Missing Data in Mutivariate Linear Models: A Comparison of Several Estimation Techniques

James D. Hosking, University of North Carolina at Chapel Hill

INTRODUCTION

One of the most common problems in applied statistics is the issue of how to deal with "missing data." The term "missing data" has been used to describe a vast assortment of topics, some of which are only distantly related. The case considered here involves:

1. missing values only in the dependent variables with independent variables assumed to be fixed,
2. values missing at random, and
3. normally distributed dependent variables.

Specifically, the topic of this paper is a comparison of alternative techniques for estimation in the General Linear Multivariate Model (GLMM) of full rank and with normality assumptions when some of the dependent variable vectors have some components missing at random. The standard technique for dealing with missing data of this type is what will be called "listwise deletion," that is, discarding any observation containing one or more missing values. When the data are missing at random, using listwise deletion will not bias the estimates of the GLMM parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$. However, listwise deletion involves discarding information (contained in the observed components of the incomplete vectors) which may be useful for improving the precision of the estimates and the power of tests. This study focuses on three other techniques which have been proposed in the last 10 years. These appear to be the only techniques applicable to the situation described above which have been proposed on analytic, rather than heuristic grounds. The three techniques were each described in a pair of related articles:

1. Hocking and Smith (1968) and Hartley and Hocking (1971)
2. Woodbury and Hasselblad (1970) and Orchard and Woodbury (1972)
3. Kleinbaum (1970, 1973)

Each of these pairs of papers presents a different approach to the problem, and demonstrates various desirable properties of the approach. However, the only analytic results which have been presented concerning the techniques have been asymtotic. No studies have been published which investigated the behavior of the techniques in finite samples. This paper presents such results.

Description of the techniques

The first of these three techniques was first proposed, with heuristic justification, by Hocking and Smith (1968). It was subsequently formalized and generalized in Hartley and Hocking (1971). It will be referred to in this paper as the HHS technique.

In their development, the N observations are first partitioned into T groups, such that all $_tN$ observations in the t-th group have the same pattern of missing values. The likelihood function for the entire sample is then expressed as the product of the T group likelihood functions. Differentiating with respect to the vector $\underset{\sim}{\mu}$ and the matrix $\underset{\sim}{\Sigma}$ produces a system of equations which can be solved iteratively to obtain the maximum likelihood estimates of $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$. Their derivations considered only estimation of a mean vector $\underset{\sim}{\mu}$ rather than a matrix $\underset{\sim}{\beta}$. However, in an appendix they present the basic formuli for estimation of $\underset{\sim}{\beta}$. Hosking (1980) contains an explicit derivation of those results.

The second of the three techniques is presented in Woodbury and Hasselblad (1970) and Orchard and Woodbury (1972). It will subsequently be referred to as the WOH technique. It also produces maximum likelihood estimates of $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$ through an iterative technique. However, the development in these papers is quite different, with the problem being formulated and solved through what Woodbury terms "the missing information principle" or MIP. The two different derivations lead to very different computational algorithms. The first step in their algorithm is to use the complete observations to compute an initial estimate of $\underset{\sim}{\beta}$. This estimate is used to compute estimated $\underset{\sim}{Y}$ values to replace those which are missing. These are used to compute new estimates of $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$, which are in turn used to compute new estimates of the missing values. This process is iterated to convergence. The development of the WOH technique assumes a vector $\underset{\sim}{\mu}$ rather than a matrix $\underset{\sim}{\beta}$. However, Hosking (1980) presents an extension of the technique to estimation of $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\beta}$.

The third technique to be considered is quite different from the first two. It was first suggested by Kleinbaum (1970) in his dissertation, and later published in abbreviated form in Kleinbaum (1973). It will subsequently be referred to as the KLN technique. Kleinbaum proposed a generalization of the standard GLMM model which he called the more general linear model (MGLM). Special cases of this model include the multiple design matrix multivariate model (Srivastava, 1966), the seemingly unrelated regressions model (Zellner, 1962), and the case of interest here, the incomplete general linear multivariate model (IGLMM) (Srivastva, 1966). Limiting considerations to this special case allows Kleinbaum's complex notation (which involves four levels of subscripts at one point) to be vastly simplified, with a corresponding reduction of computational effort.

