DOCUMENT RESUME

ED 388 686 TM 023 737

AUTHOR Velotta, Cynthia L.
TITLE Coping with Missing Data in Educational Research and Evaluation.
PUB DATE Apr 95
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Research; *Evaluation Methods; *Kindergarten Children; *Pretests Posttests; Primary Education; Research Methodology; *Statistical Analysis
IDENTIFIERS *Missing Data

ABSTRACT

The characteristics of missing data in statistical analysis are outlined, and techniques to deal with missing data are explored using a real data set that contained pretest and posttest measures of kindergarten children. In the first part of the investigation, randomly created missing values for 5%, 10%, 20%, and 25% of the sample were supplied for a sample that originally had no missing values. Four missing data techniques (listwise deletion, mean substitution, adjustment-cell mean imputation, and regression imputation) were then implemented, and the results were compared with those of analysis of variance tests. In a second part of the study, the missing data techniques were applied to a real missing data problem, using the same subsample (443 kindergarten students) with pretest data missing for 83 students. Results indicated that disparate results may be obtained with various missing data techniques. When faced with missing data, researchers should investigate whether the data are missing at random or for an identifiable reason, apply some missing data techniques, and consider the consequences carefully if different techniques lead to dissimilar conclusions. (Contains 1 figure, 12 tables, and 18 references.) (SLD)

Coping with Missing Data in Educational Research and Evaluation

Cynthia L. Velotta
Dayton Public Schools

Paper presented at the Annual Meeting of the American Educational Research Association

San Francisco, April 18-22, 1995

Abstract

Much has been written regarding missing data in statistical analyses; however, the majority of these articles focus on theoretical considerations of missing data and missing data techniques. Because my work focuses on applied statistics, the discussion is directed in a manner that would be useful to others in my field. Specifically, the paper: (a) outlines characteristics of missing data, (b) describes missing data techniques and associated problems, (c) summarizes research that compared missing data techniques, (d) presents my research comparing missing data techniques, and (e) provides a practical suggestion for dealing with missing data. My research utilized a real data set that contains pre and posttest measures of kindergarten children. My investigation contained two parts. In the first part, I used a sample that had no missing values and randomly created missing values for 5%, 10%, 20%, and 25% of the sample. Then, I implemented four missing data techniques (listwise deletion, mean substitution, adjustment-cell mean imputation, and regression imputation) and compared the results of analysis of variance tests to the results obtained using the actual values. In the second part, I applied these missing data techniques to a real missing data problem. The results of the study revealed that disparate results may be obtained using the various missing data techniques. Specifically, different conclusions may be drawn depending on the technique used to cope with missing data. Therefore, when faced with the problem of missing data, researchers should: (a) investigate whether the data are missing due to some factor or are missing at random, (b) apply a few of the missing data techniques, (c) determine if different conclusions would be drawn from the applied techniques, and (d) carefully consider the consequences when different techniques lead to dissimilar conclusions.

Coping with Missing Data in Educational Research and Evaluation

INTRODUCTION

An undeniable characteristic of educational research and evaluation is incomplete data. In survey research, data may be incomplete due to undercoverage, unit nonresponse, or item nonresponse (Madow, Nisselson, & Olkin, 1983). Additional problems arise in longitudinal studies: some subjects may quit the study, some may move, and others may miss measurements due to vacations or illnesses. Another reason for incomplete data is that not all subjects are measured on every variable. This is evident in college student personnel records where students take different entrance exams, take different courses, etc. Because statistics are predicated on sampling methods, loss of data may bias results. In addition, loss of subjects diminishes the power of statistical tests. Depending on the nature of the missing data, some statistical analyses may be inappropriate. Therefore, missing data should be carefully considered.

Much has been written regarding missing data in statistical analyses; however, the majority of these articles are published in journals such as: Journal of the American Statistical Association, Journal of the Royal Statistical Society, and Psychometrika. Such articles focus on theoretical considerations of missing data and missing data techniques. Because my work focuses on applied statistics, this discussion is directed in a manner that would be useful to others in my field. Specifically, the paper will: (a) outline characteristics of missing data, (b) describe missing data techniques and associated problems, (c) summarize research that compared missing data techniques, (d) present my research comparing missing data techniques, and (e) provide a practical suggestion for dealing with missing data.

Missing Data Characteristics

In univariate analyses, characteristics of missing data may be classified as follows. The missing data: (1) depends on variable Y and possibly variable X, (2) depends on X but not Y, or (3) is independent of X and Y (Little & Rubin, 1987). The terminology missing completely at random (MCAR) has been applied when missingness is characterized by case 3 from above. Missing at random (MAR) is used to describe instances in case 2, and in case 1, the data are not MCAR or MAR (Little & Rubin, 1987). These are important distinctions when deciding how to proceed with analyses. An example may clarify the distinctions. Let X = race/ethnicity and Y = achievement; Y contains missing values. If the probability that achievement is nonmissing varies according to achievement within ethnicity groups, then the data fall under case 1. If the probability that achievement is nonmissing varies according to ethnicity but not achievement, the data fall under case 2 (MAR). Finally, if the probability that achievement is nonmissing is the same for all subjects, then the data fall under case 3 (MCAR).

