A STUDY OF THE POWER OF MULTIVARIATE ANALYSIS OF
VARIANCE ON STANDARDIZED ACHIEVEMENT TESTING
WHEN ESTIMATORS FOR OMISSIONS UTILIZE MEAN
VALUE AND REGRESSION APPROACHES

By

STEPHEN S. SLEDJESKI

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1976

UNIVERSITY OF FLORIDA

3 1262 08552 7785

ACKNOWLEDGEMENTS

My appreciation is extended to the members of my doctoral committee for their contributions to the development of this dissertation. They are: Drs. Vynce A. Hines (Chairman), Ira J. Gordon, Zorin R. Pop-Stojanovic, and Robert S. Soar.

To Dr. Hattie Bessent, no statement can express her impact and assistance in attaining my educational goals. Words can be neither sufficient nor appropriate to express my esteem.

To Drs. Ann Bromley, Molly Harrower, and Wilson H. Guertin, I present thanks for direction and assistance in the understanding of my educational commitment.

To my sisters, Helen Brush and Ann Pendzick, and their families, I can but state our fortuitous interaction which has allowed not only educational growth but also complete dispersion while retaining faith in one another's existence.

To my mother, Helen Sledjeski, and my late father, Stephen Sledjeski, I wish to express my deepest appreciation for their successful development of a family unit filled with motivation, sincerity, trust, and love. This work is dedicated to their lives and memory.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
ABSTRACT	vi
 Chapter	
I. INTRODUCTION	1
Nature of the Study	1
The Problem and the Hypotheses	4
Significance of the Study	5
II. REVIEW OF RELATED LITERATURE	7
Introduction	7
Historical Overview	7
Problems of Missing Multiresponse Observations in Education	13
Direction of Present Research	14
III. DESIGN OF THE STUDY	15
Procedures	15
Method	17
IV. RESULTS	20
Comparison of the Mean Value and the Regression Estimated Data Sets with One Another and with the Complete Data Set at the 2½ Percent Level of Missing Subsamples	22
Comparison of the Mean Value and the Regression Estimated Data Sets with One Another and with the Complete Data Set at the 5 Percent Level of Missing Subsamples	24

TABLE OF CONTENTS—Continued

Chapter		Page
IV.	Comparison of the Mean Value and the Regression Estimated Data Sets with One Another and with the Complete Data Set at the 10 Percent Level of Missing Subsamples	26
	Comparison of the Mean Value and the Regression Estimated Data Sets with One Another and with the Complete Data Set at the 15 Percent Level of Missing Subsamples	28
	Comparison of the Mean Value and the Regression Estimated Data Sets with One Another and with the Complete Data Set at the 20 Percent Level of Missing Subsamples	30
	Further Results	32
	Summary	34
V.	DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS ..	36
	Discussion	36
	Conclusions	37
	Recommendations	39
	REFERENCES	41
	BIOGRAPHICAL SKETCH	45

LIST OF TABLES

Table		Page
1	F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of $2\frac{1}{2}$ Percent of the Complete Samples	23
2	F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 5 Percent of the Complete Samples	25
3	F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 10 Percent of the Complete Samples	27
4	F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 15 Percent of the Complete Samples	29
5	F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 20 Percent of the Complete Samples	31

Abstract of Dissertation Presented to the Graduate Council
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

A STUDY OF THE POWER OF MULTIVARIATE ANALYSIS OF
VARIANCE ON STANDARDIZED ACHIEVEMENT TESTING
WHEN ESTIMATORS FOR OMISSIONS UTILIZE MEAN
VALUE AND REGRESSION APPROACHES

By

Stephen S. Sledjeski

March, 1976

Chairman: Dr. Vynce A. Hines
Major Department: Foundations of Education

The efficacy of utilizing estimators for omissions in a multiresponse achievement data set which is analyzed using multivariate analysis of variance (MANOVA) techniques is the concern of this study. The estimates were determined employing mean value and regression methods.

Random samples of fourth- and fifth-grade students were administered the Stanford Achievement Test, Intermediate Level I and Intermediate Level II, respectively, in the spring of 1974. Each sample had a n of 193 consisting of two fixed groups as the independent variables and the achievement sub-scores as the dependent variables.

These two samples comprised the complete data sets from which random subsamples of missing data were removed

from among the dependent variables. The missing subsample consisted of $2\frac{1}{2}$, 5, 10, 15, and 20 percent of the complete samples, each percent level being investigated five times for each of the two methods of estimation.

The MANOVA results of the data sets with mean value and regression estimates were compared to one another and to the complete data set. The null hypotheses tested were:

- There is no difference in MANOVA results for the complete data set and the mean value estimated data set with the size of the missing subsample ranging from $2\frac{1}{2}$ to 20 percent of the complete data set.
- There is no difference in MANOVA results for the complete data set and the regression estimated data set with the size of the missing subsample ranging from $2\frac{1}{2}$ to 20 percent of the complete data set.
- There is no difference in MANOVA results for the mean value estimated data set and the regression estimated data set both with the size of the missing subsample ranging from $2\frac{1}{2}$ to 20 percent of the complete data set.

The hypotheses were analyzed by comparing the complement of the cumulative distribution function derived from the F-ratio of each MANOVA of the complete data set to that of the estimated data sets. No significant differences were found for the three hypotheses. Inspection of the results demonstrated that the regression estimates provide MANOVA results apparently closer to that of the complete data set than did mean value estimates.

The research concluded that, within the confines of this study, one cannot reject the use of mean value and

regression estimates for data sets with missing values which are to be analyzed using MANOVA.

CHAPTER I

INTRODUCTION

With the increased emphasis on multivariate analysis, the experimenter has been confronted with multiresponse data where measurements on all responses are not available for every experimental unit. Since the time, resources, and money involved in gathering multiple observations on experimental subjects are greater than for gathering single observations, multivariate analysis of variance (MANOVA) must give attention to missing data. It is the purpose of this study to consider missing observations in MANOVA utilizing mean value and regression estimators on a set of achievement data with subsets of randomly chosen missing data ranging in size from $2\frac{1}{2}$ to 20 percent of the complete data set. The power of MANOVA results will then be determined.

Nature of the Study

Missing data estimation has been of interest to educational and statistical researchers for several decades. Estimation of uniresponse data has been conducted for various experimental designs. Baird and Kramer (1960) investigated the balanced incomplete block design. They developed

formulas through minimization of the error sum of squares for the special case where missing values are within the same block or treatment. Their method facilitates calculations but does nothing to restore missing information.

