit parts treated separately. An extension of this concept which greatly enhances its usefulness in prediction is that a pattern can have greater predictive power than _any_ of its parts, including not only its unit parts, but also smaller configurations within the larger one. Such a pattern might be termed _hyper-configural_. These considerations lead to the following definition of an acceptable pattern:

> A pattern of responses to items A, B, . . . , R is _acceptable_ for Criterion Pattern Analysis if and only if the pattern in the items A, B, . . . , R is a better predictor of the criterion than is each of the subsets of the pattern in items A, B, . . . , R.

A pattern is a better predictor than a subpattern if the _level of discrimination_ of the pattern is greater than the _level of discrimination_ of the subpattern. The _level of discrimination_ of a pattern is the ratio of the number of times it occurs with a specified criterion category over the total number of times it occurs (irrespective of the criterion categories with which it occurs).

The following examples help clarify the application and significance of the definition. Consider four items, one, two, three, four, answered 1, 2, 2, 1, respectively. This pattern in the items can be represented 1(1) for item one answered one; 2(2) for item two answered two; 3(2) for three, two; and 4(1) for four, one. The pattern 1(1) 2(2) 3(2) 4(1) is acceptable if it predicts better than its fourteen subpatterns, given in Table 2, and

TABLE 2.--Subsets of the pattern 1(1) 2(2) 3(2) 4(1).

Subpattern	Discrimination Level for Criterion Category 1
1(1) 2(2) 3(2)	2/3
1(1) 2(2) 4(1)	2/3
1(1) 3(2) 4(1)	2/3
2(2) 3(2) 4(1)	2/3
1(1) 2(2)	2/4
1(1) 3(2)	2/4
1(1) 4(1)	3/5
2(2) 3(2)	3/5
2(2) 4(1)	2/4
3(2) 4(1)	2/4
1(1)	3/6
2(2)	3/6
3(2)	3/6
4(1)	3/6
Criterion Marginal	5/10

also if it predicts better than the marginal relative frequency of the criterion category. The marginal relative frequency of a criterion category is the frequency of that category over the frequency of all categories. The marginal relative frequency of the criterion category can be thought of as the discrimination level of a "subpattern" of zero items which is a subset of all patterns and against which they must be compared. In the special case of a "pattern" of one item, it is the only "subpattern" which is tested against.

To elaborate this illustration, let the pattern 1(1) 2(2) 3(2) 4(1) come from observations 4 and 5 of the data shown in Table 3. Also assume we are interested

TABLE 3.--Example data.

Observation	1	2	3	4	. . .	Criterion Category
1	2	2	2	2		1
2	1	1	1	1		1
3	2	1	1	2		1
4	1	2	2	1		1
5	1	2	2	1		1
6	1	2	2	2		2
7	2	2	2	1		2
8	1	2	1	1		2
9	1	1	2	1		2
10	2	1	1	2		2

in predicting criterion category one. Pattern 1(1) 2(2) 3(2) 4(1) occurs twice in association with criterion category one and not at all with category two, for a discrimination level of 2/2.

The subpatterns with their discrimination levels are shown in Table 2. In each case the discrimination level is less than 2/2. Furthermore, the marginal relative frequency of criterion category one is 5/10, which also is less than 2/2. Therefore, according to the definition, pattern 1(1) 2(2) 3(2) 4(1) satisfies our definition of an acceptable pattern.

On the other hand, none of the single-item patterns (i.e., 1(1), 2(2), 3(2), or 4(1)) is an acceptable pattern, since the discrimination levels of 3/6 are not more than the criterion marginal frequency 5/10 (see Table 2).

One consequence of the definition is that any pattern of responses that is unique is acceptable, unless a subset of the pattern is itself unique. This consequence leads to the objection that too many acceptable patterns emerge. Not only are all the unique patterns acceptable, but any pattern that has a higher discrimination level than do its subpatterns is thereby acceptable. This leads to the impossible task of recording thousands and millions of patterns which are acceptable for predicting to the criterion. Furthermore, if almost everything predicts, then very little is added by applying the method.

The difficulty can be ameliorated by stipulating a strict requirement for a pattern to be judged a better predictor than its subpatterns. Let N_1 be the frequency of occurrence of the pattern in association with the selected criterion category, and N_2 be the frequency of occurrence in association outside the criterion category. The terms n_1 and n_2 are the analogous frequencies for a subpattern. Then $\frac{N_1}{N_1+N_2}$ is the discrimination level of the pattern, and $\frac{n_1}{n_1+n_2}$ is the discrimination level of a subpattern. Previously a pattern was acceptable if $\frac{N_1}{N_1+N_2} > \frac{n_1}{n_1+n_2}$. Under the revised requirements a pattern is acceptable if for each subpattern:

$$\frac{\binom{n_1}{N_1}\binom{n_2}{N_2}}{\binom{n_1+n_2}{N_1+N_2}} + \frac{\binom{n_1}{N_1+1}\binom{n_2}{N_2-1}}{\binom{n_1+n_2}{N_1+N_2}} + \frac{\binom{n_1}{N_1+2}\binom{n_2}{N_2-2}}{\binom{n_1+n_2}{N_1+N_2}} + \ldots + \frac{\binom{n_1}{N_1+k}\binom{n_2}{0}}{\binom{n_1+n_2}{N_1+N_2}} \leq \alpha^{*}$$

where α is a preassigned positive number less than 1. This expression is the tail of the hypergeometric distribution,

---

*Actually the last term of this expression is either

$$\frac{\binom{n_1}{N_1+k}\binom{n_2}{0}}{\binom{n_1+n_2}{N_1+N_2}} \text{ or } \frac{\binom{n_1}{n_1}\binom{n_2}{k'}}{\binom{n_1+n_2}{N_1+N_2}},$$ whichever occurs first in the series.

and α is the proportion of the tail covered. The denominator of the first term, $\binom{n_1+n_2}{N_1+N_2}$, is the number of ways that the N_1+N_2 occurrences of the pattern can be chosen from the n_1+n_2 occurrences of the subpattern. The expression $\binom{n_1}{N_1}$ in the numerator is the number of ways that the N_1 occurrences of the pattern which are in the criterion can be chosen from the n_1 occurrences of the subpattern which are in the criterion. Similarly $\binom{n_2}{N_2}$ is the number of ways that the N_2 occurrences of the pattern which are outside the criterion can be chosen from the n_2 occurrences of the subpattern outside the criterion.

The complete term $\dfrac{\binom{n_1}{N_1}\binom{n_2}{N_2}}{\binom{n_1+n_2}{N_1+N_2}}$ is the probability of having N_1 and N_2 occurrences when choosing N_1+N_2 occurrences which fall into two groups of n_1 and n_2. The remaining terms are probabilities for less likely events, so that the whole expression is the probability of having N_1 occurrences or more and N_2 occurrences or less when choosing N_1+N_2 occurrences out of n_1+n_2 occurrences which fall into two groups of n_1 and n_2.

In other words the first term of the expression is the probability of the observed occurrences of the pattern among the criterion categories, given the occurrences of its subpattern among the criterion categories. The whole expression is the probability of having a pattern occur in the category under scrutiny with a frequency as great or greater than that observed, given the occurrences of its subpattern.

An approximation to the whole expression can be had by computing chi square for the following 2x2 table:

	Subpattern Minus Pattern	Total Pattern	
Within the Criterion Category	n_1-N_1	N_1	n_1
Outside the Criterion Category	n_2-N_2	N_2	n_2
	n_1+n_2 $-N_1-N_2$	N_1+N_2	n_1+n_2

The meaning of the above requirement in the analysis of data is that for every subpattern a test is made as to whether or not the pattern significantly improves the prediction of the criterion. The reasoning is that if it does not improve prediction at the level specified, then the subpattern itself might as well be used. The level of significance for each test is set at α.

Thus the hypergeometric distribution serves as a decision function for limiting the otherwise overwhelming number of acceptable patterns. Clearly the smaller α is, the fewer patterns will be accepted, and α can be set so that no patterns will emerge at all. On the other hand, for α close to 1.0, the same situation obtains as with the requirement that the pattern be just better than its subpatterns: $\frac{N_1}{N_1+N_2} > \frac{n_1}{n_1+n_2}$.

The value α does not represent the significance of a pattern in relation to the criterion. The level of significance has not been determined, and is not required for successful use of this method.

Presently, in the analysis of data, the setting of α is done by trial and error, small enough to preclude a flood of patterns and large enough to admit the cream of the acceptable patterns.

Finding Acceptable Patterns in Data

In the previous section the definition of an acceptable pattern was developed. In this section the problem of finding all acceptable patterns in a set of data is discussed.

It follows from the critique in the first chapter that all possible patterns must be considered. Even with a high speed electronic computer the job of generating each possible pattern for acceptance or rejection is overwhelming. For example, consider a very small problem of ten dichotomous items. There is a total* of 69048 different patterns to be generated. This total increases very rapidly with the number of items. With fifteen dichotomous items where are over fourteen million patterns to be generated and checked. With twenty dichotomous items there are over five billion patterns. Clearly the need to reduce the number of patterns actually handled is imperative. Previously described methods have attempted to cope with the task by placing severe restrictions on the patterns considered, thereby reducing effectiveness and leading to a demand for large samples. Criterion Pattern Analysis solves this problem by considering all possible patterns without generating and examining them one by one. A computational scheme which allows this to be accomplished will be developed and carefully programmed to make most efficient use of the computer's capabilities. In the paragraphs that follow, the procedure for finding all acceptable patterns is described in detail and related to the problem of considering all possible patterns. In discussing the solution of this problem, two closely interrelated aspects

---

*The formula for computing the total number of possible patterns for N dichotomous items is:

$$\sum_{i=1}^{N} \binom{N}{i} 2^{i}$$

are considered: (1) the order of examining patterns; (2) the judging of each pattern as it is brought up.

