Title
A Probability Based Framework for Testing the Missing Data Mechanism

Author
Lin, Johnny Cheng-Han

Publication Date
2013

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

A Probability Based Framework for Testing the Missing Data Mechanism

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Psychology

by

Johnny Cheng-Han Lin

2013

Abstract of the Dissertation

A Probability Based Framework for Testing the Missing Data Mechanism

by

Johnny Cheng-Han Lin

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2013

Professor Peter M. Bentler, Chair

Many methods exist for imputing missing data but fewer methods have been proposed to test the missing data mechanism. Little (1988) introduced a multivariate chi-square test for the missing completely at random data mechanism (MCAR) that compares observed means for each pattern with expectation-maximization (EM) estimated means. As an alternative, this manuscript proposed two new ways of testing MCAR that use estimated parameters from missingness indicators rather than moment information from observed scores. The first statistic in the probability-based (PBB) family, PBB-MCAR I, is a chi-square test of independence that tests the assumption that missingness indicators are independent among all grouping patterns. The second statistic, PBB-MCAR II, is a chi-square goodness of fit statistic that tests differences of observed versus expected probabilities conditional on ranked values of a suspect variable that drives missingness dependencies. A simulation study showed that although Little's test consistently maintained optimal Type I error rates, the empirical power of PBB-MCAR II to detect violations of MCAR was on par with Little's test under most conditions, whereas PBB-MCAR I had lower power to detect aberrations of MCAR because it tests a more restricted set of independence assumptions. These newly-developed test statistics were demonstrated in two education-based applications, a) as a way

of testing the missing data mechanism when creating longitudinal trajectories of intramural sports participation among African American students, and b) as a tool to detect departures from completely at random test-taking. Future work will involve creating an R package to promote the use of these missing data tests among education researchers, extending PBB-MCAR II to incorporate auxiliary variables, and resolving the problem of sparse missing data patterns by adopting the limited information goodness of fit test proposed by Maydeu-Olivares and Joe (2005).

The dissertation of Johnny Cheng-Han Lin is approved.

Thomas R. Belin

Li Cai

Steve S. Lee

Peter M. Bentler, Committee Chair

University of California, Los Angeles

2013

To my family and friends who have helped me through this journey.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank first and foremost Dr. Peter Bentler for his patience and guidance throughout the years. He encouraged me to think critically and independently, and without him this dissertation would not be possible. I would also like to thank Dr. Thomas Belin for his constructive feedback in making sure that my dissertation is “bullet-proof”. Drs. Li Cai and Steve Lee were especially gracious for serving on my committee and offering me important feedback about applying my techniques to the education and psychological contexts. I thank my parents for always encouraging, reminding and supporting me – never allowing me to give up or be let down. Last but not least, I give credit to my tremendous support network of friends with whom I can share in this challenging but enriching journey we call graduate school.

Part of this research is made possible by a pre-doctoral advanced quantitative methodology training grant (#R305B080016) awarded to UCLA by the Institute of Education Sciences of the US Department of Education. The views expressed in this paper are the author’s alone and do not reflect the views/policies of the funding agencies or grantees.

Vita

2011	C. Phil. in Quantitative Psychology University of California, Los Angeles
2009	M.A. in Quantitative Psychology University of California, Los Angeles
2006	B.S., Psychology Honors with Highest Distinction (Cum Laude) University of California, San Diego

Publications

Bazargan-Hejazi, S., Medeiros, S. Mohammadi, R., Lin, J.C., and Dalal, K. (2013). Patterns of intimate partner violence: A study of female victims in Malawi. *Journal of Injury and Violence Research, 5*(1). DOI: 10.5249/jivr.v5i1.139

Chavira, C., Bazargan-Hejazi, S., Lin, J.C., del Pino, H.E., Bazargan, M. (2011). Type of alcohol drink and exposure to violence: An emergency department study. *Journal of Community Health, 36*, 597 - 604. DOI: 10.1007/s10900-010-9347-1

Lin, J.C., and Bentler, P.M. (2012). A third moment adjusted test statistic for small sample factor analysis. *Multivariate Behavioral Research, 47*, 448 - 462. DOI: 10.1080/00273171.2012.673948.

Tynes, B.M., Umaña-Taylor, A.J., Rose, C.A., Lin, J.C., Anderson, C.J. (2012). Online racial discrimination and the protective function of ethnic identity and self-esteem for African American adolescents. *Developmental Psychology, 48*(2), 343 - 355. DOI: 10.1037/a0027032343

CHAPTER 1

Introduction

1.1 Introduction to Missing Data in the Social Sciences

Researchers in the social sciences regularly encounter missing data, but many do not take further steps to analyze or test the sources of missingness. Peugh and Enders (2004) did a study on the use of missing data handling techniques among researchers in education and psychology and found that 96% of randomly selected articles used listwise deletion, pairwise deletion, or some combination of the two. The relatively small number of studies that used more sophisticated missing data handling techniques may be attributed to the added complexity of implementation and unfamiliarity with tests of the missing data mechanism.

Missing data handling techniques in practice have been relegated to substandard ad hoc methods such as listwise deletion, pairwise deletion and mean imputation (McKnight, 2007). More advanced techniques such as the a) expectation maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) that take into account the missingness pattern, b) full information maximum likelihood (Anderson, 1957; Bock, Gibbons, & Muraki, 1988) for item factor analysis, and c) multiple imputation (Rubin, 2004), offer less biased parameter estimates when the data is missing completely at random (MCAR) or missing at random (MAR), which are the two ignorable types of missing data (Enders & Bandalos, 2001; Enders, 2001). Less work has been done on data that is not missing at random (MNAR) or non-ignorable. Muthén, Kaplan and Hollis (1987) proposed using multiple group structural equation models that are akin to pattern mixture models to handle non-ignorable data. More generally, Greenlees, Reece and Zieschang (1982) explored the use of non-ignorable missing models using logistic regression with covariates to improve data imputation. For a more in depth review of various "state-of-the-art" missing data handling techniques, see Schafer and Graham (2002) and Allison (2003).

Despite the existence of advanced techniques to handle missing data, many researchers continue to use simpler deletion methods. However, deletion methods are only appropriate when the data is missing completely at random (MCAR), which in practice requires a statistical test to help verify this assumption. Little (1988) proposed a multivariate test for the MCAR missingness mechanism. Currently, major software packages that implement this technique include the IBM SPSS Missing Values™20 module (IBM Corporation, 2011) and EQS 6.2 (Bentler, 2006). This manuscript will explore Little's test in detail, providing computational examples and a set of simulation studies to evaluate its performance.

1.1.1 Defining Missing Data Mechanisms

Suppose you have a random variable that is the set of two random variables, $Y = \{Y_{\text{obs}}, Y_{\text{mis}}\}$ where Y_{obs} is the variable that is observed in the data and Y_{mis} is the unobserved random variable. Typically, Y depends on parameter θ. Then define M as the pattern of missing responses where $m_{ij} = 1$ if y_{ij} is missing and $m_{ij} = 0$ if y_{ij} is observed, with M conditional on another parameter called ψ. Conditions on either observed or unobserved variables determine the type of missing data. Little and Rubin (1976; 2002) define three main types of mechanisms: a) missing completely at random, b) missing at random, and c) missing not at random.

For the case of missing completely at random (MCAR):

$$\Pr(M|Y_{\text{obs}}, Y_{\text{mis}}, \psi) = \Pr(M|\psi) \tag{1.1}$$

which means that the probability of missingness does not depend on either observed or unobserved variables but only on the parameter ψ, which is distinct from θ. This is the case of ignorability (Little & Rubin, 2002). See Figure 1.1 for a diagrammatic representation of the MCAR mechanism. In this example, MS, defined as the missing indicator of a student's math score, does not depend on the student's observed or unobserved GPA, (represented by the green and white boxes labeled GPA, respectively).

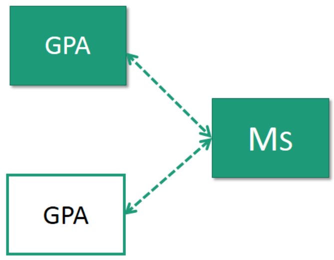


Figure 1.1: Diagram of the MCAR Missingness Mechanism

For the case of missing at random (MAR):

$$\Pr(M|Y_{\text{obs}}, Y_{\text{mis}}, \psi) = \Pr(M|Y_{\text{obs}}, \psi) \tag{1.2}$$

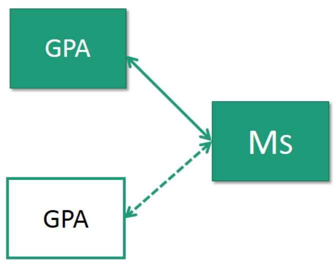


Figure 1.2: Diagram of the MAR Missingness Mechanism

which means that in addition to the dependency on ψ, the probability of missingness depends on observed variables (see Figure 1.2). In this case, the probability of a student's math score being missing depends on her GPA, but not on unobserved GPA. When the data is not missing at random (MNAR, see Figure 1.3), the probability mechanism cannot be simplified as in the previous two cases because either the data is not MAR or the θ and ψ parameters are not distinct (Little & Rubin, 2002), and so M is conditional on both θ and ψ. This means that the missingness indicators may depend on either observed or unobserved variables. In this case, a student's missing math score depends on his GPA whether or not the school has his transcript available.

Pattern mixture models and sensitivity analysis have been the two primary methods of handling MAR and MNAR data, especially in the context of longitudinal data [see Enders (2011) for a review]. The existence of MAR data is less problematic for the researcher because maximum likelihood methods can produce sound parameter estimates under this case. For MNAR data, auxiliary information is required and since the missing data is not available to the researcher, a sensitivity analysis is recommended. For example, Jamshidian and Mata (2008) examined the sensitivity of a given model to the missing data mechanism in the area of structural equation models.

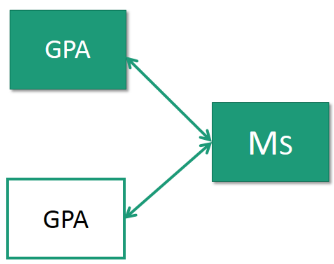


Figure 1.3: Diagram of the MNAR Missingness Mechanism

1.1.2 Review of Tests of the Missing Data Mechanism

A simple way to test for MCAR is to compare the means of observed values between a pair of missing and non-missing groups using pairwise t-tests, however this can result in multiple comparison problems. To avoid issues with Type I error, Little (1988) proposed a global chi-square statistic that uses all of the available data. Park and Davis (1993) extended Little's test for longitudinal missing data using a Wald test, an improvement over prior studies that used a weighted least squares estimation. They explored a wide variety of methods such as weighted GEE's, non-parametric estimation of conditional scores, and modeling conditional distributions. Chen and Little (1999) extended Little's test to generalized estimating equations (GEEs) which forgoes distributional assumptions. This method reduced bias when data was MAR but increased variance when data was MCAR.

Kim and Bentler (2002) extended Little's test to structural equation models by assessing homogeneity of means and covariances (HMC) using generalized least squares estimation. The rejection of HMC implies rejection of MCAR but not vice-versa. In structural equation modeling, a single mean and covariance structure is modeled, but if HMC is not true, then a single set of mean and covariance parameters will not represent the population mean and covariance. Jamshidian and Schott (2007) also tested the equality of means and covariances in structural equation models but allowed for the partition of cases that belong to more than one pattern of missingness. Not dividing groups by missing patterns means that the researcher needs to manually define the groupings a priori. Jamshidian and Jalal (2010) imputed the missing data for each group and then applied F-tests based on the Hawkins test as well as on a non-parametric test of homoscedasticity. The authors' method involves imputing the missing data for each group and then applying a complete data method to the imputed data. To perform a non-parametric test of homoscedasticity the authors use an Anderson Darling k-sample test. The simulation results showed that when the data was normally distributed, the Hawkins test had optimal Type I error rates, but both Hawkins and Kim and Bentler tests failed for non-normal data. Alternatively, the non-parametric test worked well for normal and non-normal data. The authors suggest a sequence of tests, first using the Hawkins test to assess multivariate normality and homoscedasticity. If this test is rejected and the data is non-normal, perform a non-parametric test. Then if the non-parametric test is rejected, the researcher can reject MCAR.

A brief overview of the literature showed that much of the work on missing data tests derive from Little's (1988) concept of testing equality of means or covariances across subgroups. This manuscript offers an alternative to Little's test using a new framework for testing the missing data mechanism that does not test for the equality means and covariances across groups but for the assumption of independence using probabilities of missingness – more closely aligning with Rubin's (1976) theory of missingness.

1.2 Describing Little's Test of MCAR

Since Little's test is the foundation for all HMC MCAR tests, a brief exposition to Little's method is warranted. For future reference, the author denotes Little's test of MCAR as the D^2 test statistic, and assumes that all vectors are column vectors. Begin with an index of notation to be used throughout the paper and in subsequent methods:

- i is the particular subject number
- j is an index of the variable
- y_{ij} is a particular subject's response for a particular variable
- m_{ij} is the missing response for a particular subject

- k is an index of the $K = 2^p$ missing patterns
- r_k is a vector of the particular missing response pattern
- r_{kj} is the single response within the k-th pattern for the j-th variable
- p is the number of variables
- p_k is the number of observed variables for missing pattern k
- N is the total number of observations

The procedure for Little's test of MCAR can be summarized using the following steps:

1. Use the EM algorithm to obtain the expected estimates of the mean and variance-covariance matrix.
2. Group cases according to the missing pattern to obtain observed means for each group.
3. Take the difference between observed and expected means weighted by the estimated variance-covariance matrix and the number of observations within each group to obtain a statistic that is asymptotically chi-square and perform a hypothesis test.

Step 1

The Expectation-Maximization (EM) algorithm takes into account missing data to generate a maximum likelihood estimate. Given a dataset $Y = [y_{ij}]$ that includes the set of observed and missing variables, enter the $Y : N \times p$ matrix into the EM algorithm. The R package `norm` includes the `em.norm()` function for multivariate normal data. The EM maximum likelihood estimate for the population mean vector μ and variance-covariance matrix Σ is defined as $\tilde{\mu}$ and $\tilde{\Sigma}$ respectively.

To calculate the chi-square statistic, the expected mean vectors and variance-covariance matrices need to be partitioned for every response pattern to include only the observed values. Define $\tilde{\mu}_{\text{obs}.k} = \tilde{\mu}^T D_k$ where D_k is a $(p \times p_k)$ selection matrix with only one 1 per column that selects the non-missing observations. This results in a vector of p_k means for every k-th response pattern. Similarly, define $\tilde{\Sigma}_{\text{obs}.k} = D_k^T \tilde{\Sigma} D_k$ as the $(p_k \times p_k)$ variance-covariance matrix for every k-th response pattern.

Step 2

Create a set of matrices S_k for $k = 1, \cdots, K$ where each matrix is a subset of Y consisting of all cases that are identified with particular missing pattern. Define $N^{(r_k)}$, which is the number of cases that belong to a particular missing response pattern (e.g., there are 103 cases with a pattern of $[1, 0, 0]'$). From these $K-1$ sets of matrices (since it is not possible to calculate means for the pattern $[1, 1, 1]'$), calculate the observed vector of means $\bar{y}_{\text{obs}.k}$ for each response pattern r_k.

Step 3

Take the difference of each vector of observed means estimated in Step 2 from the overall EM-estimated means estimated in Step 1 weighted by the EM-estimated variance-covariance matrix to obtain the goodness of fit statistic D^2:

$$D^2 = \sum_{k=1}^{K} N^{(r_k)} (\bar{y}_{\text{obs}.k} - \tilde{\mu}_{\text{obs}.k})^T \tilde{\Sigma}_{\text{obs}.k}^{-1} (\bar{y}_{\text{obs}.k} - \tilde{\mu}_{\text{obs}.k}) \tag{1.3}$$

where $N^{(r_k)}$ is the number of observed samples for the k-th missing response pattern, and the chi-square statistic has degrees of freedom $\sum_{k=1}^{K} p_k - p$ where p_k is the number of observed variables for all K patterns.

A caveat of testing MCAR is that the rejection of the null hypothesis may not be very informative because even under MAR, ignoring the missing data mechanism produces valid estimates (Rubin, 1976; Chen & Little, 1999). The practical advantage of an MCAR test is that it can empirically confirm whether listwise deletion is plausible for the particular analysis, which research has shown is the most widely used technique in psychology and education (Peugh & Enders, 2004).

1.3 Illustrative Example of Little's Test

The following sections will demonstrate the performance of Little's test of MCAR using a simulated example. Details about the three missing data generation methods (denoted MDM-1, MDM-2, and MDM-3) are described in the *Method* section of the *Simulation Studies* chapter. Using Little's procedure described in the previous section, obtain the table of model-obtained (EM) means and the observed means as summarized in Table 1.1.

1.3.1 Little's Test Under MDM-1

From the observed and expected means, obtain the Pearson's goodness of fit test statistic as:

$$
\begin{aligned}
D^2 &= (554)\begin{bmatrix}-0.009 & 0 & 0.023\end{bmatrix}\begin{bmatrix}1.047 & -0.062 & 0.008\\ 0.062 & 0.994 & 0.006\\ 0.008 & 0.006 & 0.951\end{bmatrix}^{-1}\begin{bmatrix}-0.009\\ 0\\ 0.023\end{bmatrix}\\
&\quad + \cdots + (27)\begin{bmatrix}-0.143\end{bmatrix}\begin{bmatrix}0.951\end{bmatrix}^{-1}\begin{bmatrix}-0.143\end{bmatrix}\\
&= 10.05
\end{aligned}
$$

Note that it is not possible to calculate the eighth pattern with completely unobserved means, so only calculations up to the seventh pattern was done. For

Table 1.1: Little's Observed and Expected Means Under MDM-1

Response Pattern	Observed Frequency	Observed Means			Expected Means		
		1	2	3	1	2	3
0,0,0	554	0.036	0.038	-0.008	0.045	0.038	-0.031
0,0,1	110	-0.024	0.106	–	0.045	0.038	–
0,1,0	125	0.193	–	-0.162	0.045	–	-0.031
0,1,1	41	-0.114	–	–	0.045	–	–
1,0,0	103	–	0.043	0.044	–	0.038	-0.031
1,0,1	30	–	-0.198	–	–	0.038	–
1,1,0	27	–	–	-0.173	–	–	-0.031
1,1,1	10	–	–	–	–	–	–

pattern $[1,1,1]$, there are three observed variables, for patterns $[1,1,0]$, $[1,0,1]$, $[1,1,0]$ there are a total of six observed variables, and for patterns $[1,0,0]$, $[0,1,0]$, $[0,0,1]$ there are a total of three observed variables. So the degrees of freedom is $(3+6+3)-3 = 12-3 = 9$, and the critical chi-square value at $\alpha = 0.05$ is $\chi^2(9) = 16.92$. Since $D^2 < \chi^2(9)$, you fail to reject the null hypothesis that the data is MDM-1.

