DOCUMENT RESUME

ED 143 686 TM 006 439

AUTHOR Fugita, Stephen S.; And Others
TITLE Some Multivariate Conceptualizations in Nonverbal Research.
PUB DATE 23 Apr 77
NOTE 25p.; Paper presented at the Ohio Academy of Sciences (Columbus, Ohio, April 23, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Behavioral Science Research; *Mathematical Models; *Nonverbal Communication; Predictor Variables; Research Criteria; *Research Design; *Research Methodology; *Statistical Analysis

ABSTRACT

According to the authors, statistical techniques should accurately reflect the research question of interest; no statistical technique should be used mechanically. A researcher should employ the statistical models most capable of answering the research question. If the true research question is concerned with predicting a variety of dependent variables simultaneously, then no univariate model is capable of reflecting that research question. If a univariate technique is used in such cases, a Type VI error is being committed, i.e., there is an inconsistency between the research question and the question as reflected by the statistical model. Several multivariate statistical procedures are outlined; these include canonical correlations, discriminative analysis, path analysis, component analysis, analysis of covariance, multiple regression analysis, and Q factor analysis. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED143686

Paper presented at Ohio Academy of Sciences, April 23, 1977, Columbus, Ohio.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS ~~COPYRIGHTED~~ MATERIAL HAS BEEN GRANTED BY

Stephen S. Fugita

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Some Multivariate Conceptualizations in Nonverbal Research

Stephen S. Fugita, Isadore Newman,
Mark Hogrebe, and Mike Smiley
University of Akron

TM006 439

ERIC

Some Multivariate Conceptualizations in Nonverbal Research

The topic of nonverbal behavior has become very popular during the past decade. Even though Darwin and Freud pointed out its significance over half a century ago, it has not been until fairly recently that it has attracted a significant amount of attention from either the lay public or social scientists. Not only have popular treatments such as journalist Julius Fast's Body Language been published, but also, for example, almost every new social psychology textbook contains a section discussing nonverbal behavior. It has become a major established research area in communication, social psychology, clinical/counseling psychology, and education.

Unfortunately, due to the subtleness, speed, and complexity of patterns involved in nonverbal communication, some difficult methodological and statistical issues must be surmounted. One of the characteristics of the previous research (as in most of the social sciences) is its nomothetic, univariate approach. As Secord (1976) has recently pointed out, such an approach has historically resulted in accounting for a relatively trivial proportion of the variance. Moreover, it is unlikely that real people respond to each other's nonverbal cues in such a manner. Occasionally, individual differences are included in the research design. As might be expected, the data indicates that there are very large individual differences with nonverbal variables. Another obvious need is research of a sequential nature. Most of the studies in nonverbal behavior are of the "snapshot" variety which capture only a "slice" of an ongoing behavior stream. Clearly, the sequential nature of social interaction is lost with this approach. Recently, Bakeman and Dabbs (1976) have suggested a number of useful ways to examine sequential patterns and cycles utilizing transitional probabilities.

An example of the utility of this approach has been provided by Stokes in Bakeman and Dabbs (1976). He studied the conversational patterns among 33 pairs of female undergraduates. Each dyad spent approximately three minutes agreeing on a topic and three minutes disagreeing on another topic. Observers recorded the onset and offset of each subject's looking at and talking with her partner. A computer program transformed this information into a form which showed the state of talking and looking within each pair at each half-second interval.

One analysis that he performed was examining the looking and talking patterns that can occur within a subject. There are four mutually exclusive and exhaustive combination events of talking and looking (look only, talk only, look and talk, neither).

Several results are clear from an analysis of the transitional probabilities:

> Several points are suggested by Stokes' data. First, subjects tended to look at their partners while listening. "Look only" (.40) was more likely than "neither" talk nor look (.15), apparently indicating a pattern of attentive listening. Second, subjects tended not to look while talking. The simple probability of "talk and look" (.25) was less than would be predicted from the probability of looking (.65) times the probability of talking (.46), and this was true for 59 out of 66 subjects ($p < .001$ by sign -

test). And third, subjects tended to look away before speaking. The "neither" state was slightly less likely to transition to "talk only" than to "look only" (.14 versus .18), but because "talk only" was much less probable than "look only" (.21 versus .40), the expected value of the transition to "talk only" was less and *z*-scores were higher for transitions to "talk only" than to "look only" for 51 out of 66 subjects ($p < .001$). This indicates a disproportionately high probability that a person who is looking away silently will begin talking at the next moment.