In Kleinbaum's approach, he estimates a matrix $\underset{\sim}{\xi}$, rather than $\underset{\sim}{\beta}$. For the IGLMM, $\underset{\sim}{\xi}$ is a reshaped form of $\underset{\sim}{\beta}$. If $\underset{\sim}{\beta}$ is q x p then $\underset{\sim}{\xi}$ is a pq x 1 vector formed by concatenating the p columns of $\underset{\sim}{\beta}$ under each other. In order to estimate $\underset{\sim}{\beta}$ in this form, it is necessary to reshape $\underset{\sim}{Y}$ into an N x 1 vector, and $\underset{\sim}{X}$ into an N x pq matrix. $\underset{\sim}{Y}$ is defined by $\underset{\sim}{Y}' = (\underset{\sim}{Y}_1'\underset{\sim}{Y}_2'\ldots\underset{\sim}{Y}_p')$, where $\underset{\sim}{Y}_s$ is the n_s x 1 vector of all observed values of variable s and N defined by:

$$N = \sum_{s=1}^{p} n_s$$

Note that N is the total number of observed dependent variable values, not the number of subjects. $\underset{\sim}{X}$ is a block diagonal matrix defined by:

$$\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{X}_1 & & & \\ & \underset{\sim}{X}_2 & & 0 \\ & & \ddots & \\ 0 & & & \underset{\sim}{X}_q \end{bmatrix}$$

where $\underset{\sim}{X}_s$ is the $n_s \times q$ subset of rows from the original design matrix corresponding to those observations having non-missing values for the s^{th} dependent variable. Note that this is perfectly valid as a representation of the standard GLMM. As is demonstrated by Kleinbaum (1970) such a representation will produce the same estimates for the elements of $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$; and any hypothesis of the form H_0: $\underset{\sim}{C}\,\underset{\sim}{\beta}\,\underset{\sim}{U} = \underset{\sim}{0}$ can be transformed into one of the form H_0: $\underset{\sim}{H}\,\underset{\sim}{\xi} = \underset{\sim}{0}$. In fact, this representation is more general. Using it, one can test hypotheses which cannot be expressed in the form $\underset{\sim}{C}\,\underset{\sim}{\beta}\,\underset{\sim}{U}$.

The primary disadvantage of this method is that one of the matrices involved in the formulas for estimation and hypothesis testing is of order N x N, which quickly becomes too large for practical computations. However, for the IGLMM, this matrix is never needed. For this special case, the matrix is block diagonal, and it is possible to exploit this fact when estimating $\underset{\sim}{\xi}$ and $\underset{\sim}{\Sigma}$. Kleinbaum derives a family of BAN, unbiased estimators for $\underset{\sim}{\xi}$ based on any consistent estimator of $\underset{\sim}{\Sigma}$. $\underset{\sim}{\Sigma}$ produced by listwise deletion is an unbiased and consistent estimator of $\underset{\sim}{\Sigma}$. Hence one estimation method, which would not be iterative, is to use that $\underset{\sim}{\Sigma}$ to estimate $\underset{\sim}{\xi}$. However, he also suggests two iterative techniques which may produce BAN estimates of $\underset{\sim}{\xi}$ which are better, in the sense of producing estimators whose variance converges on the limiting variance at a faster rate as sample size increases (Kleinbaum, 1970). Each of these iterative algorithms adjusts the initial estimates of $\underset{\sim}{\xi}$ and $\underset{\sim}{\Sigma}$ in much the same way as the maximum likelihood algorithms.

An important characteristic of Kleinbaum's work is that he derives a test statistic, asymtotically chi-square, for testing any hypothesis of the form H_0: $\underset{\sim}{H}\,\underset{\sim}{\xi} = \underset{\sim}{0}$. It reduces to the Hotelling-Lawley trace statistic when the data are complete.

Theoretical Relationships Among the Techniques

Equivalence of the WOH and HHS Techniques

As was mentioned above, the WOH and HHS techniques both produce MLEs, although by very different algorithms. This was first noted by Hartley and Hocking (1971) in their reply to the discussion by Woodbury of their paper. Hosking (1980) presents an explicit derivation of the equivalence of the two techniques. Dempster, Laird, and Rubin (1977) show that both algorithms may be regarded as special cases of their general EM algorithm.

Relationship between the KLN and MLE Techniques

It is well known (see E.G. Kendall and Stuart, 1979; Dudewicz, 1976) that sufficient conditions for an estimator to be BAN are

1) that it be an MLE, and
2) that the first and second derivatives of the likelihood function exist.