Missing Data Techniques

The missing data techniques that will be discussed here and that have received considerable attention in the literature are based on the assumption that missing data are MCAR. Techniques for missing data that are not MCAR are likelihood-based, and are discussed in articles such as: Dempster, Laird, and Rubin (1977); Gleason and Staelin (1975); Little and Rubin (1987); and Muthen, Kaplan, and Hollis (1987).

Everyone who conducts research and computes analyses makes a decision regarding missing data. This decision is sometimes an unconscious one in that the statistical software applies a default mechanism The default is frequently deletion of cases with missing data

When discarding cases bias may be introduced, power is affected, and Type II error rates are increased (Raymond, 1987). Another option is pairwise deletion of cases. This technique utilizes all available pairs of values when computing covariances. Disadvantages of pairwise deletion include: the population to which generalization is sought is no longer clear (Raymond, 1987), the sample size varies, and inconsistencies can occur. The following example provided in Norusis (1993) illustrated what could happen with pairwise deletion. Three variables height, weight, and age are correlated utilizing pairwise deletion. Age and height are found to have a high positive correlation. Age and weight also have high positive correlation. However, height and weight have high negative correlation. This may occur when different cases are used in the computation of each correlation.

The two techniques described above utilize *available* data when conducting analyses. Another class of techniques, imputation, replaces missing values by suitable estimates. Data are then analyzed as complete cases. There are many variations of imputation techniques; the more common ones are described here. Perhaps the most common imputation technique is the replacement of missing values with the variable mean that was computed using the complete cases. Limitations associated with variable mean imputation include: (a) sample size is overestimated, (b) variance is underestimated, (c) correlations are negatively biased, and (d) the distribution of new values is an incorrect representation of the population values because the shape of the distribution is distorted by adding values equal to the mean (Ford, 1983; Little & Rubin, 1990; and Raymond, 1987).

Perhaps a better imputation procedure, especially useful when one variable is a categorical variable, is adjustment-cell mean imputation. All cases are classified into cells based on similar values of a variable X. The within-cell mean of Y is then imputed for missing values. The more homogeneous the groups (several variables may be used to classify cases), the more effective this procedure will be.

Many variations of the regression technique have been proposed. A simple regression technique estimates missing data by regressing an incomplete variable onto a highly correlated variable. A multiple regression technique estimates missing data by regressing an incomplete variable onto two or more variables. Variations of these procedures include the way in which data are treated in the regression computation (listwise or mean substitution, for example), and whether an iterative solution is used (Raymond, 1987). While there are not large differences among any of these regression techniques, regression-based procedures often perform better than listwise and variable mean techniques (Ward & Clark, 1991). However, these techniques also have limitations. Raymond (1987) cautioned that "...using predictors to estimate criteria can result in inflated R^2s in subsequent analyses" (p. 4). In addition, multicollinearity may be introduced when predictors are used to estimate one another.

Hot-deck is a technique in which an observed value from the current sample is imputed for a missing value. This may be carried out by classifying the subjects into homogeneous groups. For each missing value in a particular group, an observed value is duplicated. A common hot-deck method imputes the observed value from the immediately preceding record (Bailar & Bailar, 1983). Another option is that the imputation value is randomly selected from observed values in the adjustment group The assumption is that within each group, nonrespondents follow the same distribution as respondents (Ford, 1983). If this assumption is inaccurate, results will be biased. While standard errors of estimates are less biased than those of mean substitution, they are still underestimated (Ford, 1983, Little & Rubin, 1990)

<u>Missing Data Research</u>

The missing data literature contains studies that utilized either real or simulated data to compare missing data techniques Four of the more recent studies are summarized below

Raymond and Roberts (1987) used computer-generated data matrices to investigate the effectiveness of four missing data techniques (listwise deletion, variable mean substitution, simple regression imputation, and iterative multiple regression imputation) on three different sample sizes (50, 100, and 200) and three different scenarios of missing data (2%, 6%, and 10%). The data matrices were subjected to multiple regression analyses. The regression equations were compared to equations obtained from the complete data matrices. These authors found that the regression procedures provided the most accurate regression equations. The listwise deletion was the least accurate method. However, the differences among procedures were small. Raymond and Roberts (1987) concluded that when missing data values are less than five percent of the values, the technique is of little importance. They suggested that when one variable has more than five percent of its values missing, the researcher should compare the results of at least two of the techniques.