Kramer and Glass (1960) examined the Latin square design. In the same manner as Baird and Kramer, they developed formulas through minimizing of the error sums of squares for several missing values to restore the balance of the design. The formulas are for the specific cases described and not for the completely general case.

Preece (1972) studied the two-way classification design. He developed a method of estimating block and treatment parameters from the nonmissing data plus the estimated data.

Mitra (1959) considered the effect of missing value estimates on the F-test in analysis of variance (ANOVA). He demonstrated that the numerator in F (the treatment mean square) and the denominator (the error mean square) cannot have the same expected value when missing observations exist.

An examination of various missing data procedures was performed by Wilkinson (1960). He put forth a method of solving for estimates through simultaneous equations and compares it to an iterative least squares method and a covariance method. His method is preferred since it requires fewer steps and gives the correct residual sums of squares directly.

Studies investigating multiresponse data estimators have been less numerous. The works of Kleinbaum (1970), Srivastava (1967), and Trawinski (1961) are some examples of early endeavors in multiresponse data. Kleinbaum looked at the effect of estimation upon hypothesis testing of generalized multivariate linear models. In concurrence with Mitra who investigated the uniresponse situation, he demonstrated that hypotheses are rejected with bias when utilizing estimators for missing values.

Srivastava extended the Gauss-Markov theorem to multivariate linear models.

Trawinski showed that it is not necessary to collect data on each characteristic of interest for each experimental unit. She brought out the important fact that in many situations one needs to have experiments where observations on some of the responses are missing not by accident, but by design.

The relevance and importance of missing observations were demonstrated by Srivastava and McDonald (1969, 1971). They established, under realistic conditions, the preference for the hierarchical incomplete models within the groups of general incomplete multiresponse models.

Dempster (1971) provided an overview of the problems involved. He surveyed a cross section of the developing topics in multivariate analysis of data concentrating on problems of pragmatic data analysis and not on technical and mathematical detail.

The Problem and the Hypotheses

The present investigation will attempt to determine the efficacy of two types of estimates of missing data in MANOVA. One type of estimate will be the mean value of the variable for a particular treatment; the other, the regression of one of the MANOVA dependent variables on the remaining dependent variables which then act as independent variables. The results of these MANOVAs will be compared to MANOVA results of nonmissing data. The hypotheses to be investigated are:

- H₁: There is no difference in MANOVA results for the complete data set and the mean value estimated data set with the size of the missing subsample ranging from 2½ to 20 percent of the complete data set.
- H₂: There is no difference in MANOVA results for the complete data set and the regression estimated data set with the size of the missing subsample ranging from 2½ to 20 percent of the complete data set.
- H₃: There is no difference in MANOVA results for the mean value estimated data set and the regression estimated data set both with the size of the missing subsample ranging from 2½ to 20 percent of the complete data set.

For each hypothesis, missing subsamples will be randomly chosen which will comprise 2½, 5, 10, 15, and 20 percent of the original complete sample. Each subsample percent level will be investigated five times. Estimated values will then be substituted and be subjected to MANOVA.

F-values from the MANOVA results will be compared using the cumulative distribution function to determine the power of the analyses.

Data used in the analysis will consist of achievement scores as determined on the Stanford Achievement Test collected in the spring of 1974. Two samples will be investigated: a fourth-grade sample of 193 students who were administered the Intermediate I Battery (eight variables) and a fifth-grade sample of 193 students who were administered the Intermediate II Battery (seven variables). The students in each sample were chosen at random from each of two fixed groups, an experimental group and a control group. For each MANOVA, the independent variables will be the two fixed groups.

Significance of the Study

The two types of estimators to be investigated differ from one another in an important sense. The mean value estimator considers all nonmissing values of a particular dependent variable for a specific treatment whereas the regression estimators consider only those experimental units with complete data. One approach attempts to utilize all possible data elements, and the other forms an estimation based on even less information.

Combining the fact of the two approaches with that of varying subsamples of missing data will provide a thorough look at omissions in multiresponse data taken from an educational setting. It is hoped that insights will be developed for future analysis of similar educational data.

This chapter has presented the problem to be investigated and the nature, significance, and hypotheses of the study. Chapter II contains a review of literature related to the problem of the study. The design and procedures are stated in Chapter III; the results of the study are in Chapter IV; and the discussion, conclusions, and recommendations are given in Chapter V.

CHAPTER II

REVIEW OF RELATED LITERATURE

Introduction

Missing data have posed a problem in data analysis for more than four decades. The initial investigations involving incomplete data sets concerned univariate statistical analysis. With the developments in computational technology in the past quarter century, multivariate data analysis has become feasible (Dempster, 1971) as has the investigation of missing data in multivariate analysis.

The initial focus of researchers concerned the techniques involved in the estimation of parameters when there existed missing observations in the data set. It was a question of developing the parameters and then adjusting these parameters considering the missing data. The direction taken in the review of the literature which follows is first, the estimation of the missing observations and second, the formulation of the parameters required for analysis.

Historical Overview

The first researcher to develop analysis procedures by first estimating values for the missing observations was

Wilks (1932). He examined the incomplete bivariate case of a bivariate normal distribution using sample means for the missing observations. He found that the optimum method of determining the variance between the two variables was the correlation between the two variables which included only those pairs that were complete.

Wilks' example of a sample of statistical data from a multivariate population has been popularized in many related papers. Srivastava and Zaatar (1972) summarized Wilks' example as:

[T]he situation when the experimental units are skulls that have been dug out from a certain graveyard. Since these skulls may be partly mutilated, the choice as to which characteristics should be measured on a particular unit is not entirely in the hand of the investigator. (One may suggest that in such a situation, we should restrict ourselves to those skulls on which all measurements of interest can be obtained. However, clearly this would in general not be very proper unless there were a rather large number of skulls free from any mutilation.) p. 117

Little more was published on incomplete multivariate data sets until the 1950s when papers began to appear extending the work of Wilks. Matthai (1951) developed a method to determine the correlation between two variates with missing data using the total available data set. He formulated a solution for the trivariate case using the correlation estimates. His estimates, he concluded, were inconsistent. For example, correlation coefficients could exceed unity.

Federspiel et al. (1959) and Glasser (1964) generalized this situation. They investigated the

correlation matrix of a general number of variates based on all available paired data. They studied intuitive approaches for estimating linear regression coefficients when an unspecified number and pattern of missing values exist among the independent values. It is shown that the efficacy of the approaches depends upon the correlations among the independent variables as well as the proportion of observations which are missing.