As discussed above, the WOH and HHS techniques each produce MLEs for $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$ using different computational algorithms. In both cases the marginal likelihood of the observed data is the likelihood maximized. It is clear that the first and second derivatives of that likelihood function exist, since it is simply the product of T independent "standard" multivariate normal likelihoods (one for each pattern of missing data). Hence it follows immediately that the MLE techniques are BAN. It is obvious from the derivations, however, that those estimators are not equal to Kleinbaum's BAN estimator in finite samples. For example, the MLE algorithms use a denominator of N in computing $\underset{\sim}{\Sigma}$, while Kleinbaum's algorithm uses $N_{ij} - q$ as the devisor for elements (i,j) and (j,i) of $\underset{\sim}{\Sigma}$.

It would be possible to use N as the devisor for every element of $\underset{\sim}{\Sigma}$ in the Kleinbaum technique, since $\underset{\sim}{\xi}$ would still be consistant (although biased). This would produce an alternate set of BAN estimators of $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$. However, it has not been possible to show any relationship between these estimators and the MLEs in finite samples.

DESCRIPTION OF THE MONTE-CARLO STUDY

There are three basic issues which this study was designed to address. First, do the techniques work? That is, do the algorithms converge, and if so do they converge to reasonable estimates? Second, are the techniques any better than standard techniques such as listwise deletion? More generally, how do the accuracy of the estimates compare with each other and with those produced by the techniques currently in use? Third, are the techniques practical? That is, do they require an excessive amount of memory or time to produce estimates for problems of reasonable size? As will be shown subsequently, the study described here provided fairly consistent, clear answers to each of these questions for the conditions considered.

Techniques Investigated

In all, six techniques were investigated. The four already discussed were listwise deletion and the three algorithms described above. The fifth was simply analysis of the complete data (before any values were set to missing). This was included to provide an upper bound against which to compare the performance of the other techniques. The sixth technique examined was pairwise deletion. Although little can be said about the finite sample properties of the pairwise estimators of $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$, they are consistent and unbiased. Furthermore, they are intuitively appealing in that the estimate of each element of $\underset{\sim}{\beta}$ is based on all of

the non-missing observations for the corresponding variable and the estimate of each element of $\underset{\sim}{\Sigma}$ is based on all of the non-missing observations for the corresponding pair of variables. However, the estimate of $\underset{\sim}{\Sigma}$ produced by pairwise deletion may be singular. While this does not pose a problem for point estimation, it is extremely undesirable if interval estimation, hypothesis testing, or multivariate analyses (such as factor analysis) are of interest.

Factors Varied

The three factors varied in the study were sample size, proportion of missing values, and intercorrelation among the dependent variables. Two levels were chosen for each factor, resulting in a 2 x 2 x 2 design for the study. Sample sizes of 30 and 60 were used. Data sets containing 10 and 20 percent of missing values were examined. Two patterns of intercorrelation were selected, one with an average off-diagonal correlation of .3 and the other with an average off-diagonal correlation of .6.

Other Characteristics of the Study

The three factors just discussed were selected to be varied because they seemed the most likely to affect the relative performance of the techniques. In this section, the values chosen for other characteristics of the situation are presented. Any of these factors might also affect the performance of the techniques, but they were felt to be less likely to differentially affect the techniques than those above.

The number of dependent variables was fixed at three. The variances chosen for the dependent variables were 9, 4, and 16, producing the covariance matrices presented in Table 1. The structure chosen for X was a natural polynomial model. Values of X_1 were generated as integers uniformly distributed on the interval (0,9). X_0, X_2, and X_3 were then computed as X_1 raised to the powers 0, 2, and 3, respectively. The structure chosen for $\underset{\sim}{\beta}$ is presented in Table 2. The number of samples generated for each condition was 50.

Measures of Precision of Estimates

The measures used to evaluate the performance of the techniques can be grouped into two classes: matrix valued measures and summary measures. The matrix valued measures have a separate component for each element of $\underset{\sim}{\beta}$ and each unique element of $\underset{\sim}{\Sigma}$. The summary measures combine, in various ways, the components of the matrix valued measures to provide univariate summaries of the accuracy of the techniques. The summary measures produce two numbers for each of the 50 samples within a cell, one computed across the 12 elements of $\underset{\sim}{\beta}$ and the other across the six distinct elements of $\underset{\sim}{\Sigma}$. The conclusions drawn from these two classes of dependent variables were consistent. Due to space constraints I will discuss only the summary measures here.