Kaiser and Tracy (1988) also used simulated data to investigate four different regression techniques and the mean substitution technique on three sample sizes (30, 60, and 120) with 10%, 20%, and 30% missing data. The four regression techniques included estimation with one predictor, two predictors, three predictors, and three predictors modified by a correction factor to adjust for missing values in the predictors. The authors found no systematic trend of one technique over the others across levels of sample size, or percent of missing data The corrected regression metho was consistently the least accurate followed by mean substitution

Witta and Kiser (1991) selected their sample from the General Social Survey-1984 and examined the effectiveness of four missing data techniques (listwise deletion, pairwise deletion, mean substitution, and regression imputation) on sample sizes of 25 and 50. The selected sample (n=829) was randomly divided into two subsamples of 414 and 415. One of the subsamples (n=414) was reduced to complete cases (n=283). The mean of the criterion variable in this sample was used in comparisons with the means from treated samples. Using the other subsample (n=415), five random samples of 25 cases and five random samples of 50 cases were selected. Each sample was treated with the four missing data techniques. Using Dunnett's test for contrasts, Witta and Kiser found that the mean substitution technique was the least appropriate method. The mean substitution technique differed significantly from the comparison mean in eight of the ten samples.

Perhaps one of the most significant studies to date was conducted by Ward and Clark (1991). They compared the influence of four missing data techniques (listwise, mean substitution, simple regression, and iterative regr ssion) on three published analyses of the High School and Beyond data set. Ward and Clark investigated if the missing data techniques would effect the results given in the published analyses. All three published studies compared achievement of public and private school students; however, different statistical methods were employed. Brief descriptions of these studies follow.

Coleman, Hoffer, and Kilgore (1981) used 11,990 cases in their analysis and found a positive effect for private schooling. Page and Keith (1981) used 18,058 cases and found no difference in achievement for public and private school students. Walberg and Shanahan (1983) also found no difference when using 24,159 cases. These studies were carried out with missing data values for 57.54%, 36.05%, and 14.45% of the original cases in the data set. Walberg and Shanahan utilized mean substitution to replace missing data while the others did not employ a missing data technique

When Ward and Clark (1991) reanalyzed the data in these studies, they found differences between the original analysis and the analyses with replaced data. In addition, some of the analyses changed the effect of private/public schooling on achievement. Particularly, most of the no difference findings were changed to favor private schooling.

PROCEDURES

As an evaluator in a research department in a large public school system, I often encounter missing data. Typically, I conduct analyses on available data. However, my research of the literature has made me aware of techniques that may alleviate missing data problems. Therefore, I investigated four missing data techniques that are relatively quick and easy to apply using my current statistical software. My investigation was divided into two parts, simulated missing data problems and a real missing data problem. First, using a real data set, I created missing data to compare various missing data techniques. Then, I compared the results of these analyses with results obtained from analysis of the data without missing values. Second, I compared the missing data techniques using a real missing data problem. The data set contains 1993-1994 pre- and posttest variables for 2,697 kindergarten students from a large, urban district. In an attempt to control for SES and environment, I chose four schools for my sample that had both full-day and half-day kindergarten classes within the same school and did not have missing data for the variables of interest. This sample contains 443 cases. The variables that were used included: Peabody Picture Vocabulary Test, Pre-School Language Scale, ethnicity, gender and type of kindergarten schedule.

Simulated Missing Data Problems

The analyses in this section used two independent variables: minority status (minority, nonminority) and schedule (half-day, full-day), and one dependent variable: the posttest Peabody Picture Vocabulary Test (PPVT)

The purpose was to compare the effects of various missing data techniques (listwise deletion, mean substitution, adjustment-cell mean imputation, and regression) and the effects of samples with different numbers of missing values (5%, 10%, 20%, and 25%) to analysis with the actual values The SPSS for Windows (Release 6) command for randomly selecting an approximate percentage of cases was used to generate missing data. The missing data techniques are further described below:

Listwise Deletion. Most software packages proceed with this analysis, which omits those cases with missing values. Analysis was computed on available cases.

Mean Substitution. The mean value for PPVT computed on complete cases was substituted for missing values. The SPSS command RMV (replace missing values) was utilized. This procedure replaced missing values with the variable mean.

Adjustment-Cell Mean Imputation. Each sample (5%, 10%, 20%, 25%) was subdivided into four samples based on minority status and schedule, and then mean subsample values for PPVT were substituted for missing values. For example, mean PPVT was computed for nonminority/half-day students. That value was substituted for any nonminority/half-day students with missing values. The SPSS RMV command was employed here as well.

Regression. Correlational analyses revealed that posttest PPVT was correlated with other variables in the data set. Four of the five Pre-School Language scales had moderate correlations with PPVT (0.53, 0.60, 0.49, 0.55). A regression equation was computed on available cases for each of the four missing data types (5%, 10%, 20%, and 25%). Then, the predicted values were imputed for missing data.

Means and standard deviations were computed for each sample. Next, 2 x 2 analysis of variance tests were carried out to determine if the missing data technique or amount of missing data would result in a conclusion different than what was found using the actual scores. For all analyses, alpha was set at 0.05. Because of unequal cell sizes, the regression method (unique) was utilized to calculate sums of squares in the ANOVAs.

Real Missing Data Problem

For this part of the study, I applied the four missing data techniques used in the simulation study to a real missing data problem. In exploring the same data set used in the previous analyses (n=443), I found that 83 students did not have pretest PPVT assessment scores. Therefore, the pretest PPVT variable had 18.7% missing values. For this real problem, I investigated the effects of minority status and gender on the pretest PPVT scores.

