DOCUMENT RESUME

ED 106 356 TM 004 472

AUTHOR Mandeville, Garrett K.; And Others
TITLE A Nonparametric Procedure for Demonstrating a Non-Chance Fit Among Pairs of Multivariate Responses.
PUB DATE [Apr 75]
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Comparative Analysis; Correlation; Data Analysis; Goodness of Fit; Measurement Techniques; *Nonparametric Statistics; Observation; Probability; *Rating Scales; Response Mode; *Response Style (Tests); Simulation; Test Reliability; *Test Results; Video Tape Recordings

ABSTRACT

A strategy for comparing two sets of results (one based upon early childhood recollections (ECR) and another upon video taped (VT) group behavior) from the Perceptual Characteristics Rating Scale was developed. The null distribution of the mean deviation was estimated by randomly matching an ECR response vector with a VT response vector. To evaluate the degree of relationship of ECR's and VT's for each subject, a correct pairing was made. The observed values of the mean deviations were compared to the null distributions and the associated empirical probabilities were converted to chi squares to allow summary tests of "no relationship." (Author)

19.22

A NONPARAMETRIC PROCEDURE FOR DEMONSTRATING A NON-CHANCE FIT AMONG PAIRS OF MULTIVARIATE RESPONSES

Garrett K. Mandeville
University of South Carolina

James L. Crandall
University of Alabama-Birmingham

Vana H. Meredith
University of South Carolina

Paper presented at the annual meeting of the
American Educational Research Association
Washington, D.C., March 30-April 3, 1975

INTRODUCTION AND OBJECTIVES

The primary objectives of this paper are threefold: (1) to identify various strategies for analyzing data of a specific type which occurs fairly often, (2) to compare the statistical properties of these designated strategies and (3) to introduce the educational research community to a rather simple general strategy for handling certain unusual data analysis requests. The situation in which the data arises is one in which a respondent, often a rater, is called upon to rate some kind of observation he has made on a unidimensional scale consisting of a modest number of items. Subsequently, he is asked to perform the same task for some other observation which, in some way, is associated with the first. The general problem these writers wish to address is what recommendation to make to the researcher who wishes to determine whether or not the respondents were able to implicitly identify the association, i.e., is the fit between the data obtained under the two modes better than chance?

For clarity, the specific situation which lead to the preparation of this paper was a request to suggest a strategy for comparing two sets of results from the Perceptual Characteristics Rating Scale (PCRS). One set of data was obtained based on 'behavioral predictions' derived from the subjects' early childhood recollections (ECR's); and the other set of data was obtained from the observation of video taped behavior (VT's), actual behavioral manifestations observed in a group setting. The research question involved the predictive power of ECR's regarding behavior; and, therefore, the primary statistical analysis was required to provide evidence of a non-chance fit between the two sets of responses made by the same subjects.

PERSPECTIVES AND THEORETICAL FRAMEWORK

In reveiwing the literature on rater agreement, e.g., Naylor and Dudycha (1967), Naylor and Schenck (1966), Taylor (1968) Taylor, et. al., (1970), little research was uncovered which bore directly on the problem under consideration here. (Most of the above studies dealt with psychometric properties of rating scales, usually reliability.) Although the cannonical correlation appears to be the classical technique required to attack the problem, no such applications were found. Assuming unidimensionality, the first cannonical r should contain the information on the association between the pairs of ratings. This, then, appears to be the most reasonable classical technique. Thinking that possibly alternative strategies might exist for analyzing these data, these researchers considered other approaches.

One would think, for example, that since the problem is one of association, some form of ordinary Pearson r might be appropriate. One might, for example, compute correlations between the associated response pairs and attempt to evaluate the magnitude of the computed r's. As is true in Q methodology, the responses entering into the computation of the correlations are not independent but, there are more serious objections to the use of the Pearson r for this purpose. Realizing that there would be N correlations to be evaluated, if N is the number of pairs of objects or situations rated, it would seem to be possible to compare the distribution of the computed r's with the null distribution. There are two serious drawbacks to this approach which invalidate it as a possible analytic method. They are (1) since we are thinking of situations in which there are relatively few items per scale, the degrees of freedom associated with

the null distribution of r would be exceedingly small but more importantly (2) the r distribution would be highly dependent upon the average item responses. For example if the items in a scale differed considerably in mean response (regardless of stimulus mode) the responses under the two modes could reflect an artificial correlation even though there was no actual association between the item responses made by a rater under the two modes. Thus, this approach will not be considered further. Although multidimensional scaling might initially appear to be useful, it is not appropriate for these data. (See, e.g., Kruskal, 1964a, 1964b.)