The experimental conditions of agreement and disagreement had several effects. For example, the probability that "look only" would continue as an unbroken state (transitioning to itself) was greater during disagreement than during agreement for 45 out of 66 subjects ($p < .01$). The tendency not to look while talking, as described in the preceding paragraph, was greater during disagreement; the probability of "talk and look" fell further below its predicted value during disagreement than during agreement for 48 out of 66 subjects ($p < .001$). Thus, the normal patterns of looking and talking appear to have been exaggerated by disagreement, with listeners looking more and speakers looking less.

The authors feel that a multivariate conceptualization would be potentially helpful in furthering understanding of nonverbal behavior. Multivariate conceptualization is defined, for our purposes, as the use of more than two dependent variables and more than two independent variables, simultaneously, when analyzing research. The emphasis is being placed, not on the multivariate statistical technique per se, but rather on the "research question" that requires looking at sets of variables, simultaneously, which are needed to accurately reflect the conceptual or theoretical construct. For the above reasons, multivariate techniques will be discussed mainly in terms of what they conceptually measure and problems with their interpretations.

The following will be a discussion of specifically chosen multivariate statistics which were selected for their power and potential misinterpretations. The statistical procedures that will be outlined are: (1) Canonical Correlations, (2) Discriminative Analysis, (3) Path Analysis, (4) Component Analysis, (5) Analysis of Covariance and Multiple Regression Analysis, and (6) Q Factor Analysis.

Canonical Correlation. Canonical Correlation (C_R) is the general case of multiple linear regression. One has a number of criterion variables being predicted by a number of independent variables. For example, the researcher may want to predict smiling, eye contact, and interpersonal distance (the criteria) by a set of independent variables such as reported liking, sex, and race.

When one thinks about measuring a nonverbal display, it generally involves more than one variable (criteria). Because of this, C_R often has a great deal of intuitive appeal. C_R allows one to predict a multitude of criterion variables simultaneously.

However, there is a significant problem with the interpretation of C_R. In its calculation, it "adjusts" the criterion variables in such a way that it makes them independent of each other. Conceptually, this means that it subtracts the correlation (variability overlap) between them. Thus, C_R will take out all the variability in smiling that is common to eye contact and interpersonal distance, and the variable smiling becomes smiling. This is the smiling that is left over after all the common variability with eye contact and interpersonal distance is removed. The calculation of C_R will then do the same for the eye contact. Therefore, this variable becomes eye contact, which is all the variability due eye contact after the common variability between smiling, eye contact and interpersonal distance is removed. The same procedure is then followed for interpersonal distance where interpersonal distance becomes interpersonal distance. Therefore, the criterion is conceptually made up of residuals, probably not the criterion variables that the investigator thought he was measuring.

Multivariate Analysis of Variance (MANOVA). In a way very similar to C_R, MANOVA determines if there is a significant difference between two or more groups (treatments, etc.) on two or more dependent variables simultaneously. However, if one finds significance, one usually has to revert to univariate tests to interpret the results. Therefore, the apparent advantage may not be as beneficial as it originally seemed. However, as pointed out by Hummel and Sligo (1971) using MANOVA in conjunction with ANOVA controls for experimentwise probability pyramiding.

There are a few further comments that should be made about MANOVA:

1. There are a number of tests ofsignificance. Three of them are: (a) Wilks Lambda--Roy's Largest root criterion,

(b) Hotelling's trace criterion, and (c) step down F procedure.

It is important to realize that these three tests are <u>not</u> equivalent. Therefore, you may get slightly different results from each.

2. There should never be fewer dependent (criterion) variables than there are groups (treatments). However, the more dependent variables, the more difficult the results are to interpret.

