THE EGYPTION POPULATION AND
FAMILY PLANNING REVIEW
ISSR,CAIRO UNIV. Vol.37,No.1,2004

A Comparatively Studies for Missing Data in Regression Analysis

Ahmed H. Youssef
Assistance Professor, Department of Applied
Statistics and Econometrics, Institute of Statistical
Studies And Research, Cairo University.
ahyoussef@hotmail.com

The statistical analysis with missing data is an important applied problem because missing data are commonly encountered in practice and most data analysis procedures were not designed for missing data. In this paper, we introduce a recent comparison for the methodologies that handle missing data problems. Some real examples of missing data problems, the pattern of missing data and the mechanisms that lead to missing data will be introduced in the case of regression analysis. A comparative study among the complete case (CC) method, the available case (AC) method, the least squares (LS) on imputed data and the maximum likelihood (ML) are presented. A real data and GPA scores for undergraduate students, College of Administrative Sciences at King Saud University in 1998, is used to make an artificial missing data in three different types of the mechanisms: Missing At Random (MAR), Missing Completely At Random (MCAR), and Missing Not At Random (MNAR). Then we reanalyze the new data set, using a Monte Carlo simulation, for the different methods to make some comparisons.

Key Words: Missing Data; Imputed Data; Regression Analysis; Simulation Studies.

1. Introduction

It is natural to treat the data that are not observed as missing. For example, the respondents in a household survey may refuse to report income. In an industrial experiment, some results are missing because of mechanical breakdowns unrelated to the experimental process. However, it is less natural to treat the unobserved data as missing. Such that, in an opinion survey some individuals may be unable to express a preference for one candidate over another. The lack of a response is essentially an additional point in sample space of the variable being measured, which identified as "Don't know", "Refused", "unintelligible" and so on.

Most statistical software packages allow using more than one code to identify particular type of non-response. Therefore, we can exclude units that have missing value codes for any of the variables involved in an analysis.

The literature on the statistical analysis of data with missing values has flourished since the early 1970s, spurred by advances in computer technology that made previously laborious numerical calculation a simple matter. The problem of estimating the parameters of statistical models when some observations are missing has been treated along two different types. In the first type of methods, the missing data are considered as functions of additional parameters see Beale and Little (1975); Conniffe (1983); Little and Rubin (1983, 2002); Dixon (1983, 1988); and Park, Lee and Woolson (1993).

Many data sets can be arranged on a rectangular or matrix form, where the rows correspond to observational units or participants and the columns correspond to items or variables. With rectangular data, there are several important classes of overall missing data patterns, for detail see Little & Rubin (2002), but we will consider the univariate pattern examples of missing data:

Missing values occur on an item Z but a set of P other items X_1 , X_2 ,…,X_p is complete observed data, we call this a univariate pattern. The univariate pattern is also meant to include situations in which Z represents a group of items that is either entirely observed or entirely missing for each unit.

Rubin's (1976) definitions describe statistical relationships between the data and the relationships between the data and the missingness, not causal relationships. Because we often consider real world reasons why data become missing, let us imagine that one could code all the myriad reasons for missingness into a set of variables. This set might include variables that explain why some participants were physical unable to show up (age, health status), variables that explain the tendency to say "I don't know" or "I'm not sure" (cognitive functioning), variables that explain outright refusal (concerns about privacy), and so on. This causes of missingness are not likely to be present in data set, but some of them are possibly related to X and Z.

If the participants are independently sampled from the population, then we can define the missing data as "Missing At Random" (MAR), if the missing data depends on X but not Z. However, if the missing data does not depend on his or her own values of X or Z (and by independence, does not depend on the X or Y of the other participants either), then the missing data is called "Missing Completely At Random" (MCAR). The missing data are said to be "Missing Not At Random" (MNAR), if the missing data depends on Z.

2. Methods of Estimation

We explain the methods on which the missing data are considered as functions of additional parameters, such that the least squares solution with complete case analysis, the available case analysis, and the least squares on imputed data, and the maximum likelihood.