The final approach to be considered here will be to compute various measure of the similarity of or distance between the two points (the sets of responses under the two modes) in p dimensional space where p is the number of responses under each mode. The problems with this approach are twofold: the researcher (1) must select from among various similarity and distance measures and (2) must obtain the null distribution of the selected distance measure. In addition to the correlation, the common measures that come to mind are the ordinary Euclidean distance and the Mahalanobis distance. The former seems quite reasonable but the validity of the latter is somewhat questionable for this situation. The two vector random variables entering into the Mahalanobis distance are assumed to be independent and have a common dispersion matrix. The first requirement is definitely not true although the second might be reasonably well satisfied. If the two sample covariance matrices were found to be similar possibly some overall estimate would provide useful results. For these reasons and others described later the Mahalanobis distance was not used in this research. Although not generally categorized as a distance

measure, and causing obvious mathematical difficulties, the average (absolute) deviation might be considered because of its simplicity. For this study, then, similarity-distance measures were limited to Pearson r (R), absolute average deviation (D) and Euclidean distance (E).[1]

Although the selection of a similarity-distance measure might cause the researcher some difficulty, the problem of how to evaluate the resulting measures is an even more troublesome problem. To the knowledge of these writers there is no 'reasonably well known' solution to the distribution theory problems raised here. It appeared to these researchers that if any of these similarity-distance measures were to be used the null distributions would have to be obtained before any evaluation of the subject responses would be possible.

METHOD AND TECHNIQUES

Although the initial interest to these writers was relative to the specific problem of the PCRS data, and a few results from that study will be presented here, the primary thrust in this paper is a limited methodological study to, at least partially, evaluate the usefulness of a non-parametric solution to the problem in relation to other possible strategies. The non-parametric approach was developed because of the 'distributional problem' mentioned earlier. Before going into the details of the methodological study, the reader is due a little more information concerning the original research from which this paper derived. (Taylor, 1973)

A total of 47 subjects were involved in the study being primarily graduate students in education enrolled in "Group Procedures in Guidance" at a southern university. The study was conducted during the 1972-73 school year. Twenty

---

[1]The writers are currently looking into other measures of similarity, e.g., the sum of cross products and covariance and other measures of distance, e.g., norms.

of the subjects were female and 27 male. The students were, for the purposes of instruction, divided into five 'classes' of 5, 11, 9, 9, and 13 students. These 'classes' constituted the groups which were videotaped. Each tape was approximately one hour in length. Each subject was requested to provide the six earliest instances that he or she could recall from childhood. The subject sex was also noted on the ER form. Two novices (other graduate students) and two experts (instructors) were requested to complete a modification of Combs (1969) PCRS based on the ER's. The modified PCRS contains 14 items each of which is rated on a 1-7 scale. The 14 items are considered to provide measurement on four subscales as follows: general perceptual orientation (2 items), perceptions of other people (4 items), perceptions of self (5 items) and purposes of behavior (3 items). Following a 'cooling off' period each of the four raters was asked to perform the same task based on the VT's. The raters were allowed to view the tapes as often as necessary to accomplish the task.

Since the sex of the subject could be identified under both modes, it was felt necessary to test to determine whether there were any differences which were associated with the sex of the subject. Although the affect would probably be slight, a sex difference could cause the agreement among the responses under the two modes to be better than chance without there being any true association among the data. Multivariate analysis of variance (MANOVA) was used to compare the responses for male and female subjects individually by rater. The results indicated that there was no systematic variation associated with the sex of the subject whose behavior or early recollections were being evaluated.

To carry out the 'test of association' a computer program was written which, for each rater and scale combination (4x4=16 in all) randomly matched

ER and VT responce vectors a total of 500 times. The Pearson r (R) and average absolute deviation (D) were computed for each random match and an empirical distribution was obtained for each. The same statistics were computed, then, for each of the 47 'correct' matches, i.e., when the same ER and VT vectors were matched. The 'tail probabilities' P_i were obtained for each by comparing the obtained value to the corresponding empirical 'null distribution'. The reader will note that the 'tail' of interest for R includes those values near 1.0 while the corresponding 'tails' for D and E are those values near 0.0. For each of the 16 combinations, then, the quantity

$$\chi^2 = \sum_{i=1}^{47} -2 \ln P_i$$

was obtained for each of the two statistics. Under the null hypothesis, this statistic should be approximately a chi square variable with 94 degrees of freedom. Table 1 below contains the results of this analysis for the average deviation D.