Unlike the simulated analyses, I do not know if the data are missing at random. Ideally, I would call the schools and inquire as to why specific students were not tested. It is possible that they were new students in the district, or they missed testing days. However, testing was done over a few weeks and teachers made many attempts to have their students tested. Further exploration of the data revealed the following characteristics of students with missing data: 57% male, 43% female, 18% nonminority, and 82% minority. I found that 52% of the students with missing values came from one school. These students were distributed over the six homerooms within that school. The fact that 82% of the missing values were minority students raised question about randomness of missing values because minorities made up 54.4% of the sample (n=443). However, because 43 of the 83 students with missing values came from one school and minorities made up 96% of that school, I feel comfortable that data are MCAR.

I applied four different missing data techniques (regression, mean and adjustment-cell mean imputation, and listwise deletion) to the pretest PPVT variable. For the adjustment-cell mean imputation, the sample was separated into two groups based on minority status. Then, the mean for each respective group was imputed for missing values. For the regression imputation, the student's posttest PPVT scores were used to predict pretest PPVT scores. These variables were moderately correlated, $r=.67$.

RESULTS

Simulated Missing Data

Let's begin with the complete data set where all 443 cases had PPVT scores. Table 1 provides the means and standard deviations for PPVT by minority status and schedule. An examination of the table reveals that there is virtually no difference among nonminority children for schedule. However, minority children in the full-day classrooms had a mean score about ten points higher than minority children in half-day classrooms. These results lead one to suspect an interaction effect. Table 2 contains the ANOVA summary that shows the F for interaction was significant, $F(1,439) = 8.94$, $p=.003$. A plot of the minority status by schedule interaction is contained in Figure 1

Table 1. Means and Standard Deviations for Actual Cases

	Minority status		
Schedule	**Minority**	**Nonminority**	**Total**
Full-Day (FD)			
Mean	61.10	66.46	62.98
SD	12.62	13.32	13.06
N	69	37	106
Half-Day (HD)			
Mean	50.52	65.57	57.89
SD	12.25	16.18	16.15
N	172	165	337
Total			
Mean	53.55	65.74	59.11
SD	13.23	15.67	15.61
N	241	202	443

Table 2. ANOVA Summary Table for Actual Cases

Source	SS	DF	F	Sig of F
Minority status	7818.16	1	40.00	.000
Schedule	2474.69	1	12.66	.000
Minority status by Schedule	1748.18	1	8 94	003
Within	85796.89	439		
Total	107650.80	442		

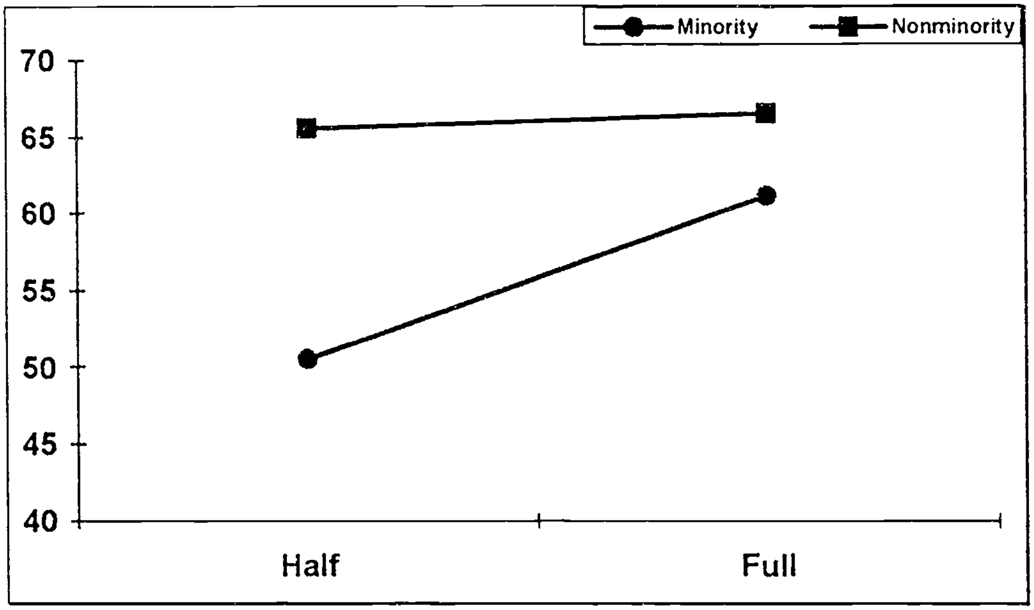

Figure 1. Interaction of Minority Status and Schedule on the Actual Data

Although the pattern is not disordinal (the lines do not cross), the schedule variable showed a greater effect on minority students than nonminority students. The main effects for minority status and schedule were also significant ($p<.000$).