2.1 The exact least squares solution with complete data

The classical linear regression model is concerned with the association between a single dependent variable Y and a collection of predictor variables $X_1, X_2, \ldots, Xp$. The regression model that we have considered treats Y as a random variable whose mean depends on fixed values of the X_i's. This mean is assumed to be a linear function of the regression coefficients $b_0, b_1, b_2, \ldots, bp$. Suppose all the variables $Y, X_1, X_2, \ldots, Xp$ are random and have joint distribution, not necessarily normal, with mean μ, an (p+1) x 1 vector and variances-covariance matrix Σ, an (p+1) x (p+1) matrix. Partitioning μ and Σ in an obvious fashion, then

$$\mu = \begin{pmatrix} \underset{(1x1)}{\mu_y} \\ \underset{(px1)}{\mu_x} \end{pmatrix},$$

and

$$\Sigma = \begin{pmatrix} \underset{(1x1)}{\sigma_{YY}} & \underset{(1xP)}{\Sigma_{YX}} \\ \underset{(Px1)}{\Sigma_{XY}} & \underset{(PXP)}{\Sigma_{XX}} \end{pmatrix}.$$

Where $\Sigma_{YX} = (\sigma_{YX_1} \quad \sigma_{YX_2} \quad \cdots \quad \sigma_{YX_P})$, and Σ_{xx} can be taken to have full rank.We focus on homoscedastic linear regression where

$$E(Y \setminus X_1, X_2, \cdots, X_P) = \beta_0 + \sum_{i=1}^{P} \beta_i X_i\,,$$

and

$$Var(Y \setminus X_1, X_2, \cdots, X_P) = \sigma^2.$$

The mean square error, for complete data, is minimized, when

$$\left.\begin{aligned} \beta_0 &= \mu_y - \sum_{j=1}^{P} \beta_j \mu_j , \\ \beta &= \Sigma_{XX}^{-1} \Sigma_{YX} , \\ \sigma^2 &= \sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} , \end{aligned}\right\} \quad (1)$$

and

$$Var(\beta) = \frac{\sigma^2}{n} \Sigma_{xx}^{-1}$$

Least squares (LS) estimates are obtained by replacing μ and Σ with the first and second sample moments.

2.2 The complete-case analysis (CC)

In this method, we discard any case which contains a missing value (i.e. we delete any row that contains a missing value from the rectangular data), and we analyze the remaining data by using the standard statistical analysis without modification. This method is also known as "list wise deletion" or "case deletion". Rubin (1976) suggested the dropping regressor variables with high levels of non-response, while Little (1993) applied this method in multiple regression case.

2.3 The available-case analysis (AC)

A natural alternative procedure for complete case analysis is to include all the available values for estimating μ and Σ. This method is called “Available-Case” and some time called “Pair-wise deletion”. We use every observed value of X_j to estimate the standard deviation of X_j, and every observed pair of values (X_j, X_k) to estimate the covariance of X_j, and X_k . That is

$$S_{jk}^{(jk)} = \frac{1}{\left(n^{(jk)} - 1\right)} \sum_{jk} \left(X_{ij} - \bar{X}_j^{(jk)}\right)\left(X_{ik} - \bar{X}_k^{(jk)}\right). \quad (2)$$

Where $n^{(jk)}$ is the number of cases with both X_j and X_k observed, and the means

$\bar{X}_j^{(jk)}$, $\bar{X}_k^{(jk)}$, and the summation in (2) are calculated over those $n^{(jk)}$ cases. Let $S_{jj}^{(j)}$, and $S_{kk}^{(k)}$ be the sample variances of X_j and X_k from available cases.

A defect of this method is that the estimated covariance matrix of the X's is not necessarily positive definite. Haitovsky (1968) found this problem is severe when the variables X's are highly correlated, while Dixon (1983) suggested the estimator of sample variances as

$$\tilde{S}_{jk}^{(jk)} = \frac{1}{\left(n^{(jk)} - 1\right)} \sum_{jk} \left(X_{ij} - \bar{X}_j^{(j)}\right)\left(X_{ik} - \bar{X}_k^{(k)}\right). \quad (3)$$

Where $\bar{X}_j^{(j)}, \bar{X}_k^{(k)}$ be the sample mean from the available cases in variables X_j and X_k, respectively. The variance covariance matrix for AC estimators is

$$Var(\tilde{\beta}) \cong \frac{\tilde{\sigma}^2}{\tilde{n}} \tilde{\Sigma}_{xx}^{-1}. \quad (4)$$

Where $\tilde{\Sigma}_{xx}$ and $\tilde{\sigma}^2$ are AC estimates of Σ_{xx} and σ^2, and $\tilde{n}$ is the harmonic mean of the sample size of the individual variables (Dixon, 1988). Little (1993) applied this method in multiple regression cases.