1.3.2 Little's Test Under MDM-2

The observed means from the second variable are a lot higher than expected for the first and second patterns. From the observed and expected means, obtain the test statistic:

Table 1.2: Little's Observed and Expected Means Under MDM-2

Response Pattern	Observed Frequency	Observed Means			Expected Means		
		1	2	3	1	2	3
0,0,0	554	0.021	0.810	0.015	0.127	0.039	-0.031
0,0,1	110	-0.074	0.954	–	0.127	0.039	–
0,1,0	125	0.193	–	-0.162	0.127	–	-0.031
0,1,1	41	-0.114	–	–	0.127	–	–
1,0,0	103	–	-0.556	-0.011	–	0.039	-0.031
1,0,1	30	–	-0.568	–	–	0.039	–
1,1,0	27	–	–	-0.173	–	–	-0.031
1,1,1	10	–	–	–	–	–	–

$$
\begin{aligned}
D^2 &= (554)\begin{bmatrix}-0.107 & 0.77 & 0.045\end{bmatrix}\begin{bmatrix}1.042 & -0.159 & -0.002\\ -0.159 & 0.994 & 0.007\\ -0.002 & 0.007 & 0.951\end{bmatrix}^{-1}\begin{bmatrix}-0.107\\ 0.77\\ 0.045\end{bmatrix}\\
&\quad + \cdots + (27)\begin{bmatrix}-0.143\end{bmatrix}\begin{bmatrix}0.951\end{bmatrix}^{-1}\begin{bmatrix}-0.143\end{bmatrix}\\
&= 387.52
\end{aligned}
$$

In the variance-covariance matrix, all elements except those related to the first variable are equivalent to the variance-covariance matrix as specified in the MDM-1 case above. Since $D^2 > \chi^2(9)$, reject the null hypothesis that the data is MCAR. The results show that the D^2 statistic can detect the case when the missing data mechanism is not MCAR and is generated under MDM-2.

1.3.3 Little's Test Under MDM-3

Generate a dataset using the generation method specified in the MDM-1 case but add to it the condition that if $y_{1j} < 0$, assign $m_{1j} = 1$. For example, when the z-score observed in the first variable y_{11} is less than 0, set the missing indicator for itself to 1. As before, the population probability of missingness is 20% for each of the three variables. The observed frequencies, observed means, and expected means calculated from the EM algorithm are shown in Table 1.3.

Table 1.3: Little's Observed and Expected Means Under MDM-3

Response Pattern	Observed Frequency	Observed Means			Expected Means		
		1	2	3	1	2	3
0,0,0	554	0.839	0.015	-0.013	0.815	0.039	-0.031
0,0,1	110	0.836	0.018	–	0.815	0.039	–
0,1,0	125	0.751	–	-0.225	0.815	–	-0.031
0,1,1	41	0.655	–	–	0.815	–	–
1,0,0	103	–	0.058	0.010	–	0.039	-0.031
1,0,1	30	–	0.056	–	–	0.039	–
1,1,0	27	–	–	-0.095	–	–	-0.031
1,1,1	10	–	–	–	–	–	–

The second and third EM-estimated means remain the same as in the MDM-1 and MDM-2 cases above except that the first mean is a lot higher. From the observed and expected frequencies, obtain the test statistic:

$$
\begin{aligned}
D^2 &= (554)\begin{bmatrix} 0.024 & -0.025 & 0.018 \end{bmatrix}\begin{bmatrix} 0.375 & -0.010 & 0.007 \\ -0.010 & 0.994 & 0.005 \\ 0.007 & 0.005 & 0.951 \end{bmatrix}^{-1}\begin{bmatrix} 0.024 \\ -0.025 \\ 0.018 \end{bmatrix} \\
&\quad + \cdots + (27)\begin{bmatrix} -0.064 \end{bmatrix}\begin{bmatrix} 0.951 \end{bmatrix}^{-1}\begin{bmatrix} -0.064 \end{bmatrix} \\
&= 7.22
\end{aligned}
$$

As before, all elements in the variance-covariance matrix relating to the first variable differ from the matrices in the MDM-1 and MDM-2 cases. Since $D^2 < \chi^2(9)$, you fail to reject the null hypothesis that the data is MCAR. We therefore incorrectly fail to reject the null hypothesis given MDM-3. See the *Simulation Studies* chapter for more details about the performance of Little's test under this case.

1.4 The Probability Based (PBB) Framework for Testing MCAR

Let Y be the matrix of observed and missing random variables consisting of $y_1, \cdots, y_p$ vectors and each $y_j = \{y_{\text{obs},j}, y_{\text{mis},j}\}$, where $y_j \sim N(\mu_j, \sigma_j^2)$ is i.i.d. for $j = 1, \cdots, p$. Let M be a matrix with vectors $m_1, \cdots, m_p$ where $m_j \sim \text{Bernoulli}(\psi_j)$ is i.i.d. for $j = 1, \cdots, p$, and ψ_j is the population probability of missingness. Assuming the data is missing completely at random (MCAR), the overall joint density of missing data in the population can be modeled as (Rubin, 1976):

$$
f(M|\psi, Y) = f(m_1, \cdots, m_p|\psi, Y) = f(m_1, \cdots, m_p|\psi) = \prod_{j=1}^{p} \psi_j \tag{1.4}
$$

where p is the number of variables and ψ is the probability of missingness in the population.

In practical settings, we do not have population random variables but rather samples drawn from the population. Drawing i.i.d. samples, $y_{1j}, \cdots, y_{nj}$ and $m_{1j}, \cdots, m_{nj}$, the missing indicator variable is defined as:

$$m_{ij} = \begin{cases} 1 & \text{when } y_{ij} \text{ is missing} \\ 0 & \text{when } y_{ij} \text{ is observed} \end{cases} \tag{1.5}$$

In succeeding sections, we describe two tests that make use of this perspective.

1.5 The MCAR Test of Independence (PBB-MCAR I)

A first test, which we call PBB-MCAR I, amounts to a chi-square test of independence on the missing data indicators. Estimate the population parameter ψ_j using the following equation:

$$\hat{\psi}_j = \frac{1}{N} \sum_{i=1}^{N} m_{ij} \tag{1.6}$$

The overall pattern of missingness can be estimated using the likelihood:

$$f(M|\hat{\psi}) = \prod_{j=1}^{p} \hat{\psi}_j \tag{1.7}$$

The probability of missingness for each response pattern r_k is of interest rather than the marginal probabilities. Let $\mathbf{r_k} = [r_{k1}, \cdots, r_{kp}]'$ be a vector of K unique patterns of M for $k = 1, \cdots, K$. For example, for a three-variable scenario where $p = 3$, a unique pattern is $\mathbf{r_1} = [0,0,0]'$, $\mathbf{r_2} = [0,0,1]'$, $\mathbf{r_3} = [0,1,0]'$,$\mathbf{r_4} = [0,1,1]'$,$\mathbf{r_5} = [1,0,0]'$,$\mathbf{r_6} = [1,0,1]'$, $\mathbf{r_7} = [1,1,0]'$, $\mathbf{r_8} = [1,1,1]'$. To obtain the expected proportion of missing for each pattern, first define

$$\psi^{(\hat{r_{kj}})} = \begin{cases} \hat{\psi}_j & \text{if } r_{kj} = 1 \\ 1 - \hat{\psi}_j & \text{if } r_{kj} = 0 \end{cases} \tag{1.8}$$

Then obtain expected probabilities for each pattern $\mathbf{r_k}$ for $k = 1, \cdots, K$ by

$$p_e^{(k)} = \prod_{j=1}^{p} \hat{\psi}^{(r_{kj})} \tag{1.9}$$

The expected probabilities for each response pattern r_k as defined by Equation 1.9 is distinguished from Equation 1.7, which is the overall probability of missingness across all patterns. Let the observed probability be

$$p_o^{(k)} = \frac{N^{(r_k)}}{N} \tag{1.10}$$

where $N^{(r_k)}$ is the number of observations belonging to the k-th missing pattern and N is the total sample size. Define $\psi_o^{(k)}$ as the population-observed proportion of missingness and $\psi_e^{(k)}$ as the population-expected probability of missingness for the k-th pattern respectively. Then the null hypothesis is defined as H_0 : $\psi_o^{(k)} = \psi_e^{(k)}$ for $k = 1, \cdots, K$. Let $\mathbf{p_o}$ be the vector of observed probabilities $(p_o^{(1)}, \cdots, p_o^{(K)})'$ and $\mathbf{p_e}$ be the vector of expected probabilities $(p_e^{(1)}, \cdots, p_e^{(K)})'$. The Pearson's chi-square test of independence can be written as:

$$X_I^2 = N(\mathbf{p_o} - \mathbf{p_e})' D^{-1} (\mathbf{p_o} - \mathbf{p_e}) \tag{1.11}$$

with degrees of freedom $df = K - p$, and $D = \text{diag}(p_e^{(1)}, \cdots, p_e^{(K)})$. Rejection of the null hypothesis implies that missingness indicators are independent among all patterns of missingness.

1.5.1 PBB-MCAR I Under MDM-1

To recreate the following examples, use the same missing data generation methods as explained for the previous examples of Little's MCAR test. To implement PBB-MCAR I, instead of estimating the expected and observed means, use the M matrix to calculate the expected and observed probabilities as given in Equations

1.9 and 1.10 respectively. Defining $\hat{\psi} = [\hat{\psi}_1, \cdots, \hat{\psi}_p]'$ as the vector of estimated expected marginal probabilities, obtain $\hat{\psi} = [0.170, 0.203, 0.191]'$ from this particular sample generated under MDM-1. Based on this estimated parameter vector, the expected probabilities along with the observed probabilities for each pattern are given in Table 1.4.

Table 1.4: PBB-MCAR I Observed and Expected Probabilities Under MDM-1

Response Pattern	Observed Frequency	Observed Probability	Expected Probability
0,0,0	554	0.554	0.535
0,0,1	110	0.110	0.126
0,1,0	125	0.125	0.136
0,1,1	41	0.041	0.032
1,0,0	103	0.103	0.110
1,0,1	30	0.030	0.026
1,1,0	27	0.027	0.028
1,1,1	10	0.010	0.007

From the observed and expected probabilities, obtain the chi-square test of independence statistic:

$$\begin{aligned}
X_I^2 &= 1000[\frac{(0.554-0.535)^2}{0.535} + \frac{(0.110-0.126)^2}{0.126} + \frac{(0.125-0.136)^2}{0.136} \\
&\quad + \frac{(0.041-0.032)^2}{0.032} + \frac{(0.103-0.110)^2}{0.110} + \frac{(0.030-0.026)^2}{0.026} \\
&\quad + + \frac{(0.027-0.028)^2}{0.028} + \frac{(0.010-0.007)^2}{0.007}] \\
&= 8.46
\end{aligned}$$

The critical chi-square value at $\alpha = 0.05$ is $\chi^2(5) = 11.07$. Since $X_I^2 < \chi^2(5)$, you fail to reject the null hypothesis and conclude that missingness indicators are independent.

1.5.2 PBB-MCAR I Under MDM-2

To show that the proposed X_I^2 statistic can detect certain violations of MCAR, generate a dataset using MDM-2 to obtain a vector of estimated probabilities of missingness, $\hat{\psi} = [0.492, 0.203, 0.191]'$. The observed and expected probabilities for each pattern generated under MDM-2 are shown in Table 1.5.

Table 1.5: PBB-MCAR I Observed and Expected Probabilities Under MDM-2

Response Pattern	Observed Frequency	Observed Probability	Expected Probability
0,0,0	286	0.286	0.328
0,0,1	56	0.056	0.077
0,1,0	125	0.125	0.083
0,1,1	41	0.041	0.020
1,0,0	371	0.371	0.317
1,0,1	84	0.084	0.075
1,1,0	27	0.027	0.081
1,1,1	10	0.010	0.019

From the observed and expected probabilities, obtain:

$$\begin{aligned} X_I^2 &= 1000[\frac{(0.286-0.328)^2}{0.328} + \frac{(0.056-0.077)^2}{0.077} + \frac{(0.125-0.083)^2}{0.083} \\ &\quad + \frac{(0.041-0.020)^2}{0.020} + \frac{(0.371-0.317)^2}{0.317} + \frac{(0.084-0.075)^2}{0.075} \\ &\quad + \frac{(0.027-0.081)^2}{0.081} + \frac{(0.010-0.019)^2}{0.019}] \\ &= 221.33 \end{aligned}$$

Since $X_I^2 > \chi^2(5)$, reject the null hypothesis that missingness indicators are independent.

1.5.3 PBB-MCAR I Under MDM-3

Although PBB-MCAR I is not intended to be a test of violations of MAR, its performance under MDM-3 is explored. Given the estimated $\hat{\psi} = [0.557, 0.203, 0.191]'$, the observed and expected probabilities under MDM-3 are given in Table 1.6.

Table 1.6: PBB-MCAR I Observed and Expected Probabilities Under MDM-3

Response Pattern	Observed Frequency	Observed Probability	Expected Probability
0,0,0	289	0.289	0.286
0,0,1	55	0.055	0.067
0,1,0	80	0.080	0.073
0,1,1	19	0.019	0.017
1,0,0	368	0.368	0.359
1,0,1	85	0.085	0.085
1,1,0	72	0.072	0.091
1,1,1	32	0.032	0.022

From the observed and expected probabilities, obtain the the chi-square test of independence statistic as:

$$\begin{aligned} X_I^2 &= 1000[\frac{(0.289-0.286)^2}{0.286} + \frac{(0.055-0.067)^2}{0.067} + \frac{(0.080-0.073)^2}{0.073} \\ &\quad + \frac{(0.019-0.017)^2}{0.017} + \frac{(0.368-0.359)^2}{0.359} + \frac{(0.085-0.085)^2}{0.085} \\ &\quad + \frac{(0.072-0.091)^2}{0.091} + \frac{(0.032-0.022)^2}{0.022}] \\ &= 14.97 \end{aligned}$$

Since $X_I^2 > \chi^2(5)$, reject the null hypothesis. The results suggest that PBB-MCAR I can discover dependencies among missing indicators even under MDM-3.

However the relatively small chi-square value means that the power may not be as high as detecting dependencies under MDM-2.

1.6 The Probability Based Goodness of Fit MCAR Test (PBB-MCAR II)

The motivation for this second test statistic in the PBB family comes from the findings that PBB-MCAR I failed to perform on-par with Little's test under a particular MAR generation condition where missingness depended completely on an observed variable (see the *Departures from Completely at Random Test-Taking* chapter). Rubin's (1976) definition of MCAR is that

$$f(M|\psi, Y) = f(M|\psi). \tag{1.12}$$

Based on this definition, PBB-MCAR I only tests an implication of MCAR, that the missingness indicators are independent across all K missing patterns. Equation 1.12 also assumes that missingness indicators are independent of observed Y variables. An alternative test statistic is proposed which supposes that under an MCAR generation model, patterns of missingness are independent given ranked values of Y_{obs}. The user picks one completely-observed vector in the dataset, y_j, and rank orders the dataset according to the observed values in this vector. The dataset minus the user-chosen column is separated into q-blocks and estimated proportions are calculated within each q-block. These proportions are then used to calculate the observed probabilities for the K patterns of missingness. The expected probabilities are calculated collapsing across all blocks and is compared to the q-observed probabilities across missingness patterns using a chi-square goodness of fit statistic. The algorithm is described more thoroughly in the following steps:

Step 1: Rank order the dataset by the j-th column

An implication of MCAR is that the indicator matrix M is independent of any permutation of dataset Y. In order to test for this implication, sort the data by the user-chosen column vector y_j and assign ranked values $r = 1, 2, \cdots, N$. The arrangement of elements y_{ij} then becomes y_{rj} for all r. If the first variable is defined as the suspect variable, the command line in R is `Y[order(Y[,1]),]`.

Step 2: Create a new indicator matrix excluding the ranked j-th column

Since estimating probabilities including the ranked j-th column can generate nonsensical proportion estimates such as $\hat{\psi}_j = 0$ or $\hat{\psi}_j = 1$, exclude the j-th column from analysis, leaving $p - 1$ columns in the remaining dataset, Y_{-j}. The resulting indicator matrix is the same as Equation 1.5, except it is ranked by r with the j-th column excluded, denoted by M_{-j}, for $r = 1, 2, \cdots, N$.

Note for the special case of $p = 2$, using the M_{-j} matrix would result in an m-vector which would be unsuitable for analysis. A more in depth discussion of the problems associated with this is discussed in the *Simulation Studies* chapter.

Step 3: Separate the new indicator matrix into q-blocks

Another implication of MCAR is that no matter how the indicator matrix is partitioned, the estimated proportions within each column will be roughly equivalent across the q-blocks given sufficiently large sample size. For each column vector m_j in the M_{-j} matrix, calculate

$$\hat{\psi_{j.q}} = \frac{1}{N} \sum_{i=1}^{N} m_{ij.q} \tag{1.13}$$

where $j = 1, \cdots, (p - 1)$ is the range of variables not excluded from the analysis in Step 2, and $q = 1, \cdots, Q$ is the range of blocks. In the default case, $Q = 2$.

Step 4: Calculate observed versus expected probabilities

After obtaining $\hat{\psi_{j.q}}$ observed proportions of missingness, calculate the overall probabilities of missingness for each pattern $\mathbf{r_k}$ for $k = 1, \cdots, K$ just as in Equation 1.8. The overall probabilities of missingness by pattern are calculated for each q-block

$$p_{o.q}^{(k)} = \prod_{j=1}^{p-1} \hat{\psi}^{(r_{kj.q})} \tag{1.14}$$

where $k = 1, \cdots, K$ across all patterns of missingness on the M_{-j} matrix and $q = 1, \cdots, Q$ across all blocks. The expected probabilities of missingness by pattern is calculated as

$$p_{e}^{(k)} = \prod_{j=1}^{p-1} \hat{\psi}^{(r_{kj})} \tag{1.15}$$

where $k = 1, \cdots, K$ across all patterns of missingness on M_{-j}.

Step 5: Calculate the goodness of fit chi-square statistic

From the estimated parameters, calculate the chi-square test statistic as

$$X_{II}{}^{2} = \sum_{k=1}^{K} \sum_{q=1}^{Q} N.q^{(r_k)} (\mathbf{p_{o.q}}^{(k)} - \mathbf{p_e}^{(k)}) D^{-1} (\mathbf{p_{o.q}}^{(k)} - \mathbf{p_e}^{(k)}) \tag{1.16}$$

where $N.q^{(r_k)}$ is the number of observed samples for the k-th missing response pattern within each q-th block, and $D = \text{diag}(p_e^{(1)}, \cdots, p_e^{(K)})$ is the denominator term in the chi-square goodness-of-fit test. The reference degrees of freedom is $K - 1$ where K is the number of observed missing patterns. Rejection of the null hypothesis indicates that the missing data generation mechanism is not independent across all patterns of missingness conditional on the new ranked dataset, Y_{-j}.