- - - - - - - - - - - - - - - - - - -

Table 1 about here

- - - - - - - - - - - - - - - - - - -

Except for subscale 4 we note the lack of consistency in the data across rater. In the original study the chi squares for each subscale were summed for 'experts' (raters 1 and 2) and novices (raters 3 and 4), tested separately for significance and compared with an F test. The details of this analysis are not presented here because they are not relevant to the primary thrust of this paper. Instead, we will describe the data which went into the above table and look at the results of applying alternative analytic methods. The following table presents some correlational data which only partially

agree with the results above.

- - - - - - - - - - - - - - - - - - - -

Table 2 about here

- - - - - - - - - - - - - - - - - - - -

The agreement between the significant chi squares and the modest average correlations between ER and VT is only partial. Obviously the correlation as a measure of similarity addresses different aspects of the data than the average absolute deviation. In order to attemp to summarize the seeming lack of association indicated in Table 2 cannonical correlation analysis was performed for each scale and rater. The VT responses were viewed as the dependent variables and were regressed onto the corresponding ER's. The results are found in Table 3.

- - - - - - - - - - - - - - - - - - - -

Table 3 about here

- - - - - - - - - - - - - - - - - - - -

We note that the cannonical correlation approach identified as significant only responses of raters 1 and 2 to subscale 4. Noting the descrepancy between the above results and wishing to investigate the relative merits of various strategies mentioned above, these researchers launched out on a simulation study.

The parameters for the simulation were selected with an eye on the descriptive data from the present study. Regarding the means and standard deviations of the responses, the most extreme values (1 and 7) were seldom used by the raters and the standard deviations for the items were generally around 1.0 or somewhat less. The question of whether the items differed in average response tendency may be addressed by viewing the graphs in

Figure 1.

- - - - - - - - - - - - - - - - - - - -

Figure 1 about here

- - - - - - - - - - - - - - - - - - - -

We note some trends which are quite general across raters. For example, for subscale 1 the second item generally receives a higher mean response than the first. Other trends may also be noted but the variation in 'rater effects' is quite large. The data do suggest, however, that spurious correlations between ER and VT responses probably occurred for some raters.

Three values for the number of variables under each response mode were chosen to cover the situations in this study and extend upward a bit. The values selected were three, five and ten variables. Since the data tended to support a common dimension within the responses under each mode, and the 'within mode' correlations were reasonably similar, it was felt that, for simplicity a common intercorrelation should be used to relate the various responses within a set. The two values chosen for this correlation were 0.5 and 0.7 and hereafter this parameter will be denoted as RIN. The extent to which the responses under the two modes were in agreement was considered next. A correlation of 0 was deemed as necessary to investigate the null properties of the procedures. Although an argument could be put forth that the two responses to the same item in a scale should be more highly related than the responses to two different items, the data did not support this notion (actually the associations were so limited, the data really gave little information on this question). For simplicity, the non-null measures of association between responses under the two modes were taken to be 0.3 and 0.5 and this parameter will be denoted as RBET.

The question of the validity of these choices and their effect on the similarity measures must be considered. Since the standard deviations were taken as constant (see below) the "within mode" covariance matrix will exhibit compound symmetry. In this situation, the Mahalanobis distance should be no more informative than the Euclidean distance since the weights (elements of the inverse of the covariance matrix) would be constant. This fact provided yet another reason for exclusion of the Mahalanobis distance function from this study.

Another issue is the use of a constant correlation matrix for the relationships between responses under the two modes. This does a disservice to the Pearson r which addresses the similarity of the two vectors on an item by item basis. Under this model, the only way that the correlation cannot distribute around zero is if the item means vary consistently for the two modes. This is a group effect, however, and not the type of effect that we are interested in identifying. Thus, if the interrelationship between the responses under the two modes is "global" rather than "item specific," the R approach should not be expected to identify it. As we view the data, we will observe the rather capricious behavior of R. It was generally felt that it would be unreasonable to expect that RBET would be as large as RIN; and , therefore, the following five combinations were selected for simulation.