2.4 The least squares on imputed conditional means

We impute (or fill in) the missing X's using the regression of Y on X's and compute the filled in data by ordinary least squares (OLS) and reanalyze the new data with the ordinary least squares (OLS) method. There are different ways to impute the data and we will concern with the following ways of imputations

A simple approach imputes missing X's by their unconditional sample mean. Haitovsky (1968) found that, assuming MCAR, this method yields an inconsistent estimate of Σ. The sample variance of X_j is biased by a factor $(n^{(j)}-1)/(n-1)$ and the sample covariance of X_j and X_k is biased by a factor $(n^{(jk)}-1)/(n-1)$. Adjustment factors $(n-1)/(n^{(j)}-1)$ for the variance of X_j and $(n-1)/(n^{(jk)}-1)$ for the covariance of X_j and X_k, simply yield the estimates of (3).

A worthwhile improvement, imputing conditional means based on X's, is to use information in the observed X's in a case to impute the missing X's. Dear (1959) and Timm (1970) based imputations on a principal component analysis. However, a more obvious approach is to impute for a missing X by linear regression on the observed X's in that case, estimated from the complete cases (Dagenais, 1973).

For univariate missing data, suppose that X_1 is observed for (m) cases and missing for (n - m) cases. Since

$$E(y_i / x_i) = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$

and

$$E(y_i / x_{i2}, x_{i3}, \dots, x_{ip}) = \beta_0 + \beta_1 x_{i1}^{*} + \sum_{j=2}^{p} \beta_j x_{ij}$$

where $x_{i1}^{*} = E(x_{i1} / x_{i2}, \dots, x_{ip})$. Thus, if conditional means x_{i1}^{*} are substituted for missing values of X_{i1}. Then (LS) on the filled-in data produces consistent estimates of the regression coefficients assuming MCAR. Let (s) signifies to the set of subscripts (2, ... p) and let $\sigma_{yy.s}$ and $\sigma_{yy.1s}$ denote the residual variances of Y given $X_2, \dots, X_p$ and Y given $X_1, X_2, \dots, X_p$ respectively. To compensate for the increased residual variance when X_1 is missing, incomplete cases should be assigned the reduced weight.

$$w^{*} = \sigma_{yy.1s} / \sigma_{yy.s} = 1 - \rho_{1y.s}^{2} \tag{5}$$

Where $\rho_{1y.s}$ is the partial correlation of X_1 and Y given $X_2, \dots, X_p$. Replacing the parameters in (5) by sample estimates yield weights proposed by (Dagenais, 1973 and Beale and Little, 1975). The imputations x_{i1}^{*} depend on the unknown regression parameters, which in practice must be estimated from the data.

Gourieroux and Montfort (1981) and Conniffe (1983) noted that estimation error in regression coefficients inflates the residual variance and introduces a correlation between the incomplete observations. This does not affect the consistency of the (WLS) estimates, but does affect the best choice of weight and consistency of estimates of standard error. Arguing by a rather loose analogy with generalized least square (GLS), these authors proposed improved weight.

$$w = \frac{\left(1-\rho_{1y.s}^{2}\right)\frac{m}{n}}{\rho_{1y.s}^{2} + \left(1-\rho_{1y.s}^{2}\right)\frac{m}{n}}, \tag{6}$$

Which is approximately(5), when the fraction of complete cases is large, but gives less weight to the incomplete cases. Dagenais (1973) proposed this weighting with imputations and weights based on the complete cases. Beale and Little (1975) studied a similar method but with imputations based on an estimate of the covariance matrix Σ_{xx} that used all the data. Analogs of the weight (6) for general pattern of missing data have not been developed.