Order of Examining Patterns

According to the critique at the end of the first chapter, acceptable predictive patterns are sought separately for each criterion category. Within each criterion category there are many alternative approaches. The one chosen here is to first find all acceptable one-item patterns, then all acceptable two-item patterns, then three, four, and so on. As will be shown below, this order of proceeding allows reduction in the number of patterns examined.

Judging the Patterns

Two judgments are made for every pattern. The first is whether the pattern is acceptable under the definition. The second judgment is whether prediction can possibly be improved by annexing another item to the pattern.

According to the strict definition, a pattern is acceptable if it predicts better than any of its subpatterns, where better is determined by the preassigned α and the hypergeometric distribution. A pattern of r items has r subpatterns of r-1 items, $\frac{r(r-1)}{2}$ subpatterns of r-2 items, $\frac{r(r-1)(r-2)}{6}$ subpatterns of r-3 items, etc. Since the number of tests to be made becomes quite large as r increases, testing all subpatterns would be too laborious.

The problem is solved here by testing only the pattern's immediate subpatterns; a pattern of r items is accepted if it predicts better than its subpatterns of r-1 items. For example, pattern 1(2) 3(1) 7(2) is accepted if it improves the prediction of all of the subpatterns of two items, i.e., 1(2) 3(1), 1(2) 7(2), and 3(1) 7(2).

Occasionally patterns can be accepted which are not better predictors than some of their remote subpatterns, and also not better than predicting from the marginals of the criterion. This is more than compensated for by the fact that it does include all of the patterns which enhance prediction, and does this in a reasonable fashion in terms of the amount of analysis required.*

The second judgment made on each pattern, whether or not it has been previously judged acceptable, is whether prediction can possibly be improved by annexing

---

*Testing only whether a pattern predicts better than the marginals of the criterion category was also tried at one point in the development of the Criterion Pattern Analysis method. The results were a multitude of "acceptable" patterns, most of which contained a subpattern which was a very good though not a perfect predictor. Almost any item affixed to this subpattern would have produced a pattern which was also a very good predictor. This kind of result not only contradicts the strict definition of an acceptable pattern, but also produced too many patterns, all very similar.

another item to the pattern. Only if a pattern can be improved will it be used in the formation of larger patterns. Improvement is measured by α and the hypergeometric distribution in the following manner. As before, $\frac{N_1}{N_1+N_2}$ is the discrimination level of a pattern for the criterion category. Annexing another item will make its greatest improvement when it results in a discrimination level of $\frac{N_1}{N_1}$. Using the hypergeometric formula, $\frac{\binom{N_1}{N_1}\binom{N_2}{0}}{\binom{N_1+N_2}{N_1}}$ is computed. If less than α, then the pattern can be improved; if greater than α, then no possible improvement can be made, and the pattern is rejected from further consideration.

When a pattern is so rejected, automatically a class of many patterns is rejected. The patterns in the class are those which contain the original pattern as a subpattern. This effects a vast reduction in the number of patterns actually handled. As an example, suppose we are searching for patterns associated with category one of the criterion, and suppose we find that pattern 3(1) 7(2) never occurs in criterion category one. Not only can we reject pattern 3(1) 7(2), but we can also reject at the same time all those patterns of three or more items in which both 3(1) and 7(2) occur (for example, 1(2) 3(1) 7(2), or 2(1) 3(1) 4(2) 5(1) 7(2)). See Table 4.

TABLE 4.--Pattern acceptance and rejection.

	Occurrence of a Pattern in the Criterion Category					
	Not at All	Few Times	Often	Many	Very Often	Exclusively
Individual Pattern	reject	reject	reject	accept	accept	accept
Large Class of Patterns Containing the Smaller Individual Pattern	reject	reject	not reject	not reject	reject	reject

If pattern 3(1) 7(2) does occur in criterion category one, but only a very small number of times, then again it can be rejected, and along with it all those patterns in which both 3(1) and 7(2) appear. If 3(1) 7(2) occurs often in category one, it may still be rejected, but larger patterns containing it cannot be rejected now. Certainly if 3(1) 7(2) occurs many times and most often in category one, it is likely to meet the requirements for acceptance, and also the larger patterns not be rejected. However, if it occurs exclusively or almost exclusively in category one, then the larger patterns which include it must be rejected, for there is no way to improve prediction under the assigned α.

If a class of patterns is not rejected, it is not thereby accepted; it is just not rejected. This means that an individual pattern can still become a part of a pattern of more items which stands a chance of being accepted. In implementing this condition, the individual patterns are saved and actually used to form trial patterns with more items. In general, patterns with r items which have been saved are used in combination to form patterns of r+1 items. For example, to form the pattern 3(1) 7(2) 8(1) the list of previously saved patterns must include 3(1) 7(2); 3(1) 8(1); and 7(2) 8(1). If any one of these patterns is missing, the new pattern should not be formed. And incidentally, even if it were formed it could not satisfy the test of improving prediction over its subpatterns, since the missing subpattern has been omitted from the saved list because it could in no way be improved by the addition of other items.

The Computational Scheme

Following the procedure above, the method is implemented as shown in Figure 1. Each criterion category is considered separately, starting with the first. Then in turn, one-item patterns, two-item patterns, three-item patterns, etc., are generated and tested as outlined above, to determine whether they exist in the data, whether to accept them as predictive of the criterion category, and whether they should be saved or not.

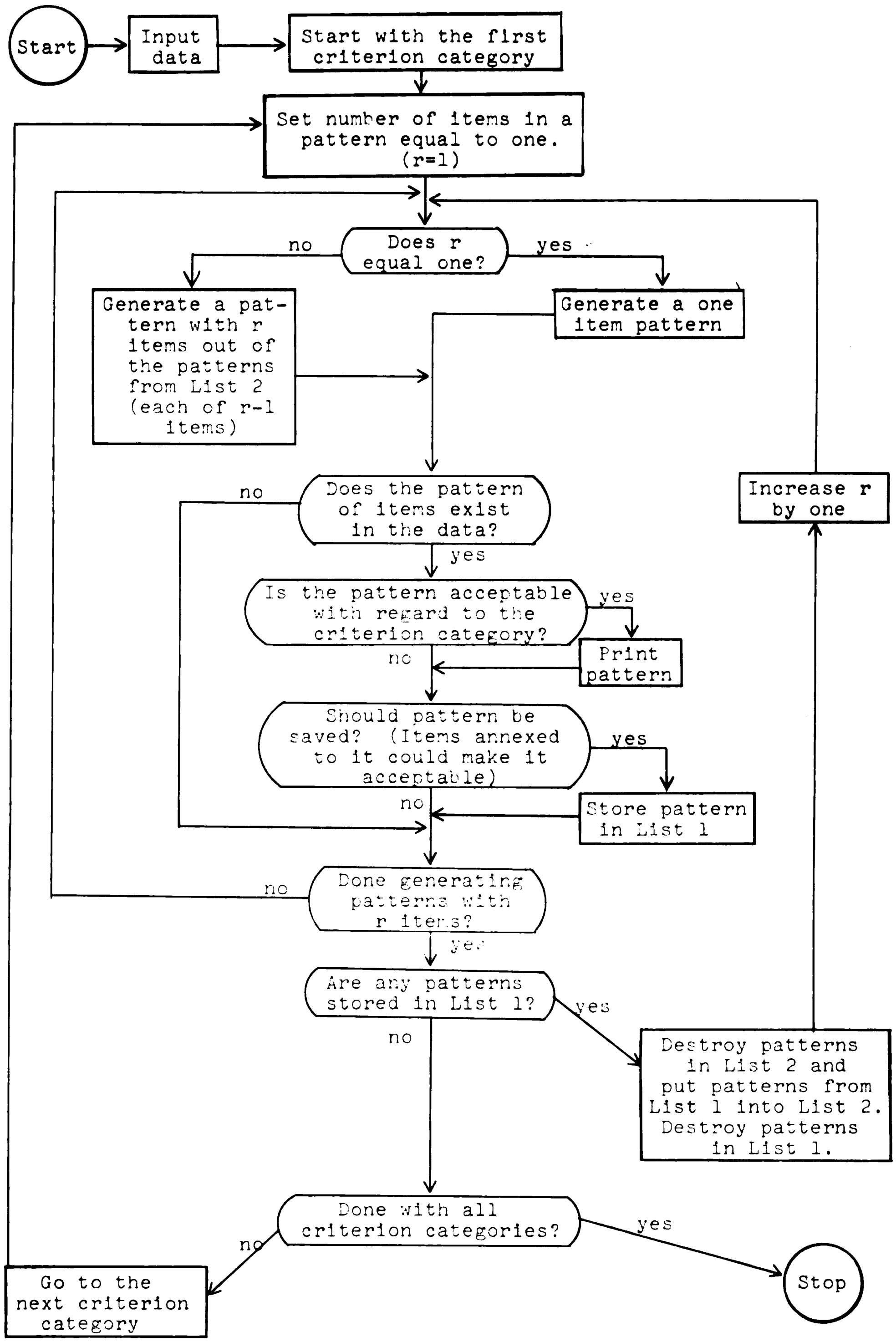

FIGURE 1

A FLOW CHART FOR PERFORMING CRITERION PATTERN ANALYSIS

One-item patterns are generated without consulting the "saved" list: first 1(1) (item 1 scored 1), then 1(2), 1(3), 1(4), as far as the response categories extend. Next 2(1), 2(2), 2(3), 2(4),...; then 3(1), 3(2), 3(3), 3(4),...; and so on for all items and response categories. Each of these one-item patterns is tested as to whether or not it exists in the data. If it does, it is tested for acceptance. If accepted, it is recorded as a predictive pattern. Next, the same pattern is checked as to whether it should be saved or not. If it is saved it is stacked in a list called LIST 1. When all one-item patterns have been generated and checked, LIST 1 contains saved one-item patterns.