3. The total number of subjects should never be less than two times as many as the number of dependent variables.

<u>Path Analysis and Component Analysis</u>. Frequently, a researcher is interested in inferring "cause" (explaining the effects) when he was unable to utilize a "true experimental" design. In this situation, there are two procedures, Path Analysis and Component Analysis, the researcher may find helpful. The purpose of these procedures is to help study the <u>relationships</u> between the independent variables of interest and the dependent variable. This is done to facilitate the explanation of the effects of the independent variable on the dependent variable.

a. <u>Path Analysis</u>. This method aids one in studying the direct and indirect effects of independent variables on the dependent variables (the assumed causal effect). It is important to keep in mind that Path Analysis cannot discover variables that are the "cause," but rather, it is used as a method to support or fail-to-support a theoretical causal model.

It is also important to keep in mind two points about Path Analysis. One is that there are important underlying assumptions that cannot be violated; such as, the relationships among the variables are generally considered additive, linear, and most

ERIC
Full Text Provided by ERIC

importantly, "causal." Two, different theoretical explanations can be given equivalent support from the same Path Analysis solution. All the Path Analysis solution can say is that the relationship between the variables is not inconsistent with a specific theoretical position. Thus, to utilize Path Analysis appropriately, it is necessary that one must first have some theoretical model. Unfortunately, this is oftentimes not the case in nonverbal research.

One of the underlying assumptions, as previously noted, is that the independent variables are orthogonal (not correlated). This is an absolutely necessary assumption for the appropriate interpretation of Path Analysis.

The path coefficients are, in reality, the beta coefficients in a regression equation. It has been well documented and widely known that beta weights are only interpretable when the variables are independent. Thus, if the variables are not independent, then the beta weights (path coefficients) are not interpretable. That is, when multicolinearity exists between independent variables, the beta weights are highly unstable and will fluctuate greatly between samples.

When Path Analysis is used, the independent variables of interest are most frequently correlated (multicolinearity exists). Wolfe (1977) states that in most cases, Path Analysis has been found to have limited usefulness. It is our belief that this "limited usefulness" is a result of the violation of the assumption of orthogonality of the independent variables. If this is the case, then multivariate procedures can be used to produce orthogonality of the independent variables. This could be done using such techniques as orthogonal factor analysis, canonical correlation, component analysis, and part and partial correlation to account for unique variance. It is believed that obtaining orthogonality would increase the usefulness and interpretability of Path Analysis, thereby aiding the researcher in understanding the phenomena under investigation.

b. <u>Component Analysis</u>. The following discussion is based on a paper presented by Newman and Newman (1975).

The Component Analysis procedure was developed in the late '60's to aid researchers in explaining and interpreting the results of statisical analyses in which the predictor variables are not independent (nonorthogonal). If the variables are interelated, as are intelligence, socio-economic status, and race, it is difficult to accurately estimate the relative importance of each predictor variable to the criterion. Darlington (1968), Mood (1969, 1971), McNeil, Kelly, and McNeil (1975), and Kerlinger and Padhazur (1973) clearly delineate the various aspects of this problem.

Component Analysis (C_p) is a procedure which divides the proportion of variance accounted for into common and unique variance. The unique variance (U_q) is the proportion of variance attributed to a particular variable when entered last in the regression equation. It is what Bottenberg and Ward (1963) and McNeil et al. (1975) call the proportion of variance attributed to a particular variable, above and beyond the variance accounted for by the other independent variables in the equation (analysis of covariance, semipartial correlation). Therefore, the unique variance accounted for is represented by a full model which contains all the independent variables tested against a restricted model in which all the predictor variables are represented except for the one(s) for which the unique variance is to be estimated.

Common variance (Cv) may be conceptually thought of as the degree the overlap of correlated variables which is predictive of the criterion. It must be independent of unique and other common variance. In an example with three predictor variables, there are three sets of unique variance [$Uq_{(1)}$, $Uq_{(2)}$, $Uq_{(3)}$], three sets of second order commonality [$C_{(1,2)}$, $Cv_{(1,3)}$, $Cv_{(2,3)}$], and one third order commonality variance [$Cv_{(1,2,3)}$].

The number of independent components in a component analysis procedure can be determined by the equation:

$$2^N - 1$$

where: N = number of predictor variables.

Therefore, if one had four predictor variables, the number of components would equal:

$$2^4 - 1 = 15.$$

Since there are four predictor variables, there will be four components of unique variance (Uq), six components of second order common variance, four components of third order common variance, and one component of fourth order common variance. These components are additive and when summed will equal the total proportion of variance accounted for by the R_f^2 of the full model.