The three summary measures used were: 1) the mean absolute deviation (MAD), 2) the mean squared deviation (MSD), and 3) the maximum absolute deviation, (MXAD). These measures were computed separately for $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$. In many cases, estimation of one of the parameters might be of much more interest than estimation of the other. For example, if the primary purpose of a study were to examine the structure of the residual correlation matrix among the dependent variables, estimation of $\underset{\sim}{\Sigma}$ would be paramount. Estimation of β might be of interest only for the contribution of $X\beta$ to the estimation of $\underset{\sim}{\Sigma}$. In other cases, point estimation of $\underset{\sim}{\beta}$ might be of primary importance, or both might be equally important. Hence the performance of the techniques is evaluated separately for $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$.

Let $\ddot{\theta}_{ij}$ be an element of $\underset{\sim}{\beta}$ or $\underset{\sim}{\Sigma}$ from sample i, and let θ_j be the corresponding parameter. Then the summary measures are defined by:

$$1)\ MAD_i = \sum_{j=1}^{J} \left|\ddot{\theta}_{ij} - \theta_j\right| / J,$$

$$2)\ MSD_i = \sum_{j=1}^{J} (\ddot{\theta}_{ij} - \theta_j)^2 / J,$$

$$3)\ MXAD_i = \max_{j=1}^{J} \left|\ddot{\theta}_{ij} - \theta j\right|.$$

Execution of the Monte-Carlo

The first step in performing the study was to generate 50 samples from each of the four combinations of sample size and level of intercorrelation. For each combination, a single $\underset{\sim}{X}$ matrix was generated. Given $\underset{\sim}{X}$ and the corresponding $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$, 50 sample data sets were generated. Each observation contained the variables X_0, X_1, X_2, X_3, and Y_1, Y_2, and Y_3 as well as a sample number and observation number. The data sets were generated by a SAS procedure PROC DATAGEN, written for this simulation. The $\mathcal{N}(\underset{\sim}{X}\underset{\sim}{\beta}, \underset{\sim}{\Sigma})$ pseudo-random deviates were generated using standard techniques.

The second step was to produce the data sets containing 10 and 20 percent missing values. A second SAS procedure, PROC DATADLT, was used to read the complete datasets and produce output data sets in which each Y value was set to missing with the desired probability. PROC DATADLT also added three variables to the output observations. Two contain strings of ones and zeroes representing the pattern of observed and missing variables (in bases 2 and 10).

The final step was to analyze the data sets using each of the six techniques. The analysis routines were written in PROC MATRIX, with several new functions written in FORTRAN and added to PROC MATRIX to enhance the efficiency of the code. The choice of PROC MATRIX as the language was made primarily to keep the programming task reasonable. Even though PROC MATRIX is a high level language for statistical computing, over 1000 lines of PROC MATRIX code were needed to program the six algorithms. Even using subroutine libraries such as IMSL, the corresponding routines would have been much longer in PLI or FORTRAN. Furthermore, the functions and operators in

PROC MATRIX are PLO and ASSEMBLER load modules and are at least as efficient as their IMSL counterparts (although completely dependent on IBM 360/370 architecture). In all, approximately 1000 lines of PLO and FORTRAN H code were written (PROC DATAGEN, PROC DATADLT, and the functions) and added to SAS and PROC MATRIX. Overall, the efficiency of the resulting programs were acceptable. With one exception (discussed later), the techniques required at most 20 seconds to compute estimates of $\underset{\sim}{\beta}$ and $\underset{\sim}{\Sigma}$ for any combination of the factors.

RESULTS OF THE MONTE-CARLO STUDY

Before discussing the details of the results, it seems useful to provide an overall summary, since the major results were remarkably consistent. The single most striking result is that the KLN and WOH algorithms do indeed "work" (i.e. they converge on reasonable estimates), but the HHS algorithm does not. Because of the need to invert $({}_{g}\underset{\sim}{X}'\ {}_{g}\underset{\sim}{X})$ for each group, the HHS algorithm reduced to listwise deletion in about a third of the samples. Even when at least one large enough group of observations with a common pattern of missing values was available, its behavior was unacceptable. In contrast to the other two algorithms which typically converged in five to ten iterations (and never required more than 50), the HHS algorithm failed to converge in 100 iterations for approximately 10 percent of the samples. Due to these problems, this technique is omitted from the discussion which follows.

In contrast to the HHS algorithm, the WOH and KLN algorithms always converged on reasonable estimates in a reasonable time. In general, these techniques were both substantially better than listwise deletion, using any measure of accuracy in any of the cells.