Analyses With 5% Missing Data

PPVT scores for a 5% random selection of cases were deleted in order to create missing data. Next, four analyses were carried out using various missing data techniques. The techniques included listwise deletion, mean substitution, adjustment-cell mean imputation, and regression imputation. Means and standard deviations are provided in Table 3. There were six cases missing from three of the four groups (minority/half-day, minority/full-day, and nonminority/half-day) and five cases missing from the fourth group (nonminority/full-day). With one exception, the differences in means and standard deviations among the four techniques were less than 1. The exception was the nonminority/full-day group. Here the difference between the listwise and mean imputation means was 1.18. For all groups, the listwise standard deviation was always the highest, and the mean and adjustment-cell mean imputation techniques always had the lowest standard deviations. The 2x2 ANOVAs computed for each technique are provided in Table 4. Depending on the technique, we may have drawn different conclusions. The interaction effect for the mean, adjustment, and regression techniques were significant ($p < .05$). However, the interaction was not significant for the listwise technique.

Table 3. Means and Standard Deviations for 5% Missing Data

	MINORITY STATUS											
	MINORITY				NONMINORITY				TOTAL			
	METHOD											
Sche	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg
FD												
M	59.40	59.35	59.40	60.07	67.59	66.41	67.59	66.99	62.16	61.81	62.26	62.49
SD	10.73	10.25	10.25	10.58	13.79	13.15	12.80	13.24	12.40	11.78	11.81	11.98
N	63	69	69	69	32	37	37	37	95	106	106	106
HD												
M	50.67	50.96	50.67	50.89	65.36	65.12	65.36	65.44	57.86	57.89	57.86	58.02
SD	12.41	12.28	12.19	12.31	16.19	15.94	15.89	15.94	16.13	15.85	15.90	15.95
N	166	172	172	172	159	165	165	165	325	337	337	337
TOT												
M	53.07	53.36	53.17	53.52	65.73	65.36	65.77	65.73	58.83	58.83	58.92	59.09
SD	12.57	12.32	12.30	12.53	15.80	15.45	15.37	15.47	15.46	15.05	15.13	15.20
N	229	241	241	241	191	202	202	202	420	443	443	443

Table 4. ANOVA Summaries for the 5% Missing Data Sample

SOURCE	SS	DF	F	Sig of F
Listwise Deletion				
Minority status	8808.19	1	45.87	.000
Schedule	2020.05	1	10.52	.001
Minority status by Schedule	707.97	1	3.69	.056
Within	79875.80	416		
Total	100171.00	419		
Mean Imputation				
Minority status	8436.34	1	45.81	.000
Schedule	1753.78	1	9.52	.002
Minority status by Schedule	944.09	1	5.13	.024
Within	80841.18	439		
Total	100171.00	442		
Adjustment-Cell Imput.				
Minority status	9804.92	1	53.89	.000
Schedule	2248.63	1	12.36	.000
Minority status by Schedule	788.09	1	4.33	.038
Within	79875.80	439		
Total	101208.49	442		
Regression				
Minority status	8631.94	1	46.48	.000
Schedule	2156.24	1	11.61	.001
Minority status by Schedule	1089.45	1	5.87	.016
Within	81534.21	439		
Total	102100.69	442		

Analyses With 10% Missing Data

PPVT scores for a 10% random selection of cases were deleted in order to create missing data. This resulted in 44 cases with missing values. Next, four analyses were carried out using the same missing data techniques that were applied above. Means and standard deviations are provided in Table 5. The differences in means and standard deviations between techniques for all four groups were less than 1. As would be expected, the two methods utilizing mean imputation (mean and adjustment-cell) consistently had the lowest standard deviations among the methods, and the listwise technique resulted in the highest standard deviations. The ANOVA summaries computed for each technique are provided in Table 6. Here, the conclusions are similar for all techniques. Specifically, the interaction effect was significant ($p<.05$).

Table 5. Means and Standard Deviations for 10% Missing Data

	MINORITY STATUS											
	MINORITY				NONMINORITY				TOTAL			
	METHOD											
Schc	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg
FD												
M	61.22	61.09	61.22	61.12	67.03	66.38	67.03	66.38	63.21	62.93	63.24	62.95
SD	12.84	12.47	12.46	12.56	13.04	12.68	12.49	13.27	13.14	12.74	12.72	12.99
N	65	69	69	69	34	37	37	37	99	106	106	106
HD												
M	49.68	50.77	49.68	50.62	65.74	65.05	65.74	65.42	57.61	57.76	57.55	57.86
SD	12.40	12.03	11.65	12.30	15.64	14.95	14.81	15.16	16.20	15.29	15.51	15.62
N	152	172	172	172	148	165	165	165	300	337	337	337
TOT												
M	53.14	53.72	52.99	53.62	65.98	65.29	65.98	65.60	59.00	59.00	58.91	59.08
SD	13.58	13.00	12.96	13.23	15.16	14.54	14.39	14.80	15.67	14.87	15.08	15.18
N	217	241	241	241	182	202	202	202	399	443	443	443