CHAPTER 2

Simulation Studies

2.1 Pilot Simulation Study

PBB-MCAR I has not been studied in terms of its Type I error rate and empirical power relative to Little's test. The purpose of this pilot simulation study is to assess these properties in detail (as of the pilot study, PBB-MCAR II had not been developed and was therefore not assessed).

2.1.1 Method

The MDM-1 generation technique described below was developed by Jamshidian, Yuan, and Le (in press). A total of $n = 500$ replications were run using $N = 1000$ students under the three missing data mechanism conditions assessed at an $\alpha = 0.05$. The null hypothesis was deemed to be rejected if $p < 0.05$.

1. Generate a random vector of three standard normal variables $y_j \sim N(0, 1)$ for $j = 1, 2, 3$ using `rnorm()` in R software.

2. Generate three independent vectors of random uniform variables $v_j \sim U(0, 1)$ for $j = 1, 2, 3$ using `runif()` in R software. Let q_j be the proportion of missing data for variable j.

3. If $v_{ij} < q_j$, then assign $m_{ij} = 1$. Otherwise, assign $m_{ij} = 0$.

4. Combine the three univariate vectors using `cbind()` to get the M $(N \times p)$ matrix.

5. Calculate the observed and expected means using the procedures described previously.

Equal probabilities were assumed across the three variables, $q_1 = q_2 = q_3 = 0.2$. To ensure that the random number generation gave consistent results across tests, a seed was set for each univariate normal generate, with a seed of 100 set for y_1, 101 for y_2, and 111 for y_3 using the `set.seed()` function. Three data mechanisms were generated for this simulation study: a) Missing Data Mechanism 1 (MDM-1), b) Missing Data Mechanism 2 (MDM-2), and Missing Data Mechanism 3 (MDM-3). MDM-1 corresponds to the MCAR procedure, described above. To generate MDM-2, add to the MDM-1 condition that if $y_{i,j} < 0$, then assign $m_{i,j-1} = 1$. In other words, generate the same missing indicator pattern as before, but for example if the z-score observed in the *second* variable is less than 0, set the missing indicator for the *first* variable to 1. The population probability of missingness is 20% for each of the three variables, but the inclusion of the MDM-2 condition changes the observed probability of the first variable. For MDM-3, add to MDM-1 the condition that if $y_{ij} < 0$, then assign $m_{ij} = 1$. This is the case when the z-score observed in the first variable y_1 is less than 0, set the missing indicator for itself to 1.

2.1.2 Results

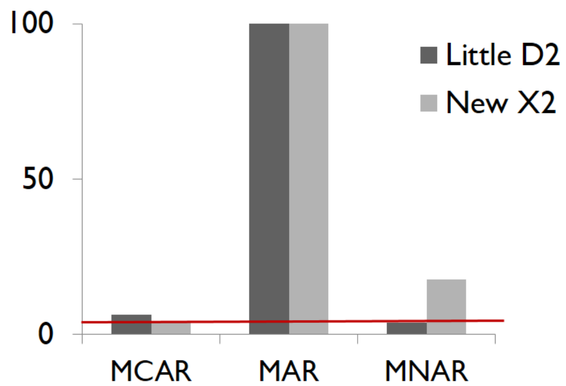
The results of this pilot simulation study showed that both Little's test and PBB-MCAR I performed relatively well under MDM-1, with Little's test showing slight over-rejection and PBB-MCAR I showing slight under-rejection given $\alpha = 0.05$ (see Table 2.1 and Figure 2.1). Both statistics also adequately rejected the null hypothesis under MDM-2 at 100%. Under the MDM-3, Little's test rejected the MCAR probability model only 4% of the time. PBB-MCAR I was slightly more powerful than Little's test, but both tests were not adequately powerful. More

Table 2.1: Results of the Pilot Simulation Study ($N = 500$ Replications)

Generating Mechanism	Little	PBB-MCAR I
MDM-1	6.5	3.8
MDM-2	100	100
MDM-3	3.8	17.8

work needs to be done to incorporate the MNAR probability mechanism directly into the PBB framework to adequately detect aberrations of MAR (Molenberghs, Beunckens, Sotto, & Kenward, 2008).

Figure 2.1: Plot of Results From the Pilot Simulation Study (Red Line Indicates 5% Rejection)



2.1.3 Discussion

The pilot simulation study showed that Little's test and PBB-MCAR I performed on par under MDM-1 and MDM-2. Both tests correctly failed to reject the null hypothesis under MDM-1 and had sufficient power to detect aberrations of MCAR under MDM-2, having rejected the incorrect model at 100% when the sample size was large ($N = 1000$). Under MDM-3 however, both tests under-rejected the null hypothesis: PBB-MCAR I rejected 17.8% of the time, and Little's test rejected 3.8% of the time. In this pilot simulation study, PBB-MCAR I detected aberrations under MDM-3 more frequently than Little's test, although the empirical power for both statistics was much lower than expected (below 25%). This is not surprising, given that the purpose of these tests is not to explicitly detect violations of MAR. The pilot simulation study demonstrated the promising large sample characteristics of PBB-MCAR I. However, this study was limited in its scope of simulation. The next section will describe a more comprehensive simulation study that varies in its range of sample sizes, number of variables, and correlations among observed variables. In addition, the performance of PBB-MCAR II will be assessed along with PBB-MCAR I and Little's test.

2.2 Full Simulation Study

The pilot simulation study showed that PBB-MCAR I performed on par with Little's test in Type I error rate and empirical power, despite being derived from different theoretical perspectives. Whereas Little's test uses differences among observed and expected means, PBB-MCAR I tests an implication of MCAR that missingness indicators are independent of each other. Anecdotal evidence from the *Departures from Completely at Random Test-Taking* chapter showed that although PBB-MCAR I failed to reject the null hypothesis adequately under a particular MCAR generation model, it did not adequately reject MCAR when the missing data generation mechanism depended strictly on the observed values of another variable. Therefore, a more thorough simulation study is warranted to assess the performance of PBB-MCAR I under a broader range of conditions, as well as to assess the viability of the second test in the PBB family, PBB-MCAR II, as an alternative to PBB-MCAR I. The hypothesis is that PBB-MCAR II will have higher empirical power than PBB-MCAR I when missingness depends on an observed variable.

The pilot simulation study was limited in its scope and applicability to the field of education. Improvements made to the current simulation study include i) setting realistic ranges of simulated values that reflect education testing outcomes, ii) generating sample sizes that mimic those found in classrooms and schools, iii) generating a range in the number of variables, iv) varying the proportions of missingness, and v) generating correlations among observed variables that mimic longitudinal outcomes in testing.

2.2.1 Method

Table 2.2 describes the variables and parameters tested in this full simulation study. Little's test, PBB-MCAR I, and PBB-MCAR II were each assessed for their Type I error rates and empirical power under MDM-1, MDM-2 and MDM-3. All parameters were held constant across the Number of Variables, including Probability of Missingness and Correlation Among Variables. A range of three MDM generation conditions, six sample sizes, three variable sizes, two probabilities of missingness and two correlations (fixed across variables) generated a total of 216 conditions. Due to computing constraints, 500 replications were run across each condition to balance computation speed and statistical efficiency. Each condition was then plotted with sample size on the x-axis and percentage of rejection on the y-axis (see *Appendix A* for non-collapsed plots of each of the 216 conditions). Preliminary simulation runs showed that PBB-MCAR II is problematic for variables of size two (see Figures 2.5, 2.8, and 2.11), as it over-inflates the Type I error rate and generates artificially high power as a result. More discussion is mentioned under the *Number of Variables* section. Other limitations are discussed in the *Discussion and Limitations* section of this chapter.

Table 2.2: Table of Parameters Tested in the Full Simulation Study (Total of 216 Conditions)

Variable	Parameters
Missingness mechanism	MCAR, MAR, MNAR
Sample size	35, 50, 100, 250, 500, 1000
Number of variables	2, 3, 5
Probability of missingness	0.2, 0.6
Correlation among variables	0, 0.9

2.2.2 Missing Data Generation Method

1. Generate a matrix of p-multivariate normal variables where $y_j \sim N(350, 150)$ for $j = 1, \cdots, p$ using `mvrnorm()` in the R library `MASS`.

2. Determine real world cut-off ranges. If $y_{ij} < L$, then assign $y_{ij} = L$, and if $y_{ij} > U$ then assign $y_{ij} = U$. Since this study focuses on applicability to educational outcomes, a minimum score of $L = 0$ and a maximum score of $U = 600$ were assigned to reflect realistic scores on the California Standards Test (CST).

3. Generate p-independent vectors of random uniform variables $v_j \sim U(0, 1)$ for $j = 1, \cdots, p$ using `runif()` in R and combine into a matrix V.

4. For MDM-1, if $v_{ij} < q_j$, then assign $m_{ij} = 1$. Otherwise, assign $m_{ij} = 0$, where q_j are probabilities of missingness. This corresponds to the MCAR generation mechanism.

5. For MDM-2, add to the MDM-1 condition that if $y_{i,(j-1)} < C$ and $v_{ij} < q_j$, then assign $m_{ij} = 1$. For this particular study, a cut-off of $C = 350$ corresponds to the cut-off for basic proficiency on the CST. For example, if Student 28's first year CST score is less than 350 and the random number

generated for $v_{(28,2)} < 0.2$, then set the missing indicator $m_{(28,2)} = 1$. For the special case of m_1 (i.e., the first column in the M-matrix), assign the missingness to be MDM-1 as in Step 4. Note that under this generation condition, there may be a special case where missingness on m_{ij} depends on a missing y_{ij} that is generated by the MCAR condition, making the condition MNAR (although it is expected that this would be for the minority of cases). For this reason, the MDM-2 condition is not technically an MAR generation condition but can be said to be predominately MAR.

6. For MDM-3, add to MDM-1 that if $y_{ij} < C$ and $v_{ij} < q_j$, then assign $m_{ij} = 1$. For example, if Student 130's second year CST score is less than 350 and the random number generated for $v_{(130,2)} < 0.2$, then set the missing indicator $m_{(130,2)} = 1$. For the special case of m_{i1}, assign the missingness to be MDM-1. This condition corresponds to the MNAR generation condition.

2.2.3 Results

For ease of interpretation, rejection percentages for variables were collapsed to generate plots of main effects. The main effects were plotted for the Missingness Mechanism, Number of Variables, Probability of Missingness, and Correlation Among Variables for Little's test, PBB-MCAR I and PBB-MCAR II. Due to the problematic nature of PBB-MCAR II's performance with two variables, all main effects except for the *Number of Variables* section excluded data with two variables for the three tests considered in this simulation study.

2.2.4 Describing Main Effects: Missingness Mechanism

Figures 2.2, 2.3, and 2.4 depict rejection rates for the three statistics, plotted against sample size under MDM-1, MDM-2 and MDM-3 respectively, collapsing across all variables including Number of Variables (e.g., 3 and 5), Probability of

Missingness (e.g., 0.2 and 0.6), and Correlation Among Variables (e.g., 0, and 0.9). Under MDM-1, sample size did not affect the Type I error performance of the three tests, although Little's test had the best Type I error rates at $N = 1000$ compared to the PBB-MCAR tests.

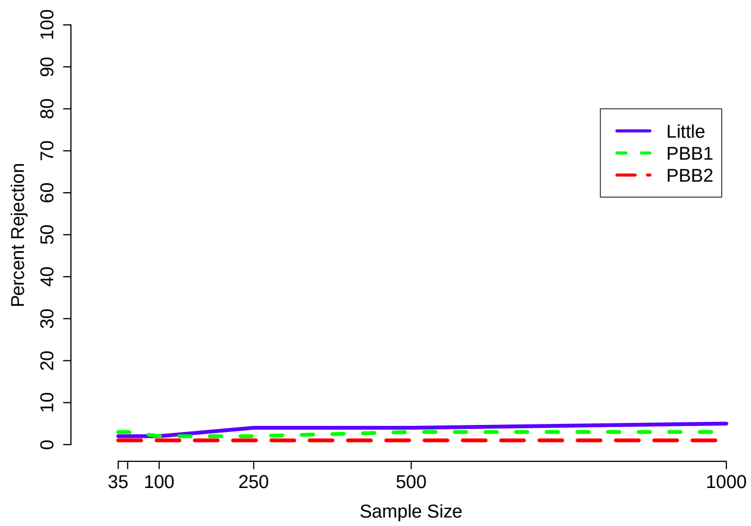


Figure 2.2: Full Simulation Study: MDM-1 Collapsed Across All Other Variables

Under MDM-2, Little's test continued to have the highest empirical power, which was inconsistent with results from the pilot simulation study. PBB-MCAR II showed an improvement over PBB-MCAR I in empirical power, especially at samples below 250. At sample sizes greater than 250, PBB-MCAR II performed on par with Little's test, given that two blocks were used for the analysis and only one variable was chosen as the sorting index.

Under MDM-3, PBB-MCAR I outperformed PBB-MCAR II at all levels and performed slightly better than Little's test at the smallest sample size ($N \leq 50$). Theoretically, there is little justification for this performance advantage, which warrants further investigation as to why Little's test would still offer optimal empirical power rates under MDM-3.

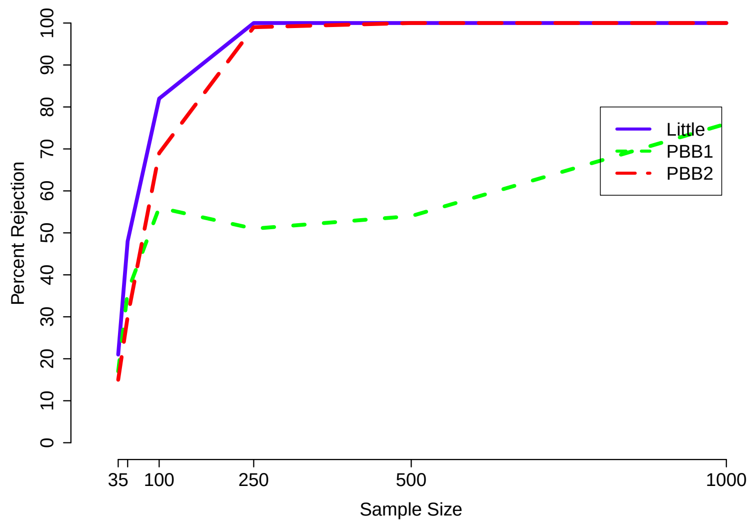


Figure 2.3: Full Simulation Study: MDM-2 Collapsed Across All Other Variables

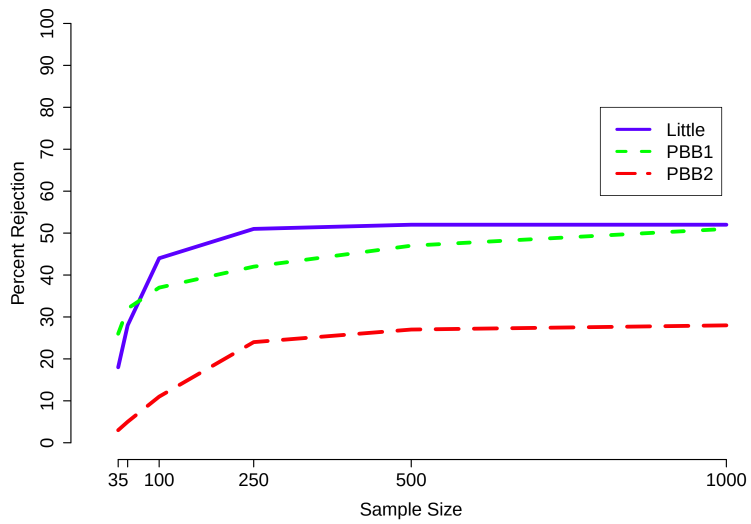


Figure 2.4: Full Simulation Study: MDM-3 Collapsed Across All Other Variables

2.2.5 Describing Main Effects: Number of Variables

Figures 2.5 through 2.7, Figures 2.8 through 2.10, and Figures 2.11 through 2.13 depict rejection rates by the number of variables under MDM-1, MDM-2 and MDM-3 respectively.

As noted previously, the Type I error rates of PBB-MCAR II under the two-variable condition is especially problematic. Although Little's test and PBB-MCAR I performed on par under MDM-1, PBB-MCAR II over-rejected the null hypothesis, nearing 100% rejection for $N > 100$. This problem may be due to a weakness in the PBB-MCAR II algorithm, which discarded the j-th variable from the analysis, leaving a single m-vector which gave inaccurate marginal estimates of ψ. The problem was remedied when there were at least $p = 3$ variables in the dataset, leaving two variables in the M_{-j} matrix for subsequent analysis. Under this case, PBB-MCAR II's performance matched the patterns seen previously, although PBB-MCAR II tended to under-reject as sample size increased.

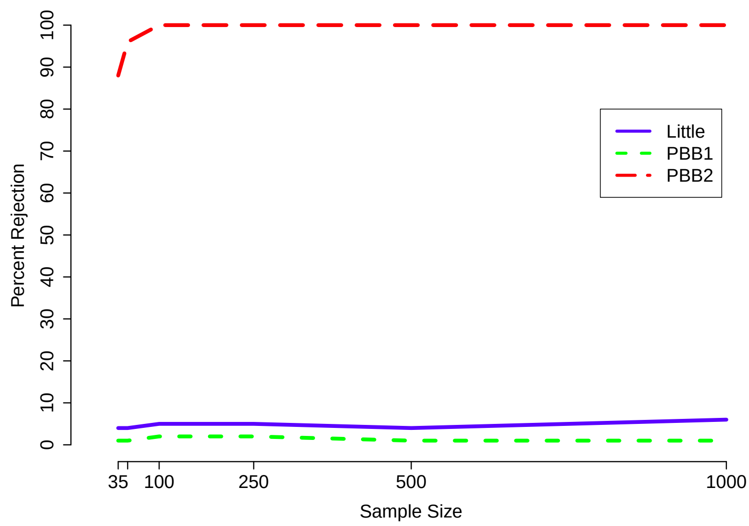


Figure 2.5: Full Simulation Study: MDM-1, Number of Variables = 2, Collapsed Across All Other Variables

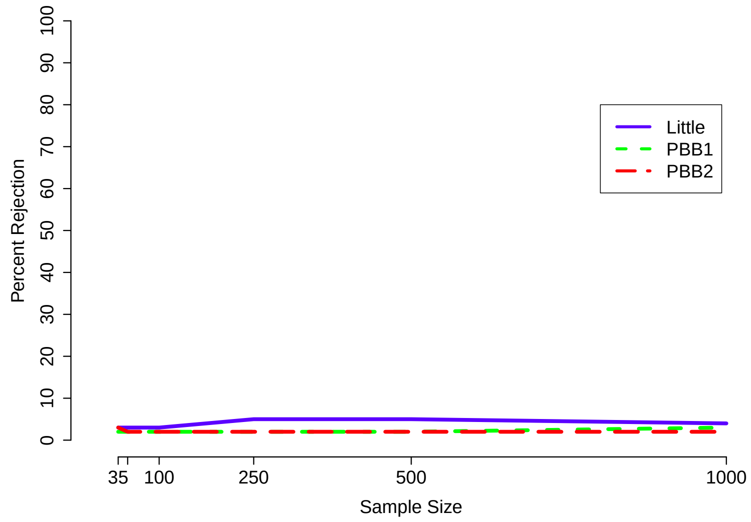


Figure 2.6: Full Simulation Study: MDM-1, Number of Variables = 3, Collapsed Across All Other Variables

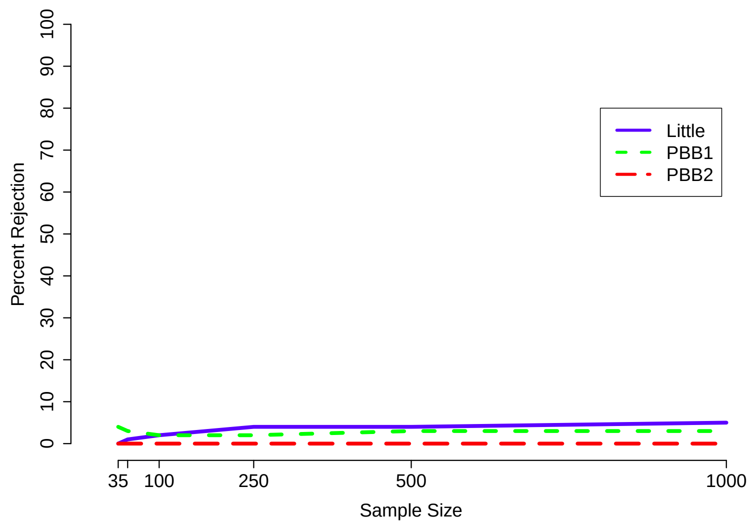


Figure 2.7: Full Simulation Study: MDM-1, Number of Variables = 5, Collapsed Across All Other Variables

Under MDM-2, PBB-MCAR II slightly out-performed Little's test under the two-variable condition, although the author suspects this may be due to the artificially raised rates of rejection. Under three variables, PBB-MCAR II performed slightly better than Little's test, and II performed on par with Little's test at five variables. Compared to both Little's test and PBB-MCAR II, PBB-MCAR I under-performed under MDM-2 for the two-, three- and five-variable conditions.