Combinations of RIN and RBET

Used in the Simulation

Simulation	RIN	RBET
1	.5	.0
2	.5	.3
3	.7	.0
4	.7	.3
5	.7	.5

Thus, two null and three non-null conditions were simulated. Although some variation was evident in the data regarding the item means, all were taken to be 4.0 and the item standard deviations for all items were set at 1.0 in reasonable accord with the data. Although the data were not checked in this regard, a pseudo random normal generator GAUSS, an SSP subroutine, was used to generate the data. In keeping with the size of the original study, data sets for 50 subjects were generated in the following fashion:

1. The parameters of number of variables (NVAR) means, standard deviations, RIN and RBET, were read in from control cards
2. The full 2xNVAR correlation matrix was developed and factored using a Cholesky factorization algorithm
3. Standard normal pseudo random vectors were generated and multiplied by the factored correlation matrix
4. Finally, the results were rescaled, translated and rounded off to whole numbers in the range 1-7.

The above procedure was repeated 50 times and the cannonical correlation routin , CANOR of the SSP Subroutine Library, was used to compute the first cannonical r and evaluate its significance with an approximate chi square procedure described by Cooley and Lohnes (1962). These results were tabulated to give power estimates for .05 and .01 percent level tests. There was no permutation of the data in this phase of the study. The results will be found in the summary table at the end of this section.

The principle underlying this section is a rather simple one and represents, what these researchers believe to be, a rather unique analytic strategy for the problem at hand. Given the sample of 50 pairs of vectors, $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_i$ say, the problem is to determine whether they are associated. The null hypothe-

of no association suggests that the match between $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_i$, the two vectors for the ith subject, should be no better on the average than the match between $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_{i'}$, ($i \neq i'$). The task, then, is to generate the empirical 'null' distribution by randomly matching an $\underset{\sim}{X}$ with a $\underset{\sim}{Y}$, computing the selected measures of similarity or distance, and casting these into a distribution. Although for 50 subjects there are 50x50=250 random matches, a sample of 500 such random matches was judged to be sufficient to obtain the empirical distributions. The distributions were formed by sorting the generated values and developing percent distributions based on the unique values. Attempts to apply a fixed number of class intervals proved not to be feasible due to the variability in the 'richness' of the distributions under consideration. (We use the term richness to indicate the degree of continuity in the distribution.) Some data on this will be mentioned later. After the empirical distributions were obtained, the statistics for the 'correct' matches, i.e., $\underset{\sim}{X}_i$ and $\underset{\sim}{Y}_i$ were computed and their percentile ranks in the null distributions were computed. Because of the discrete nature of these distributions, especially for D which does not produce 'rich' distributions, it is important to develop the distributions in terms of the unique values generated rather then to use class intervals. Percentile ranks were calculated in the usual fashion taking the percent of cases below the value plus one-half of the percent of cases at the specific value. This was felt to be in reasonable agreement with the so called 'significance level' being the probability of a result this far into the tail or further. (Of course, an argument for a more conservative approach could be made.)

A well known method of combining the results of independent experiments is to sum the values of $-2x\ln(P_i)$ where the P_i are the tail probabilities associated with the independent experiments (Johnson and Johnson, 1959). Each

of the quantities in the sum is approximately a chi square random variable with 2 degrees of freedom. Although the independence of the various statistics calculated for the 50 subjects in this study would be difficult to justify, this approach was used so that the resulting statistics, each based on $\Sigma -2x\ln(P_i)$, were compared to the chi-square distribution with 100 degrees of freedom.

With only 3 variables, there is little variability in the data and, therefore, few unique values for any of the three statistics calculated. There are typically about 12 unique values of D and 24 values of E. The number of unique values of R varies from 15 to 20 for RIN = .5 but is reduced to about 7 unique values when the restrictions concomitant with RIN = .7 are applied to the data. Statistics D and E are quite consistent for the cases considered. There are approximately 3xNVAR unique values of D and twice as many unique values of E. Because of the way the information is processed in obtaining R, it is the 'richest' distribution for 5 and 10 variables but its distribution is reduced when RIN = .7 relative to when RIN = .5. For example, there were 333 unique values of R among the 500 permutation of 10 variables for RIN = .5, RBET = 0.