Results from the Summary Measures

Overall, the KLN and WOH techniques are clearly better than listwise deletion. The three measures, eight cells, and two parameters produce 48 comparisons. The KLN measures are better than the corresponding listwise measures in all 48. The WOH measures are better in all 24 comparisons involving $\underset{\sim}{\beta}$ and in 21 of 24 comparisons involving $\underset{\sim}{\Sigma}$. Most of these mean differences are at least three standard errors in magnitude. This can also be clearly seen in Table 3 which presents marginal means for each summary measure, broken down by method versus number of subjects, proportion of missing values and average intercorrelation.

The comparisons among the KLN, WOH, and pairwise techniques are less consistent. For estimation of $\underset{\sim}{\Sigma}$ the KLN technique is consistently better than either the WOH or pairwise techniques, which are roughly comparable. For estimation of $\underset{\sim}{\beta}$ the WOH technique is somewhat better than the KLN and pairwise techniques, which are roughly comparable.

The marginal effects of the number of subjects and the probability of missing values were consistent, large and as expected; precision improved with more subjects and worsened with more missing data. The accuracy improved with increasing average intercorrelation for $\underset{\sim}{\beta}$ but did not change for $\underset{\sim}{\Sigma}$.

The only major "interaction" revealed by these tables is between method and average intercorrelation. For the low level of intercorrelation, pairwise is about as good as the KLN and WOH methods, but for high intercorrelation the WOH and (especially) the KLN algorithms are better.

Cost Comparisons Among the Techniques

The relative cost of the techniques is measured by two variables, CPU second equivalent per sample and maximum main storage required per condition. CPU second equivalent is a weighted linear combination of actual CPU seconds, unit record EXCPs, disk EXCPs, tape charges (none in this study), and memory charges (constant across methods and conditions in this study) used by the computer center to compute job costs.

The pattern of results for CPU equivalent is quite interesting. The figures for listwise deletion were about 20 percent greater than for analysis of the complete data. Pairwise deletion was about 10 percent more than listwise. Neither technique was affected by the proportion of missing values or the average intercorrelation among the factors. The WOH algorithm was about 35 percent more expensive than listwise deletion when 10 percent of the data were missing. When 20 percent of the data were missing, the WOH algorithm was over twice as expensive as listwise deletion. The effect of average intercorrelation was negligible, as was the effect of number of observations. The KLN algorithm was much more expensive than any of the other techniques, ranging from two to five times as expensive as the WOH algorithm. In contrast to all of the other techniques, the number of subjects had a dramatic effect on the cost of the KLN algorithm. The conditions with 60 observations were roughly twice as expensive as those with 30 observations. The proportions of missing data had a substantial effect on the KLN algorithm but in the opposite direction from its effect on the WOH algorithm. The conditions with 20 percent missing data were about 10 percent less expensive than those with 10 percent missing data. The average intercorrelation again had little effect.

The differences among the pairwise, listwise and complete data algorithms in maximum core used are again small. The listwise algorithm used about the same amount as the complete data algorithm, while the pairwise algorithm used about three percent more than the other two. The WOH algorithm used about six percent more core than the listwise algorithm. For all four of these techniques, the number of subjects had an effect of about five percent, while the proportion of missing data and the average intercorrelation had no effect. The KLN algorithm required much more

core than any of the other techniques. The number of subjects had the greatest influence on the core requirements. For the 30 subject conditions, the KLN technique required about 20 percent more core than the listwise algorithm; for the 60 subject conditions, the KLN technique required twice as much core. The conditions with 30 subjects and 20 percent missing data used about 10 percent less storage than the conditions with 30 subjects and 10 percent missing data. In the 60 subject conditions, those with 20 percent missing data required about 25 percent less core. Once again, the average intercorrelation had no substantial effect.

Overall, two conclusions are apparent from the cost comparisons. First, both the new techniques are cheap enough to be usable, although considerably more expensive than standard techniques. Second, the WOH technique is clearly much less expensive than the KLN technique.

DISCUSSION

Overall, the results of the study were encouraging. Although the HHS algorithm does not seem usable, both the KLN and WOH technique offered substantial improvement over listwise deletion. They were at least as effective as pairwise deletion; more effective when the average intercorrelations among the dependent variables is high. Neither showed any problems with convergence for any of the conditions studied. Both had time and core requirements which were feasable for general use, although much greater than the listwise or pairwise techniques. In summary, the results obtained suggest that these may be valuable techniques for use whenever the cost of collecting additional data exceeds their additional computational expense.

REFERENCES

Dudewicz, E.J. Introduction to Statistics and Probability. New York: Holt, Rineholt & Winston, 1976.