2.5 The maximum likelihood method

Many method of estimation for incomplete data can be viewed as maximizing the likelihood function under certain modeling assumptions. Anderson (1957) introduced the important idea of factoring the likelihood to obtain explicitly ML solutions for special pattern of missing data. Gourieroux and Montfort (1981) applied Anderson's method to regression with missing one independent variable for data pattern in figure (1), in the case of multivariate normal distribution, with mean μ and variance covariance matrix Σ. The distribution of X_1 and Y given the other X's can be defined as

$$f\left(x_1, y / x_2, \cdots, x_p; \theta\right) = f\left(x_1 / y, x_2, \cdots, x_p; \phi_1\right) . f\left(y / x_2, \cdots, x_p; \phi_2\right)$$

The corresponding likelihood of ϕ_1 and ϕ_2 are

$$L\left(\phi_1, \phi_2\right) = L_1\left(\phi_1\right) . L_2\left(\phi_2\right). \tag{7}$$

Where L_1 is the product of the normal density of X_1 given $X_2, \ldots, X_P$ and Y over (m) complete observations, and L_2 is the product of the normal density of Y given $X_2, \ldots, X_P$ over all (n) observations. Since ϕ_1 and ϕ_2 are distinct sets of parameters their ML estimates are obtained by maximizing L_1 and L_2 separately. If $\hat{\phi}$ is the resulting ML estimate of ϕ, then the ML estimate of any function, $\theta(\phi)$, of ϕ is obtained. The parameter $\phi = \left(\sigma_{yy.s}, \beta_{yj.s}, \sigma_{11.sy}, \beta_{1y.sy}, \beta_{1j.sy}\right)'$ is one to one monotone function of the original parameter $\theta = \left(\beta_{y1.1s}, \beta_{yj.1s}, \sigma^2\right)'$ of the joint distribution of X_1 and Y. In particular, the parameters of θ can be expressed as the following functions of the components of ϕ:

$$\beta_{y1.1s} = \frac{\beta_{1y.sy}\sigma_{yy.s}}{\sigma_{11.sy} + \beta_{1y.sy}^2\sigma_{yy.s}}, \tag{8}$$

$$\beta_{yj.1s} = \frac{\beta_{yj.s}\sigma_{11.sy} - \beta_{1y.sy}\beta_{1j.sy}\sigma_{yy.s}}{\sigma_{11.sy} + \beta_{1y.sy}^2\sigma_{yy.s}}, \tag{9}$$

and

$$\sigma^2 = \frac{\sigma_{11.sy}\sigma_{yy.s}}{\sigma_{11.sy} + \beta_{1y.sy}^2\sigma_{yy.s}}. \tag{10}$$

Where the parameters $\beta_{1y.xy}, \beta_{1j.xy} (j = 2,...,p)$, are the slope coefficients of Y, and $X_j (j = 2,...,p)$, and $\sigma_{11.xy}$ is the residual variance for the regression of X_1 on X_2, ..., X_p, Y from (m) complete cases. Appling Anderson's method to the missing data, we get

$$f\left(x_1, y / x_2,\cdots,x_p;\theta\right) = \prod_{i=1}^{m} f\left(x_1, y / x_2,\cdots,x_p;\theta\right) . \prod_{i=m+1}^{n} f\left(y / x_2,\cdots,x_p;\theta\right)$$

$$= \prod_{i=1}^{m} f\left(x_1 / x_2,\cdots,x_p;\theta\right) f\left(y / x_2,\cdots,x_p;\theta\right) . \prod_{i=m+1}^{n} f\left(y / x_2,\cdots,x_p;\theta\right)$$

$$= \prod_{i=1}^{n} f\left(y / x_2,\cdots,x_p;\phi_1\right) . \prod_{i=1}^{m} f\left(x_1 / x_2,\cdots,x_p;\phi_2\right) \qquad (11)$$

The maximum likelihood estimates of ϕ can be obtained by independently maximizing the likelihoods corresponding to these two components. Substituting ML estimates of ϕ in (8)-(10), we can now be obtained the ML estimate of θ.