Unfortunately, in order to fully investigate the permutation approach under discussion, the simulation experiment described above would have to be carried out a number of times in order to gain insight into its properties. Although this was done in a limited fashion (50 replications) for the cannonical correlation, a similar number of replications for the permutation approach was not practical. For each set of parameters, however, two independent replications were run and for those cells where the results of the two runs were inconsistent, further attempts were made to more clearly identify the

properties of the procedure. The following table, then, presents a summary of the main analyses of the procedures described above.

- - - - - - - - - - - - - - - - - - - -

Table 4 about here

- - - - - - - - - - - - - - - - - - - -

RESULTS

First of all looking at the results for the chi square test of the first cannonical correlation, we observe that in all cases the results for the null models are in close agreement with the nominal alpha values of .05 and .01. For the three non-null models we see that the empirical power is lowest for the (.7, .3) models, larger for the (.5,.3) models and reaches a higher (and quite acceptable) level for the (.7,.5) models. In virtually every case, increasing the number of variables reduces the power. The power for alpha of .05 is quite unacceptable (less than .5) for all (.7,.3) models and the (.5,.3) models for 5 or more variables. For the other four simulations, i.e., the (.7,.5) models and the (.5,.3) model for 3 variables, the power is in the .60 or better range for 5% level tests.

The permutation results tend to follow the same patterns although comparisons are rather difficult due to the fact that only two replications are replications are reported whereas we have 50 replications for the cannonical correlation analysis. The results reported in the table are tail probabilities associated with the accumulated chi square statistic and are reported separately for each of the two replications. Therefore, they are not comparable to the power results for the cannonical correlation. The magnitudes of the tail probabilities for D and E are quite similar but may be quite divergent from the correspon-

ding value of R for the same data. In no cases were tail probabilities less then .1 for D and E for 12 independent replications of null models whereas the R probability is essentially 0 for two null 3 variable models. One reason for the instability of R for 3 variables appears to be that fact that a large number of extreme correlations are generated. For replication #2 for the (.5,0) 3 variable model nearly 40 percent of the permuted R's were -1.0. Thus the 17 'correct' matches which produced a -1.0, the most unsatisfactory value from a standpoint of positive association, were assigned a tail probability of .80. This is simply the largest value possible for the permution distribution generated. For the same data, the permutation distribution contained only 9% of R values of .0 so that the correct matches yielding this R value received a tail probability of about .045. The average tail probability assigned to the 50 correct matches was, therefore, .42 and the resulting chi square highly significant.

The R statistic was capable of correctly identifying a non-null association for the (.5,.3) 3 variable model. The only other non-null model for which R was significant was one of the two replications of the 5 variable, (.7,.5) model. Thus, the R statistic does not appear to have the desired properties.

Looking at non-null models we observe that D and E are quite similar in performance with E having the smaller of the tow probabilities about 80% of the time (11 of 13 where any difference exists). The E procedure provided consistent significant (.05 level) results for three of the nine non-null models and the D statistic performed almost as well. For each of the three 10 variable non-null models and the (.5,.3) 5 variable simulation E was significant on one of the two replications and D was generally also significant or

close. The cannonical r power at the 5% level for these four situations were estimated to be .22, .20, .68 and .42 so that the four significant replications in eight suggests the possibility of some improved performance of the permutation approach over the cannonical r for a large number of variables (upon closer scrutinity, they appear to be about the same). The permutation approach does not appear to identify the association very well when the association is strong (.7,.5 and .7,.3) and there are only 3 variables. This most likely has to do with the 'richness' issue since these two cells exhibit the minimum number of unique values (about 7,12 and 23 for R,D and E respectively). Three extra runs were made for the four cells in which the results from the two replications for E were inconsistent. The three additional runs for the 5 variable (.5,.3) cell produced one significant (.05) result for D and E so that 2 of 5 or 40% of the replications for E produced significance. This is in good agreement with the power estimate of .42 for the cannonical r approach. Additional runs for the 10 variable (.5,.3) model produced no further significant results and so the estimate for that cell is .20 based on five runs. (The cannonical correlation approach yielded a power estimate of .22 .)

For the 10 variable (.7,.3) model none of the three additional runs were significant (although all of the P's were less than .20) so, once again, the estimate is .20 based on five replications, the same as the cannonical estimate. For the 10 variable (.7,.5) model, two of the three additional runs were significant for both D and E yielding a power estimate of .60 based on five replications. This compares quite closely with the estimate based on cannonical r of .68. Thus, although of a very limited nature, these results are in relatively good agreement with those of the cannonical r.