Lord (1955) demonstrated the solutions for the trivariate case when the dependent variable is recorded for all experimental units in the sample. Either of the two independent variables is recorded for all experimental units, but not both. He showed that, in this instance, means and regression coefficients can be estimated accurately.

The trivariate case was studied by Edgett (1956) in the opposite sense of Lord. He gave attention to the instance when the dependent variable has missing values and the two independent variates were complete. Nicholson (1957) extended Edgett's work to any number of independent variables. Edgett and Nicholson demonstrated that a maximum likelihood function for a plausible probability distribution could provide as good population parameter estimates as could least squares estimates.

A mode of estimation different from Wilks' method was provided by Dear (1959). He substituted for each

missing observation of an independent variate the division of the sum of the value of all observed independent variables by the sum of the number of observations for all observed independent variables. This somewhat corresponds to the grand mean of all the independent variables. It is clear that serious difficulties would be incurred when the independent variables are measured on different scales.

Walsh (1959) and Buck (1960) considered omission estimates in respect to paired simple linear regression. Walsh studied the utilization of all data available for a pair of variables in the simple linear regression computation. Those experimental units for which no data were missing were looked at by Buck in the paired regression analysis. Both Walsh and Buck determined that the average of values obtained from the simple linear regression provided suitable estimates for missing responses.

Anderson (1957) investigated a particular pattern of missing observations called a monotone sample. This is a sample in which the observations on each variate is a subset of another variate, i.e., each variate is nested within another variate. He set forth a method of estimation very similar to Edgett's although greatly simplified in the amount of necessary mathematical manipulation. Several writers (Bhargava, 1962; Afifi and Elashoff, 1966, 1967) have gone beyond the monotone trivariate case of Anderson and determined solutions for the general variate case.

In addition, Bhargava developed the likelihood ratio tests for hypotheses dealing with the linear model and equality of covariance matrices with multivariate monotone samples.

Trawinski and Bargmann (1964) examined a considerably more complicate pattern of missing data than Anderson (1957), Bhargava (1962), and Afifi and Elashoff (1966, 1967). The concern of Trawinski and Bargmann was with observations that were missing not by accident, but by design. They found that correlation coefficients were logically consistent estimates to use with incomplete multivariate data.

In deference to data missing by accident or design, Hocking and Smith (1968) assumed neither in developing their analytic procedures. They formulated a procedure to compute maximum likelihood estimates for parameters but only in the case of large samples.

Anderson, Trawinski and Bargmann, and Hocking and Smith used estimates of groups of data. They did not estimate specific missing observations.

The design of experiments which involve multiresponses and omissions was considered by Srivastava (1968). He pointed out that an experimenter must give attention to whether or not each response on each experimental unit is to be measured. He provides a discussion of what he calls the lack of need of a regular design. (A regular design is one where all responses are sought on all experimental units.) Before data collection, a researcher should set up his design such that the only data collected will be somewhat convenient or useful.

Haitovsky (1968) compared the methods of Buck and Walsh. He carried out a simulated data analysis, first using only complete data, discarding incomplete experimental units and second, using all available observations to estimate correlations. He found the former procedure superior. This is the case when the number of missing entries is not high.

A comparison of a complete data set and an incomplete data set which is a subset of the complete set was conducted by Morrison (1971). He determined that when the correlations between the complete and incomplete variates of the data set are small, the multivariate missing value estimates are less accurate in the estimation of the mean square error term than the multivariate data set with no estimates.

An extension of the work of Walsh and Buck was conducted by Dagenais (1971). He developed a more generalized method which not only corrects for data omissions but also provides for additional corrections during data analysis. His estimates are consistent when the independent variable is fixed; each observation contains a value for the dependent variable and at least one of the independent variables; and some observations are complete.

Srivastava and Zaatar (1972) dealt with the problem of classifying a future multiresponse observation into one of two populations given two incomplete multiresponse

samples, one from each population. They developed a rule for the classification given the fact that the observation did come from one of the populations.

Investigations of entire sections of missing data were performed by Hartwell and Gaylor (1973) and Rubin (1974). The former examined missing cells employing the method of unweighted means. He provides a method of cell estimation using estimated variances. Rubin looked at complete blocks of missing data by decomposing the original estimation problem into smaller estimation problems using a technique he denotes as "factorization." This consists of discovering those subject responses that are complete and using these response patterns to estimate missing observations of subjects with a similar response pattern.

Problems of Missing Multiresponse Observations in Education

In a paper which is an overview of multivariate data in education, Pruzek (1971) brought both the educational community and other areas of research face to face with the problem of incomplete multiresponse data sets and their investigation employing multivariate analysis of variance (MANOVA). He outlined two procedures regarding the phenomenon of missing data in MANOVA applications. The first is the situation where several scattered responses are missing for each dependent variable, and the second is where whole vectors of responses are missing. No proven method of estimations for omissions is provided.

Raffeld (1973) and Lord (1974) considered missing item responses and their estimates. Lord examined ability and item parameters. His emphasis was on the inappropriateness of scoring an item as incorrect if it were omitted by the subject. He uses probability methods to estimate the omitted data from a minimum of two or three thousand other subjects. Raffeld pursued estimates of items on standardized achievement tests using mean value estimates. He concluded that for omitted items on a standardized achievement test it is better to assign a value which is the mean of the alternatives for that item rather than assigning the mean response for the group omitting the item. Neither Lord nor Raffeld concerned himself with subscore estimates.

Direction of Present Research

The above review was concerned either with estimates of missing data and their parameters or estimates of missing data without concern for analysis. The intention of this study is to forego parametric concerns, apply simple methods of data estimation, analyze the estimated data sets, examine the results of the analysis, and provide results directly related to educational research. It will use a frequently employed educational measurement, the achievement test with several subscores, and investigate estimation methods understood by most researchers and students of research.

CHAPTER III

DESIGN OF THE STUDY

The research conducted in this study focused on the usefulness of the inclusion of multiresponse data, which consists of several subscores, in a multivariate analysis of variance as dependent variables when random missing subscores were estimated using mean value and regression techniques. The analyses of the data sets formed by the two methods of estimation were compared to each other and to the analysis of the complete data set.

The underlying focus of the research concerned the efficacy of the above method when applied to educationally related data. Thus the data sets investigated consisted of achievement scores collected on elementary school students.