The variance covariance matrix of θ can be found as follows:

$$Var\left(\hat{\theta}\right) \cong Var\left(\tilde{\theta}\right) - \hat{\theta}_2' \left[I_m^{-1}\left(\hat{\phi}_2\right) - I_n^{-1}\left(\hat{\phi}_2\right) \right] \hat{\theta}_2 . \qquad (12)$$

Where

$$\hat{\theta}_j = \frac{\partial \hat{\theta}}{\partial \hat{\phi}_j}, \qquad j = 1,2,$$

and the information matrix

$$I\left(\hat{\phi}_j\right) = E\left(-\frac{\partial^2 \log L\left(\hat{\phi}_j\right)}{\partial \hat{\phi}_j} \right), \qquad j = 1,2$$

where the symbols " $\tilde{}$ "and " $\hat{}$ " represent the regression estimate from (m) and (n) observations, respectively. Note that, the variance in (12) does not involve partial derivatives with respect to ϕ_1, and the last term in the right hand is positive which represents the reduction in variance from including the incomplete cases. For more detail about more complex missing data patterns, see Little and Rubin (2002).

3. Simulation Study

Although missing data studies are beginning to reappear in the literature, no recent studies have examined the performance of ML estimation in the multiple regression contexts. Although there is a substantial body of missing data research in the area of multiple regression, much of this literature is dated /or is quite limited by current simulation standards. For example, most simulation studies used only 50 or fewer iterations (Little, 1988; Raymond & Roberts, 1987), and several studies used only 10 iterations (Beale & Little, 1975; Haitovsky, 1968). Clearly, such studies are inadequate by current computing standards and should be viewed with some caution.

The goal of this study was to investigate the performance of the ML estimator relative to three missing data method (CC, AV, and IMC) within the context of four predictors multiple regression model. A comprehensive Monte Carlo simulation study was designed to address four research questions: How do missing data techniques differ with respect to (i) regression coefficient bias, (ii) Mean square error, (iii) Relative efficiency of the parameters estimates, and (iv) Closeness to the normal distribution?.

Using four predictor multiple regression model, our simulation was manipulated within the missing data patterns (MCAR, MNAR, and MAR), and the missing data technique (CC, AV, IMC, and ML). To illustrate missing data problems and methods, we used the real data, in 1998, from undergraduate students in King Saud University (KSU), Riyadh, Saudi Arabia. A total of 1352 raw data were obtained within four predictor variables. X_1 is the percentage score that the student had taken from high school. X_2 is the number of credited hours that the student had taken in the first term in KSU, X_3 is number of credited hours that the student had passed in the first term, X_4 is the high school type (a binary variable indicating whether the student had taken his high school certification in science =1 or not =2), and Y is the GPA scores for the student.

An artificial variable $U = \delta_1 X_1 + \delta_2 Y + \zeta$ was created, where ζ the random number is generated from normal distribution with mean zero and variance one. Values of X_1 were deleted when U was positive (this yield data set with half values of X_1 was missing). Three mechanisms were simulated by the following choices of δ_1, and δ_2: (a) MCAR was selected when $\delta_1 = \delta_2 = 0$, (b) MNAR was selected when $\delta_1 = 1$, and $\delta_2 = 0$, and (c) MAR was selected when $\delta_1 = 0$, and $\delta_2 = 1$.

The LS regression yield coefficient and standard errors estimates, these results will be compared with those from incomplete data methods after values of X_1 have been deleted in various ways. The deletion of values from a complete data set is rather artificial, but it allows the mechanism of deletion to be varies and provides comparisons with the regression results before deletion.

- The regression coefficients bias was defined as the differences between averages coefficient estimates in missing data technique from the 1000 run and its least square estimate from complete data (true parameters β_{ols}), i.e. $\hat{\beta}_i - \beta_{ols}, i = 0,1,2,3,4$, and mean squares deviation, which equivalent to mean squares error when $\beta_{ols} = \beta$, can be calculated as efficiency of estimator from the following equation:

$$MSD = \frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{\beta}_{ij} - \beta_{ols}\right)^2, \qquad i = 0,1,2,3,4$$

- The relative efficiency can be computed as follows:

$$\frac{\hat{\sigma}_i^2}{\sigma_{ols}^2}, \qquad i = 1,2,3,4$$

- The Kolmogorov goodness of fit test can be used to confirm if the data arose from the normal distribution. With this test, it is necessary to know the mean and variance of the normal distribution being tested for fit. So, the parameters of the distribution, from LS estimated, must be specified. A method to compute the p-values for Kolmogorov test is asymptotic only.