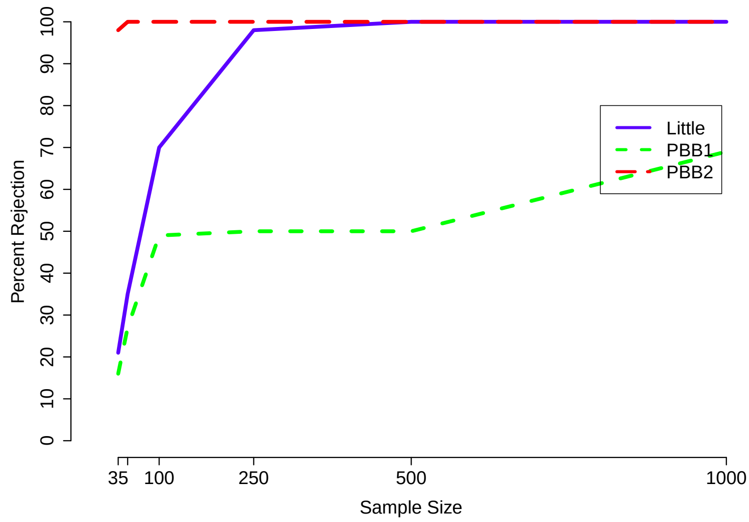


Figure 2.8: Full Simulation Study: MDM-2, Number of Variables = 2, Collapsed Across All Other Variables

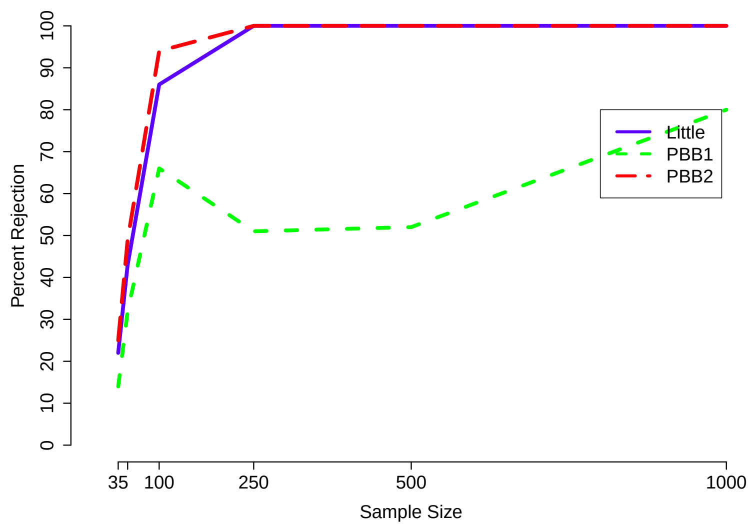


Figure 2.9: Full Simulation Study: MDM-2, Number of Variables = 3, Collapsed Across All Other Variables

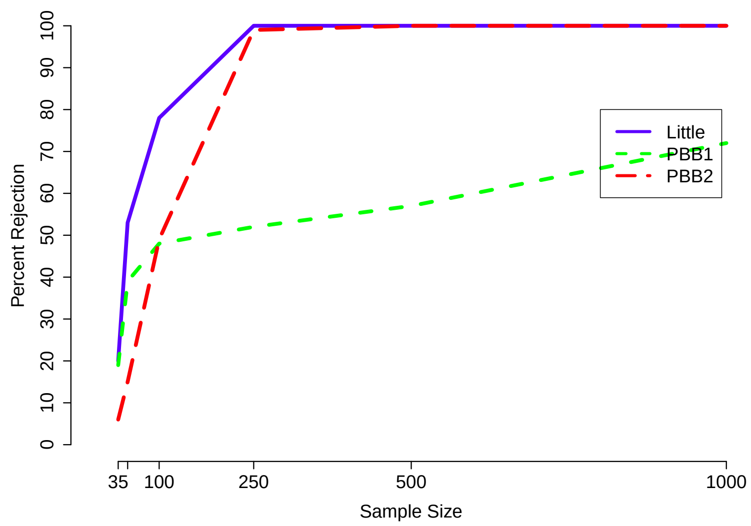


Figure 2.10: Full Simulation Study: MDM-2, Number of Variables = 5, Collapsed Across All Other Variables

Under MDM-3, PBB-MCAR II continued to over-reject the null hypothesis relative to the other statistics under the two-variable case.

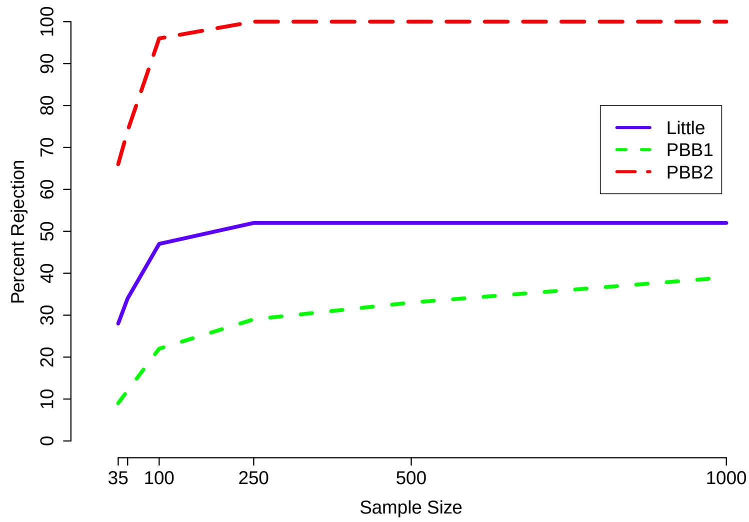


Figure 2.11: Full Simulation Study: MDM-3, Number of Variables = 2, Collapsed Across All Other Variables

For the three-variable case, Little's test rejected the null hypothesis more frequently than both PBB-MCAR I and II, which was consistent with the main effect trends described in the *Missingness Mechanism* section of this chapter. For the five-variable condition at $N = 35$, PBB-MCAR I rejected the null hypothesis at 32%, Little's test at 11% and PBB-MCAR II by 1%, with I having shown the best performance compared to the other two statistics. However, these benefits disappeared at large sample sizes. It is not clear why increasing the number of variables had a positive effect on PBB-MCAR I under MDM-3 even though these advantages did not arise under MDM-2.

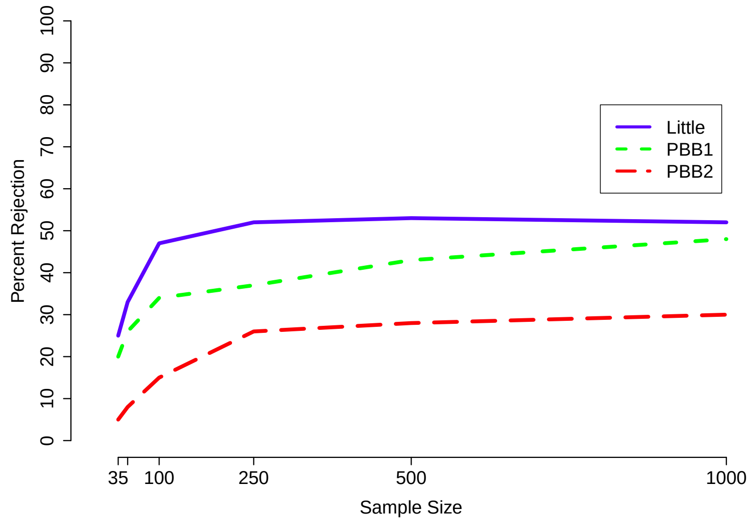


Figure 2.12: Full Simulation Study: MDM-3, Number of Variables = 3, Collapsed Across All Other Variables

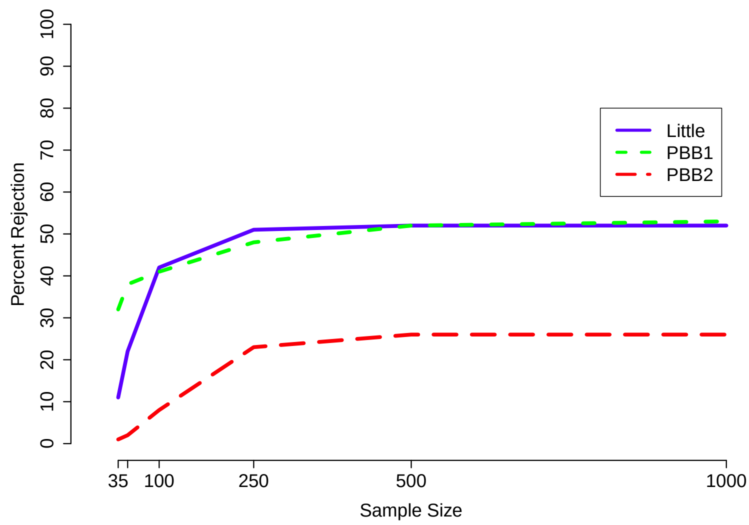


Figure 2.13: Full Simulation Study: MDM-3, Number of Variables = 5, Collapsed Across All Other Variables

2.2.6 Describing Main Effects: Probability of Missingness

Figures 2.14 & 2.15, Figures 2.16 & 2.17, and Figures 2.18 & 2.19 depict rejection rates for 0.2 and 0.6 probabilities of missing under MDM-1, MDM-2 and MDM-3 respectively. The results show that the probabilities of missingness appear to have little effect on the Type I error rate.

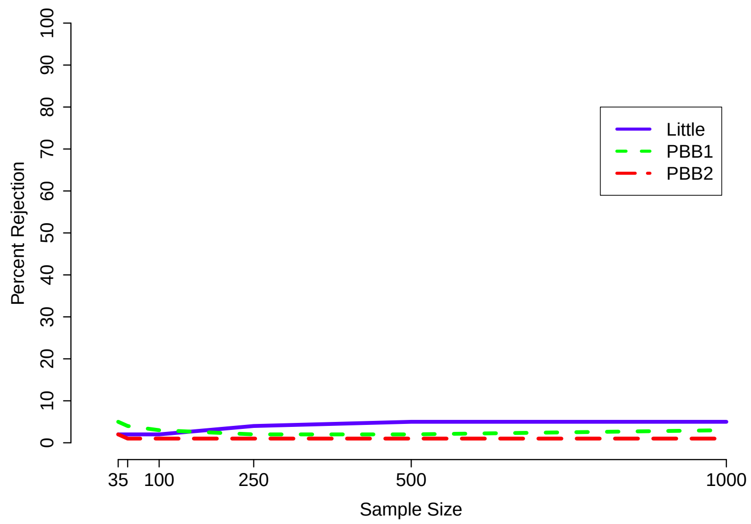


Figure 2.14: Full Simulation Study: MDM-1, Probability of Missingness = 0.2, Collapsed Across All Other Variables

However, there were striking differences in empirical power among tests under MDM-2. For Probability of Missingness = 0.2, PBB-MCAR II slightly outperformed Little's test and greatly outperformed PBB-MCAR I. In particular at $N = 50$, PBB-MCAR II rejected the null hypothesis at 24%, Little's test at 16% and PBB-MCAR I at 1%. However, when the probability of missingness was 0.6, PBB-MCAR I performed better than II. For example, at $N = 50$, Little's test rejected the null hypothesis 74% of the time, PBB-MCAR I rejected 63% of the time, and PBB-MCAR II rejected 35% of the time. The results suggest a relationship between the probability of missingness and the independence among indicator variables.

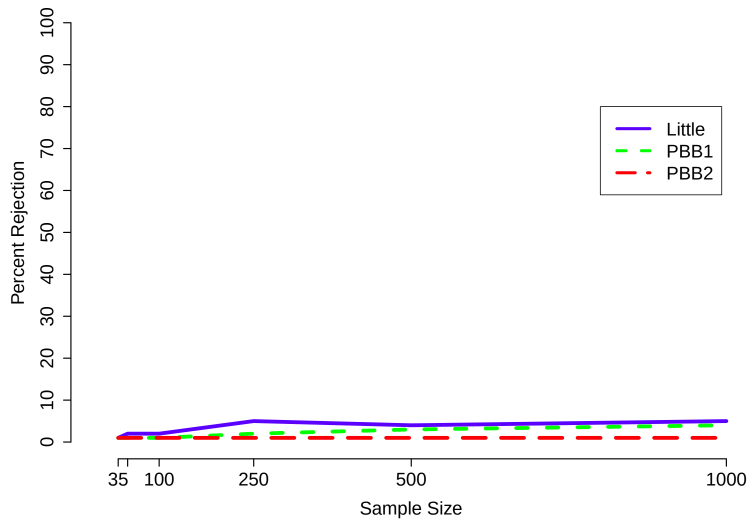


Figure 2.15: Full Simulation Study: MDM-1, Probability of Missingness = 0.6, Collapsed Across All Other Variables

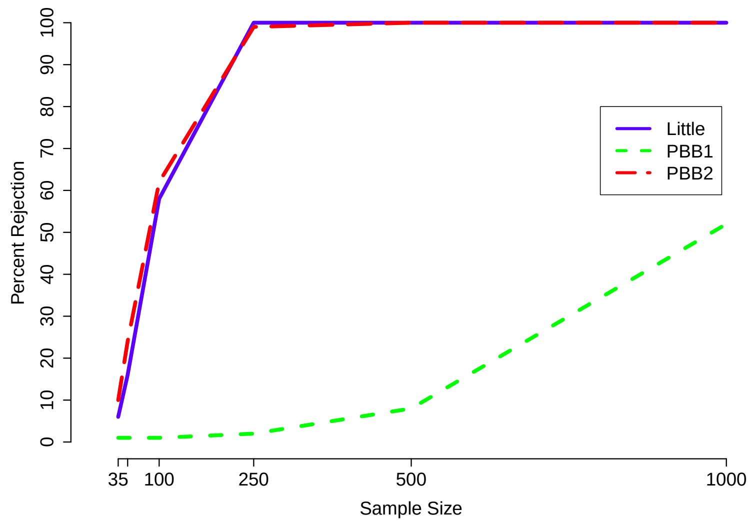


Figure 2.16: Full Simulation Study: MDM-2, Probability of Missingness = 0.2, Collapsed Across All Other Variables

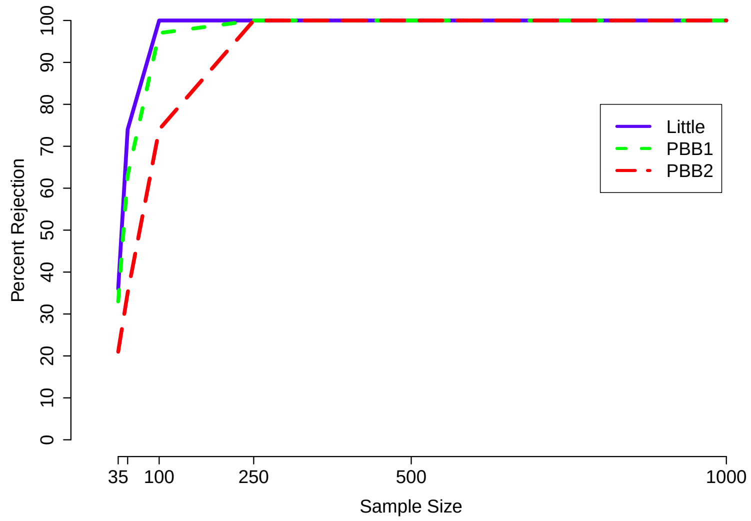


Figure 2.17: Full Simulation Study: MDM-2, Probability of Missingness = 0.6, Collapsed Across All Other Variables

Under MDM-3, the probability of missingness made no difference in rejection rates among the three tests when the probability was low, but made a big difference when the probability was high. PBB-MCAR II under-rejected the null hypothesis for an average 3.8% of the time, when it should have rejected the null hypothesis 100% of the time.