Procedures

Two random samples were drawn from two fixed groups. The first sample consisted of 193 fourth-grade students and the second of an equal number of fifth-grade students. Both were administered the Stanford Achievement Test Battery in the spring of 1974. The fourth-grade sample was given the Intermediate I Battery and the fifth-grade sample the Intermediate II Battery providing raw scores for analysis.

In preparing the data for analysis, random subsamples were drawn comprising 2½, 5, 10, 15, and 20 percent of each of the two original complete data sets. The number of subjects in each of these subsamples was 5, 10, 20, 29, and 39, respectively. The subjects in these subsamples were considered as having missing data. One achievement subscore was randomly discarded for each subject in each of the missing subsamples. This procedure was conducted five times for each of the five percent levels, obtaining five different random subsamples.

Utilizing the subjects without randomly chosen missing subscores, means on each achievement test variable were formed. These means were substituted for the randomly discarded subscore for each subject in each of the missing subsamples.

Likewise, the subjects without randomly chosen missing subscores were subjected to multiple linear regression analysis. One achievement test subscore was randomly chosen as the dependent variable, and the remaining subscores were the independent variables. The nondiscarded subscores of each of the subjects with a missing subscore were substituted in the corresponding resulting regression equation. The value obtained from the regression equation was substituted for the randomly discarded subscores.

Method

In testing the hypotheses, multivariate analysis of variance (MANOVA) was conducted on each of the 100 adjusted samples with missing data and on the complete original sample with no missing data. The two fixed groups were the independent variables, and the achievement test subscores were the dependent variables in each case. The MANOVA results of the mean value estimates and the multiple linear regression estimates were compared to the MANOVA results of the complete original sample and to each other.

The comparisons of the resulting F-ratios were determined by the evaluation of the complement of the cumulative distribution function of the variance ratio distribution. The method consists of the following series expansion. Let n and m be the first and second number of degrees of freedom, respectively, and let

$$\alpha = \tan^{-1} \sqrt{nF/m}$$

where F is the F-ratio value. Then if n is even, the complement P is defined as

$$P(n, m, F) = \cos^m \alpha \left[1 + \frac{m}{2} \sin^2 \alpha + \frac{m(m+2)}{(2)(4)} \sin^4 \alpha + \dots + \frac{m(m+2) \dots (m+n-4)}{(2)(4) \dots (n-2)} \sin^{n-2} \alpha \right]$$

If m is even,

$$P(n,m,F) = 1 - \sin^n \alpha \left[1 + \frac{n}{2} \cos^2 \alpha + \frac{n(n+2)}{(2)(4)} \cos^4 \alpha + \dots + \frac{n(n+2) \dots (n+m-4)}{(2)(4) \dots (m-2)} \cos^{m-2} \alpha \right]$$

If n and m are both odd,

$$\begin{aligned} P(n,m,F) = & \frac{2}{\pi} \frac{(2)(4) \dots (m-1)}{(1)(3) \dots (m-2)} \cos^m \alpha \sin \alpha \\ & \times \left[1 + \frac{m+1}{3} \sin^2 \alpha + \frac{(m+1)(m+3)}{(3)(5)} \sin^4 \alpha + \dots + \frac{(m+1)(m+3) \dots (m+n-4)}{(3)(5) \dots (n-2)} \sin^{n-3} \alpha \right] \\ & - \frac{2 \sin \alpha \cos \alpha}{\pi} \left[1 + \frac{2}{3} \cos^2 \alpha + \frac{(2)(4)}{(3)(5)} \cos^4 \alpha + \dots + \frac{(2)(4) \dots (m-3)}{(3)(5) \dots (m-2)} \cos^{m-3} \alpha \right] + 1 - \frac{2\alpha}{\pi} \end{aligned}$$

where, if $n = 1$, the first series is to be taken as zero, and if $m = 1$, the second series is to be taken as zero and the factor $\frac{(2)(4) \dots (m-1)}{(3)(5) \dots (m-2)}$ is to be taken as unity (Hopper, 1970).

If the complement of the complete data set is greater than 0.05 and the complement of a data set with an estimated missing subsample is less than or equal to 0.05, then the

MANOVA results are considered significantly different from one another. Likewise, if the complement of the complete data set is less than or equal to 0.05 and the complement of a data set with an estimated missing subsample is greater than 0.05, then the MANOVA results are considered significantly different from one another. If both results are either greater than 0.05 or less than or equal to 0.05, then the MANOVA results are not considered significantly different from one another.*

---

*This method is contingent upon the level of significance chosen and relies on the fact that the point of significance is immutable.

CHAPTER IV

RESULTS

It has been the experience of the researcher that when conducting data analysis on achievement tests, he obtains a list of scores which contains missing subscores. The data on experimental units with missing subscores must then be discarded and results in a loss of information.

The present study questioned the applicability of using estimates for multiresponse data in multivariate analysis of variance (MANOVA) when one response of an experimental unit is missing. Both mean value and regression estimates were employed for missing data in the manner reported in Chapter III.

There were three specific questions investigated in this study: Do mean value estimates provide different MANOVA results from that obtained when analyzing the total data set? Do regression estimates provide different MANOVA results from that obtained when analyzing the complete data set? and thus, Do mean value estimates provide different MANOVA results from regression estimates? Each of these inquiries was looked at for varying percent levels of missing data ($2\frac{1}{2}$, 5, 10, 15, and 20 percent of the total sample). The five different levels were employed on five different

random subsamples of missing data. This was performed on two different data sets of fourth- and fifth-grade elementary school students for the two types of estimates. This resulted in $5 \times 5 \times 2 \times 2$ random incomplete samples, or a total of 100 incomplete samples, that were studied and compared to the two complete data sets of fourth- and fifth-grade students.

The presentation of results in this chapter is according to each of the five percent levels of missing data for the three aforementioned questions. These three questions represent the three hypotheses which are stated as follows:

- H₁: There is no difference in MANOVA results for the complete data set and the mean value estimated data set with the size of the missing subsample ranging from 2½ to 20 percent of the complete data set.
- H₂: There is no difference in MANOVA results for the complete data set and the regression estimated data set with the size of the missing subsample ranging from 2½ to 20 percent of the complete data set.
- H₃: There is no difference in MANOVA results for the mean value estimated data set and the regression estimated data set both with the size of the missing subsample ranging from 2½ to 20 percent of the complete data set.

The MANOVA F-ratios and the corresponding complement of the cumulative distribution function of the variance ratio distribution are provided in response to these hypotheses.