Two-item patterns are considered; but before doing so, the patterns in LIST 1 are transferred into another list, LIST 2, making sure that nothing is in LIST 2 beforehand, and that nothing remains in LIST 1 after the transfer.

Two-item patterns are generated out of the one-item patterns found in LIST 2. Again the various tests are made, acceptable patterns recorded, and two-item patterns to be saved are stacked into LIST 1. When all two-item patterns have been generated from LIST 2 and tested, the contents of LIST 1 are again transferred into LIST 2 and three-item patterns are considered. This process continues until LIST 1 contains nothing to be transferred to LIST 2. At this point the next criterion category is considered, starting over again with one-item patterns. When all criterion categories have been examined for acceptable patterns, the process is completed.

Help in reducing computational time is obtained by adding the stipulation that no pattern be stored in LIST 1 unless its frequency of occurrence in the criterion category is larger than some preassigned constant. The requirement prevents loading LIST 1 with patterns that account for only a small percentage of people in the criterion category. Patterns with frequencies in the criterion smaller than the constant can still show themselves to be acceptable if their subpatterns have frequencies in the criteria larger than the constant. This somewhat artificial procedure is particularly useful when attacking data with a large number of observations. All the acceptable patterns will not emerge, but the ones that do will, in general, be those which occur most often. Of course, for small problems the constant can be set to one.

Computer Implementation

The foregoing procedure was intended and conceptualized for a high speed computer. In order to facilitate consideration of all possible patterns in a reasonable amount of time, the computational scheme was carefully programmed to make most efficient use of the computer's capabilities. It might be informative to indicate some of the principles utilized in developing the algorithm as a computer problem.

The task of finding predictive patterns is of course an enormous one. If one were to try each possible pattern in turn, the task would be virtually impossible: if a thousand patterns could possibly be generated and checked in one second, it would still take over two months of computer time to complete the job of analyzing the more than five billion patterns associated with twenty dichotomous items.

While generating and checking all possible patterns has the disadvantage of taking too much time, it does have the feature of requiring very little of the computer's memory space. By appropriately changing the method so that more memory is used, the amount of time required can be decreased. This was ultimately accomplished by using LIST 1 and LIST 2 of saved patterns. Although the processing of lists takes more time per pattern than does simple generating and checking, fewer patterns are actually processed (see page 33), and the net result is a saving in computation time. Hence, the first principle applied to the problem of checking all patterns is the reciprocity of computer time and space. If a problem takes too much computer time, it may be possible to reduce the time by using more of the computer's memory.

Conversely, when a problem overflows the available memory, it may still be handled by reprogramming using more computation time.

One way of extending the computer's memory capacity is to use "packed storage." This amounts to storing many numbers in a computer word (memory location), where only one number would be stored ordinarily. This is possible where many small integers are to be processed. It is not readily applicable with large integers or fractions. This procedure was applied to the present problem and helped reduce computation time. In addition it was possible to operate on all the numbers packed in a single computer word at once instead of separately. This was done when counting the occurrences of a given pattern throughout the observations. In counting the occurrence of a four-item pattern over two hundred people, only twenty computer words were used in the computation instead of $4 \times 200 = 800$ words. In this way computation time was grossly reduced. Hence, the second principle applied might be termed "aggregate operation." When a computer operation is performed on one computer word, it becomes tantamount to several operations on several computer words.

To expedite the principle of "aggregate operation" it is sometimes necessary and always convenient to use machine or assembly language operation codes. This means that the basic operations built into the computer are directly selected by the programmer for his program. Such basic operation codes differ from one kind of computer to another and require a new program to be written for each computer. By using a compiler based language such as FORTRAN, ALGOL, or COBOL, it would be possible to write programs easily transferable to a different computer. However, these languages result in a program which is not as efficient with respect to computer time as one written in a basic language (although they are usually very efficient in terms of the time required to write a program). Much of the method in this thesis was programmed in a basic machine language with a saving of computation time and computer memory. Thus the third principle applied was keeping to basic operation codes, especially in those parts of the program which would be repeated many times.

These three programming principles, the reciprocity of computer time and space, aggregate operation, and efficient basic operation codes, were employed in solving the problem of identifying all acceptable predictive patterns. Without these principles the method in this thesis could not have been developed into a practical tool.

Predicting from Patterns

Once the major predictive patterns have been found they can be used for predicting the criterion category of a person for whom only the responses to the predictive instrument are known.

Direct Prediction Method

The method of predicting directly from patterns starts by checking the new person's responses for the patterns which were previously determined. If no patterns are found for this person, no prediction can be made. If one pattern is found, the prediction is made for the criterion group with which that pattern is associated. If many patterns are found and all are associated with the same criterion group, again that group is predicted. A problem arises when patterns are found, some associated with one criterion group and some with another.

This problem is solved as follows: each pattern, as it is extracted by the method of Criterion Pattern Analysis, has associated with it a fraction, $\frac{N_1}{N_1+N_2}$, termed the level of discrimination. The denominator of the fraction is the total number of people from the original data who have that pattern; the numerator is the number who have that pattern and who are in the criterion category with which that pattern is associated. The level of discrimination multiplied by 100 gives the percentage of people having that pattern who are in the criterion category.

These considerations are now applied to the new person who is found to have patterns in more than one criterion category. For each criterion category, the one pattern is selected which has the highest level of discrimination. Then with one pattern per criterion category, the prediction is made to that criterion category with the highest discrimination level. For example, suppose the six patterns given in Table 5 were found for

TABLE 5.--An example of patterns to be used in prediction.

Observed Pattern	Level of Discrimination	Criterion Category
3(2) 4(2)	.63	1
2(1) 3(2) 6(1)	1.00	1
5(2) 7(1)	.95	2
1(2)	.65	3
3(2) 8(2)	.78	3
1(2) 8(2) 5(2)	.80	3

a person: two patterns associated with criterion one, one pattern associated with criterion two, and three patterns associated with criterion three. On the basis of pattern 2(1) 3(2) 6(1), with a discrimination level of 1.00, the prediction is made to criterion one.

It is advisable in practical situations to consider the highest predictions made for each criterion category: in the former example, 1.00 for criterion one, .95 for criterion two, and .80 for criterion three (see Table 5). If action on the basis of the prediction to criterion category one is precluded, the next highest prediction, to criterion category two, might be chosen by the user of this method. In other words, this method is not an imperative for choosing one criterion category over another in an applied situation. By providing a listing of alternative predictions and their relative levels, Criterion Pattern Analysis uniquely provides additional information that can be relevant for practical use.

The procedure of predicting the criterion category whose pattern has the highest discrimination level may again lead to no prediction. This would occur, for example, if all criterion categories have the same discrimination levels associated with their best patterns. No prediction is, in one sense, a kind of information. And no prediction because patterns have the same level of discrimination is different information from no prediction because there were no patterns at all.

Combination Prediction Methods

While one method of predicting is using the patterns directly, as was done above, other methods can be developed.

Viewing Criterion Pattern Analysis as an extension of item analysis suggests treating every pattern as an item, and scoring people as having or not having each pattern. Every person may then be redefined by a new set of items, each item corresponding to one of the patterns. In this way any configural information present in a pattern is accounted for through a single score. Consequently linear methods, such as multiple regression, can be applied to this new set of data. Configural acumen will thereby be combined with the mathematical strength of linear procedures.

CHAPTER III

DATA AND RESULTS

Data

In this chapter two sets of data are used. The first set is from Crego* (1966). A sample of 99 college-age women responded to 23 items of the I-E scale developed by Rotter (1966) (see Appendix A), and to the Hidden Figures Test (Test CF from the Educational Testing Service Battery).

The I-E scale measures whether the subject believes the locus of control of reinforcement is in an external or an internal site. In each item the subject selects one of two alternative statements. The usual score is the total number of external responses. However, for our purpose here, the items will be considered as a set of predictor variables.

The Hidden Figures Test measures field dependence-independence (Witkin *et al.*, 1962). Thirty-two complex

---

*The writer wishes to express appreciation to Dr. Clyde Crego for permission to use his data and for his encouragement in applying configural methods to it.

patterns are presented and the subject is to determine which one of several simple designs is present in each complex pattern. The total number of identified embedded figures is the score given, and indicates the degree of field independence. On the present sample the scores ranged from 2 to 25 with a median of 11.2. The subjects were given a score of 1 or 2 according to whether they scored above or below the median. These scores yielded the criterion categories to be predicted by the items of the I-E scale.

All subjects were divided into two samples by a table of random numbers. The first sample of 50 subjects is the analysis sample, on which all methods are initially applied. The second sample of 49 subjects is the cross validation sample.

The second set of data is from the roll-call voting record of the seventeenth session of the United Nations General Assembly (United Nations, 1964a, 1964b).* A total of 110 nations voted on 44 issues (see Appendix B). Where some nations were not yet admitted on the first few votes, or where the data were incomplete in minor ways, the procedure recommended by Wrigley (Olin, 1964) for estimating missing votes was followed. Then responses

---

*The writer wishes to thank Dr. Charles Wrigley for his generosity in welcoming use of data he has assembled.

to each issue were dichotomized into "yes" votes versus all others, including "no," "abstain," etc.

One of the issues was selected to be the criterion item. Selection was based on two considerations: one, a vote near the end of the session, and two, a vote having a low correlation with other votes, taken one at a time. The vote on issue 38 fulfilled both of these requirements, being the sixth vote from the last and having the lowest average correlation with other votes.

The 110 nations were then divided into two samples using a table of random numbers, as before. The first group of 55 nations is the analysis sample; the second group of 55 nations is the cross validation sample.

Linear Methods

In addition to Criterion Pattern Analysis, two linear methods were applied for comparison. They are multiple regression and a multivariate-normal maximum likelihood procedure.