When there are four predictor variables, there will be 15 components. One can easily see the horrendous number of R^2's that have to be calculated for just four predictor variables in the full model. However, in using multiple regression, the investigator frequently has many more than four predictor variables.

Therefore, the number of components can easily become impractical to handle. This problem will be discussed later. For further details on how to calculate Component Analysis, see Mood (1969; 1971), Kerlinger (1973), and Houston and Bolding (1975).

If one has a variety of nonorthogonal predictor variables and a variety of F-tests are used to determine if any one or set of these predictor variables are significant, then one is violating the underlying assumption of independence. Therefore, the probability associated with the F-test is inappropriate. That is, one would actually find more significant F's than is indicated by the probability associated with that specific F. Component analysis divides the sum of variances into independent partitions. Therefore, the F of any of these partitions is independent.

The following are some of the limitations one should be sensitive to when using component analysis:

1. An integral part of component analysis is the concept of Uq. Uq is operationally defined as:

   variance accounted for by a variable when entered last in a multiple regression equation.

Therefore, the Uq depends upon and is affected by the variables that are already under investigation. Even though the Uq is independent, <u>in the set of variables</u> for that <u>sample</u>, the variable is not independent.

2. As the number of predictor variables increases, the number of components generated increases rapidly. So, if one has a large number of predictor variables, it may become impractical to calculate component analysis.

3. As the number of predictor variables increases, the number of higher order commonality components also increases. Just as it is difficult to interpret higher order than third order interactions in traditional analysis of variance, it is also difficult to interpret higher order than third order commonalities.

4. In examining some of the formuli for calculating the commonality components, one becomes sensitive to the possibility that some of the components can easily account for a negative proportion of variance. When this situation is encountered, it becomes very difficult to interpret or make conceptual sense out of the analysis.

5. With any non-manipulative research technique, "causation" cannot be assumed. A causal relationship can only be assumed in situations that have a true experimental design, i.e., a situation in which the experimenter has clear control of the independent variable. Since one of the major purposes for calculating component analysis is to attempt to improve the explanation of ex post facto research designs, this can lead one to <u>mistakenly</u> believe that the Uq accounted for by an independent variable with a criterion is of a causal nature.

6. Mood (1971) stated an important limitation one should consider. The unique variance (Uq) accounted for by an independent variable can change radically from situation to situation. However, the Uq attributed to a factor that the variable is a part of is not likely to change. Therefore, Mood suggests that the variables should be grouped based on the underlying concept they seem to be measuring. This would produce a more stable estimate. This grouping process will also have a side benefit of reducing the

total number of predictor variables making the component analysis much more manageable. However, if one uses the procedure suggested by Mood, the weighting of each variable becomes a problem. Do the factors account for the same 100% of the proportion of variance accounted for when each variable is used separately? If not, one is losing potentially significant information. Finally, it is difficult to decide which variables should go together. Quite often, variables that look as if they are measuring the same underlying construct are not.

Factor multiple regression is a procedure that may circumvent some of these problems (Massy, 1965; Duff, Houston, & Bloom, 1971; Connett, Houston, & Shaw, 1972; Newman, 1972). It is a method that enables one to empirically determine the factors with which the variables are associated. If one calculates the factor scores for each factor and uses an orthogonal rotation such as Varimax, then by definition, the predictor set of variables will be orthogonal. Thus, one can easily determine the relative importance of each factor by examining its beta weight.

Duff et al. (1971) found that principle component factor analysis with Varimax rotation and an eigen value of one as a factor cutoff produced empirically determined factors which were very similar to the factors they subjectively determined. These subjective factors were formed by selecting subsets of their predictor variables which seemed to be measuring the same underlying constructs. (This is similar to what was suggested by Mood, 1971.) The advantages of using empirically determined factors as predictor variables in a regression equation are discussed by Connett et al. (1972) and some limitations of this procedure are discussed by Newman (1972).