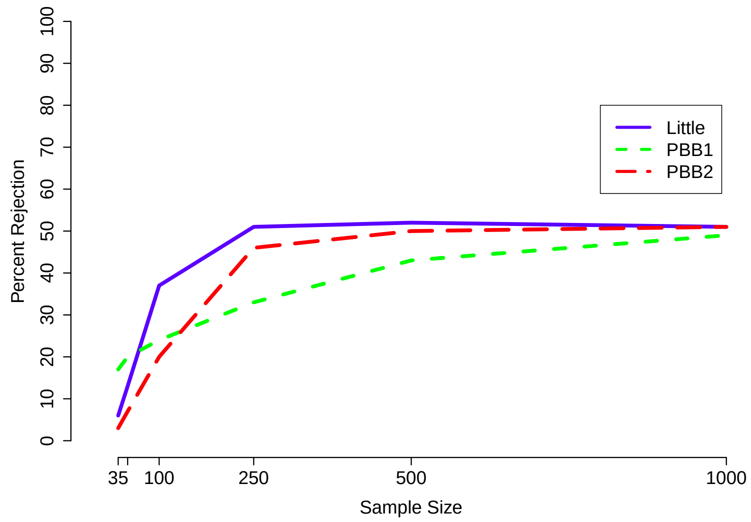


Figure 2.18: Full Simulation Study: MDM-3, Probability of Missingness = 0.2, Collapsed Across All Other Variables

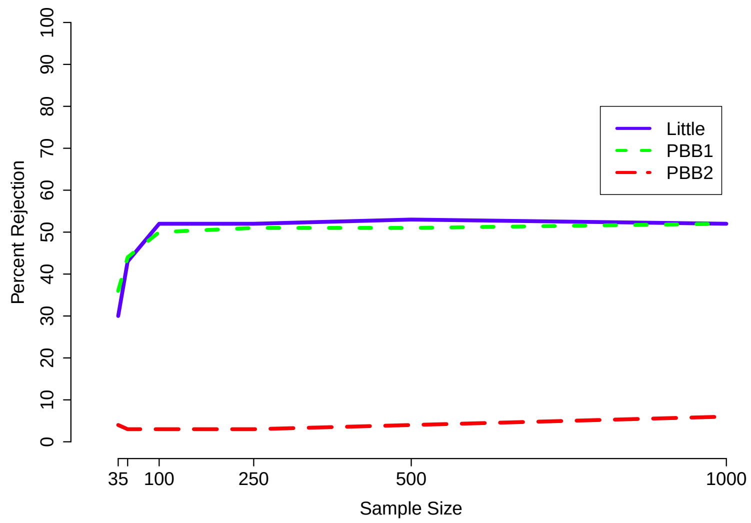


Figure 2.19: Full Simulation Study: MDM-3, Probability of Missingness = 0.6, Collapsed Across All Other Variables

2.2.7 Describing Main Effects: Correlation Among Variables

Figures 2.20 and 2.21, Figures 2.22 and 2.23, and Figures 2.24 through 2.25 depict rejection rates for correlations among observed variables of 0 and 0.9 under MDM-1, MDM-2, and MDM-3 respectively. The purpose was to assess the performance of the three test statistics under longitudinal outcomes where scores in the current year may be correlated with scores from the prior year. For this study, only compound symmetry was assumed where all off-diagonal correlations are the same. Under MDM-1, the correlation structure did not affect the correct failure to reject the null hypothesis for any of the three tests.

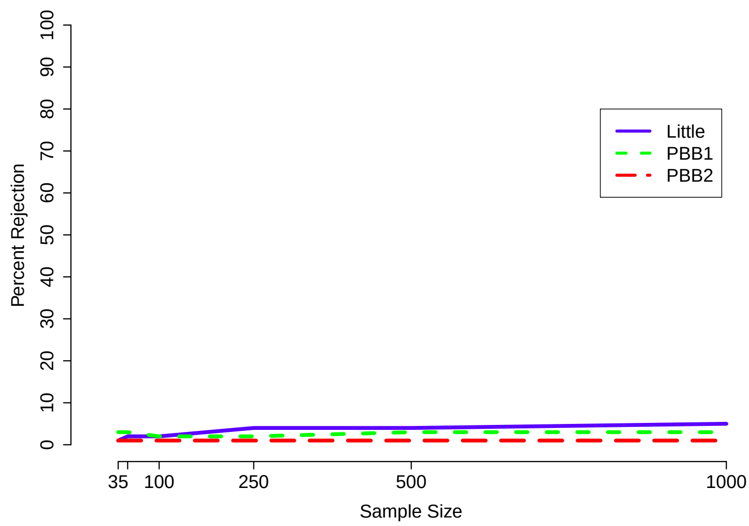


Figure 2.20: Full Simulation Study: MDM-1, Correlation Among Variables = 0, Collapsed Across All Other Variables

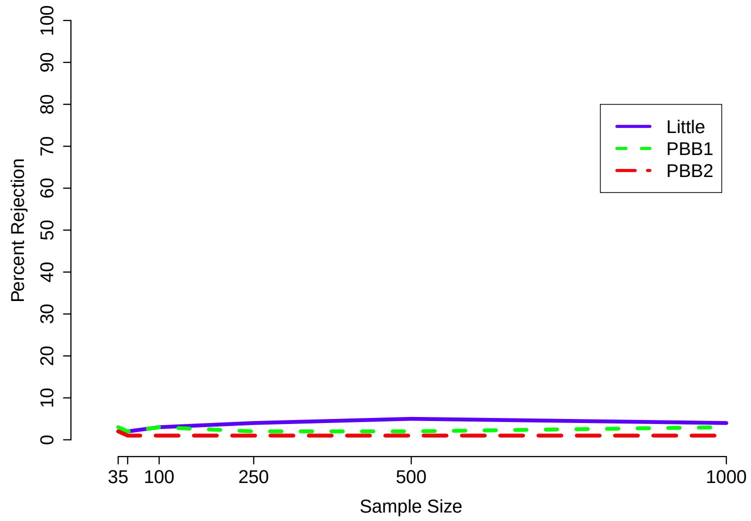


Figure 2.21: Full Simulation Study: MDM-1, Correlation Among Variables = 0.9, Collapsed Across All Other Variables

Under MDM-2 and at smaller sample sizes, a high correlation among observed variables in general had a positive effect on the correct rejection of the null hypothesis. At $N = 50$, Little's test rejected the null hypothesis 52% of the time, PBB-MCAR I rejected 45% of the time, and PBB-MCAR II rejected 32% of the time. At $N = 100$, Little's test rejected 81% of the time, PBB-MCAR II rejected 78% of the time, and PBB-MCAR I rejected 50% of the time. Therefore, increased sample size gave a smaller net increase in rejection rates for PBB-MCAR I compared to II. There was some benefit for the PBB-MCAR II test when there was a high compound symmetric relationship among variables.

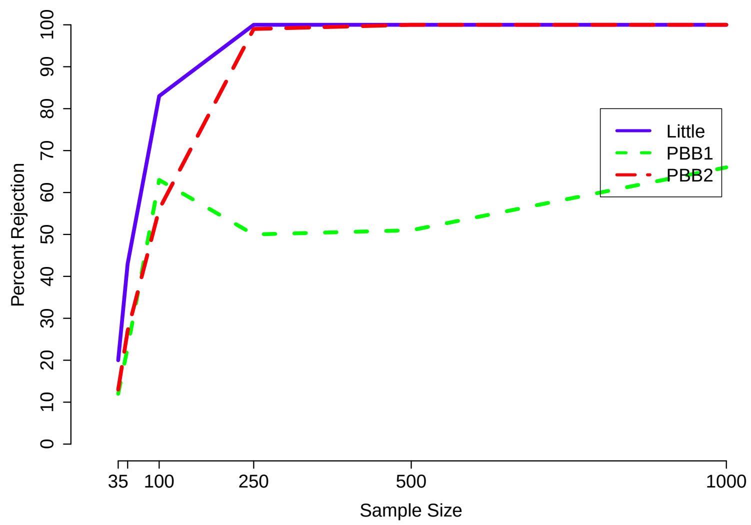


Figure 2.22: Full Simulation Study: MDM-2, Correlation Among Variables = 0, Collapsed Across All Other Variables

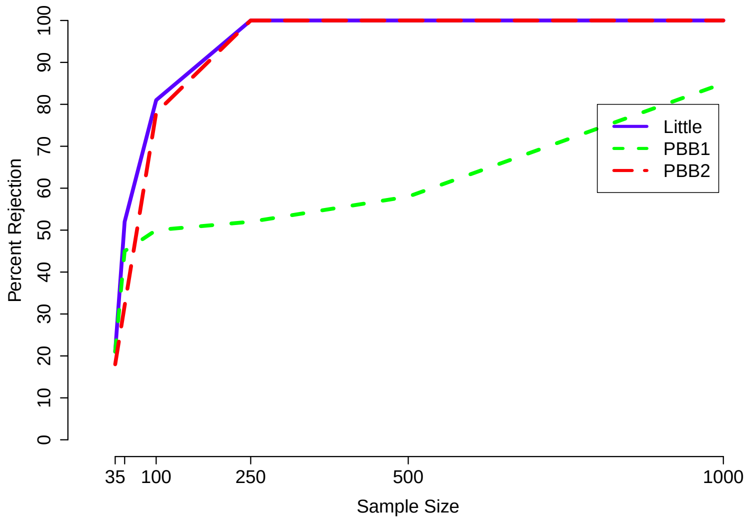


Figure 2.23: Full Simulation Study: MDM-2, Correlation Among Variables = 0.9, Collapsed Across All Other Variables

The compound symmetry among variables had a huge effect on the performance of the three test statistics when the data was generated under MDM-3. With complete independence among observed variables, all three tests underrejected the null hypothesis under MDM-3. This pattern was consistent with the pilot simulation study, except that PBB-MCAR I performed slightly better than Little's test in small samples under MDM-3. However, when the compound symmetry structure was defined, Little's test, PBB-MCAR I and II rejected the null hypothesis an average of 79%, 74% and 32% of the time, respectively across all sample sizes.

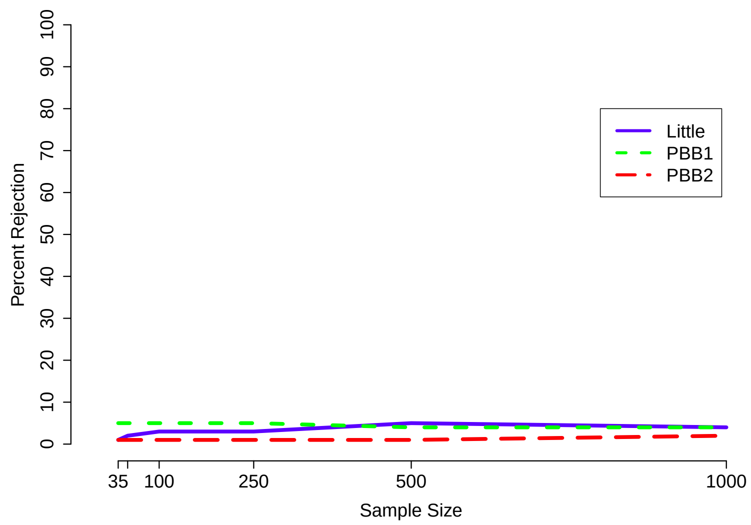


Figure 2.24: Full Simulation Study: MDM-3, Correlation Among Variables = 0, Collapsed Across All Other Variables

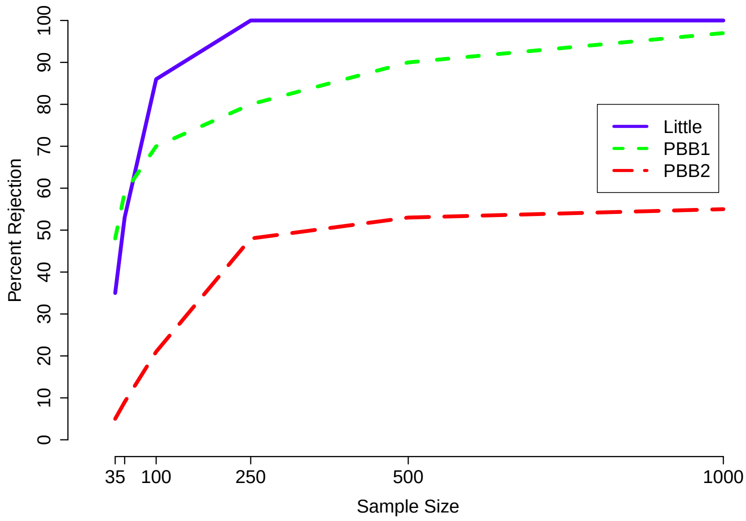


Figure 2.25: Full Simulation Study: MDM-3, Correlation Among Variables = 0.9, Collapsed Across All Other Variables

2.2.8 Discussion and Limitations

Table 2.3 summarizes the comparisons among the three tests for each of the main effects plotted. Each cell represents the ranked performance for each of the three tests (e.g., {L / I / II} indicates that Little's test outperformed PBB-MCAR I, which outperformed PBB-MCAR II). Overall across main effects, Little's test maintained optimal Type I error rates. Both PBB-MCAR I and II under-rejected the null hypothesis, with II under-rejecting more frequently than I. Although PBB-MCAR II rejection rates were inflated due to the problematic over-rejection of the null hypothesis under MDM-1 for the two-variable cases, II showed enhanced performance under the three-variable case and with low probabilities of missingness. Even in instances of {L / II / I}, Little's test had only a slight performance advantage over II, which suggests PBB-MCAR II may be a viable alternative to Little's test. Under MDM-2, PBB-MCAR II outperformed PBB-MCAR I, suggesting that taking into account ranked values of the observed variable is an important characteristic of testing for aberrations from MDM-1. The picture was not as clear under MDM-3: although Little's test outperformed the other two tests, II outperformed I only when the probability of missingness was low, and I outperformed II when the correlation among variables was high (i.e., under high compound symmetry). There is no theoretical justification why any of the three tests would reject the null hypothesis under MDM-3. Just as in the pilot simulation study, the author hypothesized that there would be severe under-rejection of the null hypothesis because there may be instances when an MNAR generation mechanism produces the same observed means as those under MCAR (Enders, 2010), although this hypothesis was not confirmed in this study.

Restrictions set on the parameters of the full simulation study prevents the generalization of the current findings to real world situations. For example, correlation values of independence and compound symmetry at $\rho = 0.9$ may not be realistic, but was chosen to magnify potential main effects. The number of variables were limited in its range, and chosen so that zero-cell counts of missingness in each category would be minimized (this is discussed further in the *Future Research* chapter). Additionally, the computational expense of running a condition even with five variables is costly. On a computer equipped with an AMD A6-3500 2.1 Ghz Triple Core processor running Windows 8, even a single run under the five-variable condition took 40 to 60 minutes at 500 replications. Simulating a run with more than 10 variables, although more realistic to education research, would be prohibitively slow with current computing technology.

Despite the inherent limitations in the design of the simulation study, PBB-MCAR II in particular showed great promise as an alternative to Little's test, especially as a test to detect the aberrations of MCAR under MDM-2. However in order for PBB-MCAR II to become a viable alternative to Little's test, the algorithm must be improved in order to give optimal Type I error rates and higher empirical power. For the two-variable case, there seems to be a direct relationship between improving Type I error under MDM-1 and increasing empirical power under MDM-2. Nevertheless, the empirical power of PBB-MCAR II is already quite promising given the stated limitations. One foreseeable enhancement to PBB-MCAR II is to change the number of blocks from which the various ψ's are estimated and to use as the suspect-rank variable the observed variable with the highest dependencies on missingness.

Table 2.3: Summary Table of Test Comparisons in the Full Simulation Study

Main Effect		MDM-1 $\alpha = 5\%$	MDM-2 target $\pi = 100\%$	MDM-3 target $\pi = 100\%$
Missingness mechanism		L / I / II	**L / II / I**	L / I / II
Number of variables	2	L / I / II	**II / L / I**	**II / L / I**
	3	L / I / II	**II / L / I**	L / I / II
	5	L / I / II	**L / II / I**	**I / L / II**
Probability of missingness	0.2	L / I / II	**II / L / I**	**L / II / I**
	0.6	L / I / II	L / I / II	L / I / II
Correlation among variables	0	L / I / II	**L / II / I**	**L / II / I**
	0.9	L / I / II	**L / II / I**	**I / L / II**

Note: (L) Little's MCAR, (I) PBB-MCAR I, (II) PBB-MCAR II. A / B means that A has a closer rejection rate to the target rejection rate compared to B. Patterns that differ from L / II / I are bolded.

CHAPTER 3

Education Application 1: Creating Longitudinal Profiles of Sports Participation

3.1 Introduction

Prior research shows that sports participation has positive effects on students' academic attitudes and achievement by increasing their interest in school and their need to maintain good grades (Snyder & Spreitzer, 1990). Extracurricular activities reinforce educational goals (Hanks & Eckland, 1976) while increasing students' attitudes and commitment toward school and their contact with teachers (Crain, 1981; Trent & Braddock, 1992; Jordan, 1999). Sports participation has also been positively associated with academic self-esteem, curriculum placement, grades, and college plans, with Black male athletes receiving social rewards both at school and among their peers (Braddock et al., 1981). Braddock et al. (1991) found that interscholastic and intramural sports participation leads to academic resilience and that athletes had higher educational aspirations than non-athletes. While Melnick et al. (1992) found no relationship between sports participation and academic achievement among a sample of 3,686 Black and Latino students, sports participation was associated with lower dropout rates (Trudeau & Shephard, 2008). This chapter studies the link between intramural sports participation and academic achievement among African American students by creating longitudinal profiles of sports participation and using tests of the missing data mechanism to determine whether these profiles are generalizable to the population of both respondents and non-respondents.

3.2 Method

3.2.1 Education Longitudinal Study of 2002

The data to be analyzed comes from the Education Longitudinal Study of 2002 (ELS:2002) public use dataset, which is a longitudinal and multilevel survey of high school sophomores, teachers, parents and administrators starting from spring of 2002 (Base Year), and assessed again in 2004 when most of the students were high school seniors (Followup 1). A total of 750 schools were sampled nationwide, with questionnaires administered to principals, head librarians, teachers, parents and students. For the purposes of this study, only the 2,020 students who self-reported 'Black or African American, non-Hispanic' on the BYRACE variable were selected for the analysis.

3.2.2 Variables Used to Assess Intramural Sports Participation

Among students who self-identified as African American, students were assessed for their participation in intramural sports at Base Year (BY) and Followup 1 (F1). Intramural sports are defined as participation in any type of sport that is not classified under the junior or varsity league. In the BY Student Questionnaire, variables BYS39A-G asked students whether they participated in intramural baseball, softball, basketball, football, soccer, another type of intramural team sport, or an individual intramural sport. These seven variables were then combined to create a new variable (BY IM Sport) that indicates a count of the student's involvement in intramural sports as high school sophomores. Student responses classified as 'Missing','Nonrespondent','Survey component legitimate skip/NA' were coded as NA, 'School doesn't have intramural team' was coded as 0, and 'Multiple response' was coded as 1. At F1, a single question assessed the student's overall participation in intramural sports (F1S26A, renamed F1 IM Sport). Responses considered missing in BY were also coded as NA, except that 'Don't know' was coded 0 and 'Participated as officer/leader/captain ' was coded 1. From these variables, four distinct longitudinal profiles of intramural sports participation were created: 1) Never Participants, 2) Persisters, 3) Late-Joiners and 4) Fallouts. Additionally, students who did not respond to the intramural sports questions were counted at each time point (see Table 3.1). This table in conjunction with Table 3.2 show the importance of considering non-responses in the analysis, as 23% of students had missing data. Excluding the missing responses, Persisters comprised the largest group (779/1561), followed by Late-Joiners (390/1561), Never-Participants (248/1561) and Fallouts (144/1561). In this sample, 84% of students who responded reported some form of intramural sports participation from their sophomore to senior years in high school. To be able to generalize these findings to the entire population of African-American students, it is important to assess whether the 23% of non-respondents consist of students that differ significantly from respondents.