In order to demonstrate that the power of the permutation procedure based on D and E, both of which appeared to have reasonable properties, was related to the size of the original sample, two samples of size 100 were generated for each of the three non-null (.7,.5) models. In each case both D and E produced results significant at the .01 level for each replication. Thus, although the power for samples of size 50 surely differs somewhat from that for the cannonical r, the permutation procedure (along with cannonical r, surely) appears to have reasonable power properties as n increases. In general E appears to be somewhat more powerful than D.

SUMMARY

Although in the specific situation described here a classical procedure for analyzing the data existed, the non parametric procedure outlined here appears to provide results that compete reasonably well with the cannonical r. One of our primary reasons for presenting these data, however, is to illustrate the relative simplicity of the technique. It is rather likely that most statistical consultants are requested, from time to time, to provide recommendations for data analyses where no such classical method is available. For these situations it may be possible that a procedure along the lines of the one presented here could be used. The programming is quite simple requiring a uniform random number generator to select random indices, a subroutine to compute the statistics desired and possibly another to form the empirical frequency distribution(s). The same subroutine can then be used to compute the statistics for the correct matches. Then their percentile ranks need to be calculated, transformed and accumulated. We feel that the procedure has sufficient merit to be included, possibly toward the bottom, but certainly in your 'bag of tricks'.

REFERENCES

Cooley, W.W. and Lohnes, P.R. Multivariate Procedures for the Behavioral Sciences, NY: John Wiley, 1962

Combs, A.W. Florida Studies in The Helping Professions, University of Florida Monographs, Social Sciences No. 37. Gainsville: University of Florida Press, 1969

Johnson, P. and Jackson, R. Modern Statistical Methods: Descriptive and Inductive, Chicago: Rand McNally and Co., 1959, 142-143.

Kruskal, J.B. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis", Psychometrika ; 29,1964a,1-27.

Kruskal, J.B., "Nonmetric Multidimensional Scaling: A Numerical Method", Psychometrika, 29, 1964b, 115-129.

Naylor, J.C. and Schenck, E.A., "Rho_m as an 'Error Free' Index of Rater Agreement", Educational and Psychological Measurement, 26, 1966, 815-834

Naylor, J.C., Dudycha, A.L. and Schenck, E.A., "An Empirical Comparison of Rho_a and Rho_m as Indices of Rater Policy Agreement", Educational and Psychological Measurement, 27, 1967, 7-20

Scientific Subroutine Package, IBM, System/360 Scientific Subroutine Package, Version III Programmers' Manual, 360-A-CM-03X, IBM, Fifth Edition , 1970

Taylor, J.B., "Rating Scales as Measures of Clinical Judgement: A Method for Increasing Scale Reliability and Sensitivity", Educational and Psychological Measurement, 1968, 28, 747-766.

Taylor, J.B., Hoefele, E., Thompson, P. and O'Donoghue, C., "Rating scales as Measures of Clinical Judgement II: The Reliability of Example Anchored Scales under Conditions of Rater Heterogeniety and Divergent Behavior Sampling", Educational and Psychological Measurement, 30, 1970, 301-310.

Taylor, Jane A., "Early Recollections as a Projective Technique", Unpublished Ph.D. Dissertation, University of South Carolina, 1973

The writers wish to express thanks to:

(1) Dr. Jane Taylor for making her data available for use in this paper and

(2) The University of South Carolian Computer Services Division for providing computer time for the simulations

TABLE 1

CHI SQUARES COMPUTED FROM THE PROBABILITIES
OBTAINED BY COMPARING THE DEVIATIONS BETWEEN
TWO SETS OF RESPONSES ON THE PERCEPTUAL CHARACTERISTICS
RATING SCALE TO THE COMPUTER-GENERATED
"RANDOM-MATCHED" DISTRIBUTION

Subscales	Perceptual Characteristics Rating Scale Categories	Chi Squares				
		Rater 1	Rater 2	Rater 3	Rater 4	df
1	General Perceptual Orientation (2 items)	113.29	173.94**	197.65**	86.88	94
2	Perceptions of Other People (4 items)	146.97**	62.58	96.69	115.79	94
3	Perceptions of Self (5 items)	96.39	128.54*	102.14	92.26	94
4	Purposes of Behavior (3 items)	131.98*	163.98**	138.06**	135.91**	94