Hartley, H.O. & Hocking, R.R. The analysis of incomplete data (with discussion). Biometrics, 1971, 27, 783-808.

Hocking, R.M. & Smith, W.B. Estimation of parameters in the multivariate normal distribution with missing observations. Journal of the American Statistical Association, 1968, 63, 159-173.

Hosking, J.D. Missing Data in Multivariate Linear Models: A Comparison of Several Estimation Techniques. Ph.D. Dissertation, University of North Carolina, 1980.

Kleinbaum, D.G. Estimation and hypothesis testing for generalized multivariate linear models. University of North Carolina Institute of Statistics Mimeo Series, No. 669, 1970.

Kleinbaum, D.G. Testing linear hypotheses in generalized multivariate linear models. Communications in Statistics, 1973, 1, 433-457.

Kendall, M.G. & Stuart, A. The Advanced Theory of Statistics (3 vols). New York: Hafner, 1979.

Orchard, T. & Woodbury, M.A. A missing information principal: Theory and applications. Sixth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1972, 697-715.

Srivastava, J.N. Some generalizations of multivariate analysis of variance. In P.R. Krishnaiah (Ed.) Multivariate Analysis, New York: Academic Press, 1966.

Woodbury, M.A. & Hasselblad, V. Maximum likelihood estimates of the variance-covariance matrix from the multivariate normal. Paper presented at the SHARE national meeting, Denver, March 1970.

Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 1962, 57, 348-368.

Table 1
Population Covariance Matrices Used in the Monte-Carlo Study

Average Intercorrelation	Dependent Variables Y_1	Y_2	Y_3
Low	9.0	1.2	4.8
		4.0	2.4
			16.
High	9.0	4.2	6.0
		4.0	4.8
			16.

Table 2
Population Value of β for the Monte-Carlo Study

Independent Variable	Dependent Variable Y_1	Y_2	Y_3
X_0	1	4	10
X_1	1	3	-2
X_2	1	2	0
X_3	1	1	-2

Table 3

Marginal Means of Summary Accuracy Measures by Estimation Method

Factor	Estimation Method	Beta MAD	Beta MSD	Beta MXAD	Sigma MAD	Sigma MSD	Sigma MXAD
Overall							
	LISTWISE	0.85	2.84	3.03	2.61	8.16	5.58
	PAIRWISE	0.63	1.13	2.31	2.37	6.54	4.94
	KLNEST	0.62	1.13	2.27	2.29	5.80	4.71
	WOHEST	0.61	1.07	2.23	2.42	6.44	5.07
Number of Observations							
30	LISTWISE	1.05	4.54	3.80	2.98	10.98	6.73
	PAIRWISE	0.74	1.55	2.78	2.69	8.67	5.77
	KLNEST	0.73	1.56	2.71	2.56	7.37	5.41
	WOHEST	0.72	1.47	2.68	2.64	7.90	5.70
60	LISTWISE	0.66	1.13	2.27	2.23	5.35	4.43
	PAIRWISE	0.52	0.71	1.85	2.05	4.42	4.11
	KLNEST	0.51	0.70	1.83	2.23	4.22	4.01
	WOHEST	0.50	0.67	1.79	2.21	4.98	4.43
Proportion of Missing Data							
.1	LISTWISE	0.71	1.40	2.50	2.49	7.38	5.30
	PAIRWISE	0.60	1.01	2.17	2.34	6.39	4.83
	KLNEST	0.59	0.97	2.13	2.27	5.70	4.63
	WOHEST	0.58	0.95	2.09	2.30	5.67	4.73
.2	LISTWISE	0.99	4.27	3.57	2.73	8.95	5.86
	PAIRWISE	0.66	1.25	2.46	2.40	6.70	5.04
	KLNEST	0.65	1.29	2.41	2.31	5.89	4.80
	WOHEST	0.64	1.20	2.54	2.54	7.13	5.40
Average Intercorrelation							
.3	LISTWISE	0.91	3.89	3.35	2.58	7.99	5.65
	PAIRWISE	0.65	1.23	2.42	2.28	5.87	4.87
	KLNEST	0.66	1.32	2.44	2.32	5.63	4.66
	WOHEST	0.65	1.24	2.42	2.53	6.68	5.09
.6	LISTWISE	0.79	1.78	2.71	2.64	8.34	5.51
	PAIRWISE	0.61	1.03	2.20	2.46	7.22	4.99
	KLNEST	0.59	0.94	2.10	2.27	5.96	4.76
	WOHEST	0.58	0.90	2.05	2.32	6.20	5.04