A computer program, using Minitab software, was made for the comparatively studies. After the missing data patterns were created, regression models were estimated using different missing data techniques. We repeated the simulation 1000 times, and the simulation results can be found from tables (1) to (8).

4. Analytical Results

As noted by Little & Rubin (2002), the performance of a given missing data technique is largely dependent on the mechanism that cause the missing data. Rubin (1976) was the first to explicate formally missing data theory and the mechanisms that cause missing data. According to Rubin (1976), the MCAR and MAR conditions are considered ignorable missing data mechanism in the sense that unbiased parameters estimates can be obtained using standard ML estimation. That is, ML requires only that the weaker MAR assumption hold. In contrast, CC and AV require the strict MCAR assumption. Analytical results of missing data techniques from our simulation studies will be discussed in this section.

MCAR Data

The missing data techniques (CC, AV, IMC, and ML) yield unbiased point estimates of the parameter β as shown in tables (1) – (4). In addition, the mean square deviation of the parameters was approximately zero for all methods of estimation. Therefore, any obvious result of implementing CC is that a great deal of complete data is potentially lost. Obviously, we know that the missing data, which will be discarded as result of using CC, will reduce statistical power and leading to increased sampling variability of parameter estimates. So, the sampling variance of each method was examined relative to that of LS in complete data. Relative efficiency (RE) values less that unity reflected situations in which methods less sampling variability, i.e. greater efficiency. As seen from table (1) to table (4) that IMC performed better relative efficiency than any other techniques. ML yield efficiency gains relative to CC and AV, but the sampling variability of AV was actually lower than CC technique.

Finally, we used P-value of the Kolmogrove test (KT) to see the closeness distribution of parameter estimates from the distribution of LS estimates in complete data. We found that, the P-value of parameter estimates in AV, IMC, and ML are highly significant except in the predictor variables, which has missing data. The P-value of CC indicated that there is no significant different which mean that the parameter estimates are coming from different normal distribution parameters.

MAR Data

The missing data techniques (CC, AV, IMC, and ML) yield unbiased estimates for the parameter β, see tables (5 - 8), while ML is approximately unbiased estimated. The mean square deviation is good criteria to choose among methods, not bias. Therefore, we found that ML is the smallest MSD among missing data techniques, while CC is the largest. We can consider all technique have zero MSD in two decimal digital numbers.

The relative efficiency values appeared that CC is better, in sense of lower values, than any other techniques. ML performed less variability than IMC and AV. The P-values of Kolmogrove test showed that all techniques came from the same original distribution.

MNAR Data

In MNAR data, we satisfied bias estimators for β and the MSD have approximately zero in all missing data techniques. In relative efficiency values, we found that IMC has less variability than ML, which has gain of efficiency than AV and CC. All the parameter estimated came from the same original distribution as shown by Kolmogrove test, see tables below.

Selecting an Approach:

We hoped that these guidelines would be helpful in supplementing and organizing the data and the manner in which it was collected. The first decision must an investigator make is whether the data may be assumed to be randomly missing or not. If not, the investigator is placed in the difficult situation of knowingly having a biased sample and any of the methods involving substituting of an imputed value for the missing value will only aggravate the problem by perhaps adding further bias to the sample. If data are randomly missing, and when the desired goal is parameter estimation, the investigator can delete cases or choose some method to treat missing data.

From the missing data techniques (MCAR, MAR, and MNAR), our results indicated that ML estimation was superior to the three ad hoc techniques (CC, AV, and IMC) across the simulation studies. ML parameter estimates generally had less bias, less MSD, and less sampling variability than the three ad hoc methods. In addition, IMC estimation was superior to CC and AV techniques in the sense that it has less bias, MSD, and sampling variability. It is simple to implement that uses all-available information does not reduce the sample size.