MANOVA performed on the complete data set of fourth graders resulted in a $F = 2.8851$ with 8 and 185 df (degrees

of freedom); for the fifth graders, there resulted a $F = 3.3229$ with 7 and 185 df. Determining the complement of the cumulative distribution function, the P value obtained for the fourth-grade data set was 0.004745 and that for the fifth-grade data set was 0.002341.

Comparison of the Mean Value and the Regression
Estimated Data Sets with One Another and
with the Complete Data Set at the 2½
Percent Level of Missing Subsamples

The values of the F-ratio and complement of the cumulative distribution function for fourth- and fifth-grade mean value and regression estimated data sets at the 2½ percent level are presented in Table 1. For the fourth-grade sample, no F-ratio of the mean value estimated data sets differed from the complete data set's F-ratio by more than 0.1267. Likewise, for the regression estimated data sets, no F-ratio differed from the complete data set's F-ratio by more than 0.0675. Equivalent ranges for the fifth-grade sample were 0.0329 and 0.0397, respectively.

Examining the complement of the cumulative distribution function for the fourth-grade sample, no P of the mean value estimated data sets differed from the complete data set's complement by a value greater than 0.001388. Likewise, for the regression estimated data sets, no complement differed from the complete data set's complement by a value greater than 0.000798. Equivalent ranges for the fifth-grade sample were 0.000196 and 0.000245, respectively.

TABLE 1. F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of $2\frac{1}{2}$ Percent of the Complete Samples

	Grade Four			Grade Five				
	Mean Value		Regression	Mean Value		Regression		
	F	P		F	P			
							F	P
Sample 1	2.9708	0.003756	2.9228	0.004282	3.3265	0.002323	3.2832	0.002589
Sample 2	2.8974	0.004589	2.9338	0.004155	3.3126	0.002406	3.2907	0.002541
Sample 3	2.8796	0.004817	2.9096	0.004440	3.3462	0.002212	3.2865	0.002568
Sample 4	3.0118	0.003357	2.9526	0.003947	3.2983	0.002493	3.2852	0.002576
Sample 5	2.9590	0.003878	2.9490	0.003988	3.3558	0.002158	3.2953	0.002512

Since the complement of the complete data set for both the fourth and fifth grades was less than 0.05 while at the same time the five complements of the mean value and the regression estimated data sets were less than 0.05, the three null hypotheses are not rejected at the $2\frac{1}{2}$ percent level of missing subsamples.

Comparison of the Mean Value and the Regression
Estimated Data Sets with One Another and
with the Complete Data Set at the 5
Percent Level of Missing Subsamples

The values of the F-ratio and complement of the cumulative distribution function for fourth- and fifth-grade mean value and regression estimated data sets at the 5 percent level are presented in Table 2. For the fourth-grade sample, no F-ratio of the mean value estimated data sets differed from the complete data set's F-ratio by more than 0.1859. Likewise, for the regression estimated data sets, no F-ratio differed from the complete data set's F-ratio by more than 0.0302. Equivalent ranges for the fifth-grade sample were 0.1268 and 0.1226, respectively.

Examining the complement of the cumulative distribution function for the fourth-grade sample, no P of the mean value estimated data sets differed from the complete data set's complement by a value greater than 0.001893. Likewise, for the regression estimated data sets, no complement differed from the complete data set's complement

TABLE 2. F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 5 Percent of the Complete Samples

	Grade Four				Grade Five			
	Mean Value		Regression		Mean Value		Regression	
	F	P	F	P	F	P	F	P
Sample 1	2.9982	0.003484	2.9094	0.004418	3.3587	0.002143	3.3830	0.002016
Sample 2	2.8943	0.004628	2.8848	0.004750	3.2744	0.002647	3.2745	0.002647
Sample 3	2.8706	0.004937	2.8771	0.004851	3.3053	0.002450	3.2786	0.002619
Sample 4	3.0710	0.002852	2.9153	0.004370	3.2904	0.002543	3.3363	0.002267
Sample 5	2.9555	0.003916	2.8999	0.004558	3.1961	0.003219	3.2003	0.003186

by a value greater than 0.000375. Equivalent ranges for the fifth-grade sample were 0.000875 and 0.000842, respectively.

Since the complement of the complete data set for both the fourth and fifth grades was less than 0.05 while at the same time the five complements of the mean value and the regression estimated data sets were less than 0.05, the three null hypotheses are not rejected at the 5 percent level of missing subsamples.

Comparison of the Mean Value and the Regression
Estimated Data Sets with One Another and
with the Complete Data Set at the 10
Percent Level of Missing Subsamples

The values of the F-ratio and complement of the cumulative distribution function for fourth- and fifth-grade mean value and regression estimated data sets at the 10 percent level are presented in Table 3. For the fourth-grade sample, no F-ratio of the mean value estimated data sets differed from the complete data set's F-ratio by more than 0.5650. Likewise, for the regression estimated data sets, no F-ratio differed from the complete data set's F-ratio by more than 0.1607. Equivalent ranges for the fifth-grade sample were 0.1006 and 0.0801, respectively.

Examining the complement of the cumulative distribution function for the fourth-grade sample, no P of the mean value estimated data sets differed from the complete data set's complement by a value greater than 0.003977. Likewise, for the regression estimated data sets, no

TABLE 3. F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 10 Percent of the Complete Samples

	Grade Four				Grade Five			
	Mean Value		Regression		Mean Value		Regression	
	F	P	F	P	F	P	F	P
Sample 1	3.0076	0.003395	2.9488	0.003988	3.4235	0.001821	3.4030	0.001917
Sample 2	2.9682	0.003782	2.9043	0.004504	3.2743	0.002648	3.2802	0.002609
Sample 3	2.8678	0.004975	2.8713	0.004928	3.3378	0.002259	3.2773	0.002628
Sample 4	3.4501	0.000998	3.0458	0.003057	3.2941	0.002520	3.3524	0.002177
Sample 5	3.0149	0.003328	2.8983	0.004578	3.2814	0.002601	3.2859	0.002572

complement differed from the complete data set's complement by a value greater than 0.001688. Equivalent ranges for the fifth-grade sample were 0.000523 and 0.000427, respectively.

Since the complement of the complete data set for both the fourth and fifth grades was less than 0.05 while at the same time the five complements of the mean value and the regression estimated data sets were less than 0.05, the three null hypotheses are not rejected at the 10 percent level of missing subsamples.