Multiple regression is a well-known statistical procedure (Walker and Lev, 1950, Chapter 13), which works as follows: each of N people has a score on each of r items; let X_1, X_2, ..., X_r be a set of scores for one person. Also, all N people have a score on another item; let Y be the score for any one person. The purpose is to predict Y from X_1, X_2, ..., X_r for every person. Suppose for each person we find a weighted sum

W of the X's: $W = b_0+b_1X_1+b_2X_2+ \ldots +b_rX_r$. Now for each person there are two scores: Y and W. The correlation between Y and W can be found across all people. The method of multiple regression selects the weights b_0, b_1, b_2, ..., b_r so that the correlation between Y and W is maximized. This correlation is called the coefficient of multiple correlation; it estimates how well Y can be predicted from the X's on the sample of N people. The weights b_0, b_1, b_2, ..., b_r are called regression weights, or if the X's are first expressed in terms of standard scores, the weights are called beta weights.

A related method called discriminant function analysis (Tatsuoka and Tiedeman, 1957) also finds a set of weights b_0, b_1, b_2, ..., b_r for the X's with which to form a sum W for each person. With discriminant function analysis there is no criterion variable Y; instead each person has previously been classified into one of several groups. The weights b_0, b_1, b_2, ..., b_r are selected so that the mean W score of each group is most different from one group to another. In general, there is more than one set of weights found by discriminant function analysis. In the usual case if there are K different groups of people, then there will be K-1 sets of weights. When there are two groups, then there is just one set of weights and discriminant function analysis gives the same results as does multiple regression with a dichotomous criterion variable (Welch, 1939). Hence for purposes of this thesis, only the results of multiple regression will be presented.

A second, entirely different, method applied to the data is a maximum likelihood procedure using the multivariate-normal density function (Cooley and Lohnes, 1962, Chapter 7). This procedure, while not yet widely applied, offers an alternative to discriminant function analysis. Again N people have responded to r items, $X_1, X_2, \ldots, X_r$. Each person has previously been classified into one of several groups. For each group the parameters for the multivariate-normal density function are calculated. The density is then calculated for each person using his responses $X_1, X_2, \ldots, X_r$. The higher the density, the closer that person is to being at the center of the distribution for that group.

This method is applied by computing for each person his density for each group: $p_{G_1}, p_{G_2}, p_{G_3}, \ldots, p_{G_k}$. The person is predicted to be in that group for which the density is highest.

Results from Crego Data

Criterion Pattern Analysis

The analysis sample of the Crego data was subjected to Criterion Pattern Analysis: thirteen patterns were found using α = .05. Table 6 lists these patterns: eight for criterion group one, field dependence, and five for criterion group two, field independence. The first pattern in Table 6, 8(2), means item 8 answered with the second response alternative; 2(1) 17(2) means item 2 answered with the first alternative along with item 17 answered with the second alternative. Also listed in Table 6 are the total number of people having each pattern, the number of people in the criterion group having the pattern, and the discrimination level for the pattern.

These patterns were used to predict for the analysis sample. The results are shown in Table 7. The phi of .663 shows the correlation between the actual and the predicted groups.

The same patterns were then used to predict in the cross validation sample. Table 8 shows these results. The phi dropped to .230 and the chi square of 2.58 has a probability of .12.

Interpretation of Patterns

In criterion group one, field dependence, all 26 people have at least one of the eight patterns extracted. In criterion group two, field independence, only 17 of the 24 people have at least one of the five patterns. This may mean more heterogeneity among field independent people. Looking at the patterns for group one, we find patterns 20(2) 23(2) and 17(2) 20(1) 21(2). These define two exclusive kinds of field dependent people (since no one person responds both ways to item 20). In criterion group two, pattern 20(1) 21(1) defines one set of people, and pattern 6(1) 8(1) defines another set of people.

TABLE 6.--Results of Criterion Pattern Analysis on the analysis sample of the Crego data.

Pattern	Total Number of People Having the Pattern	Number of People in the Criterion Having the Pattern	Discrimination
Patterns Associated with Field Dependence			
8(2)	30	20	.67
2(1) 17(2)	28	20	.71
10(1) 16(1)	5	5	1.00
20(2) 23(2)	5	5	1.00
1(1) 8(2) 9(2)	15	13	.87
17(2) 20(1) 21(2)	10	9	.90
11(2) 12(2) 17(2) 21(2)	10	10	1.00
13(1) 14(2) 18(2) 19(1)	7	7	1.00
Patterns Associated with Field Independence			
2(2)	16	11	.69
8(1)	20	14	.70
2(2) 4(1)	10	9	.90
6(1) 8(1)	8	8	1.00
20(1) 21(1)	8	7	.88

TABLE 7.--Results of predicting to the Crego analysis sample from patterns from Criterion Pattern Analysis.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	25	8	33
Field Independent	1	16	17
	26	24	50

$\phi = .663$

TABLE 8.--Results of predicting to the Crego cross validation sample from patterns.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	19	12	31
Field Independent	5	10	15
No Prediction	1	2	3
	25	24	49

$\phi = .230$*

$\chi^2 = 2.58$

$p < .12$

*The ϕ and χ^2 were computed on the following table:

19	13	32
6	11	17
25	24	49

> In contrast to the patterns described above for group one, these two sets of people are not mutually exclusive, since a few people belong to both sets. An attempt to find the meaning of these patterns from the item responses leads to the conclusion that no obvious underlying concept is responsible for these types. The external-internal dimension built into the test items does not help in interpretation, since both kinds of responses occur in most patterns. Additional testing or information on the people in these types might reveal what is being responded to in the items.

These results are now compared with multiple regression and the maximum likelihood procedure.

Multiple Regression

The analysis sample of the Crego data was subjected to multiple regression. The resulting regression coefficients and their significances are shown in Table 9. The levels of significance are the result of testing whether the regression coefficients are different from zero (Walker and Lev, 1953, pp. 337-339). Also shown in Table 9 are the correlations between the items and the criterion variable. The multiple regression yielded a multiple correlation coefficient of .697.

The prediction of the criterion groups using regression coefficients is shown in Table 10 for the analysis sample, resulting in a phi of .639.

In Table 11 the results are shown for predicting to the cross validation sample using the regression coefficients. The phi of .183 is less than that produced through predicting directly from patterns. However the

TABLE 9.--Results of multiple regression on the analysis sample of the Crego data.

Item	Regression Coefficient	Significance of Regression Coefficient	Correlation of the Item with the Criterion
1	.16	.52	.07
2	.25	.22	.28*
3	.01	.91	-.11
4	-.02	.88	-.21
5	-.25	.26	-.07
6	.01	.91	-.11
7	.39	.06	.20
8	-.19	.42	-.36**
9	-.04	.82	-.07
10	-.16	.44	-.07
11	.04	.81	-.03
12	-.40	.13	-.16
13	-.14	.59	-.08
14	-.02	.90	.00
15	-.04	.85	.02
16	.05	.80	.13
17	-.31	.24	-.22
18	.23	.22	.21
19	-.12	.62	.17
20	-.18	.49	-.04
21	-.17	.37	-.16
22	-.04	.81	-.02
23	-.42	.13	-.16

*Significant at the .05 level.

**Significant at the .01 level.

TABLE 10.--Results of predicting to the Crego analysis sample from regression coefficients.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	22	5	27
Field Independent	4	19	23
	26	24	50

$\phi = .639$

TABLE 11.--Results of predicting to the Crego cross validation sample from regression coefficients.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	16	11	27
Field Independent	9	13	22
	25	24	49

$\phi = .183$

$\chi^2 = 1.63$

$p < .20$

difference in prediction between the two methods is not significant ($p < .42$), using a test proposed by Lubin (1950, p. 104).*

Interpretation

Item 7 has the most significant regression weight ($p < .07$) and hence makes the highest independent contribution to the prediction. Items 2 and 8 are the most highly correlated with the criterion, being the only items significant at the 5% level or less.

Although having the most significant regression weight, item 7 never appears in the patterns from Criterion Pattern Analysis. Items 2 and 8 both appear as single-item patterns; as such they would have the same interpretation. As for item 7, this is one of several items relating to the subject's belief about fate and luck. It is not clear why this one item behaves more independently in predicting the criterion.

---

*The statistical test is the difference between correlated proportions, and was incorrectly given by Lubin. If P_1 is the proportion of correct classifications by method 1, and P_2 is the proportion of ccrrect classifications by method 2, then the difference P_1-P_2 can be tested by:

$$Z = \frac{(P_1-P_2)\, N^{\frac{1}{2}}}{\sqrt{P_1Q_1+P_2Q_2-2r_{12}(P_1Q_1P_2Q_2)^{\frac{1}{2}}}}$$

where N is the total number of people classified; r_{12} is the correlation between method 1 and method 2 on their correct and incorrect classifications. When N is large, Z is normally distributed.

Maximum Likelihood Procedure

Multivariate-normal distribution parameters were estimated for the two criterion groups in the analysis sample of the Crego data. Predicting to the criterion produced the results shown in Table 12 for the analysis sample, and in Table 13 for the cross validation sample. While there was perfect prediction of the criterion groups in the analysis sample, prediction fell to near zero in the cross validation sample. The difference between this result and that of predicting directly from patterns is significant at only $p < .15$ using a one-tailed test.

Interpretation

> The interpretation of the parameters computed is essentially the same as in the case of multiple regression, for the following reasons: for each criterion group the mean score for each item has been calculated. The greater the differences between means, the more important is the item for indicating differences between the groups. The same information is contained in the correlation coefficient between the item and the criterion (for which see Table 9).

Resumé

In all, the prediction of the Crego data was not very successful upon cross validation under these methods. However, using patterns directly for prediction produced slightly better results than did using multiple regression and considerably better results than did using the maximum likelihood method.