It is the authors' opinion that the factor regression approach may be more appropriate than component analysis where one is interested in determining the unique variance accounted for, especially when the number of predictor variables is relatively large, and there is a minimum of 10 subjects for every variable. However, if one is interested in commonality, the factor regression procedure is not appropriate. In this case, if there is a large number of variables and subjects, it is possible to use factor analysis with oblique rotation. This procedure will condense the large number of variables into factors which can be used as a new set of predictor variables. Since these factors may be oblique (correlated), one may then wish to perform a component analysis which will yield estimates of the unique and common variance attributed to the factors. Obviously, the oblique solutions lack many of the desirable characteristics which make the orthogonal solution easier to interpret. However, there are times when a researcher may be interested in the common proportion of variance attributed to factors which are theoretically and empirically related.

Analysis of Covariance (ANCOV)

ANCOV is generally used when the design cannot or did not control for a specific attribute such as subject eye contact, talking, etc. Frequently, in nonverbal research, it is difficult if not impossible to control, for example, all of the confederate's nonverbal responses. This is because some semblence of "naturalness" is necessary. Hence, ANOCOV may be useful to help take some of these factors into account. Investigators tend to use ANCOV to "equalize" subjects (Groups, Treatments) on a specific attribute, i.e., ANCOV is used to "adjust" for the effects of initial differences.

There are certain underlying assumptions (in addition to all the assumptions of ANOVA) such as all subjects being randomly assigned to treatments. However, one wants to use ANCOV most when it is not possible to randomly assign subjects to treatments. The key assumption is homogeneity of the regression slope. This simply means that there is no interaction between the covariate and the independent variable(s). If there is interaction, one should not covary because the results could be very misleading. One should look at the interaction instead (simple effects).

Regression

With the increasing use of multiple linear regression, which is unfamiliar and/or misunderstood by many, it is important to clarify the reasons why multiple linear regression is an appropriate, and in many cases, a preferable procedure. This section will present some of the general arguments which can support the use of regression.

The F-test, which is the analysis of variance, is a statistical technique (a test of significance) and is calculated on the basis of a least square solution. It has unfortunately been confused with what has become known as traditional one-way or factorial analysis of variance. The traditional analysis of variance approach tends to confound, in the researcher's thinking, the statistical procedures with the research design. However, if one separates the two, some of the advantages of the regression hypothesis testing procedures become more obvious. A few of these advantages are:

a. Multiple Linear Regression (MLR) is the general case of the least sum of squares solution. Chi squares, t and F tests are all calculated on the basis of one least squares solution.

b. A significant F in a factorial design is more difficult to interpret. It may not reflect your specific hypothesis. With the regression procedure, one states the hypothesis and then writes the regression model to test that hypothesis. Thus, every test of significance is a test of a specific hypothesis.

c. Regression is more flexible because it allows the researcher to write models that specifically reflect the research hypothesis.

d. With traditional analysis of variance, one can only ask interaction questions that have categorical variables interacting with categorical variables. With regression, one can ask interaction questions between categorical variables, between categorical and continuous variables, or between continuous variables. Since regression can deal with both categorical and continuous variables, it is more flexible in its ability to reflect actual behavioral processes. With regression, there is no need to categorize variables that are continuous in nature as required by traditional ANOVA; therefore, one does not lose degrees of freedom or power (McNeil, Kelly, McNeil, 1975; Kerlinger, 1973; Newman, 1976).

e. All analyses of covariance procedures are really regression procedures because the covariate(s) are always held constant by regressing it on the criterion. The multiple linear regression procedure makes the covariance procedures easier to calculate and interpret (Kerlinger, 1973).

f. Regression also facilitates the calculation and interpretation of trends (functional relaitonships). Trends which are continuous in nature must be categorized when traditional analysis of variance is used. Since regression can deal with continuous variables, no artificial categories must be imposed.

g. The researcher, particularly the applied variety, must often deal with unequal N's and nonorthogonal designs. When these problems occur and one is using traditional analyses of variance, a correction is required. All of the corrections that produce the exact solutions are regression procedures (Newman, Deitchman, Burkholder, Sanders, Ervin, 1976). In other words, if one has unequal N's, and traditional analysis of variance is being used, regression will have to be used regardless of whether the researcher is aware of it or not. Once again, regression is more flexible.