Table 3.1: Profiles of Intramural Sports Participation

	F1 (Yes)	F1 (No)	F1 (NA)	Total
BY(Yes)	Persisters 779	Fallouts 144	193	1116
BY (No)	Late Joiners 390	Never Participants 248	109	747
BY (NA)	91	29	37	157
Total	1260	421	339	2020

3.2.3 Variables Used to Assess Academic Outcomes

The academic outcomes used for this analysis were grade point average (F1RGPP2, renamed F1 GPA), and the math standardized T-score (F1TXMSTD, renamed

F1 Score). The students' self-reported grade point average for all courses from 9th to 12th grade was collected at F1 and categorized into seven bins, ranging from a GPA of 0.0 to 4.0. The majority (61.1%) of Black males scored between 1.51 and 3.00 (Ingels, Pratt, Rogers, Siegel, & Stutts, 2004). The math standardized score is the score obtained from a low stakes examination given to students at both BY and F1, which is calculated as the T score transformation of the IRT theta (ability) estimate, and has a mean of 50 and standard deviation of 10 for the weighted subset of 12th graders in the sample.

3.3 Results and Conclusions

The purpose of this analysis is to demonstrate the viability of testing for the missing data mechanism in an education research context and to show that it can play an integral part in the analysis of outcomes such as test scores and extracurricular activities. The results showed that missingness patterns in sports participation trajectories were related to academic outcomes (GPA and math scores). You can see the correspondence between Table 3.2 and Figure 3.1. For example, most of the data block (in dark blue) is completely observed (i.e., $[0, 0]$). The next block starting from the left consists of blue on top and red on bottom, which corresponds to the $[1, 0]$ pattern making up 6% of the entire block. The final observed block consists of the red on top and blue on bottom, which corresponds to the $[0, 1]$ pattern, making up 15% of the entire block.

The omnibus tests described (which includes Little's test and PBB-MCAR I) may not lead to accurate conclusions because missingness may depend on certain variables but not others. The advantage of PBB-MCAR II is its ability to detect particular suspect variables driving the MAR mechanism. For convenience, let the j-th suspect variable be the first column in the data matrix. If you suspect for example that the second variable is driving the missingness dependencies, switch

Table 3.2: Patterns of Missingness for Base Year and Followup 1 Sports Participation Variables

Pattern	BY	F1	Count	%
1	0	0	1561	77
2	0	1	302	15
3	1	0	120	6
4	1	1	37	2

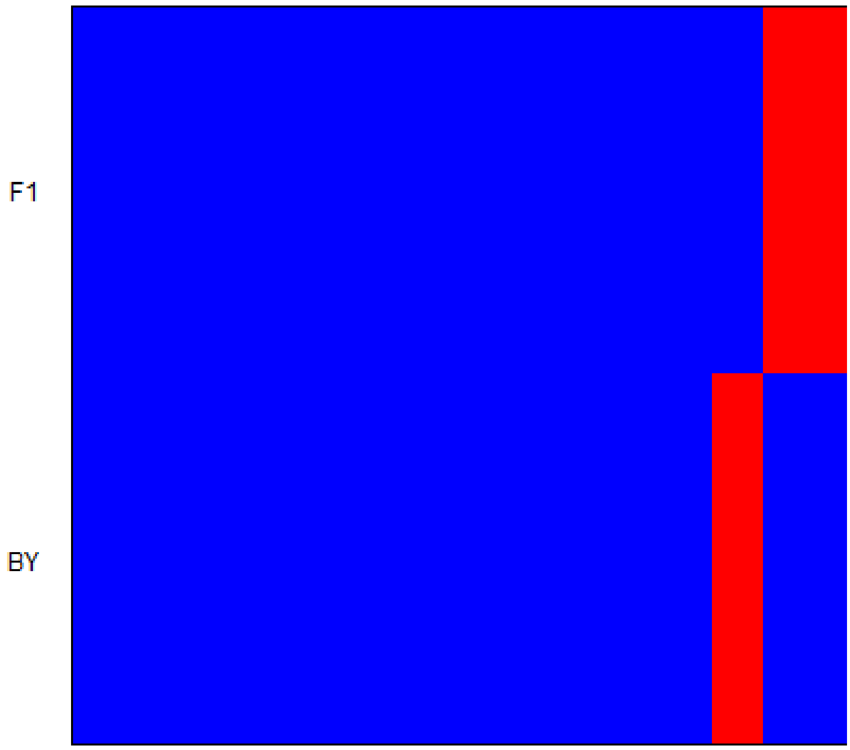


Figure 3.1: Missing Patterns for Base Year and Followup 1 Intramural Sports Variables; Red = 1 (Missing), Blue = 0 (Observed)

its column position in the dataset with the first variable. In this case, GPA was the first suspect variable and was switched with the second variable, math score. Note in Table 3.3 that PBB-MCAR II's conclusion of rejecting MCAR was the same as Little's test and PBB-MCAR I only when math score was used as the suspect ranking variable (the first 30 participants are presented in the table for demonstration purposes only). However, when GPA was used as the ranking variable, PBB-MCAR II failed to reject the null hypothesis. Further analysis using Little's test excluding the math score resulted in non-significance, $\chi^2(9) = 16.92, p = 0.79$ which would have led to the failure to reject MCAR. Additionally, the listwise correlation between math score and GPA was only $r = 0.46$, which may explain the discrepant hypothesis test results. Although it is not obvious why missingness on the math score would have an influence on missingness in sports participation, a closer inspection of Students 18 and 30, and an analysis of missingness patterns for the F1 math score and the F1 sports participation variable revealed that 16% of students had missing scores on both. The results demonstrate the utility of PBB-MCAR II and highlight the importance of considering non-respondents in your analysis, as non-response in one variable may be related to non-response in other variables. This finding would not have been discovered if we did not use PBB-MCAR II, and if we only considered the completely observed subset of students. There is also an implication that the MCAR assumption is a function of the subset of variables you consider in your model. As you increase the number of variables under consideration, the chance that one of the missingness indicators is correlated with another variable becomes greater. This suggests that any dataset could become MAR if enough variables are entered into consideration.

Table 3.3: First 30 African American Students in ELS:2002 Presented With Missing Data Tests

SID	F1 GPA	F1 Score	BY IM Sport	F1 IM Sport
1	2.51-3.00	45.59	1	1
2	1.01-1.50	42.77	0	0
3	3.01-3.5	57.5	0	0
4	1.51-2.00	42.02	1	0
5	1.01-1.50	32.69	1	0
6	1.51-2.00	29.93	0	0
7	2.51-3.00	44.84	1	1
8	2.01-2.50	41.43	0	0
9	1.51-2.00	40.11	0	0
10	2.51-3.00	43.23	NA	0
11	2.01-2.50	43.4	1	1
12	2.01-2.50	NA	0	NA
13	2.51-3.00	NA	NA	0
14	1.01-1.50	30.93	0	0
15	1.01-1.50	NA	0	NA
16	1.51-2.00	34.44	1	0
17	2.51-3.00	33.59	0	0
18	2.51-3.00	NA	0	NA
19	2.01-2.50	NA	0	0
20	1.51-2.00	34.05	0	0
21	2.01-2.50	35.07	0	0
22	1.01-1.50	47.29	1	0
23	1.01-1.50	32.16	NA	0
24	3.01-3.5	41.6	0	1
25	1.51-2.00	30.1	NA	0
26	1.51-2.00	36.8	0	0
27	1.51-2.00	26.13	0	0
28	2.51-3.00	35.24	0	0
29	1.51-2.00	39.12	0	0
30	0.00-1.00	NA	0	NA
Little		$\chi^2(25) = 220.79$, $p < 0.01$		
PBB1		$\chi^2(12) = 21.03$, $p < 0.01$		
PBB2 (y_1 = Math score)		$\chi^2(7) = 322.96$, $p < 0.01$		
PBB2 (y_1 = GPA)		$\chi^2(7) = 5.81$, $p = 0.562$		

*Analysis was done on full set of n=2020 particpants

CHAPTER 4

Education Application 2: Departures from Completely at Random Test-Taking

4.1 Introduction

Given educational policy reforms such as the No Child Left Behind Act, Race to the Top, and merit pay, teachers are faced with increasing pressure to improve student achievement on high stakes tests. In the minority of classrooms, teachers may resort to cheating as a way of artificially raising average test scores. It has been reported that 4 to 5% of classrooms cheat (Jacob & Levitt, 2003). A recent news article in USA Today (Upton, Amos, & Ryman, 2011) describes Johanna Munoz, a teacher assistant who helped her fourth-grade students cheat on a statewide achievement test by erasing wrong answers and whispering corrections into the student's ears. Amrein-Beardsley et al. (2010) define three degrees of cheating. The third degree is more benign and is not premeditated, and can involve teaching to the test or selecting questions from previous year's examination as practice problems. The second degree of cheating is subtle but premeditated, and can involve leaving the multiplication table on the board to help fourth graders solve fraction problems. The first degree of cheating is willful and premeditated, and can arise from deleting certain poor performing students from rosters or from teachers telling students to stay home sick on test day. This is what Amrein-Beardsley termed "illusions arising from exclusions". Using a hypothetical case study, the author demonstrates in this chapter how it is possible to detect departures from completely at random test-taking using the three tests of the missing data mechanism.

4.2 Hypothetical Case Study: Mr. A and Mr. B

As a case in point, consider the classrooms of Mr. A and Mr. B, each with 30 students perfectly matched by their math score and GPA across the two classrooms (the complete hypothetical dataset is shown in Table 4.1). The mean GPA is 2.39 (SD=0.81) and mean math score is 326.8 points (SD=139.36) for both teachers. In the real world, student records may be incomplete – either the student somehow missed the math exam on test day or there is no record of his GPA (see the *Creating Longitudinal Profiles of Sports Participation* chapter for an example in ELS:2002). Under completely random missingness (MCAR), you would expect that missingness on GPA is independent of missingness on math score. In Table 4.2, the MCAR missing data mechanism was separately generated for each teacher in the same manner as in the *Simulation Studies* chapter. As expected, Mr. A's classroom mean GPA is 2.26 (SD=0.71) and classroom mean math score is 322.3 (SD=143), not unlike Mr. B's mean GPA and math score of 2.26 (SD=0.80) and 330 (SD=139) respectively. Using tests of the missing data mechanism, Little's test and PBB-MCAR I both failed to reject the null hypothesis that the data was generated from an MCAR generation mechanism for both Mr. A and Mr. B ($p > 0.05$ in all cases). However, PBB-MCAR II rejected the null hypothesis of MCAR, a problem that is consistent with the over-rejection results for the two-variable case discussed in the *Simulation Studies* chapter.

Suppose that the missingness patterns are not generated completely randomly but from an informative missingness mechanism (MAR or MNAR). In the real world, there may be many forms of informative missingness, but as a case in point, suppose that Mr. A committed the first degree of teacher-assisted cheating

Table 4.1: Mr. A and Mr. B's Classroom: Complete Case Scenario

Mr. A			Mr. B		
SID	GPA	Score	SID	GPA	Score
1	3.18	447	31	3.18	447
2	2.08	234	32	2.08	234
3	3.30	476	33	3.30	476
4	1.31	121	34	1.31	121
5	2.39	385	35	2.39	385
6	4.00	600	36	4.00	600
7	2.94	401	37	2.94	401
8	3.23	422	38	3.23	422
9	2.69	390	39	2.69	390
10	2.52	357	40	2.52	357
11	1.99	223	41	1.99	223
12	2.05	216	42	2.05	216
13	2.63	387	43	2.63	387
14	2.80	417	44	2.80	417
15	1.06	124	45	1.06	124
16	3.46	460	46	3.46	460
17	3.25	459	47	3.25	459
18	1.46	194	48	1.46	194
19	1.46	225	49	1.46	225
20	3.53	522	50	3.53	522
21	1.21	129	51	1.21	129
22	1.61	191	52	1.61	191
23	2.91	491	53	2.91	491
24	1.53	145	54	1.53	145
25	3.17	457	55	3.17	457
26	2.22	296	56	2.22	296
27	2.38	404	57	2.38	404
28	2.34	290	58	2.34	290
29	1.88	238	59	1.88	238
30	1.21	103	60	1.21	103
Mean	2.39	326.80	Mean	2.39	326.80
SD	0.81	139.36	SD	0.81	139.36

Table 4.2: Mr. A and Mr. B's Classroom: MCAR Scenario

Mr. A			Mr. B		
SID	GPA	Score	SID	GPA	Score
1	3.18	447	31	3.18	447
2	2.08	234	32	NA	234
3	NA	NA	33	3.3	476
4	NA	NA	34	1.31	NA
5	2.39	385	35	NA	385
6	NA	600	36	NA	600
7	2.94	401	37	NA	NA
8	NA	422	38	NA	422
9	NA	390	39	2.69	390
10	NA	NA	40	NA	357
11	1.99	NA	41	1.99	NA
12	2.05	216	42	NA	216
13	2.63	387	43	2.63	387
14	NA	417	44	2.8	417
15	NA	124	45	1.06	124
16	NA	NA	46	3.46	NA
17	3.25	NA	47	3.25	459
18	1.46	194	48	1.46	194
19	1.46	225	49	1.46	225
20	3.53	522	50	NA	522
21	1.21	129	51	1.21	129
22	1.61	191	52	NA	NA
23	NA	491	53	2.91	NA
24	NA	145	54	1.53	145
25	3.17	457	55	3.17	457
26	2.22	NA	56	2.22	296
27	2.38	404	57	2.38	404
28	2.34	290	58	2.34	290
29	1.88	238	59	1.88	238
30	1.21	103	60	1.21	103
Mean	2.26	322.26	Mean	2.26	329.88
SD	0.71	143.06	SD	0.80	138.50
Little	$\chi^2(2) = 1.30$, $p = 0.52$		Little	$\chi^2(2) = 1.30$, $p = 0.38$	
PBB1	$\chi^2(2) = 1.65$, $p = 0.44$		PBB1	$\chi^2(2) = 0.04$, $p = 0.98$	
PBB2	$\chi^2(3) = 28.30$, $p < 0.01$		PBB2	$\chi^2(3) = 19.83$, $p < 0.01$	

by telling his lower performing students not to take the test during exam day. This effect is mimicked in Table 4.3 where 'NA' appears for students with GPA less than 2.9. Rank ordering the students by their complete GPA records reveals the pattern that students with GPA's between 1.06 – 2.8 are missing their math scores, but doing the same for Mr. B's students revealed no such pattern. Looking at the means in Table 4.3, Mr. A's mean math score is 474 (SD=55), which is above the threshold of Basic Proficiency on the California Standards Test, whereas Mr. B's mean math score of 330 (SD=139) falls slightly below the 350 cutoff. As a way to test for departures from completely at random test-taking, Little's test, PBB-MCAR I, and PBB-MCAR II were conducted on both Mr. A and Mr. B's classrooms. The analysis showed some discrepancies between the conclusions derived from the three tests. Although Little's test and PBB-MCAR II rejected the null hypothesis of MCAR, PBB-MCAR I did not. This example highlights a flaw in PBB-MCAR I that the test of independence among missingness indicators does not capture the case when missingness indicators depend completely on observed scores (MAR). The conclusion of PBB-MCAR II and Little's test is that the missingness pattern from Mr. A's class was not generated from an MCAR mechanism, signaling a departure from completely at random test-taking. Mr. B's classroom on the other hand, showed such departures based on the confluence of Little's test and PBB-MCAR I results.

The second case of informative missingness (MNAR) is much more difficult to assess in practice. The assumption of MNAR is that the probability of missingness (on math score) depends on either observed or unobserved variables (GPA, see Figure 1.3 in the *Introduction* chapter). Suppose that Mr. A not only told his lower-performing students not to show up to class during exam day, but also erased the student's transcript information from administrative records. Although extreme, this case is especially problematic for the data analyst because there is no way to distinguish whether the student's GPA is completely random or

Table 4.3: Mr. A and Mr. B's Classroom: MAR Scenario

Mr. A			Mr. B		
SID	GPA	Score	SID	GPA	Score
15	1.06	NA	2	NA	234
21	1.21	NA	5	NA	385
30	1.21	NA	6	NA	600
4	1.31	NA	7	NA	NA
18	1.46	NA	8	NA	422
19	1.46	NA	10	NA	357
24	1.53	NA	12	NA	216
22	1.61	NA	20	NA	522
29	1.88	NA	22	NA	NA
11	1.99	NA	15	1.06	124
12	2.05	NA	21	1.21	129
2	2.08	NA	30	1.21	103
26	2.22	NA	4	1.31	NA
28	2.34	NA	18	1.46	194
27	2.38	NA	19	1.46	225
5	2.39	NA	24	1.53	145
10	2.52	NA	29	1.88	238
13	2.63	NA	11	1.99	NA
9	2.69	NA	26	2.22	296
14	2.8	NA	28	2.34	290
23	2.91	491	27	2.38	404
7	2.94	401	13	2.63	387
25	3.17	457	9	2.69	390
1	3.18	447	14	2.8	417
8	3.23	422	23	2.91	NA
17	3.25	459	25	3.17	457
3	3.3	476	1	3.18	447
16	3.46	460	17	3.25	459
20	3.53	522	3	3.3	476
6	4	600	16	3.46	NA
Mean	2.39	473.50	Mean	2.26	329.88
SD	0.81	55.72	SD	0.80	138.50
Little	$\chi^2(1) = 12.60$, $p < 0.01$		Little	$\chi^2(2) = 1.30$, $p = 0.38$	
PBB1	$\chi^2(2) < 0.01$, $p = 1.00$		PBB1	$\chi^2(2) = 0.04$, $p = 0.98$	
PBB2	$\chi^2(1) = 22.50$, $p < 0.01$		PBB2	$\chi^2(3) = 19.83$, $p < 0.01$	

Table 4.4: Mr. A and Mr. B's Classroom: MNAR scenario

Mr. A			Mr. B		
SID	GPA	Score	SID	GPA	Score
15	NA	NA	2	NA	234
21	NA	NA	5	NA	385
30	NA	NA	6	NA	600
4	NA	NA	7	NA	NA
18	NA	NA	8	NA	422
19	NA	NA	10	NA	357
24	NA	NA	12	NA	216
22	NA	NA	20	NA	522
29	NA	NA	22	NA	NA
11	NA	NA	15	1.06	124
12	NA	NA	21	1.21	129
2	NA	NA	30	1.21	103
26	NA	NA	4	1.31	NA
28	NA	NA	18	1.46	194
27	NA	NA	19	1.46	225
5	NA	NA	24	1.53	145
10	NA	NA	29	1.88	238
13	NA	NA	11	1.99	NA
9	NA	NA	26	2.22	296
14	NA	NA	28	2.34	290
23	2.91	491	27	2.38	404
7	2.94	401	13	2.63	387
25	3.17	457	9	2.69	390
1	3.18	447	14	2.8	417
8	3.23	422	23	2.91	NA
17	3.25	459	25	3.17	457
3	3.3	476	1	3.18	447
16	3.46	460	17	3.25	459
20	3.53	522	3	3.3	476
6	4	600	16	3.46	NA
Mean	3.30	473.50	Mean	2.26	329.88
SD	0.31	55.72	SD	0.80	138.50
Little	NA		Little	$\chi^2(2) = 1.30$, $p = 0.38$	
PBB1	NA		PBB1	$\chi^2(2) = 0.04$, $p = 0.98$	
PBB2	NA		PBB2	$\chi^2(3) = 19.83$, $p < 0.01$	

informatively missing because there is no evidence in the data to suggest either direction. Looking at Table 4.4, Mr. A's subset of students that are not missing have both GPA and math score, which renders any form of statistical testing of the missingness inconclusive. There is no way to test for differences among missingness patterns when the remaining subset of students is completely observed. It is also known that means generated under MNAR may be identical to means generated under MCAR (Enders, 2010).