Finally, we hope that this guideline will help an investigator to make choices, from different methods of estimation in various missing data techniques, in their own research. This simulation has been limited in scope, as all simulations must be. We hope that it will shed some light on performance of different methods of estimation in the case of missing values in regression analysis.

5. References

1. Anderson, T. W. (1957) Maximum Likelihood Estimation for the Multivariate Normal Distribution When Some Observations are Missing, **Journal of the American Statistical Association**, 52, 200 - 203.
2. Beale, E. M. L., and Little, R. J. A. (1975) Missing Values in Multivariate Analysis, **Journal of Royal Statistical Society**, Ser. B, 37, 129 - 145.
3. Conniffe, D. (1983) Comments on the Weighted Regression Approach to Missing Values, **Economic and Social Review**, 14, 259 - 272.
4. Dagenais, M. G. (1973) The Use of Incomplete Observations in Multiple Regression Analysis: A Generalized Least Squares Approach, **Journal of Econometrics**, 1, 317-328.
5. Dear, R. E. (1959) A Principal Components Missing Data Method for Multiple Regression Models, SP-86, Santa Monica, CA: **Systems Development Crop**.
6. Dixon, W. J. (1983) *BMDP Statistical Software*, **University of California Press**: Berkeley, CA.

7. Dixon, W. J. (1988) *BMDP Statistical Software Manual*, (Vol. 2), **University of California Press**: Berkeley, CA.
8. Gourieroux, C., and Montfort, A. (1981) On the Problem of Missing Data in Linear Models, **Review of Economic Studies**, XL VIII, 579 - 586.
9. Haitovsky, Y. (1968) Missing Data in Regression Analysis, **Journal of the Royal Statistical Society**, Ser. B, 30, 67 - 82.
10. Little, R. J. A. (1988) Robust Estimation of the Mean and Covariance Matrix from Data with Missing Values, **Applied Statistics**, 37, 23-38.
11. ____________ (1992) Regression with Missing X's: A Review, **Journal of the American Statistical Association**, 87, 1227-1237.
12. ____________ (1993) Pattern-Mixture Models for Multivariate Incomplete Data, **Journal of the American Statistical Association**, 88.
13. Little, R. J. A., and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edition, New York: **John Wiley**.
14. Park, T. ; Lee, S. and Woolson, R. F. (1993) A Test of the Missing Data Mechanism for Repeated Measures Data, **Communication. Statistical Theory Method**, 22, 2813 - 2829.
15. Raymond, M. R., and Roberts, D. M. (1987) A Comparison of Methods for Treating Incomplete Data in Selection Research, **Educational and Psychological Measurement**, 47, 13-26.
16. Rubin, D. B. (1976) Comparing Regressions When Some Predictor Variables Are Missing, **Technometrics**, 18, 201 - 205.
17. Timm, N. H. (1970) The Estimation of Variance-Covariance and Correlation Matrices from Incomplete Data, **Psychometrika**, 35, 417-437.

Table (1) Biased, MSD, P-value in KT, and RE for B_1= 0.0111 and STD=0.002354

METHODS	MCAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.0111	*	0.0000	1.0000
CC	0.0000	0.0000	0.1531	1.4180
AV	0.0000	0.0000	0.1531	1.4243
IMC	0.0000	0.0000	0.0383	1.1183
MLE	0.0000	0.0000	0.3209	1.4062

Table (2) Biased, MSD, P-value in KT, and RE for B2=-0.1124 and STD=0.008414

METHODS	MCAR			
	Bias	MSD	KT	RE
TRUE VALUE	-0.1124	*	0.0000	1.0000
CC	-0.0003	0.0001	0.0000	1.4170
AV	0.0000	0.0000	0.0000	1.1159
IMC	0.0000	0.0000	0.0000	1.0071
MLE	0.0000	0.0000	0.0000	1.0123