TABLE 12.--Results of predicting to the Crego analysis sample using maximum likelihood.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	26	0	26
Field Independent	0	24	24
	26	24	50

$\phi = 1.00$

TABLE 13.--Results of predicting to the Crego cross validation sample using maximum likelihood.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	10	9	19
Field Independent	15	15	30
	25	24	49

$\phi = .026$

$\chi^2 = .032$

$p < .90$

<u>Combination Method</u>

The method for predicting which combines patterns and linear procedures (described at the end of Chapter II) was applied to the Crego data. The patterns previously extracted from the analysis sample were used to score every person from both the analysis and the cross validation samples. A person was given scores of 2 for patterns which he did have, and scores of 1 for patterns which he did not. Of the thirteen patterns extracted, only 8(2) of criterion category one was not used since it is a duplication (except for direction) of pattern 8(1) of criterion category two. The regression coefficient corresponding to each pattern is shown in Table 14 along with the significance level of the coefficient and the pattern's correlation with the criterion. The coefficient of multiple correlation is .780.

Using the regression coefficients to predict to the analysis sample produced the results shown in Table 15. The correlation between the actual group and the predicted group is .726. Using the same regression coefficients to predict to the cross validation sample produced the results shown in Table 16. The phi of .309 with an associated chi square of 4.69 is significant at the .05 level. This phi is higher than that obtained when predicting directly from patterns alone (phi = .230) and when predicting from multiple regression of the original items (phi = .183).

TABLE 14.--Results of multiple regression applied to the patterns in the analysis sample of the Crego data.

Pattern	Regression Coefficient	Significance of the Regression Coefficient	Correlation of the Pattern with the Criterion
2(1) 17(2)	-.30	.11	-.44**
10(1) 16(1)	-.08	.68	-.32*
20(2) 23(2)	-.18	.43	-.32*
1(1) 8(2) 9(2)	-.27	.11	-.45**
17(2) 20(1) 21(2)	-.17	.36	-.38**
11(2) 12(2) 17(2) 21(2)	-.23	.20	-.48**
13(1) 14(2) 18(1) 19(1)	-.31	.10	-.39
2(2)	-.12	.62	.28*
8(1)	-.16	.29	.36**
2(2) 4(1)	.13	.58	.42**
6(1) 8(1)	.29	.12	.45**
20(1) 21(1)	.17	.29	.35*

*Significant at the .05 level.

**Significant at the .01 level.

TABLE 15.--Results of predicting to the analysis sample of the Crego data using regression coefficients from patterns.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	21	2	23
Field Independent	5	22	27
	26	24	50

$\phi = .726$

TABLE 16.--Results of predicting to the Crego cross validation sample using regression coefficients from patterns.

Predicted Group	Actual Group: Field Dependent	Actual Group: Field Independent	
Field Dependent	19	11	30
Field Independent	6	13	19
	25	24	49

$\phi = .309$

$\chi^2 = 4.69$

$p < .05$

Interpretation

Pattern 13(1) 14(2) 18(1) 19(1) has the most significant regression weight and hence makes the highest independent contribution to the prediction. Pattern 11(2) 12(2) 17(2) 21(2) has the highest correlation with the criterion. Although these two patterns are the largest, there seems to be no real relationship between number of items in the pattern and its predictive power. There are smaller patterns with similar predictive values.

It may be noted that the significances of the regression coefficients here appear smaller (i.e., more significant) than the corresponding coefficients from multiple regression on the original items. Similarly, the correlation of the patterns with the criterion are higher than are those for the original items. Of course, this is not surprising since the patterns were chosen for their high association with the criterion.

The interpretation of the regression coefficients here poses the same problem as with the regression coefficients from the original items. Fortunately since the patterns themselves are all highly correlated with the criterion, they can be used as the basis of interpretation, as was done when the patterns were used directly in the prediction.

Resumé

Using this combination of pattern and linear methods led to a successful prediction of the cross validation sample ($p < .05$). Therefore this method seems better than the previously applied procedures. However, in predicting to the cross validation sample the difference between multiple regression on patterns and multiple regression on the original items is significant at only $p < .25$.

Results from UN Data

Criterion Pattern Analysis

The analysis sample of the UN data was subjected to Criterion Pattern Analysis, with the results shown in Table 17. There are eight patterns relating to "no" on vote 38 and twelve patterns to "yes" on vote 38 using an assigned α of .05.

Using these patterns to predict back to the analysis sample produced the results shown in Table 18. The phi coefficient between actual and predicted vote is .631. When the patterns were used to predict to the cross validation sample, the results shown in Table 19 were produced. The phi dropped to .374, and the associated chi square of 7.72 shows that this correlation is very significant ($p < .006$).

Interpretation of Patterns

> The criterion, vote 38, involved offering technical assistance for national projects of population study. Inspection of the nations whose patterns characterize each of the two criterion groups (see Appendix C) reveals that this issue does not divide the nations into a communist-non-communist dichotomy. Patterns predicting to vote "no" include two types of nations: in the first is the USSR and its close satelites and some Asian and African nations (q.v. nations defined by pattern 25(1)); in the second group is the USA and some Latin American, Asian, and African nations (q.v. nations defined by pattern 4(1)).
>
> Those patterns predictive of a "yes" vote on the criterion issue do not divide the data into such well-marked groups. Perhaps nations in this criterion group follow policies independent of the USA or the USSR. Item 19(2), which is a vote

TABLE 17.--Results of criterion pattern analysis on the analysis sample of the UN data.

Pattern	Nations	Total Number of Nations Having the Pattern	Number of Nations in the Criterion Having the Pattern	Discrimination
	Patterns Associated with NO on Vote 38			
4(1)	1 2 4 5 6 7 8 9 10 11 12 13 14 16 19 21 23 27 28 30 32 35 39 43 44 45 46 47 48 50 52	32	26	.8125
19(1)	3 20 22 24 25 26 28 36 45 46 43	11	11	1.0000
25(1)	3 14 16 20 22 24 25 26 33 36 41 45	12	12	1.0000
37(2)	1 4 6 9 10 12 13 14 16 17 19 23 28 43 44 46 47 50	18	17	.9444
3(2) 9(1)	3 6 20 22 24 25 26 36 43 45 51	11	11	1.0000
5(2) 9(1)	3 6 20 22 24 25 26 36 43 45 51	11	11	1.0000
20(1) 39(2)	4 8 9 12 14 27 28 33 34 38 39 41 45 46 48 50 51 53	18	17	.9444
29(1) 41(2)	4 8 12 14 27 33 34 38 44 45 46 51	12	12	1.0000
	Patterns Associated with YES on Vote 38			
4(2) 19(2)	17 18 29 31 33 34 37 38 40 41 42 49 51 54 55	15	10	.6667
4(2) 20(2)	17 18 31 54	4	4	1.0000
4(2) 25(2)	17 18 29 31 34 37 38 40 43 49 51 53 54 55	14	10	.7143
20(2) 37(1)	2 5 6 7 11 15 18 21 31 32 35 54	12	8	.6667
1(2) 9(2) 39(1)	29 30 37 40 42 49	6	5	.8333
4(2) 9(2) 39(1)	29 37 40 42 49	5	5	1.0000
4(2) 19(2) 39(1)	29 37 40 42 49 55	6	6	1.0000
9(2) 21(1) 39(1)	29 37 40 42 49 52	6	5	.8333
19(2) 21(1) 39(1)	29 37 40 42 49 52 55	7	6	.8571
19(2) 22(1) 39(1)	29 30 37 40 42 49 52 55	8	6	.7500
19(2) 26(2) 40(1)	29 30 37 39 40 42 49 52 54	9	7	.7778
21(1) 25(2) 39(1)	29 37 40 42 49 52 55	7	6	.8571

TABLE 18.--Results of predicting to the analysis sample of the UN data from patterns.

Predicted Vote	Actual Vote		
	No	Yes	
No	37	6	43
Yes	2	10	12
	39	16	55

$\phi = .631$

TABLE 19.--Results of predicting to the cross validation sample of the UN data from patterns.

Predicted Vote	Actual Vote		
	No	Yes	
No	34	11	45
Yes	2	7	9
No Pre-diction	1		1
	37	18	55

$\phi = .374$*

$\chi^2 = 7.72$

$p < .006$

*ϕ and χ^2 were computed on this table:

34	11	45
3	7	10
37	18	55

ov "yes" to monitoring atmospheric radioactivity, enters into several patterns here. This might indicate a general willingness for UN expenditures relating to world problems, among which issue 38 is one.

In summary, the patterns selected by Criterion Pattern Analysis allow prediction to the criterion even though the criterion is not closely related to a major underlying difference between nations; at the same time these patterns define types which reflect that difference.

These results are now compared with multiple regression and the maximum likelihood procedure.

Multiple Regression

The analysis sample of the UN data produced the regression coefficients shown in Table 20. Note also their levels of significance and the correlations of the votes with the criterion vote. So that the mathematical requirements of multiple regression could be fulfilled, votes 12, 13, 31, and 42 were eliminated from the data since votes 12 and 13 were identical to vote 11, vote 31 was identical to vote 29, and vote 42 was identical to vote 44.* The multiple correlation produced by applying the regression coefficients was .885.

Table 21 shows the results of predicting to the analysis sample using the regression coefficients. Since there is only one misclassification, the phi is very high at .958. When the prediction was made to the cross validation sample, the phi dropped to .308 and the chi square of 5.22, is significant at $p < .025$ (see Table 22).

---

*None of these votes entered into the patterns extracted by Criterion Pattern Analysis.

TABLE 20.--Results of multiple regression on the analysis sample of the UN data.