One of the problems with multiple linear regression is that multiple linear regression hypotheses testing procedures may be confused with non-hypothesis testing ones and with stepwise regression. Most of the critical comments leveled against regression are for the non-hypothesis testing procedures which tend to produce either inflated R^2s and/or more significances by chance than the stated alpha level. These spurious results tend to be generalized even though they cannot be replicated. This problem is less severe with the hypotheses testing regression procedure, and basically becomes nonconsequential when cross-validation and multiple correction procedures are employed. Unfortunately, these procedures are rarely used.

Before a research project is actually initiated, one should ask which and how many criterion are appropriate to the problem. The researcher must also know what reliability and validity estimates exist for the criterion and if the estimates meet minimum requirements. A reliability estimate of .65 for group prediction and .85 for individual prediction is a minimum. There are no easy rules of thumb for validity estimates. Each situation must be examined separately.

In nonverbal research, if a regression approach were utilized, it would be much easier to designate which nonverbal cues add a significant amount of information in a particular situation. This would be preferable to the usual listing of cues which were simply statistically significant as indicated by ANOVA.

Q Factor Analysis. This technique, also known as profile or segmentation analysis, is perhaps one of the potentially most fruitful with regard to nonverbal behavior. It groups people on the basis of the similarity of their responses. Profile similarity can be calculated on the basis of three types of information: level, dispersion, and shape. If profile similarity is calculated on the basis of level, this would indicate that the profiles are similar with regard to the mean score of the variables used in the profiles. If dispersion is used to indicate profile similarity, this would indicate that the profiles are similar in terms of the amount of scatter around the average level (similarity between standard deviations). The third method involves analyzing the similarity of the shape of the profile. The shape is defined by the rank order of scores for each variable the individual has on the profile.

The major problem with using levels as a means of indicating similarities in profiles is that two people can be said to have similar profiles when their individual scores over a set of variables are totally different, but because of averaging, they have mean scores that approximate each other.

The major weaknesses with the dispersion method of profile analysis are that the dispersion method does not give an indication of level, and it is difficult to interpret profile dispersion for people because dispersion depends upon the correlation among the profile variables.

> If high positive correlation exists among the variables, people in general will tend to have small dispersions. If the correlations among variables are low, the dispersion will tend to be relatively large. If some of the correlations are positive and others are negative, the dispersion will even be larger (Nunnally, 1967, p. 374-375).

The difficulty that arises when one tries to calculate profile similarity on the basis of shape is that it is possible that two profiles can be identical and each individual's absolute rating can be quite different. This is true as long as each individual has the same rank order for the variables, since a perfect correlation will exist if the rank orders are the same even though the absolute values may be different.

The most desirable and accurate method for calculating profile similarity would be to take all three types of information into account simultaneously. This can be done by factor analyzing the cross-product matrix between the individuals on each variable. However, this demands a large computer storage capacity.

In nonverbal research, examining similar levels may be appropriate, for example, if one were looking at potential subcultural differences. However, if one were conceptually interested in something like overall activity or reactiveness, probably dispersion would be a more meaningful measure. Finally, shape profile anlaysis would probably be most appropriate if one were attempting to detect predicted changes in the patterning of nonverbal cues, for example, in deception situations.

One of the major problems with Q-factor Analysis is the argument that the profiles tend to be sample specific and, therefore, cannot be generalized to other samples or to the population from which the sample comes. This problem can be somewhat alleviated by cross-validating the profiles. This requires obtaining two independent samples and checking to see if the same profiles replicate in each sample--to the extent that they do, there is no problem. It is suggested that one only uses profiles that are stable (stable meaning the ability to be replicated).

Another frequently mentioned problem when using profile analysis is that the larger the sample, the more profiles (typologies) obtained. So, the number of typologies is a function of sample size. As a rule of thumb, very rarely will there be more than five or six typologies that are likely to be replicable. Again, it is recommended that some type of cross-validation procedure be used to identify the most stable profiles before using and/or interpreting the reuslts.

The thrust of this paper has been that the statistical techniques should accurately reflect the research question of interest. No statistical technique should be used mechanically. A researcher should write the statistical models most capable of answering the research question. If the "true" research question is concerned with predicting a variety of dependent variables simultaneously, then no univariate model is capable of reflecting that research question. If a univariate technique is used in such cases, a Type VI error is being committed. A Type VI error is the inconsistency between the research question and the question as reflected by the statistical model (Newman, Deitchman, Burkholder, Sanders, & Ervin, 1976).