Comparison of the Mean Value and the Regression
Estimated Data Sets with One Another and
with the Complete Data Set at the 15
Percent Level of Missing Subsamples

The values of the F-ratio and complement of the cumulative distribution function for fourth- and fifth-grade mean value and regression estimated data sets at the 15 percent level are presented in Table 4. For the fourth-grade sample, no F-ratio of the mean value estimated data sets differed from the complete data set's F-ratio by more than 0.3063. Likewise, for the regression estimated data sets, no F-ratio differed from the complete data set's F-ratio by more than 0.1386. Equivalent ranges for the fifth-grade sample were 0.2364 and 0.0412, respectively.

Examining the complement of the cumulative distribution function for the fourth-grade sample, no P of the mean value estimated data sets differed from the complete data set's complement by a value greater than 0.002696. Likewise,

TABLE 4. F-Ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 15 Percent of the Complete Samples

	Grade Four				Grade Five			
	Mean Value		Regression		Mean Value		Regression	
	F	P	F	P	F	P	F	P
Sample 1	2.9470	0.004008	2.9765	0.003697	3.5593	0.001294	3.3263	0.002325
Sample 2	2.8829	0.004775	2.8880	0.004708	3.2797	0.002612	3.3013	0.002475
Sample 3	2.8862	0.004731	2.8830	0.004773	3.4280	0.001801	3.2971	0.002501
Sample 4	3.1914	0.002049	3.0237	0.003249	3.2777	0.002625	3.2899	0.002547
Sample 5	3.1742	0.002146	2.9796	0.003666	3.3087	0.002430	3.2817	0.002599

for the regression estimated data sets, no complement differed from the complete data set's complement by a value greater than 0.001496. Equivalent ranges for the fifth-grade sample were 0.001050 and 0.000255, respectively.

Since the complement of the complete data set for both the fourth and fifth grades was less than 0.05 while at the same time the five complements of the mean value and the regression estimated data sets were less than 0.05, the three null hypotheses are not rejected at the 15 percent level of missing subsamples.

Comparison of the Mean Value and the Regression
Estimated Data Sets with One Another and
with the Complete Data Set at the 20
Percent Level of Missing Subsamples

The values of the F-ratio and complement of the cumulative distribution function for fourth- and fifth-grade mean value and regression estimated data sets at the 20 percent level are presented in Table 5. For the fourth-grade sample, no F-ratio of the mean value estimated data sets differed from the complete data set's F-ratio by more than 0.3305. Likewise, for the regression estimated data sets, no F-ratio differed from the complete data set's F-ratio by more than 0.1237. Equivalent ranges for the fifth-grade sample were 0.2711 and 0.0479, respectively.

Examining the complement of the cumulative distribution function for the fourth-grade sample, no P of the mean value estimated data sets differed from the complete

TABLE 5. F-ratios and Complements (P) of the Cumulative Distribution Function for Fourth- and Fifth-Grade Samples Having Mean Value and Regression Estimated Subsamples Consisting of 20 Percent of the Complete Samples

	Grade Four				Grade Five			
	Mean Value		Regression		Mean Value		Regression	
	F	P	F	P	F	P	F	P
Sample 1	2.9608	0.003859	2.9272	0.004231	3.5940	0.001185	3.3024	0.002468
Sample 2	2.8703	0.004941	2.8637	0.005031	3.3104	0.002419	3.2750	0.002643
Sample 3	2.9036	0.004513	2.8916	0.004663	3.5476	0.001333	3.3119	0.002410
Sample 4	3.0312	0.003183	2.9180	0.004339	3.3004	0.002480	3.3196	0.002364
Sample 5	3.2156	0.001915	3.0088	0.003384	3.3048	0.002453	3.2770	0.002630

data set's complement by a value greater than 0.002830. Likewise, for the regression estimated data sets, no complement differed from the complete data set's complement by a value greater than 0.001361. Equivalent ranges for the fifth-grade sample were 0.001159 and 0.000299, respectively.

Since the complement of the complete data set for both the fourth and fifth grades was less than 0.05 while at the same time the five complements of the mean value and the regression estimated data sets were less than 0.05, the three null hypotheses were not rejected at the 20 percent level of missing subsamples.

Further Results

To determine which method of estimation investigated was the stronger, an inspection of the values of the F-ratios and complements of the cumulative distribution function was conducted. The closeness of these values of the incomplete data sets to that of the appropriate complete data set was observed. For each group of five incomplete data sets at each percent level, the range of values was found and examined for largeness of width.

The largest range at each percent level of missing data for the fourth-grade sample with mean value estimates varied from 0.001388 to 0.003977, whereas, for the regression estimated samples, it varied from only 0.000375 to 0.001688. For the fifth-grade samples with mean value

estimates, the range varied from 0.000196 to 0.001159. For regression estimates, it was 0.000245 to 0.000842. Only at the $2\frac{1}{2}$ percent level of missing values did the mean value complement range not exceed that of the regression complement range.

A closer examination of the results revealed additional information. One might presume that as the percent of estimated data elements decreased, the smaller the range would be between the value of the F-ratio of the complete data set and the most distant value of the F-ratio of the data sets with estimated values. This was neither consistent within the fourth- and fifth-grade samples nor within the method of estimation. Considering the percent level of missing data with the shortest range to the level with the longest range, the order for the fourth-grade sample with mean value estimates is $2\frac{1}{2}$, 5, 15, 20, 10; for the fourth-grade sample with regression estimates, 5, $2\frac{1}{2}$, 20, 15, 10; for the fifth-grade sample with mean value estimates, $2\frac{1}{2}$, 10, 5, 15, 20; and for the fifth-grade sample with regression estimates, $2\frac{1}{2}$, 15, 20, 10, 5. The exact results hold for the complement of the cumulative distribution function.

Another presumption might be that the value of the F-ratio of the complete data set would be within the range of the values of the F-ratios at a particular percent level of missing data. This is consistent for the fourth- and fifth-grade samples within a method of estimation but not

between methods of estimation. For both the fourth- and fifth-grade samples having mean value estimates, the value of the F-ratio of the complete data set is within the range of the values of the F-ratios for all percent levels of missing data. For regression estimated samples, this is not the case. The fourth-grade samples have F-ratios not inclusive, range-wise, of the complete data set's F-ratio at the $2\frac{1}{2}$ percent level; for the fifth grade, it is at the $2\frac{1}{2}$ and 20 percent levels. The value of the F-ratio of the complete data set exceeds the values of the F-ratio in the fifth-grade sample and precedes the values in the fourth-grade sample.