Vote	Regression Coefficient	Significance of Regression Coefficient	Correlation of the Vote with the Criterion
1	-.14	.73	.22
2	-.01	.92	.04
3	-.23	.58	-.15
4	.31	.33	.27*
5	-.30	.57	-.05
6	-.73	.17	-.03
7	-.28	.66	.00
8	.36	.52	-.02
9	.05	.89	.16
10	-.92	.11	-.17
11	.99	.15	.12
14	.05	.88	.06
15	-.36	.40	-.01
16	.05	.90	.11
17	.04	.89	.04
18	-.20	.62	.08
19	.54	.24	.32*
20	.26	.51	.24
21	-.31	.38	-.03
22	.14	.81	.06
23	-.48	.53	.02
24	.50	.42	-.03
25	.14	.69	.34*
26	.50	.38	.10
27	.00	.95	-.08
28	.83	.26	.02
29	-1.20	.30	.03
30	.02	.92	.15
32	-.11	.78	.06
33	.16	.80	.07
34	-1.45	.29	.02
35	.15	.80	.08
36	-.17	.59	-.14
37	-.28	.38	-.36**
39	.09	.80	-.04
40	.16	.79	-.02
41	-.07	.88	-.06
43	.11	.87	.03
44	.43	.73	.08

*Significant at the .05 level.

**Significant at the .01 level.

TABLE 21.--Results of predicting to the analysis sample of the UN data from regression coefficients.

Predicted Vote	Actual Vote No	Actual Vote Yes	
No	38		38
Yes	1	16	17
	39	16	55

$\phi = .958$

TABLE 22.--Results of predicting to the cross validation sample of the UN data from regression coefficients.

Predicted Vote	Actual Vote No	Actual Vote Yes	
No	28	8	36
Yes	9	10	19
	37	18	55

$\phi = .308$

$\chi^2 = 5.22$

$p < .025$

Compared with predicting directly from patterns, multiple regression did better on the analysis sample but suffered more on the cross validation sample. However, the difference is not statistically significant.

Interpretation

> Issue 10 has the most significant regression weight ($p < .11$), and hence makes the highest independent contribution to the prediction. Issues 4, 19, 25, and 37 are the most highly correlated with the criterion, being the only items significant at the 5% level or less.
>
> Issue 10 never appears in the patterns from Criterion Pattern Analysis. Issues 4, 19, 25, and 37 all appear as single-item patterns, and would have the same interpretation. Like the criterion issue, issue 10 is one of the few issues on which the US and the USSR voted alike.

Maximum Likelihood Procedure

As usual the analysis sample was used to compute estimates of the parameters of the multivariate-normal distribution. For mathematical reasons, the number of issues used to predict had to be reduced from 43 to 13.* A random selection of 13 issues which satisfied the mathematical requirements was made. They were issues 2, 4, 11, 15, 16, 17, 20, 26, 34, 35, 36, 39, and 43.

After the parameters were computed for both criterion groups, predictions were made on the analysis sample (see Table 23). The phi of .728 indicates the degree of relationship between the actual and predicted

---

*Since the smallest criterion group had only 16 nations, the number of issues in computing the sum of squares and cross-product matrix for the multivariate normal had to be less than 16 to avoid singularity.

TABLE 23.--Results of predicting to the analysis sample of the UN data using maximum likelihood.

Predicted Vote	Actual Vote No	Actual Vote Yes	
No	38	5	43
Yes	1	11	12
	39	16	55

$\phi = .728$

votes. When predictions were made to the cross validation sample (Table 24), the phi dropped to .152 and the chi square is 1.27, which is significant at only the .26 level.

TABLE 24.--Results of predicting to the cross validation sample of the UN data using maximum likelihood.

Predicted Vote	Actual Vote No	Actual Vote Yes	
No	33	14	47
Yes	4	4	8
	37	18	55

$\phi = .152$

$\chi^2 = 1.27$

$p < .25$

The difference between predicting directly from patterns and the maximum likelihood result on cross validation is significant at the .07 level.

Combination Method

Using the patterns* extracted by Criterion Pattern Analysis, each nation in both the analysis sample and the cross validation sample was given scores of 2 for patterns it had and scores of 1 for patterns it did not have. The new data for the analysis sample were then subjected to multiple regression. The results in Table 25 show the regression coefficient and its significance for each pattern, along with the pattern's correlation with the criterion. The resulting multiple correlation is .872.

Predicting to the analysis sample yielded a phi of .893 between actual and predicted votes (see Table 26). Predicting to the cross validation sample resulted in a phi of .584 (see Table 27) and the associated chi square of 18.7 shows that the prediction is very significant ($p << .001$). For comparison, in predicting to the cross validation sample from patterns directly, the phi was .374, and in predicting with multiple regression with the original items, the phi was .308. In this latter comparison the difference is significant at the .06 level.

*Not used were five patterns which duplicated the nations predicted by other patterns. Mathematically, the matrix of intercorrelations of the patterns was made non-singular.

TABLE 25.--Results of multiple regression applied to the patterns in the analysis sample of the UN data.

Pattern	Regression Coefficient	Significance of the Regression Coefficient	Correlation of the Pattern with the Criterion
4(1)	.14	.61	-.27*
19(1)	-.04	.84	-.32*
25(1)	-.01	.97	-.34*
37(2)	-.25	.18	-.36**
3(2) 9(1)	-.17	.26	-.32*
20(1) 39(2)	-.04	.79	-.36**
20(1) 41(2)	-.11	.45	-.34*
4(2) 19(2)	-.03	.90	.51**
4(2) 20(2)	.67	.02	.44**
4(2) 25(2)	.004	.99	.54**
20(2) 37(1)	.19	.39	.43**
4(2) 9(2) 39(1)	-.34	.37	.49**
4(2) 19(2) 39(1)	1.51	.001	.55**
19(2) 22(1) 39(1)	-.68	.04	.42**
19(2) 26(2) 40(1)	.34	.15	.47**

*Significant at the .05 level.

**Significant at the .01 level.

TABLE 26.--Results of predicting to the analysis sample of the UN data using regression coefficients from patterns.

Predicted Vote	Actual Vote		
	No	Yes	
No	36	0	36
Yes	3	16	19
	39	16	55

$\phi = .893$

TABLE 27.--Results of predicting to the cross validation sample of the UN data using regression coefficients from patterns.

Predicted Vote	Actual Vote		
	No	Yes	
No	33	6	39
Yes	4	12	16
	37	18	55

$\phi = .584$

$\chi^2 = 18.7$

$p << .001$

Interpretation

> Pattern 4(2) 19(2) 39(1) has not only the largest and most significant regression coefficient, but also has the greatest correlation with the criterion. Hence it seems to be the best all-round single predictor. However, it is difficult to assign any unique substantive interpretation to the pattern. Substantive interpretation can again be based on patterns since they are all highly related to the criterion (all correlations are significant at least at the .05 level).

Resumé

The combination of patterns and multiple regression greatly increased the prediction to the cross validation sample of the UN data. In comparison with multiple regression on the original items, with a phi of .308, the combination procedure with a phi of .584 is better at a significance level of .06.

Summary

Predicting from patterns extracted by Criterion Pattern Analysis did not, in general, do as well on the analysis samples as did the linear methods. More importantly, on the cross validation samples, predicting directly from patterns consistently produced better results. This is in contrast to what is usually observed (e.g., Lubin, 1954; Lee, 1957), and indicates the inherent predictive strength of patterns as found by Criterion Pattern Analysis.

When both sets of data were considered together by combining the significances of differences in cross validation prediction,* predicting directly from patterns was better than multiple regression at the .33 significance level, and was better than the maximum likelihood procedure at the .06 level.

When a linear procedure, multiple regression, was used in combination with the patterns extracted by Criterion Pattern Analysis, prediction to the cross validation samples was enhanced. Combining probabilities as before yielded a significance level of .08, indicating the degree of superiority on cross validation prediction of the combination method over multiple regression on the original items.

While no quantitative measurements of interpretability were made, it was clear that Criterion Pattern Analysis displayed more directly information which could be related to the substantive material.

---

*The combining of independent probabilities is the P_λ test (Rao, 1953, p. 44).

CHAPTER IV

DISCUSSION AND CONCLUSION

Discussion

Applying Criterion Pattern Analysis

In the previous chapter, Criterion Pattern Analysis was applied to two sets of data. The two sets of data are formally similar in that both have a dichotomous criterion and both have dichotomous responses in the predictor items. It should be made clear that Criterion Pattern Analysis does not require dichotomous data; it requires discrete data. The method can only be applied to continuous data when discrete categories are imposed on the predictor items and on the criterion. The judgment as to how many categories to use for each predictor item and for the criterion is a difficulty not unique to the method here. When too many categories are used no patterns will be found; when too few categories are used, information is lost. Roughly, the more observations, the more categories that can be allowed.

The use of Criterion Pattern Analysis can become more elaborate when there are more than two criterion categories. Usually patterns are found for each category of the criterion; that is, patterns are found which distinguish each criterion category from the remaining categories. However it may be of interest to find patterns which distinguish a combination of categories from those remaining. The set of patterns found for the combination will be different from the set of patterns found for each category separately. Indeed, the search for predictive patterns can not be considered complete until patterns are found for all combinations of the criterion categories.

Possible Improvements

Criterion Pattern Analysis produces predictive and interpretable patterns. The procedure for identifying these few patterns from among many possible patterns is a complicated one, and the question can be asked if the present procedure can be improved.

Improvements might be realized in the way the α level is set for the hypergeometric distribution. At present, this setting is done by trial and error. The .05 level was used with both sets of demonstration data. On other sets of data the level has ranged from .1 to .0001. If the level is set too low, too many patterns are extracted; if set too high, too few are extracted. No doubt the level is influenced by the number of items and people, and also by how well-structured the data are. An all-over probability level, if it could be developed, would help with this problem. With this in hand, the researcher could select some probability, say .01, that would indicate the desired significance for the results. The .01 in combination with other parameters, such as number of people and items, would determine the α level for the hypergeometric distribution. Of course, if the data were non-predictive to begin with, no patterns would result; if the data contained much predictive information, a multitude of patterns would result. Thus the all-over probability level, although very helpful, will not alone solve the problem of setting α.

A related problem is that the same α level is used for testing small patterns as for testing large patterns. Since there are more tests to be made for a five-item pattern, for example, than there are for a two-item pattern, keeping α the same tends to favor patterns with fewer items. On the other hand, in any set of data there are many more five-item patterns than two-item patterns, so perhaps these two factors tend to cancel each other out. Still, what is going on is not yet completely understood, and improvements should be realizable.