Table (3) Biased, MSD, P-value in KT, and RE for B3= 0.1573 and STD=0.002575

METHODS	MCAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.1573	*	0.0000	1.0000
CC	0.0001	0.0000	0.0662	1.4176
AV	0.0000	0.0000	0.0000	1.1161
IMC	0.0000	0.0000	0.0000	1.0282
MLE	0.0000	0.0000	0.0000	1.0327

Table (4) Biased, MSD, P-value in KT, and RE for B4= 0.0998 and STD=0.026190

METHODS	MCAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.0998	*	0.0000	1.0000
CC	0.0005	0.0006	0.8136	1.4161
AV	0.0003	0.0001	0.0000	1.1160
IMC	0.0002	0.0001	0.0000	1.0537
MLE	0.0002	0.0001	0.0000	1.0557

Table (5) Biased, MSD, P-value in KT, and RE for B1= 0.0111 and STD=0.002354

METHODS	MAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.0111	*	0.0000	1.0000
CC	-0.0076	0.0001	0.0000	0.7858
AV	-0.0104	0.0001	0.0000	1.3874
IMC	-0.0076	0.0001	0.0000	1.1028
MLE	-0.0003	0.0000	0.0000	0.8027

Table (6) Biased, MSD, P-value in KT, and RE for B2=-0.1124 and STD=0.008414

METHODS	MAR			
	Bias	MSD	KT	RE
TRUE VALUE	-0.1124	*	0.0000	1.0000
CC	0.0445	0.0020	1.0000	0.8436
AV	-0.0027	0.0000	0.0000	1.1168
IMC	-0.0031	0.0000	0.0000	1.0065
MLE	-0.003	0.0000	0.0000	0.9989

Table (7) Biased, MSD, P-value in KT, and RE for B3= 0.1573 and STD=0.002575

METHODS	MAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.1573	*	0.0000	1.0000
CC	-0.0365	0.0013	0.0000	0.8061
AV	0.0025	0.0000	0.0000	1.1100
IMC	0.0019	0.0000	0.0000	1.0136
MLE	-0.0002	0.0000	0.0000	0.9863

Table (8) Biased, MSD, P-value in KT, and RE for B4= 0.0998 and STD=0.026190

METHODS	MAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.0998	*	0.0000	1.0000
CC	-0.0701	0.0051	0.0000	0.7945
AV	-0.0363	0.0013	0.0000	1.1124
IMC	-0.0269	0.0007	0.0000	1.0419
MLE	-0.0036	0.0001	0.0000	0.9783

Table (9) Biased, MSD, P-value in KT, and RE for B1= 0.0111 and STD=0.002354

METHODS	MNAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.0111	*	0.0000	1.0000
CC	-0.0088	0.0001	0.0000	1.3636
AV	-0.0091	0.0001	0.0000	1.4579
IMC	-0.0088	0.0001	0.0000	1.1392
MLE	-0.0085	0.0001	0.0000	1.3629

Table (10) Biased, MSD, P-value in KT, and RE for B2=-0.1124 and STD=0.008414

METHODS	MNAR			
	Bias	MSD	KT	RE
TRUE VALUE	-0.1124	*	0.0000	1.0000
CC	0.0254	0.0007	1.0000	1.2922
AV	-0.0023	0.0000	0.0000	1.1297
IMC	-0.0022	0.0000	0.0000	1.0094
MLE	-0.0021	0.0000	0.0000	1.0100

Table (11) Biased, MSD, P-value in KT, and RE for B3= 0.1573 and STD=0.002575

METHODS	MNAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.1573	*	0.0000	1.0000
CC	-0.0021	0.0000	0.0000	1.3391
AV	0.0022	0.0000	0.0000	1.1233
IMC	0.0021	0.0000	0.0000	1.0216
MLE	0.0021	0.0000	0.0000	1.0182

Table (12) Biased, MSD, P-value in KT, and RE for B4= 0.0998 and STD=0.026190

METHODS	MNAR			
	Bias	MSD	KT	RE
TRUE VALUE	0.0998	*	0.0000	1.0000
CC	-0.0702	0.0053	0.0000	1.3248
AV	-0.0354	0.0012	0.0000	1.0858
IMC	-0.0347	0.0012	0.0000	0.9845
MLE	-0.0343	0.0012	0.0000	0.9783