* P<.05
** P<.01

TABLE 2

AVERAGE INTER-ITEM CORRELATIONS WITHIN AND BETWEEN MODES OF PRESENTATION BY RATER AND SUBSCALE OF PCRS*

	Average Within ER				Average Within VT				Average Between ER & VT			
Rater / Subscale	1	2	3	4	1	2	3	4	1	2	3	4
1 (2 items)	.30	.42	.47	-.16	.23	.23	.55	.25	.18	.06	-.05	-.03
2 (4 items)	.77	.21	.72	.51	.70	.23	.53	.66	.04	.03	-.07	-.15
3 (5 items)	.64	.47	.50	.18	.72	.50	.47	.56	.04	.06	-.05	-.08
4 (3 items)	.57	.32	.36	.24	.39	.28	.21	.51	.19	.13	.05	-.11

* Arithmetic averages were felt to be sufficient for the purposes of this presentation although, in general, averages based on Fisher's z transformation are preferred.

TABLE 3

FIRST CANNONICAL CORRELATIONS AND SIGNIFICANCE BY RATER AND SUBSCALE OF PCRS

		Rater			
Subscale		1	2	3	4
1 (2 items)	CANR	.30	.29	.16	.25
	χ^2(4df)	5.0	4.6	1.2	3.0
	Prob	.29	.34	.88	.56
2 (4 items)	CANR	.45	.46	.45	.37
	(16df)	14.0	16.8	16.3	11.5
	Prob	.60	.40	.43	.78
3 (5 items)	CANR	.60	.49	.45	.47
	(25df)	34.1	25.6	15.5	19.8
	Prob	.10	.43	.93	.76
4 (3 items)	CANR	.55	.49	.40	.35
	(9df)	19.4	17.2	12.7	9.8
	Prob	.02	.04	.17	.36

TABLE 4

Results of Statistical Significance for Four Approaches to Analyzing Data for Samples of 50 Subjects for Five Different Data Models