4.3 Conclusion

Using the hypothetical case of Mr. A and Mr. B, the author demonstrated how a particular type of teacher-assisted cheating can signal departures from completely at random test-taking. In the case of Table 4.3, missingness on the math score was a function of the student's observed GPA. Using the tests of missing data, there was statistical evidence that Mr. A's classroom consisted of students who departed from completely at random test-taking. Although not definitive, it does raise some red flags that may be corroborated by evidence of teacher-assisted cheating in other domains. Statistically, these examples point again to some of the limitations of the current PBB-based tests: 1) PBB-MCAR I may not be a comprehensive test of aberrations from MCAR (only testing for independence among missing indicators), and 2) PBB-MCAR II over-rejects the null hypothesis when only two variables are considered. Future research needs to address the second limitation more thoroughly.

CHAPTER 5

Future Research

5.1 Developing a Comprehensive R Package

Currently, major software packages that implement Little's test include the IBM SPSS Missing Values^{TM}20 module (IBM Corporation, 2011) and EQS 6.2 (Bentler, 2006). EQS generates output for GLS tests of homogeneity in means, covariances and means/covariances automatically as part of the `MISSING=ML` specification (Kim & Bentler, 2002). Alternatively, an R package will be made available that implements Little's test as well as the PBB-MCAR tests. There is the hope that through open source software, modifications can be made to the PBB-MCAR tests to improve either their speed of implementation or their statistical properties. As of the completion of this manuscript, generalized algorithms for Little's test and PBB-MCAR I/II have been developed and are implemented in R software (see *Appendix A* for full source code). However, before being able to upload this implementation to the Comprehensive R Archive Network (CRAN), documentation should be written and bug-testing needs to be performed.

5.2 Extending the PBB-MCAR II Test Using Auxiliary Variables

The main advantage of the PBB framework is its ability to incorporate modifications to the testing mechanism that align directly with Rubin's (1976) theory of missingness. From PBB-MCAR I and II, the author shows that it is possible to extend the probability based testing framework to a number of foreseeable mechanisms of missingness. Whereas PBB-MCAR I tests only the assumption of independence among missingness patterns, PBB-MCAR II takes this a step further by assuming independence among missingness indicators given rank-ordering of a suspect variable. These suspect variables can be generalized to what are known as auxiliary variables. Auxiliary variables are components in the Saturated Correlates Model (Figure 5.1), and are variables that are not part of your prediction model, but assumed to be correlated with the probabilities of missingness. For example, if the primary model is to assess the teacher effect on a student's score, including the student's seventh-grade and eight-grade GPA (which are variables assumed to be predictive of missingness on the student's score) may increase the accuracy of estimates.

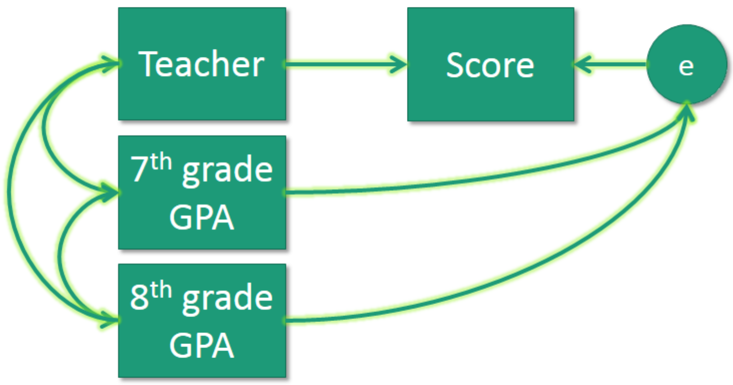


Figure 5.1: Saturated Correlates Model

This concept can be applied to the case of testing for the missing data mechanism under the PBB framework. An extension of PBB-MCAR II is to incorporate a matrix of auxiliary variables X into the ranking algorithm. A preliminary modification of the PBB-MCAR II statistic with auxiliary variables involves sorting the observed data Y simultaneously using the $i \times l$ matrix $W = \{y_j, X\}$, where

$l = 1, \cdots, L$ is the number of auxiliary variables plus one. The resulting rank index will be stored in an $(N \times 1)$ vector $\mathbf{r}$, and the arrangement of y_{ij} then becomes y_{rj}. More research needs to be done on the implementation of this algorithm.

5.3 Extending PBB Tests to Sparse Missing Data Patterns

A major disadvantage of working within the PBB framework (as with any contingency table chi-square test) is the possible existence of small counts in the contingency table as the number of variables increases. Table 5.1 shows that even with a seemingly small set of eight variables, there are $2^8 = 256$ total cells in the contingency table. This means that only 17% of the data would be expected to have complete cases, and almost zero percent would be expected to have completely missing cases. As such, improvements in both the computational efficiency of the algorithm, as well as statistical solutions to the zero cell count problem can be considered.

Table 5.1: Demonstration of Sparse Missing Pattern Counts for an Eight-Variable Dataset

	Pattern	Expected Probability		
1	OOOOOOOO	(0.8)(0.8)(0.8)(0.8)	(0.8)(0.8)(0.8)(0.8)	0.17
2	OMOOOOOO	(0.8)(0.2)(0.8)(0.8)	(0.8)(0.8)(0.8)(0.8)	0.04
$\cdots$	$\cdots$		$\cdots$	$\cdots$
256	MMMMMMMM	(0.2)(0.2)(0.2)(0.2)	(0.2)(0.2)(0.2)(0.2)	2.6E-06

A future extension of the PBB family of tests involves solving the problem of sparse missing patterns by adopting a technique borrowed from item response theory known as the M_2 statistic. The M_2 statistic is part of a limited-information family of chi-square statistics described in Maydeu-Olivares and Joe (2005), which is a weighted discrepancy between the sample and expected marginal probabilities.

Without loss of generality, the following definitions pertain only the case when $p = 3$. Define the $(K-1)$ dimensional column vector $\dot{\pi} = [\dot{\pi}_1, \dot{\pi}_2, \dot{\pi}_3]'$ where $\dot{\pi}_1 = [P(m_{i1} = 1), P(m_{i2} = 1), P(m_{i3} = 1)]'$, $\dot{\pi}_2$ the $\binom{3}{2}$ dimensional column vector of bivariate probabilities, and $\dot{\pi}_3 = [P(m_{i1} = 1)P(m_{i2} = 1)P(m_{i3} = 1)]'$ the trivariate joint probability of missingness. To capture the marginal probabilities, generate a $(K-1) \times K$ matrix $\mathbf{T}$ which is a transformation matrix consisting of zeros in the first column. If we partition this full $\mathbf{T}$ matrix and restrict it only to the second order marginals, then we get the $\mathbf{T_2}$ matrix with dimensions $s \times K$, where $s = \sum_{x=1}^{p} \binom{p}{x}$; in this case the dimension is 7×8. Let $\mathbf{p_o}$ be the vector of observed probabilities $(p_o^{(1)}, \cdots, p_o^{(K)})'$ and $\mathbf{p_e}$ be the vector of expected probabilities $(p_e^{(1)}, \cdots, p_e^{(K)})'$. Then $\dot{\pi}_e = \mathbf{T}\mathbf{p_e}$.

Begin by calculating the residual error terms as

$$\hat{\mathbf{e}} = \mathbf{p_o} - \mathbf{p_e} \tag{5.1}$$

If we only want the two-order marginals of the error terms, define

$$\hat{\mathbf{e_2}} = \mathbf{T_2}\hat{\mathbf{e}} \tag{5.2}$$

Next, define the first order derivative with respect to the parameter vector ψ as

$$\boldsymbol{\Delta} = \frac{\partial \mathbf{p_e}}{\partial \psi} \tag{5.3}$$

If we are concerned with the two-order marginals of the partial derivatives, then consider

$$\boldsymbol{\Delta_2} = \mathbf{T_2}\boldsymbol{\Delta} \tag{5.4}$$

which will generate a $s \times q$ matrix where q is the number of parameters estimated. Maydeu-Olivares and Joe (2005) take the orthogonal complement of Equation

5.4, which they call $\mathbf{\Delta_{2c}}$. Practically, this matrix can be obtained using the QR-decomposition method.

Finally, define the equations

$$\mathbf{\Gamma} = \mathrm{diag}(\mathbf{p_e} - \mathbf{p_e}\mathbf{p_e})' \tag{5.5}$$

$$\mathbf{\Xi_2} = \mathbf{T_2}\mathbf{\Gamma}\mathbf{T_2}' \tag{5.6}$$

and we can define the limited information statistic as

$$M_2 = N(\mathbf{\Delta_{2c}}'\hat{\mathbf{e_2}})'(\mathbf{\Delta_{2c}}'\mathbf{\Xi_2}\mathbf{\Delta_{2c}})^{-1}(\mathbf{\Delta_{2c}}'\hat{\mathbf{e_2}}) \tag{5.7}$$

5.3.1 Calculating the M_2 Statistic

For the particular example described in the *Introduction* chapter with three variables m_1, m_2, m_3, the 7×1 vector of univariate, bivariate and trivariate probabilities is $\dot{\pi} = [\dot{\pi}_1, \dot{\pi}_2, \dot{\pi}_3, \dot{\pi}_{12}, \dot{\pi}_{13}, \dot{\pi}_{23}, \dot{\pi}_{123}]'$. This vector will not be used in the current example, but is given for illustrative purposes. Given the ordering described in the tables of observed and expected probabilities described in previous sections, we can create the matrix

$$\mathbf{T} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.8}$$

As described previously, partition this full $\mathbf{T}$ matrix and restrict it only to the univariate and bivariate marginal probabilities; then we get the 6×8 $\mathbf{T_2}$ matrix. This allows us to calculate the M_2 statistic using the following procedure:

1. Find the residual terms by subtracting the observed and expected probabilities of missingness as described in Equation 5.1 and obtain the two order marginals using Equation 5.2.

2. Take the numerical partial derivatives with respect to ψ given in Equation 5.3 using the `numericDeriv()` package in R and get the two-order partial derivatives using Equation 5.4.

3. Find its orthogonal complement using `qr()`.

4. Obtain the two-order $\boldsymbol{\Xi_2}$ matrix using Equation 5.6.

5. Calculate the M_2 statistic using Equation 5.7.

CHAPTER 6

Conclusion

6.1 General Conclusions and Discussion

Little (1988) developed a test to distinguish between completely random (MCAR) and informative mechanisms of missingness (MAR and MNAR) by using observed and expected means – but there may be instances when these two categories of missingness may generate the same observed mean patterns and lead the researcher to conclude that the data is missing completely at random when it is actually missing informatively. As an alternative, the manuscript proposed a framework to test for the missing data mechanism that uses probabilities of missingness rather than observed means and covariances. The advantages of this framework are that it a) directly aligns with Rubin's (1976) theory of missingness, b) does not assume multivariate normality of observed variables, and c) allows extensions to a family of statistical tests. Currently, two statistics have been developed within this probability based (PBB) framework, the MCAR Chi-Square Test of Independence (PBB-MCAR I) and the MCAR Goodness of Fit Test (PBB-MCAR II).

This manuscript introduced the PBB framework and its family of test statistics (i.e., PBB-MCAR I and II), which each targets a different aspect of the missing data mechanism. PBB-MCAR I tests the assumption of independence among missing pattern indicators, whereas PBB-MCAR II tests the assumption that missing pattern indicators are independent given ranked-ordering of a suspect variable driving the missingness mechanism. The *Departures from Completely at Random Test-Taking* chapter provided the motivation for developing an alternative to PBB-MCAR I, because PBB-MCAR I failed to detect cases of MAR when the violations of MCAR were due to the dependency of missing indicators on an observed variable (GPA). PBB-MCAR II achieves goals more in line with the author's original objectives, by testing for more realistic departures from MCAR, one that depends on ranked values of the observed variable.

Simulation results showed that under MDM-2, PBB-MCAR II was a powerful alternative to Little's test. In general, the empirical power of PBB-MCAR I was lower than both Little's test and PBB-MCAR II under MDM-2, but was on par with Little's test under MDM-3. More research needs to be done to understand why Little's test and PBB-MCAR I were able to detect these violations, and why PBB-MCAR II did not perform as well under MDM-3. A foreseeable improvement of PBB-MCAR II can be made as discussed in the *Future Research* chapter, to offer a more comprehensive test of aberrations from MCAR by allowing researchers to enter in more than one variable suspected to generate missingness dependencies. As shown in the *Creating Longitudinal Profiles of Sports Participation* chapter, the rejection decision of the statistical test was sensitive to the researcher's choice of the suspect variable.

6.1.1 Missing Data Tests in the Context of Education Research

The application of the three tests of the missing data mechanism to education were demonstrated in Chapters 3 and 4. Chapter 3 considered the relationship between longitudinal profiles of sports participation and academic outcomes among African American students as part of the Education Longitudinal Study of 2002. Since 23% of the 2,020 students considered had some form of missing data across the Base Year (BY) and Followup 1 (F1) assessment points, the three tests of missing data were conducted to assess whether the missing patterns were informatively missing. When F1 Math score and GPA were entered into the analysis, Little's test and PBB-MCAR I rejected the assumption of MCAR. The results suggest that whether a student persists or drops out of an intramural sports program may be influenced either by the student's cumulative GPA or by his or her performance on a low-stakes math exam conducted at F1. Further analysis using PBB-MCAR II showed that when the F1 Math Score (rather than cumulative GPA) was entered as the suspect ranking variable, the assumption of MCAR was rejected. The complete-data correlation of GPA and F1 Math score among completely observed participants was only $r = 0.46$, suggesting that GPA may be related to longitudinal profiles of sports participation differently from F1 Math Score. The analysis demonstrated the practical advantage of PBB-MCAR II because it allows specificity in detecting suspected sources of missingness dependencies. As such, whether a student persists or drops out of intramural sports may be due his or her achievement at a particular point in time (F1) and not on the student's cumulative GPA across all years. The rejection of the null hypothesis among these tests tapped into three aspects of the missingness mechanism: 1) Little's test confirmed that observed values among GPA, Math Score and Sports Participation differ from expected values across all missingness patterns, 2) PBB-MCAR I confirmed that dependencies exist among the missingness indicators, and 3) PBB-MCAR II confirmed that these dependencies among missing indicators depend on F1 Math Score but not on GPA.

Chapter 4 described how tests of the missing data mechanism were used to detect departures from completely at random test-taking. The assumption is that in the typical classroom, missing test scores on a statewide exam does not depend on the student's background characteristics such as GPA. In rare and extreme cases of teacher-assisted cheating, teachers may intently withhold certain poor performing students from taking the exam during test day. The chapter described a hypothetical case study in which teacher Mr. A's classroom test performance was higher than Mr. B's. Using the three tests of the missing data mechanism, we concluded that Mr. A's classroom deviated from completely at random test-taking. Caution must be given when generalizing the statistical results of these tests. Not completely at random test-taking may be observed due to other factors, such as the student's own reluctance to show up during test day. Nevertheless, these tests may help in corroborating evidence from other sources of cheating such as the presence of systematic patterns in test responses or an unexpected number of perfect scores. The three tests highlighted again the fact that there may be three aspects of missingness stemming from a) differences in observed means from expected means, b) dependencies among missingness indicators, and c) dependencies of missing indicators on an observed variable.

The two applications described in this manuscript are not the only areas where missing data testing may be applied in education research. For example, in large-scale testing programs such as the National Assessment of Educational Progress (NAEP), not every item is given to every student assessed. This design is known as Balanced Incomplete Block spiraling (BIB-spiraling), which divides test items within a subject area into blocks and assigned in such as a way that each block appears in the same number of booklets and every pair of blocks of a certain type appears together in at least one booklet, but no booklet contains all items (Johnson, 1992). The BIB-spiraling design is an example of planned missingness, where the missing blocks of items are intended by design and assumed to be missing completely at random. The three tests discussed in this manuscript can test the validity of the MCAR assumption in the BIB-spiraling design and whether missingness is due to observed variables such as student background characteristics or due to the missing test items themselves.