Summary

In summary, this chapter has presented the statistical analysis of the data. The results of the study indicated that no significant differences exist among the MANOVA results of data sets having missing subscores estimated by mean values, data sets having missing subscores estimated by regression, and the complete data set with no missing values. This was demonstrated for 100 samples with estimated subscores. The estimated subsamples consisted of $2\frac{1}{2}$, 5, 10, 15, and 20 percent of the complete samples of fourth- and fifth-grade students.

Since inspection showed that the regression estimated values provided MANOVA and complement results at each

percent level closer, in all instances, to that of the complete data set, it is apparently the stronger of the two estimation procedures. Both methods of estimation, though, were demonstrated to provide MANOVA results not significantly different from the results of the complete data sets.

CHAPTER V

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Discussion

The intention of this study was to examine the effect of different estimators for missing multiresponse data on multivariate analysis of variance (MANOVA) results. Mean value and regression techniques were used in determining estimates. The MANOVA results for the data sets which employed the different estimation techniques were compared to each other and to MANOVA results of the complete data set.

Specifically investigated were the achievement test scores of a fourth-grade sample and a fifth-grade sample. Fifty MANOVAs were conducted on each grade; 25 analyzed the incomplete data sets with mean value estimates and 25 with regression estimates. The 25 analyses were subgrouped into five sets of analyses. Each set contained a different percent level of missing data. These levels were $2\frac{1}{2}$, 5, 10, 15, and 20 percent of the complete sample. Five samples with different missing subsets of data were analyzed at each level.

The results of Chapter IV demonstrated that the MANOVA results of both estimation techniques did not differ

significantly from one another nor from the results obtained from the complete data set. Inspection of the F-ratios and complements implied that the regression method was apparently the stronger estimation technique.

The latter result was determined by the closeness of the values of the F-ratios and the complements of the cumulative distribution function for the estimated samples to that of the complete data set.

In addition, two a posteriori results were observed. It was found that as the percent of estimated data elements decreased, it did not follow that the smaller the range would be between the value of the F-ratio of the complete data set and the most distant value of the F-ratio of the data sets with estimated values. The non sequitur held for both grades of students and both methods of estimation. This was likewise true for the complement of the cumulative distribution function.

A second finding was that the F-ratio of the complete data set was not within the range of the values of the F-ratios at all percent levels of missing data estimated by regression techniques. It did hold for mean value estimated data sets. The same findings occurred among the complements of the cumulative distribution function.

Conclusions

Three conclusions were drawn from the present study:

1. Achievement data with up to 20 percent missing subscores that are estimated by mean value techniques when analyzed by MANOVA provide results which do not differ significantly from MANOVA results of the same achievement data without any missing subscores.
2. Achievement data with up to 20 percent missing subscores that are estimated by regression techniques when analyzed by MANOVA provide results which do not differ significantly from MANOVA results of the same achievement data without any missing subscores.
3. Achievement data with up to 20 percent missing subscores that are estimated by mean value techniques when analyzed by MANOVA provide results which do not differ significantly from MANOVA results of achievement data with up to 20 percent missing subscores that are estimated by regression techniques.

The above conclusions seem to suggest that there exist for educators alternatives in data analysis other than discarding incomplete multiresponse observations. The alternatives provided here are the two methods of estimation: mean value and regression. In addition, the mean value method of estimation was demonstrated to be as appropriate in MANOVA as the regression method as proven by the non-rejection of the third hypothesis. Further data considerations revealed that for all levels of missing data, the F-ratio of the complete data set was located within the range of the F-values determined for the data sets with missing subsamples estimated by the mean value methods. This did not hold for the regression method.

Since the mean value method is straightforward and has been proved to be an appropriate estimation

technique, data formerly lost to analysis can be retained. No longer must estimates for omissions be evaded because of complicated data manipulations, time, money, and resources.

Recommendations

The present study has operated under various limitations which need to be investigated in order to extend the inferences of this research. Bracht and Glass (1968) stated:

The intent (sometimes explicitly stated, sometimes not) of almost all experimenters is to generalize their findings to some group of subjects and set of conditions that are not included in the experiment. To the extent and manner in which the results of an experiment can be generalized to different subjects, settings, experimenters, and, possibly, tests, the experimenter possesses external validity. pp. 437-438

The external validity of this study is restricted by the lack of reported research dealing with statistical analyses which employ data estimates without parametric estimates. Areas which require further investigation in reference to inferential conclusions are presented in the following list:

1. The samples consisted of fourth and fifth graders. Other educational levels need to be examined.
2. Achievement scores for two levels of one standardized achievement test were analyzed. Other standardized achievement tests need to be investigated.
3. In addition to achievement tests, other types of tests which measure not only the cognitive domain but also the affective domain need to be studied such as those dealing with self-concept and social acceptance.

4. Other methods of estimation need to be considered in a manner similar to the present investigation and compared to mean value methods for accuracy and simplicity.
5. Missing subsamples were determined randomly. Actual missing subsamples need to be investigated for possible commonalities.
6. The levels of missing data should be expanded in order to determine maximum levels of missing subsamples.
7. More than one missing subscore per experimental unit needs inspection.
8. Experimental designs requiring analyses different from multivariate analysis of variance need probing.

These recommendations are listed not only to provide closure to the present study but also to indicate the multidirectional approaches involved in this specific area of research. Closure is provided with respect to confining the present research's inferences to the subset of investigations outside of the above listing. The expanse of additional approaches is suggested by the list itself. No one item of the list is more worthy of study than the other. All need investigation in order to advance to the universal set of estimators for omissions of multiresponse data.

REFERENCES

- Afifi, A. and Elashoff, R. M. "Missing observations in multivariate statistics I. Review of the literature." Journal of the American Statistical Association, 1966, 61, 595-604.
- Afifi, A. and Elashoff, R. M. "Missing observations in multivariate statistics II. Point estimation in simple linear regression." Journal of the American Statistical Association, 1967, 62, 10-29.
- Anderson, T. W. "Maximum likelihood estimates for a multivariate normal distribution when some observations are missing." Journal of the American Statistical Association, 1957, 52, 200-203.
- Baird, H. R. and Kramer, C. Y. "Analysis of variance of a balanced incomplete block design with missing observations." Applied Statistics, 1960, 9, 189-198.
- Bhargava, R. Multivariate tests of hypotheses with incomplete data. Applied Mathematics and Statistical Laboratories, Technical Report 3, 1962.
- Bracht, G. H. and Glass, G. V. "The external validity of experiments." American Educational Research Journal, 1968, 5, 437-474.
- Buck, S. F. "A method of estimation of missing values in multivariate data suitable for use with an electronic computer." Journal of the Royal Statistical Society, Series B, 1960, 22, 302-307.
- Dagenais, M. G. "Further suggestions concerning the utilization of incomplete observations in regression analysis." Journal of the American Statistical Association, 1971, 66, 93-98.