Table 6. ANOVA Summaries for the 10% Missing Data Sample

SOURCE	SS	DF	F	Sig of F
Listwise Deletion				
Minority status	8230.06	1	43.15	.000
Schedule	2826.05	1	14.82	.000
Minority status by Schedule	1805.55	1	9.47	.002
Within	75331.04	395		
Total	97763 00	398		
Mean Imputation				
Minority status	7174.55	1	40.51	.000
Schedule	2541.72	1	14.35	.000
Minority status by Schedule	1513.55	1	8.55	.004
Within	77754.33	439		
Total	97763.00	442		
Adjustment-Cell Imput.				
Minority status	8960.33	1	52.22	.000
Schedule	3076.81	1	17.93	000
Minority status by Schedule	1965.75	1	11 46	.001
Within	75331.04	439		
Total	100481.17	442		
Regression				
Minority status	7537 66	1	41.06	.000
Schedule	2459 17	1	13.40	000
Minority status by Schedule	1708.66	1	9 31	002
Within	80592.52	439		
Total	101809.37	442		

Analyses With 20% Missing Data

PPVT scores for a 20% random selection of cases were deleted in order to create missing data. This resulted in 89 cases with missing values. The same missing data techniques applied above were employed here as well. Means and standard deviations are provided in Table 7. There are greater differences in mean scores between techniques for this 20% missing data analysis compared to the 10% and 5% missing data analyses. For all but one cell (minority/full-day) of Table 7, the listwise technique resulted in the highest standard deviations. In contrast, the adjustment-cell and mean imputation techniques resulted in the lowest standard deviations. The ANOVA summaries are provided in Table 8. The conclusions are *not* similar for all techniques. Specifically, the interaction effect was not significant for listwise deletion (p>.05).

Table 7. Means and Standard Deviations for 20% Missing Data

	MINORITY STATUS											
	MINORITY				NONMINORITY				TOTAL			
	METHOD											
Sched	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg
FD												
M	60.21	59.73	60.21	60.41	67.78	65.33	67.78	66.71	62.97	61.69	62.85	62.61
SD	11.94	9.85	9.82	11.95	14.02	12.59	11.91	12.91	13.17	11.15	11.15	12.60
N	47	69	69	69	27	37	37	37	74	106	106	106
HD												
M	50.86	52.09	50.86	51.18	64.81	63.71	64.81	64.80	57.59	57.78	57.69	57.85
SD	12.65	11.96	11.61	12.41	16.43	15.04	14.86	15.29	16.16	14.73	15.01	15.46
N	145	172	172	172	135	165	165	165	280	337	337	337
TOT												
M	53.15	54.28	53.53	53.82	65.31	64.00	65.36	65.15	58.71	58.71	58.93	58.99
SD	13.09	11.89	11.89	12.95	16.06	14.61	14.38	14.87	15.72	14.05	14.34	14.95
N	192	241	241	241	162	202	202	202	354	443	443	443

Table 8. ANOVA Summaries for the 20% Missing Data Sample

SOURCE	SS	DF	F	Sig of F
Listwise Deletion				
Minority status	6380.14	1	31.49	.000
Schedule	2090.35	1	10.32	001
Minority status by Schedule	563.10	1	2.78	.096
Within	70914 87	350		
Total	87218.61	353		
Mean Imputation				
Minority status	5547.11	1	32.97	.000
Schedule	1608.73	1	9.56	.002
Minority status by Schedule	679.49	1	4.04	.045
Within	73867.64	439		
Total	87218.61	442		
Adjustment-Cell Imput.				
Minority status	8677.15	1	53.72	000
Schedule	2842.93	1	17.60	.000
Minority status by Schedule	765.84	1	4.74	.030
Within	70914.87	439		
Total	90853 89	442		
Regression				
Minority status	7434.97	1	40 60	.000
Schedule	2326.73	1	12.71	.000
Minority status by Schedule	1003.84	1	5.48	002
Within	80386.25	439		
Total	98800.72	442		

Analysis with 25% Missing Data

PPVT scores for a 25% random selection of cases were deleted in order to create missing data. This resulted in 111 cases with missing PPVT values. Like previous analyses, listwise deletion, mean imputation, adjustment-cell mean imputation and regression imputation missing data techniques were used. Means and standard deviations for the 25% missing data sample are provided in Table 9. Once again, the listwise deletion procedure resulted in the highest standard deviations while the two mean imputation techniques resulted in the lowest standard deviations. Particularly, the adjustment-cell mean technique had the lowest standard deviations for six of nine cells. The ANOVA summaries (see Table 10) show that the mean imputation technique did not have a significant interaction effect (p>.05); however, the other three techniques did show this effect.