Further Comparisons

The value of the predictive information inherent in patterns of items is widely recognized. In Chapter I various efforts to obtain this information directly from patterns were reviewed. Other efforts have also been made to incorporate some of the properties of patterns into linear methods. Among them are several multiple regression procedures which employ first order interaction effects (Saunders, 1956; Ghiselli, 1960). Applying first order interaction effects in a discriminant function analysis has also been tried (Stellwagon, 1960), but without any improvement over linear prediction. This is not surprising, since similar results were obtained from the point of view of analyzing patterns as reviewed in Chapter I.

While these methods capitalize on the interaction properties which characterize patterns, a method proposed by Weiss (1964) attempts to utilize the non-linear properties of each predictor item. This is done by assessing its curvilinear relationship with the criterion. Since Weiss found that curvilinear relationships are indeed present in data, the question arises as to how Criterion Pattern Analysis would compare with this method. In this regard, Hoffman (1960) asserted that patterns not only account for interaction effects,* but also assess relationships which imply a scale transformation. Since

---

*Meehl (1954, p. 134) states that for continuous variables x_1, x_2 which predict y, patterning exists when $\frac{\partial^2 y}{\partial x_1 \partial x_2} \neq 0$. This is tantamount to asserting that patterning equals interaction in prediction.

curvilinearity is one type of scale transformation, patterns most certainly would account for it. Furthermore, pattern methods do not require that the nature of the curvilinearity be specified.

In general all linear methods which make use of any of the properties of patterns must specify beforehand which of those properties will be used. Once specified, they are incorporated into a linear variable. In other words a linear variable which displays the required properties is constructed and inserted. Obviously, if the particular properties chosen are inconsequential for the data at hand no improvement in prediction will be made. On the other hand if patterns are sought which predict to a criterion without restriction as to <u>how</u> they should predict, then any number of interaction and scale transformation properties may be utilized.

Thus it can be seen that Criterion Pattern Analysis fills a need by supplying patterns that incorporate many types of relationships. In this way Criterion Pattern Analysis functions as an extension of item analysis. Now, however, the relationship with the criterion is not restricted to a linear one.

Moreover, once the patterns are in hand, they can be combined with another method as was done in Chapter III. There, they were combined with multiple regression, with the result of greater enhancement of the predictions to the cross validation samples. It would be expected that the cross validation predictions of the maximum likelihood method would be improved as well.

Furthermore, it is interesting that the relatively simple scheme of predicting directly from patterns compares favorably with the sophisticated mathematical power inherent in the standard application of linear methods. However, the comparison is complicated by a further consideration. Most methods are constructed to maximize the prediction on the analysis sample. There is no way at present to apply the methods so that results will be maximum upon cross validation. Hence, favorable results of predicting directly from patterns may be due either to the utilization of configural properties or to the selection of item clusters which tend to maximize cross validation predictions. Of course, both influences may be at work. The previous failures of configural methods upon cross validation can be interpreted as failures to maximize cross validation prediction, and not as a failure of the configural approach. It is suggested that a general approach toward maximizing cross validation prediction be developed. When this is applied to both linear and configural methods, then a comparison between the two will be more informative.

Interpreting the Results

Another problem with the linear techniques has been interpreting the results (e.g., Hoffman, 1960, 1962; Ward, 1962; Gibson, 1962). For example, the regression coefficient in general does not reflect how well its associated item predicts the criterion. The coefficient is an indication of how well the item predicts when all other items are held constant (Lee, 1961). Only when the items are independent of each other will the coefficients be easily interpretable. Furthermore, sets of items sometimes become highly predictive in multiple regression analysis, but it is almost impossible for the user to realize when a combination is playing a special role in prediction.

In contrast, the interpretation of patterns from Criterion Pattern Analysis seems clear. Each pattern can be directly referred back to a combination of items and responses. Problems of interpretation, when they arise, will relate to item contents, rather than to obscurities grounded in the information given by the method. When multiple regression is combined with Criterion Pattern Analysis interpretation can still rely on the patterns.

Often, help in interpreting a pattern is obtained by referring to the set of subjects who have the pattern. This was particularly helpful with the UN data, where identification of nations was possible. In fact, much of the interpretation was in terms of groupings or typings of nations.

Typal Theory

Patterns associated with criterion one of the UN data, by defining two distinct sets of nations, demonstrate that Criterion Pattern Analysis is relevant to a theory of types. Such types, when used, are defined with respect to their ability to differentiate one criterion category from others. These types are not necessarily the same as types defined over all categories or defined within a category. If important category-differentiating types exist, Criterion Pattern Analysis will locate them. A set of types so identified may be mutually exclusive, with each person found in only one type, or the types may overlap, with some persons in several types.

On the other hand, Criterion Pattern Analysis is not dependent upon a theory of types in its assumptions. Therefore, its operation is not at all contingent upon differential typal structure in the data. If patterns which define a homogeneous set of people predict best, Criterion Pattern Analysis will use them rather than choosing a differential typal structure. In fact, the method may be used without reference to a theory of types. This is demonstrated by using each pattern of responses to define a new item, as was done in the combination prediction method.

In general, types are identified in data by a collection of people having similar characteristics. It is clear that Criterion Pattern Analysis is unique in that it isolates types which are characterized by their predictiveness of a criterion. Now, Lykken (1956, p. 102) says, ". . . similarity is not a general quality. It is possible to discuss similarity only with respect to specified dimensions." If instead of the term "specified dimensions" the term "*a priori* categories" is substituted, this statement would seem to apply to the present method. Lykken goes on to say that without a reference structure, types only reflect geometrical configurations and in such cases are rarely psychologically relevant. If Lykken is right, then meaningful types are defined only when their reference structure is specified. Criterion Pattern Analysis would then be the only method providing types which are psychologically meaningful.

Conclusion

The method of Criterion Pattern Analysis is now measured against the requirements set down at the end of the first chapter.

1. All major patterns which predict the criteria should be found:

Criterion Pattern Analysis finds all patterns which improve the prediction of their respective subpatterns. All possible patterns are considered, within limits of computational feasibility. These patterns are both valid and reliable, since they predict well in both analysis and cross validation samples.

2. The patterns should be found separately for each criterion category:

Criterion Pattern Analysis meets this requirement.

3. The method should isolate non-configural as well as configural relationships:

No restrictions are made on the number of items in the pattern or on the nature of the relationship with the criterion.

4. The patterns extracted by the method should be capable of being used directly for prediction.

A predictive method utilizing the discrimination level for each pattern has been developed for predicting directly from patterns. While doing better than linear methods on cross validation, using patterns directly did not utilize all predictive information; multiple regression in conjunction with pattern-scored data predicted more accurately.

5. The method should be capable of predicting better than linear methods on the analysis sample:

This capability has not been demonstrated, for the emphasis was on doing better upon cross validation. It is doubtful whether a method which attempts to maximize prediction on the initial data can possibly do as well on other samples. The linear methods in some cases predict almost perfectly on the analysis sample and fail completely upon cross validation. It would be folly to develop a pattern method which does the same thing.

6. The method should predict better than linear methods on a cross validation sample:

Criterion Pattern Analysis was consistently better than the linear methods upon cross validation. This superiority was not demonstrated to be clearly significant in every case. However, the contention that configural methods capitalize on chance and hence do not stand up under cross validation has been seriously challenged.

7. The method should be applicable to small as well as to large samples of people:

The demonstration data were chosen to test this. Criterion Pattern Analysis did as well or better than linear techniques on these relatively small samples.

8. The results of a configural prediction method should be readily interpretable:

The nature of patterns themselves offer opportunities for substantive interpretation, whether used directly for prediction or incorporated into more complex methods. The discrimination level associated with each pattern is a precise statement about how well the pattern predicts the criterion.

9. The results of a configural prediction method should be readily obtainable:

Criterion Pattern Analysis was developed to search among all possible patterns without examining each, one by one. The resulting reduction in computational time enabled the method to be readily applied with the help of a computer.

APPENDICES

APPENDIX A

I-E SCALE

I-E SCALE

1. a. Many of the unhappy things in people's lives are partly due to bad luck.

   b. People's misfortunes result from the mistakes they make.

2. a. One of the major reasons why we have wars is because people don't take enough interest in politics.

   b. There will always be wars, no matter how hard people try to prevent them.

3. a. In the long run people get the respect they deserve in this world.

   b. Unfortunately, an individual's worth often passes unrecognized no matter how hard he tries.

4. a. The idea that teachers are unfair to students is nonsense.

   b. Most students don't realize the extent to which their grades are influenced by accidental happenings.

5. a. Without the right breaks one cannot be an effective leader.

   b. Capable people who fail to become leaders have not taken advantage of their opportunities.

6. a. No matter how hard you try some people just don't like you.

   b. People who can't get others to like them don't understand how to get along with others.

7. a. I have often found that what is going to happen will happen.

   b. Trusting to fate has never turned out as well for me as making a decision to take a definite course of action.

8. a. In the case of the well prepared student there is rarely if ever such a thing as an unfair test.

   b. Many times exam questions tend to be so unrelated to course work that studying is really useless.

9. a. Becoming a success is a matter of hard work, luck has little or nothing to do with it.

   b. Getting a good job depends mainly on being in the right place at the right time.

10. a. The average citizen can have an influence in government decisions.

   b. This world is run by the few people in power, and there is not much the little guy can do about it.

11. a. When I make plans, I am almost certain that I can make them work.

   b. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.

12. a. In my case getting what I want has little or nothing to do with luck.

   b. Many times we might just as well decide what to do by flipping a coin.

13. a. Who gets to be the boss often depends on who was lucky enough to be in the right place first.

   b. Getting people to do the right thing depends upon ability, luck has little or nothing to do with it.