- Dear, R. E. "A principal-component missing-data method for multiple regression models." SP-86, Systems Development Corporation, Santa Monica, California, 1959.
- Dempster, A. P. "An overview of multivariate data analysis." Journal of Multivariate Analysis, 1971, 1, 316-346.
- Edgett, G. L. "Multiple regression with missing observations among the independent variables." Journal of the American Statistical Association, 1956, 51, 122-131.
- Federspiel, C. F., Monroe, R. J., and Greenberg, B. G. "An investigation of some multiple regression methods for incomplete samples." University of North Carolina, Institute of Statistics, Mineo Series, No. 236, August 1959.
- Glasser, M. "Linear regression analysis with missing observations and the independent variables." Journal of the American Statistical Association, 1964, 59, 834-844.
- Haitovsky, Y. "Missing data in regression analysis." Journal of the Royal Statistical Society, Series B, 1968, 30, 67-82.
- Hartwell, T. D. and Gaylor, D. W. "Estimating variance components for two-way disproportionate data with missing cells by the method of unweighted means." Journal of the American Statistical Association, 1973, 68, 379-383.
- Hocking, R. R. and Smith, W. B. "Estimation of parameters in the multivariate normal distribution with missing observations." Journal of the American Statistical Association, 1968, 63, 159-173.
- Hopper, M. J., comp. Harwell Subroutine Library: A Catalogue of Subroutines. London: Her Majesty's Stationery Office, State House, 49 High Holborn, 1970.
- Kleinbaum, D. G. Estimation and hypothesis testing for generalized multivariate linear models. Doctoral dissertation, University of North Carolina, Chapel Hill, North Carolina, 1970.
- Kramer, C. Y. and Glass, S. "Analysis of variance of a Latin square design with missing observations." Applied Statistics, 1960, 9, 43-50.

- Lord, F. M. "Estimation of parameters from incomplete data." Journal of the American Statistical Association, 1955, 50, 870-876.
- Lord, F. M. "Estimation of latent ability and item parameters when there are omitted responses." Psychometrika, 1974, 39, 247-264.
- Matthai, A. "Estimation of parameters from incomplete data with applications to design of sample surveys." Sankhyā, 1951, 2, 145-152.
- Mitra, S. K. "Some remarks on the missing plot analysis." Sankhyā, 1959, 21, 337-344.
- Morrison, D. F. "Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data." Journal of the American Statistical Association, 1971, 66, 602-604.
- Nicholson, G. E., Jr. "Estimation of parameters from incomplete multivariate samples." Journal of the American Statistical Association, 1957, 52, 523-526.
- Preece, D. A. "Query and answer: Non-additivity in two-way classifications with missing values." Biometrics, 1972, 28, 574-577.
- Pruzek, R. M. "Methods and problems in the analysis of multivariate data." Review of Educational Research, 1971, 41, 163-190.
- Raffeld, P. C. The effects of Guttman weights on the reliability and predictive validity of objective tests when omissions are not differentially weighted. Doctoral dissertation, University of Oregon, 1973.
- Rubin, D. B. "Characterizing the estimation of parameters in incomplete-data problems." Journal of the American Statistical Association, 1974, 69, 467-474.
- Srivastava, J. N. "On the extension of Gauss-Markov theorem to complex multivariate linear models." The Annals of the Institute of Statistical Mathematics, 1967, 19, 417-437.

- Srivastava, J. N. "On a general class of designs for multi-response experiments." The Annals of Mathematical Statistics, 1968, 39, 1825-1843.
- Srivastava, J. N. and McDonald L. "On the costwise optimality of hierarchical multiresponse randomized block designs under the trace criterion." The Annals of the Institute of Statistical Mathematics, 1969, 21, 507-514.
- Srivastava, J. N. and McDonald, L. "On the costwise optimality of certain hierarchical and standard multi-response models under the determinant criterion." Journal of Multivariate Statistics, 1971, 1, 118-128.
- Srivastava, J. N. and Zaatar, M. K. "On the maximum likelihood classification rule for incomplete multivariate samples and its admissibility." Journal of Multivariate Analysis, 1972, 2, 115-126.
- Trawinski, I. M. Incomplete-variable designs. Doctoral dissertation, Virginia Polytechnic Institute, Blacksburg, Virginia, 1961.
- Trawinski, I. M. and Bargmann, R. E. "Maximum likelihood estimation with incomplete multivariate data." The Annals of Mathematical Statistics, 1964, 35, 647-657.
- Walsh, J. E. "Computer-feasible general method for fitting and using regression functions when data are incomplete." SP-71, System Development Corporation, Santa Monica, California, 1959.
- Wilkinson, G. N. "Comparison of missing value procedures." Australian Journal of Statistics, 1960, 2, 53-65.
- Wilks, S. S. "Moments and distributions of estimates of population parameters from fragmentary samples." The Annals of Mathematical Statistics, 1932, 3, 163-195.

BIOGRAPHICAL SKETCH

Stephen S. Sledjeski was born November 27, 1942, in Greenport, New York. He graduated from Southold High School, Southold, New York; the Diocesan Preparatory Seminary, Buffalo, New York (A.A.); St. Bonaventure University, St. Bonaventure, New York (B.S.); and the University of Florida, Gainesville, Florida (M.Ed., Ed.S., Ph.D.).

His educational employment experience consists of working as a middle school mathematics teacher with the Alachua County Board of Public Instruction, Gainesville, Florida; a research associate with Santa Fe Community College, Gainesville, Florida; supervisor of data processing as a graduate research assistant with the Florida Parent Education Model of Project Follow Through, University of Florida, Gainesville, Florida; and Research Specialist at P. K. Yonge Laboratory School, Gainesville, Florida. In addition, he has been a statistical and computer consultant for doctoral students, the Florida State Department of Health and Rehabilitation Services, and the Career Opportunities Program, Richmond, Virginia.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Vynce A. Hines, Chairman
Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ira J. Gordon
Graduate Research Professor of
Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert S. Soar
Professor of Foundations of
Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Z. R. Pop-Stojanovic
Associate Chairman and Professor
of Mathematics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Hattie Bessent
Assistant Professor of Foundations
of Education

This dissertation was submitted to the Graduate Faculty of the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

March, 1976

Dean, College of Education

_____
Dean, Graduate School