Table 9. Means and Standard Deviations for 25% Missing Data

	MINORITY STATUS											
	MINORITY				NONMINORITY				TOTAL			
	METHOD											
Sched	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg
FD												
M	61.00	60.60	61.00	61.03	67.93	65.77	67.93	66.44	63.34	62.41	63.42	62.92
SD	12.86	11.49	11.46	12.10	12.37	11.39	10.72	12.32	13.15	11.67	11.64	12.39
N	55	69	69	69	28	37	37	37	83	106	106	106
HD												
M	50.16	52.79	50.16	51.80	64.67	63.41	64.67	64.54	57.62	57.99	57.26	58.04
SD	11.55	10.50	9.67	11.31	16.66	14.85	14.66	15.33	16.11	13.85	14.33	14.85
N	121	172	172	172	128	165	165	165	249	337	337	337
TOT												
M	53.55	55.03	53.26	54.45	65.26	63.84	65.27	64.88	59.05	59.05	58.74	59.21
SD	12.96	11.33	11.31	12.25	15.99	14.28	14.05	14.82	15.58	13.48	13.97	14.44
N	176	241	241	241	156	202	202	202	332	443	443	443

Table 10. **ANOVA Summaries for the 25% Missing Data Sample**

SOURCE	SS	DF	F	Sig of F
Listwise Deletion				
Minority status	6571.39	1	33.52	.000
Schedule	2841.10	1	14.49	.000
Minority status by Schedule	822.48	1	4.20	.041
Within	64308.09	328		
Total	80339.23	331		
Mean Imputation				
Minority status	4664.72	1	29.84	.000
Schedule	1936.43	1	12.39	.000
Minority status by Schedule	556.89	1	3.56	.060
Within	68631.72	439		
Total	80339.23	442		
Adjustment-Cell Imput.				
Minority status	8611.78	1	58.79	.000
Schedule	3723.25	1	25.42	.000
Minority status by Schedule	1077.86	1	7 36	.007
Within	64308.09	439		
Total	86261.10	442		
Regression				
Minority status	6163.13	1	35 97	.000
Schedule	2320.02	1	13.43	.000
Minority status by Schedule	1003.62	1	5.81	.016
Within	75854.83	439		
Total	92128.47	442		

Real Missing Data Problem

Pretest PPVT scores for 18.7% of the 443 cases were missing. Table 11 provides the means and standard deviations for pretest PPVT for each of the four solutions to missing data. Except for two instances listwise deletion resulted in the highest standard deviations . Regression resulted in the lowest means, with the exception of one cell (nonminority/male/list). The 2x2 ANOVAs computed using the four missing data techniques revealed similar results. Table 12 contains the ANOVA summaries. The interaction of minority status and gender was not significant for any of the missing data techniques ($p > .05$). Minority status was significant in all four analyses ($p < .05$), and gender was not significant.

Table 11. Means and Standard Deviations for the Real Missing Data Problem

	MINORITY STATUS											
	MINORITY				NONMINORITY				TOTAL			
	METHOD											
Gender	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg	List	Mean	Adjus	Reg
Male												
M	37.34	38.14	37.50	35.84	41.81	41.74	41.95	41.93	39.71	39.95	39.53	38.63
SD	14.96	12.52	12.41	13.27	14.05	13.50	14.05	13.64	14.62	13.05	13.09	13.75
N	87	126	126	126	98	106	106	106	185	232	232	232
Female												
M	38.38	39.01	38.25	37.39	45.72	45.37	45.57	44.90	42.11	41.90	41.58	40.81
SD	12.12	10.52	10.47	11.24	13.88	13.42	13.37	13.90	13.52	12.31	12.40	13.04
N	86	115	115	115	89	96	96	96	175	211	211	211
Total												
M	37.86	38.71	37.86	36.58	43.67	43.46	43.67	43.34	40.88	40.88	40.51	39.66
SD	13.60	11.59	11.51	12.34	14.97	13.55	13.53	13.81	14.13	12.73	12.79	13.44
N	173	241	241	241	187	202	202	202	360	443	443	443

Table 12. ANOVA Summary Table for Real Missing Data Problem

SOURCE	SS	DF	F	Sig of F
Listwise Deletion				
Minority status	3122.86	1	16.39	.000
Gender	550.27	1	2.89	.090
Minority status by Gender	185.37	1	.97	325
Within	67835.29	356		
Total	71626.62	359		
Mean Imputation				
Minority status	2553.10	1	16.37	.000
Gender	484.51	1	3.11	.079
Minority status by Gender	255.94	1	1.64	.201
Within				
Total				
Adjustment-Cell Input.				
Minority status	3790.50	1	24.51	.000
Gender	523.50	1	3.39	066
Minority status by Gender	226.66	1	1.47	.227
Within	67901.33	439		
Total	72302.025	442		
Regression				
Minority status	5072.20	1	29.98	.000
Gender	558.45	1	3 30	.070
Minority status by Gender	56.14	1	.332	565
Within	74277.70	439		
Total	79892.18	442		

DISCUSSION

A few conclusions may be drawn from the simulation study. In a sample of this size (n=443), effects of minority status and schedule on PPVT scores were similar for the four missing data techniques when data were missing for 10% of the cases. When missing data values were expanded to 20%, the interaction effect was not detected when using listwise deletion. Further, for the 25% missing data sample, the interaction effect was not detected using mean imputation. Unexpectedly, the interaction effect also was nonsignificant for the listwise deletion technique when only 5% of the PPVT values were missing. For all samples of missing data (5%, 10%, 20%, and 25%), adjustment-cell mean and regression imputation techniques resulted in similar conclusions. And, these conclusions were like those of the actual data set.