14. a. As far as world affairs are concerned, most of us are the victims of forces we can neither understand, nor control.

   b. Be taking an active part in political and social affairs the people can control world events.

15. a. Most people don't realize the extent to which their lives are controlled by accidental happenings.

   b. There is really no such thing as "luck."

16. a. It is hard to know whether or not a person really likes you.

   b. How many friends you have depends upon how nice a person you are.

17. a. In the long run the bad things that happen to us are balanced by the good things.

    b. Most misfortunes are the result of lack of ability, ignorance, laziness, or all three.

18. a. With enough effort we can wipe out political corruption.

    b. It is difficult for people to have much control over the things politicians do in office.

19. a. Sometimes I can't understand how teachers arrive at the grades they give.

    b. There is a direct connection between how hard I study and the grades I get.

20. a. Many times I feel that I have little influence over the things that happen to me.

    b. It is impossible for me to believe that chance or luck plays an important role in my life.

21. a. People are lonely because they don't try to be friendly.

    b. There's not much use in trying too hard to please people, if they like you, they like you.

22. a. What happens to me is my own doing.

    b. Sometimes I feel that I don't have enough control over the direction my life is taking.

23. a. Most of the time I can't understand why politicians behave the way they do.

    b. In the long run people are responsible for bad government on a national as well as on a local level.

APPENDIX B

UNITED NATIONS GENERAL ASSEMBLY,
SEVENTEENTH SESSION:

ISSUES DECIDED BY ROLL CALL VOTE

Vote	Issue
1	Note agreement over West New Guinea.
2	Put Hungarian question on agenda.
3	Censure Great Britain and urge establishment of constitutional government in Southern Rhodesia.
4	Replace Nationalist China with Communist China.
5	(Same text as vote 3)
6	Vote separately on sections of South Africa condemnation.
7	Condemn South Africa's racial policies.
8	Is test-ban conduct proposal fair?
9	Condemn all nuclear weapons tests.
10	Should set deadline for nuclear test halt.
11	Testing by U. S., U. K., and U. S. S. R. should stop by deadline.
12	Use proposed basis for test-ban negotiation.
13	Negotiate test-ban in spirit of mutual understanding.
14	Stop underground tests on interim basis by deadline.
15	Reconvene Disarmament Committee.
16	Disarmament Committee should reach a treaty.
17	<u>All</u> disputing parties under convention on marriage must refer case to International Court of Justice.
18	Specify territories covered by convention on marriage.

Vote	Issue
19	Implement plan for monitoring atmospheric radioactivity.
20	Specify which nations will be invited to atomic energy conference.
21	Continue work on Hong Kong refugee problem.
22	Study <u>gradual</u> tariff reduction at conference.
23	Oman should be declared independent.
24	Foreign forces should be withdrawn from Oman.
25	Differences over Oman should be settled peacefully.
26	Nations have right to expropriate property.
27	Support national strengthening of sovereignty over natural resources.
28	Condemn Portugese colonial policies.
29	Vote separately on decolonization deadline need.
30	Continue committee on decolonization.
31	Enlarge committee on decolonization.
32	Decolonization deadlines needed.
33	Committee on decolonization should set deadlines.
34	(Resolution containing issues 30-33 as a whole)
35	Vote separately on Angola resolution sections.
36	Condemn Portugal's Angola policies.
37	Need 2/3 majority on population and economic growth resolution.

Vote	Issue
38	Support giving technical assistance to national projects for population study and other aspects of social and economic development.
39	Accept International Court of Justice decision on financing peacekeeping.
40	Continue work on Palestine Arab refugee problem.
41	Continue Conciliation Commission for Palestine.
42	Resolutions on Hungary haven't been implemented.
43	Dissolve Office of UN Representative on Hungary.
44	(Resolution containing issues 42 and 43 as a whole)

APPENDIX C

UN DATA: COUNTRIES IN ANALYSIS SAMPLE

No.	Country	No.	Country
1	United States of America	23	Spain
2	Canada	24	Byelorussian Soviet Socialist Republic
3	Cuba	25	Ukranian Soviet Socialist Republic
4	Dominican Republic	26	Union of Soviet Socialist Republics
5	Nicaragua	27	Iran
6	Panama	28	Lebanon
7	Guatemala	29	Iraq
8	Honduras	30	Jordan
9	Trinidad and Tobago	31	India
10	Argentina	32	Thailand
11	Chile	33	Burma
12	Colombia	34	Cambodia
13	Bolivia	35	Malaysia
14	Ecuador	36	Mongolia
15	New Zealand	37	United Arab Republic
16	France	38	Ethiopia
17	United Kingdom of Great Britain and Northern Ireland	39	Libya
18	Norway	40	Morocco
19	Austria	41	Somalia
20	Hungary	42	Guinea
21	Greece	43	Central African Republic
22	Romania		

No.	Country	No.	Country
44	Chad		
45	Congo (Brazzaville)		
46	Dahomey		
47	Madagascar		
48	Togo		
49	Mali		
50	Senegal		
51	Sierra Leone		
52	Mauritania		
53	Tanganyika		
54	Burundi		
55	Uganda		

BIBLIOGRAPHY

BIBLIOGRAPHY

Alf, Edward Franklin, Jr., 1957. Configural scoring and prediction. Diss. Abstr., 17, 887.

Cochran, William G., and Hopkins, Carl E., 1961. Some classification problems with multivariate qualitative data. Biometrics, 17, 10-32.

Cooley, W. W., and Lohnes, P. R., 1962. Multivariate Procedures for the Behavioral Sciences. John Wiley and Sons, New York.

Crego, C. A., 1966. A pattern analytic approach to the measure of modes of expression of psychological differentiation. Unpbl. Ph.D. Dissertation. Mich. State Univ.

Forehand, G. A., Jr., and McQuitty, Louis L., 1959. Configurations of factor standings as predictors of educational achievement. Educ. Psychol. Measmt., 19, 31-43.

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

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

Gaier, E. L., and Lee, Marilyn C., 1953. Pattern analysis: The configural approach to predictive measurement. Psychol. Bull., 50, 401-413.

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

Gibson, W. A., 1962. Orthogonal predictors: A possible resolution of the Hoffman-Ward controversy. Psychol. Rep., 11(1), 32-34.

Hoffman, P. J., 1960. The paramorphic representation of clinical judgment. Psychol. Bull., 57, 116-131.

Hoffman, P. J., 1962. Assessment of independent contribution of predictors. Psych. Bull., 59(1), 77-80.

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

Ledley, R. R., and Lusted, L. B., 1959. Reasoning foundations of medical diagnosis. Science, 130, 892.

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

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

Loevinger, Jane, 1959. Theory and techniques of assessment. Annu. Rev. Psychol., 10, 287-316.

Lubin, A., 1950. Linear and non linear discriminating functions. British J. Psychol., Statistical Section, 3, 90-104.

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

Lubin, A., and Osburn, Hobart G., 1957. A theory of pattern analysis for the prediction of a quantitative criterion. Psychometrika, 22, 63-73.

Lykken, D. T., 1956. A method of acturial pattern analysis. Psychol. Bull., 53, 102-107.

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

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

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

McQuitty, L. L., 1959. Differential validity in some pattern analytic methods. In Bass, B. M., and Berg, I. A., (eds.), Objective Approaches to Personality Assessment, Princeton, D. VanNostrand Co., Inc.

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

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

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

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

McQuitty, L. L., Yarnold, J. K., Eckstrom, V., 1956. Patterns of high school grades in the prediction of college achievement. Mimeographed.

Meehl, P. E., 1950. Configural scoring. J. Consult. Psychol., 14, 165-171.

Meehl, P. E., 1954. Clinical vs. Statistical Prediction. Minneapolis, University of Minnesota Press.

Olin, Wayne, 1964. Voting records of the United Nation's General Assembly, Plenary Meetings, Sessions 1-18, 1946-1963. Computer Institute for Social Science Research, Michigan State University.

Osburn, H. G., and Lubin, A., 1957. The use of configural analysis for the evaluation of test scoring methods. Psychometrika, 22, 359-371.

Rao, C. R., 1952. Advanced Statistical Methods in Biometric Research. New York, John Wiley and Sons.

Rotter, J. B., 1966. Generalized expectancies for internal versus external control of reinforcement. Psychol. Monog., 80, no. 609, 1.

Saudners, D. R., 1955. Item analysis of interaction variance to produce a criterion key for item pairs. Bull. No. 55-28, Educational Testing Service, Princeton, N. J.

Saunders, D. R., 1956. Moderator Variables in Prediction. Educ. Psychol. Measmt., 20, 675-684.

Sells, S. B. et al., 1955. Symposium on Pattern Analysis. Randolph Field, Texas: USAF School of Aviation Medicine. 62 pp.

Stellwagon, W. R., 1960. An examination of the use of linear and non-linear discriminant functions in the classification of college students into academic groups. Diss. Abstr., 20, 4435.

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

Tatsuoka, Maurice M., and Tiedeman, David V., 1954. Discriminant analysis. Rev. Educ. Res., 24, 402-420.

United Nations, 1964a. Official Records of the General Assembly, Seventeenth Session, Annexes. Vol. I-III, New York, United Nations.

United Nations, 1964b. Official Records of the General Assembly, Seventeenth Session, Plenary Meetings. Vol. I-III, New York, United Nations.

Wainwright, Gorden E., 1966. Configural information in item responses. Diss. Abstr., 27(1-B), 297.

Walker, H. M., and Lev, J., 1953. Statistical Inference. New York, Henry Holt and Co.

Ward, Joe H., Jr., 1962. Comments on "The paramorphic representation of clinical judgment." Psychol. Bull., 59(1), 74-76.

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

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

Witkin, H. A., Dyk, R. B., Goodenough, D. R., Karp, S. A., 1962. Psychological Differentiation. New York, John Wiley and Sons.

Zubin, J. A., 1938. A technique for measuring like mindedness. J. Abn. and Soc. Psychol., 33, 508-516.