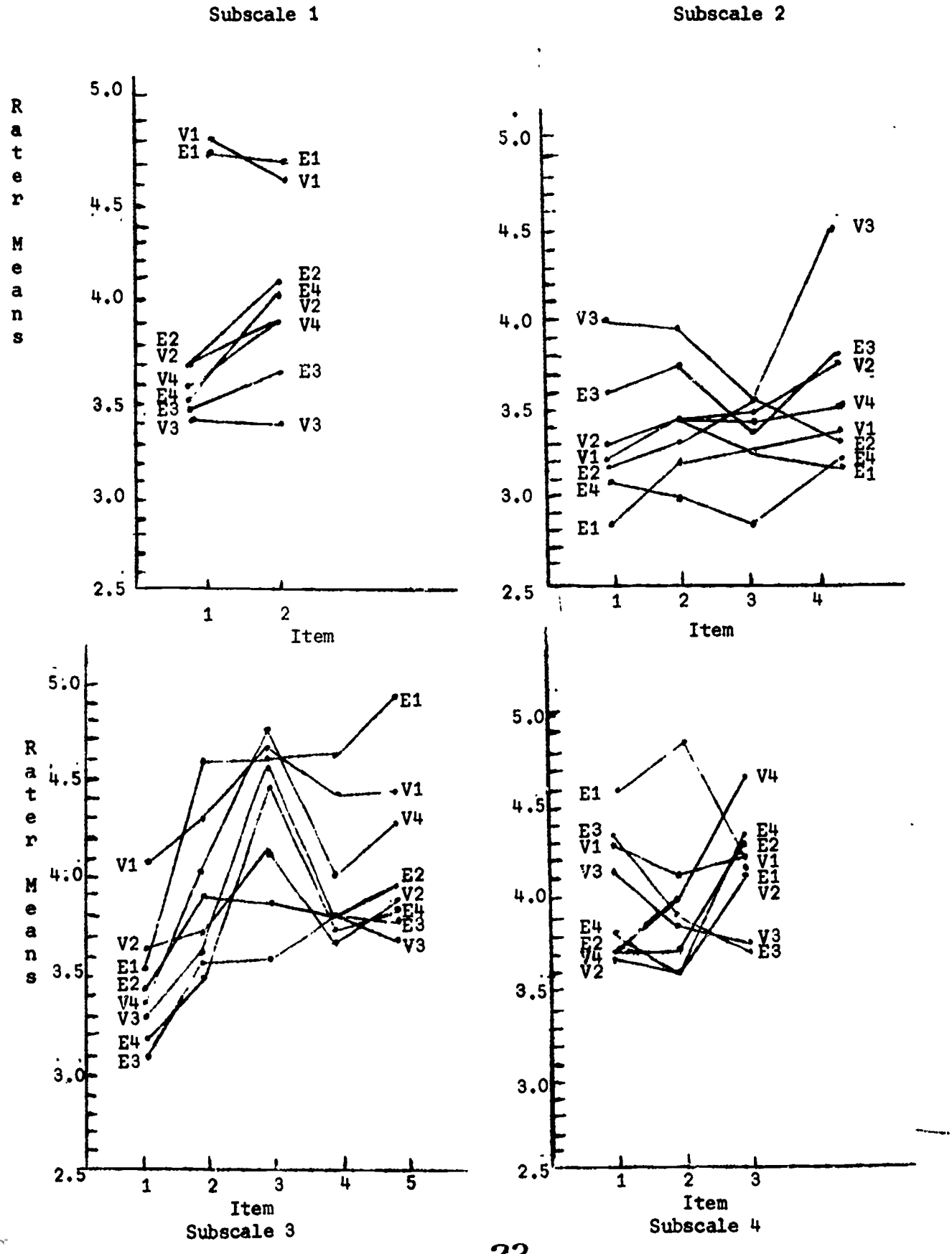
DATA MODEL	RIN	RBET	Number of Variables											
			3											
			CAN r**	+500 Permutations			CAN r	500 Permutations			CAN r	500 Permutations		
				R	D	E		R	D	E		R	D	E
	.5	.0	08 (05)	43	94	93	06 (05)	48	87	75	06 (05)	29	21	26
			02 (01)	00	65	61	00 (01)	25	22	22	02 (01)	29	15	16
	.5	.3	64 (05)	00	02	02	42 (05)	30	50	34	22 (05)	34	01	01
			20 (01)	02	04	04	22 (01)	36	06	05	12 (01)	78	41	36
	.7	.0	08 (05)	47	72	63	04 (05)	62	23	18	06 (05)	38	95	96
			00 (01)	00	77	79	02 (01)	49	89	90	00 (01)	81	10	10
	.7	.3	38 (05)	13	46	40	28 (05)	25	06	05	20 (05)	71	71	64
			12 (01)	19	51	44	12 (01)	50	06	03	06 (01)	32	00	02
	.7	.5	94 (05)	13	25	18	88 (05)	03	00	00	68 (05)	69	03	03
			80 (01)	75	06	09	66 (01)	93	07	03	42 (01)	97	25	14

* Decimal point omitted in body of table.

** Results for cannonical correlation coefficient; values reported are numbers of significant results using $\alpha = .05$ and $\alpha = .01$ for the chi square test of the first cannonical correlation.

+ Results for two replication, each based on 500 permutation of the raw data for 50 subjects.

FIGURE 1
SUBSCALE MEAN PROFILE BY RATER AND MODE OF PRESENTATION OF PCRS



the null distribution of r would be exceedingly small but more importantly (2) the r distribution would be highly dependent upon the average item responses. For example if the items in a scale differed considerably in mean response (regardless of stimulus mode) the responses under the two modes could reflect an artificial correlation even though there was no actual association between the item responses made by a rater under the two modes. Thus, this approach will not be considered further. Although multidimensional scaling might initially appear to be useful, it is not appropriate for these data. (See, e.g., Kruskal, 1964a, 1964b.)

The final approach to be considered here will be to compute various measure of the similarity of or distance between the two points (the sets of responses under the two modes) in p dimensional space where p is the number of responses under each mode. The problems with this approach are twofold: the researcher (1) must select from among various similarity and distance measures and (2) must obtain the null distribution of the selected distance measure. In addition to the correlation, the common measures that come to mind are the ordinary Euclidean distance and the Mahalanobis distance. The former seems quite reasonable but the validity of the latter is somewhat questionable for this situation. The two vector random variables entering into the Mahalanobis distance are assumed to be independent and have a common dispersion matrix. The first requirement is definitely not true although the second might be reasonably well satisfied. If the two sample covariance matrices were found to be similar possibly some overall estimate would provide useful results. For these reasons and others described later the Mahalanobis distance was not used in this research. Although not generally categorized as a distance