There were a few consistent characteristics of the missing data techniques utilized in this study. The two techniques utilizing mean imputation resulted in the lowest standard deviations. In most cases, listwise deletion resulted in the highest standard deviations. In general, the regression and adjustment-cell techniques resulted in means most consistent with actual means.

In the real missing data study, all four missing data techniques produced similar results. Therefore, even though values were missing for 18.7% of the sample, one may feel confident in making conclusions. Specifically, for the pretest PPVT assessment, nonminority students scored significantly higher than minority children. There was no difference between males and females on this measure.

The present research supports Raymond and Roberts (1987) conclusions regarding missing data studies: (a) estimation by regression appears beneficial when the data set has 10% to 20% missing data and the variables are moderately correlated, (b) listwise deletion has frequently been the least effective technique, and (c) substituting missing values with the variable mean can have deceptive results because of the tendency to attenuate variance and covariance estimates. However, in this study, the regression and adjustment-cell mean techniques were effective at producing results similar to the actual cases when missing data made up 5%, 10%, 20%, or 25% of the sample.

An important outcome of this study was that when only 5% of the values were missing, listwise deletion did not produce results that were found when using actual values. Researchers should therefore consider applying three or four of the missing data techniques even when there is missing data for a small percentage of the cases.

The generalizability of this study's results are limited. We must keep in mind that the effects of the minority status and schedule variables on PPVT were relatively large. We may have found additional discrepancies between techniques and/or for the different missing data samples (5%, 10%, 20%, 25%) if the effects were not as large. Furthermore, smaller sample sizes may be affected differently, particularly when using listwise deletion of missing data. A final consideration is that in this study, only one variable with missing data was utilized. Different results may have been obtained if many variables had missing data and multivariate analyses were employed.

In conclusion, when faced with the problem of missing data, researchers should first investigate whether the data are missing due to some factor or are missing at random. Norusis (1993) suggested dividing the data into two groups (those with missing values and those without missing values) and examining the distributions of other variables across these two groups. Next, if the data appear to be missing completely at random, apply a few of the missing data techniques. Finally, determine if different conclusions would be drawn when utilizing the missing data techniques. The researcher must carefully consider the consequences when different techniques lead to dissimilar conclusions.

References

Bailar, B.A. & Bailar, J.C. (1983). Comparison of the biases of the hot-deck imputation procedure with an equal-weights imputation procedure. In W.G, Madow, H. Nisselson, and I. Olkin (Eds.). Incomplete Data in Sample Surveys Volume 3. New York: Academic Press.

Coleman, J.S., Hoffer, T., & Kilgore, S. (1981). Contractors report: Public and private schools: An analysis of High School and Beyond. National Center for Educational Statistics, Washington, D.C.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39:1-39.

Ford, B.L. (1983). An overview of hot-deck procedures. In W.G, Madow, H. Nisselson, and I. Olkin (Eds.). Incomplete Data in Sample Surveys Volume 2. New York: Academic Press.

Frane, J.W. (1976). Some simple procedures for handling missing data in multivariate analysis. Psychometrika, 41(3) 409-415.

Gleason, T.C. & Staelin, R. (1975). A proposal for handling missing data. Psychometrika, 40, 229-252.

Kaiser, J. & Tracy, D.B. (1988). Estimation of missing values by predicted scores. Paper presented at the Annual Meeting of the American Statistical Association, New Orleans. (ERIC Document Reproduction Service No. 298 146)

Little, R.J.A., & Rubin, D.B. (1987). Statistical analysis with missing data. New York: John Wiley & Sons.

Little, R.J.A., & Rubin, D.B. (1990). The analysis of social science data with missing values. In J. Fox and J.S. Long (Eds.), Modern methods of data analysis. Newbury Park, CA: Sage Publications Inc.

Madow, W.G., Nisselson, H., & Olkin, I. (1983). Incomplete data in sample surveys Volumes 1-3. New York: Academic Press.

Muthen, B., Kaplan, D., & Hollis, M. (1987) On structural equation modeling with data that are not missing completely at random Psychometrika, 52, 431-462.

Norusis, M.J. (1993). SPSS for Windows: Advanced Statistics Release 6.0. SPSS Inc.

Page, E.B., & Keith, T.Z. (1981). Effects of U.S. private schools: A technical analysis of two recent claims. Educational Researcher, 10, 7-17.

Raymond, M.R. (1987). An interactive approach to analyzing incomplete multivariate data. Paper presented at the Annual Meeting of the American Educational Association. Washington, D.C. (ERIC Document Reproduction Service No. 281 854)

Raymond, M.R. & Roberts, D.M. (1987). A comparison of methods for treating incomplete data in selection research Educational and Psychological Measurement, 47, 13-26.

Walberg, H.J. & Shanahan, T (1983). High school effects on individual students. Educational Researcher, 12, 4-9.

Ward, T. J., & Clark, H.T (1991) A reexamination of public- versus private-school achievement: The case for missing data Journal of Educational Research, 84(3), 153-163.

Witta, L. & Kaiser, J. (1991). Four methods of handling missing data with the 1984 General Social Survey. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Lexington, KY. (ERIC Document Reproduction Service No. 339 755)