The authors would like to reemphasize that they are suggesting multivariate conceptualization and not necessarily the use of multivariate techniques. It is suggested that each researcher takes a critical look at his research question of interest, be sure that he knows what he is truly interested in ascertaining, and then select the statistical techniques (models) that are capable of reflecting the "true" question of interest.

References

Bakeman, R., & Dabbs, J. M., Jr. Social Interaction Observed: Some Approaches to the Analysis of Behavior Streams. Personality and Social Psychology Bulletin, 1976, 2, 335-345.

Bottenberg, R., and Ward, J. Applied Multiple Linear Regression, U. S. Deaprtment of Commerce, Office of Technical Services, AD413128, 1963.

Connett, W., Houston, S., and Shaw, D. The Use of Factor Regression in Data Analysis. Multiple Linear Regression Viewpoints, March, 1972, 46-49.

Darlington, R. B. Multiple Regression in Psychological Research and Practice. Psychological Bulletin, 1968, 69, 161-182.

Houston, S. and Bolding, J. Part, Partial and Multiple Correlation In Commonality Analysis of Multiple Linear Regression Models. Multiple Linear Regression Viewpoints, January, 1975, 36-40.

Hammel, T. J., & Sligo, J. R. Empirical Comparison of Univariate and Multivariate Analysis of Variance Procedures. Psychological Bulletin, 1971, 76, 49-57.

Kerlinger, F. Foundations of Behavioral Research, New York: Holt, Rinehard and Winston, 1973.

Kerlinger, F., & Pedhazur, 1973

McNeil, K., Kelly, F., and McNeil, J. Testing Research Hypotheses Using Multiple Linear Regression. Carbondale, Ill.: Southern Illinois University Press, 1975.

Massey, W. F. Principal Components Regression in Exploratory Statistical Research. Journal of American Statistical Association, March, 1965, 234-256.

Mood, A. M. Macro-Analysis of the American Educational System. Operations Research, 1969, 17, 770-784.

Mood, A. M. Partitioning Variance in Multiple Regression Analyses as a Tool for Developing Learning Models. American Educational Research Association, March, 1971, 191-202.

Newman, I. Some Further Considerations of Using Factor Regression Analysis. Multiple Linear Regression Viewpoints, October, 1972, 39-41.

Newman, I. Brief Note on the Justification for Using Multiple Linear Regression. Multiple Linear Regression Viewpoints, 1976, 6(4), 50-53.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R., and Ervin, L. "Type VI Error: Inconsistency between the Statistical Procedure and the Research Question." Paper presented to the World Population Society, Western Regional Meeting, Long Beach, California, Jan., 1976.

Newman, I. and Newman, C. A discussion of Component Analysis: Its intended purposes, strengths and weaknesses. A paper presented at the American Educational Research ASsociation, 1975; Multiple Linear Regression Special Interest Group.

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

Secord, P. F. Person Attributes in Social Psychology Theory and Paradigms. Representative Research in Social Psychology, 1976, 2, 140-146.

Wolfe, L. M. Path Analysis and causal models as regression techniques: A comment. Multiple Linear Regression Viewpoints, 1977, 7, 33-42.

Some Multivariate Conceptualizations

in Nonverbal Research

Stephen S. Fugita, Isadore Newman,

Mark Hogrebe, and Mike Smiley

University of Akron

ABSTRACT

The thrust of this paper has been that the statistical techniques should accurately reflect the research question of interest. No statistical technique should be used mechanically. A researcher should write the statistical models most capable of answering the research question. If the "true" research question is concerned with predicting a variety of dependent variables simultaneously, then no univariate model is capable of reflecting that research question. If a univariate technique is used in such cases, a Type VI error is being committed. A Type VI error is the inconsistency between the research question and the question as reflected by the statistical model (Newman, Deitchman, Burkholder, Sanders, & Ervin, 1976).

The authors would like to reemphasize that they are suggesting <u>multivariate conceptualization</u> and <u>not necessarily</u> the <u>use of multivariate techniques</u>. It is suggested that each researcher takes a critical look at his research question of interest; be sure that he knows what he is truly interested in ascertaining, and then select the statistical techniques (models) that are capable of reflecting the "true" question of interest.