CHAPTER 7

Appendix

7.1 Appendix A: Full Source Code

```
###PRESETS AND FUNCTIONS###
##define the following parameters##
#nreps (number of reps)
#N (sample size)
#pmissvec (marginal probabilities of missingness)
#meanvec (vector of variable means)
#corrvec <-(vector of correlations among variables)
#sdvec (vector standard deviations among variables)
#lowvec (lower floor for range of variable)
#highvec (upper ceiling  range of variable)
#load libraries & set path
library(gregmisc)
library(norm)
library(mi)
library(MASS)
library(MBESS)

##functions##
gendat <- function(size,meanvec,corrvec,sdvec){
dat.mean <- meanvec
```

```
dat.Sigma <- corr2Sigma(corrvec=corrvec,sdvec=sdvec,
nvars=length(dat.mean))
mvrnorm(size,mu=dat.mean,Sigma=dat.Sigma)
} #end gendat() function

identity <- function(nvars){
diag(rep(1,nvars))
} #end identity()

corr2Sigma <- function(corrvec,sdvec,nvars){
#add covariances to the offdiagonals
dat.corr <- identity(nvars)
#set upper triangle
dat.corr[upper.tri(dat.corr)] <- corrvec
#set lower triangle
dat.corr[lower.tri(dat.corr)]
<- dat.corr[upper.tri(dat.corr)]
dat.corr
dat.Sigma <- cor2cov(dat.corr,sd=sdvec)
return(dat.Sigma)
} #end corr2Sigma()

fixrange <- function(dat.Y,low,high){
dat.Y.high <- dat.Y
dat.Y.low <- dat.Y
dat.Y.fixed <- dat.Y
for (j in 1:dim(dat.Y)[2]){
#cap minimum score
```

```
dat.Y.high[,j] <- ifelse(dat.Y[,j] > high[j],
high[j],dat.Y[,j])
dat.Y.fixed[,j] <- dat.Y.high[,j]
# minimum score
dat.Y.low[,j] <- ifelse(dat.Y.fixed[,j] < low[j],
low[j],dat.Y.fixed[,j])
dat.Y.fixed[,j] <- dat.Y.low[,j]
} #end for loop
return(dat.Y.fixed)
} #end fixrange()

genmiss <- function(dat,size,pmiss, misstype,
   misslow = misslow) {
x <- matrix(NA,nrow=size,ncol=length(pmiss))
for (j in 1:length(pmiss)){
x[,j] <- runif(size)
} #end for loop
if (misstype == "MCAR") {
print("MCAR")
for (j in 1:length(pmiss)){
dat[which(x[,j] < pmiss[j]),j] <- NA
} #end for loop
return(list(dat=dat,x=x,misstype=misstype))
} #end MCAR

else if (misstype == "MAR") {
print("MAR")
for (j in 1:dim(dat)[2]){
```

```
if (j == 1) dat[which(x[,j] < pmiss[j]),j] <- NA
#if (j == 1) dat[which(dat[,dim(dat)[2]] < misslow),j] <- NA
both <- intersect(which(dat[,(j-1)] < misslow),
which(x[,j] < pmiss[j]))
if (j != 1) dat[both,j] <- NA
} #end for loop
return(list(dat=dat,x=x,misstype=misstype))
} #end MAR

else if (misstype == "MNAR") {
print("MNAR") #Jamshidian, Yuan and Le (in press)
for (j in 1:dim(dat)[2]){
both <- intersect(which(dat[,j] < misslow),
which(x[,j] < pmiss[j]))
dat[both,j] <- NA
} #end for loop
return(list(dat=dat,x=x,misstype=misstype))
} #end MNAR

} #end genmiss()

makeRmat <- function(Y){
R <- ifelse(is.na(Y),1,0)
return(R)
} #end makeRmat()

patterns <-function(ndim){
dat <- as.data.frame(permutations(n=2,r=ndim,v=0:1,
```

```
repeats.allowed=TRUE))
for (i in 1:2^ndim){
row.names(dat)[i] <- paste(dat[i,],sep="",collapse="")
}
return(dat)
} #end patterns()

match.patterns <-function(R,npatt){
pattvec <- NULL
patt <- patterns(npatt)
for (j in 1:dim(patt)[1]){
for (i in 1:dim(R)[1]){
if(isTRUE( identical(as.numeric(R[i,]),
as.numeric(patt[j,])))){
pattvec[i] <- (row.names(patt[j,]))
} #end if statement
} #end for loop
} #end for loop
return(as.factor(pattvec))
} #end match.patterns()

extractSj <- function(Sj){
v <- lapply(strsplit(names(Sj),""),as.numeric)
names(v) <- names(Sj)
return(v)
} #end extractSj()

createDj <- function(ndim,mj,Sj){
```

```
Dj.temp <- list()
vec <- NULL
howmany0s <- NULL
pattern.finder <- NULL
pattern.finder <- extractSj(Sj)
position0 <- list()

for (i in 1:length(mj)){
howmany0s[i] <- length(which(pattern.finder[[i]]
[1:ndim]==0))
position0[[i]] <- which(pattern.finder[[i]]==0)
position0[[length(mj)]] <- 0
} # end for loop
for (i in 1:length(mj)){
for (j in 1:length(ndim)){
if (howmany0s[i] > 0){
Dj.temp[[i]] <- matrix(0,
ncol=howmany0s[i],nrow=ndim)
for (k in 1:length(position0[[i]])){
Dj.temp[[i]][position0[[i]][k],k] <- 1 #end if statement
} #end for loop
} #end if statment
else if (howmany0s[i] == 0){
Dj.temp[[i]] <- matrix(0,ncol=1,nrow=ndim)
}# end else if statement
} #end inner for loop
} #end outer for loop
names(position0) <- names(pattern.finder)
```

```
return(Dj.temp)
} #end createDj()

matrixmult <- function(x,y)  {
x999 <- ifelse(is.na(as.matrix(x)),999,as.matrix(x))
y999 <- ifelse(is.na(as.matrix(y)),999,as.matrix(y))
as.matrix(x999)%*%as.matrix(y999)
} #end matrimult()

is.wholenumber <- function(x, tol = .Machine$double.eps^0.5)
{abs(x - round(x)) < tol} #end is.wholenumber()

jointpfinder <- function(Sj, mj,nvars,phij){
pattern.finder <- list()
pattern.finder <- extractSj(Sj)
outp <- matrix(NA,ncol=nvars,nrow=length(mj))
for (i in 1:length(mj)){
for (j in 1:nvars){
if( pattern.finder[[i]][j]  == 0 )
outp[i,j] <- 1-phij[j]
else if( pattern.finder[[i]][j]  == 1 )
outp[i,j] <- phij[j]
} #end inner for loop
} #end outer for loop
rownames(outp) <- names(mj)
return(outp)
}#end jointpfinder()
```

```
prepwork <- function(dat){
Rdat <- makeRmat(dat)
#calc pct missing, should be equal to obs. prob.
length(which(is.na(dat)))/(dim(dat)[1]*dim(dat)[2])
#find patterns of missingness
pattern.id <- match.patterns(R=Rdat,npatt=dim(Rdat)[2])
mj <- NULL
mj <- table(pattern.id)
Sj <- list()
Sj <- split(x=as.data.frame(dat),f=pattern.id)
#create Dj matrix
N <- dim(Rdat)[1]
nvars <- dim(Rdat)[2]
Dj <- list()
Dj <- createDj(ndim=nvars,mj=mj,Sj=Sj)
names(Dj) <- names(Sj)
#create yobs matrix
J <- length(levels(pattern.id)) #distinct no. of patterns
yobs <- list()
sum(mj)== N #check that the sum is equal to n
#split Sj patterns into numeric form
nnames <- as.numeric(unlist(strsplit(names(Sj),"")))
times <- as.factor(rep(seq(1:(length(nnames)/nvars)),
 each=nvars))
pj <- NULL
pj <- nvars - unlist(lapply(split(x=nnames,f=times),sum))
names(pj) <- names(Sj)
names(mj) <- names(Sj)
```

```
return(list(Rdat=Rdat,Sj=Sj,Dj=Dj,mj=mj,pj=pj,
J=J,pj=pj,nvars=nvars))
} #prepwork()

#### B) LITTLE'S MCAR TEST####
littlemcar <- function(dat){
outprep <- prepwork(dat)
Sj <- outprep$Sj
Dj <- outprep$Dj
mj <- outprep$mj
pj <- outprep$pj
J <- outprep$J
nvars <- outprep$nvars

yobs <- list()
for (i in 1:(J-1)){
yobs[[i]] <- matrixmult(Sj[[i]],Dj[[i]])
}
ybar <- list() #get ybar
for (i in 1:((length(mj))-1)){
ybar[[i]] <- apply(yobs[[i]],2,sum)/as.vector(mj)[i]
} # end for loop
names(ybar) <- names(Sj)[1:(length(mj)-1)]
s <- prelim.norm(as.matrix(dat)) #EM algorithm
thetahat <- em.norm(s,showits=FALSE)
out <- getparam.norm(s,thetahat,corr=FALSE)
mu.mle <- out$mu
sigma.mle <- out$sigma
```

```
sigma.tilde <- (N/(N-1))*sigma.mle
muobsj <- list() #find muobsj sigmaobsj
sigmaobsj <- list()
sigmatildeobsj <- list()

for (i in 1:(J-1)){
muobsj[[i]] <- mu.mle%*%Dj[[i]]
sigmaobsj[[i]] <- t(Dj[[i]])%*%sigma.mle%*%Dj[[i]]
sigmatildeobsj[[i]] <- t(Dj[[i]])%*%sigma.tilde
%*% Dj[[i]]
}

#find d0sq
d0sq <- list()
for (i in 1:(J-1)){
d0sq[[i]] <- mj[i]*(ybar[[i]] - muobsj[[i]])
%*%solve(sigmaobsj[[i]])%*%t(ybar[[i]]
- muobsj[[i]])
}
d0sqsum <- sum(unlist(d0sq))
dsq <- list() #p.1200 in Little (1988) sigma unknown
for (i in 1:(J-1)){
dsq[[i]] <- mj[i]*(ybar[[i]] - muobsj[[i]])
%*%solve(sigmatildeobsj[[i]])%*%t(ybar[[i]] - muobsj[[i]])
}
little.dsqsum <- sum(unlist(dsq))
little.df <- sum(pj)-nvars #df p.1200 in Little (1988)
little.chicrit <- qchisq(.95,little.df)
```

```
little.pval <- 1-pchisq(little.dsqsum,little.df)
print(c("little.pval", little.pval),quote=FALSE)
print(c("little.dsqsum", little.dsqsum),quote=FALSE)
print(c("little.df", little.df),quote=FALSE)
if(little.dsqsum < little.chicrit){
little.rejectH0 <- 0 else little.rejectH0 <- 1
} #end if statement

return(list(little.dsqsum=little.dsqsum,little.df=little.df,
little.chicrit=little.chicrit,little.pval=little.pval,
little.rejectH0=little.rejectH0))

} #end littlemcar()

### C)PBB-MCAR 1 ###
pbb1mcar <- function(dat){
outprep <- prepwork(dat)
Rdat <- outprep$Rdat
Sj <- outprep$Sj
Dj <- outprep$Dj
mj <- outprep$mj
pj <- outprep$pj
J <- outprep$J
nvars <- outprep$nvars
N <- dim(dat)[1]
phi.j <- apply(Rdat,2,sum)/dim(Rdat)[1]
expected.phi.j  <- jointpfinder(Sj=Sj,mj=mj,
nvars=nvars,phij = phi.j)
```

```
expected.phi.j
expected.phi.j.vec <- apply(expected.phi.j,1,prod)
expected.phi.j.vec
phi.j.vec <- expected.phi.j.vec
observed.p.vec <- NULL
observed.p.vec <- mj/N
p0a.j.vec <- observed.p.vec
D <- diag(c(expected.phi.j.vec))
pbb1.csqsum <- list()
pbb1.csqsum <- N*t(p0a.j.vec  - phi.j.vec)
%*%solve(D)%*%(p0a.j.vec  - phi.j.vec)
pbb1.df <- 2^nvars - nvars   #this may be incorrect
pbb1.crit <- qchisq(.95,pbb1.df)
pbb1.pval <- 1-pchisq(pbb1.csqsum,pbb1.df)
print(c("PBB1 pvalue",pbb1.pval))
if(pbb1.csqsum < pbb1.crit) pbb1.rejectH0 <- 0
else pbb1.rejectH0 <- 1

return(list(phi.j=phi.j,expected.phi.j=expected.phi.j,
observed.p.vec=observed.p.vec,pbb1.csqsum=pbb1.csqsum,
pbb1.df=pbb1.df,pbb1.crit=pbb1.crit,pbb1.pval=pbb1.pval,
pbb1.rejectH0=pbb1.rejectH0))

} #end pbb1mcar()

### D) PBB-MCAR 2 ###
pbb2mcar <- function(dat,nq=2){
Rdat <- makeRmat(dat)
```

```
Rdat.r <- Rdat[order(dat[,1]),]
if(dim(Rdat.r)[2] >= 3) Rdat.rcut <- Rdat.r[,2:dim(dat)[2]]
if(dim(Rdat.r)[2] < 3) Rdat.rcut <- Rdat.r
if (is.wholenumber(N/nq)==TRUE) quads
<- rep(1:nq,each=N/nq)
if (is.wholenumber(N/nq)==FALSE) quads
<- cut(1:N,nq)
Rdat.rcut.split <- split(x=as.data.frame(Rdat.rcut),
f=as.factor(quads))
sumeach <- t(sapply(Rdat.rcut.split,apply,2,sum))
neach <- sapply(Rdat.rcut.split,dim)[1,]
phi.j <- apply(Rdat.rcut,2,sum)/dim(Rdat.r)[1]
phi.j.split <- sumeach/neach
#the two should be equal
round(apply(phi.j.split,2,mean),3) == phi.j
if(dim(Rdat.r)[2] >= 3) outprep <- prepwork(dat[,-1])
if(dim(Rdat.r)[2] < 3) outprep <- prepwork(dat)
Sj <- outprep$Sj
Dj <- outprep$Dj
mj <- outprep$mj
pj <- outprep$pj
J <- outprep$J
nvars <- outprep$nvars
N <- dim(dat)[1]
observed.phi.j  <- list()
for (i in 1:nq){
observed.phi.j[[i]] <-
jointpfinder(Sj=Sj,mj=mj,nvars=nvars,
```

```
phij = phi.j.split[i,])
} #end for loop
observed.phi.j.vec <- t(sapply(observed.phi.j,apply,1,prod))
expected.phi.j  <- jointpfinder(Sj=Sj,mj=mj,
nvars=nvars,phij = phi.j)
expected.phi.j.vec <- apply(expected.phi.j,1,prod)
phi.j.vec <- expected.phi.j.vec
p0b.j.vec <- observed.phi.j.vec
D <- diag(c(expected.phi.j.vec))
difference <- matrix(NA,nrow=nq,ncol=length(expected.phi.j.vec))
for (i in 1:nq){
difference[i,] <- p0b.j.vec[i,] - expected.phi.j.vec
} #end for loop
pbb2.csq <- c() #find chi-square for PBB MCAR
for (i in 1:nq){
pbb2.csq[i] <- neach[i]*t(difference[i,])%*%solve(D)
   %*%(difference[i,])
} #end for loop
pbb2.csqsum <- sum(pbb2.csq)
pbb2.csqsum
pbb2.df <- length(mj) - 1 #goodness of fit df
pbb2.crit <- qchisq(.95,pbb2.df)
pbb2.pval <- 1-pchisq(pbb2.csqsum,pbb2.df)

if(pbb2.csqsum < pbb2.crit) pbb2.rejectH0 <- 0
else pbb2.rejectH0 <- 1

return(list(nq=nq,mj=mj,phi.j=phi.j,
```

```
expected.phi.j=expected.phi.j,
observed.phi.j.vec=observed.phi.j.vec,
expected.phi.j.vec=expected.phi.j.vec,
pbb2.csq=pbb2.csq,pbb2.csqsum=pbb2.csqsum,
pbb2.df=pbb2.df,pbb2.crit=pbb2.crit,
pbb2.pval=pbb2.pval,pbb2.rejectH0=pbb2.rejectH0))

} #end pbb2mcar()
```

7.2 Appendix B: Figures of All 36 Conditions in the Full Simulation Study

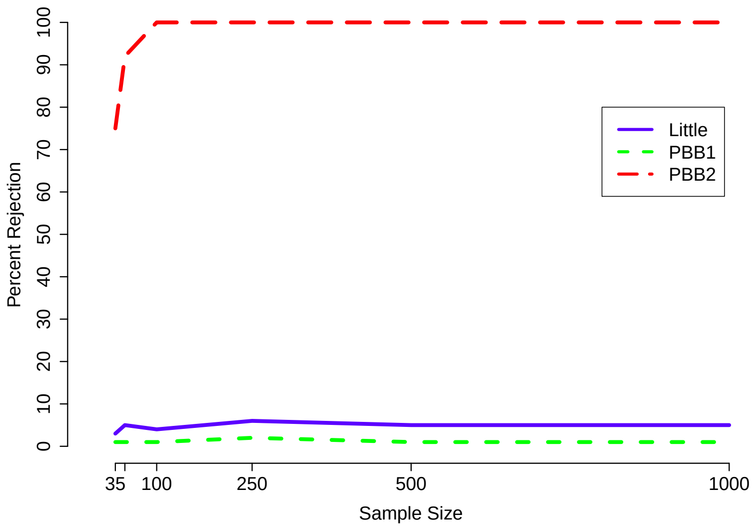


Figure 7.1: Full Simulation Study: MDM-1, Number of Variables = 2, Probability of missingness = 0.2, No Correlation Among Variables

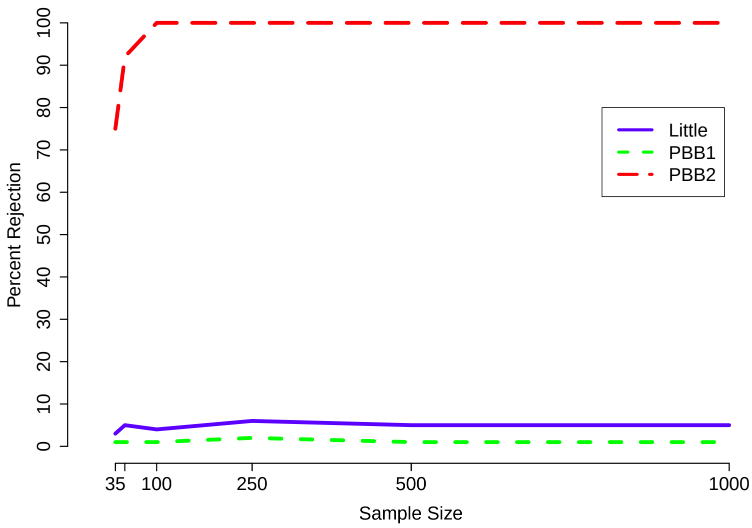


Figure 7.2: Full Simulation Study: MDM-1, Number of Variables = 2, Probability of missingness = 0.2, No Correlation Among Variables

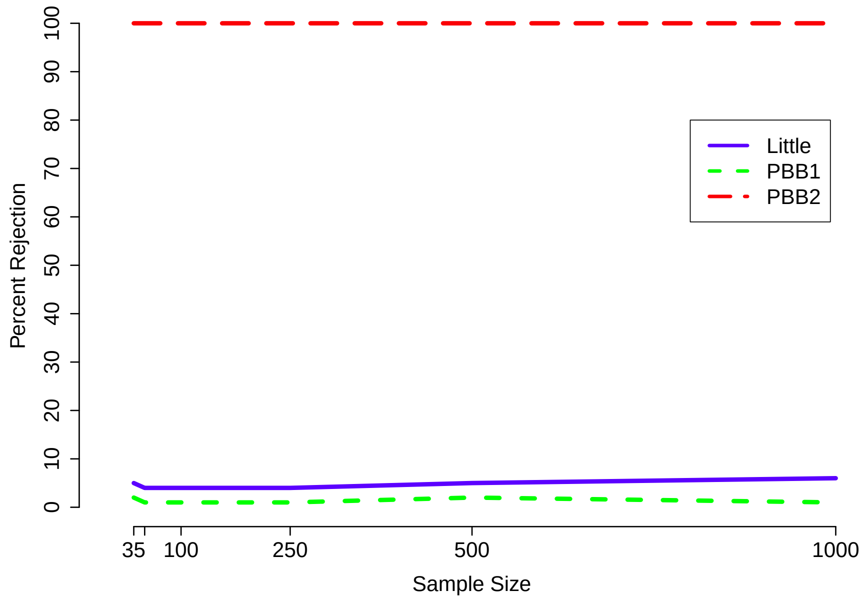


Figure 7.3: Full Simulation Study: MDM-1, Number of Variables = 2, Probability of missingness = 0.6, No Correlation Among Variables

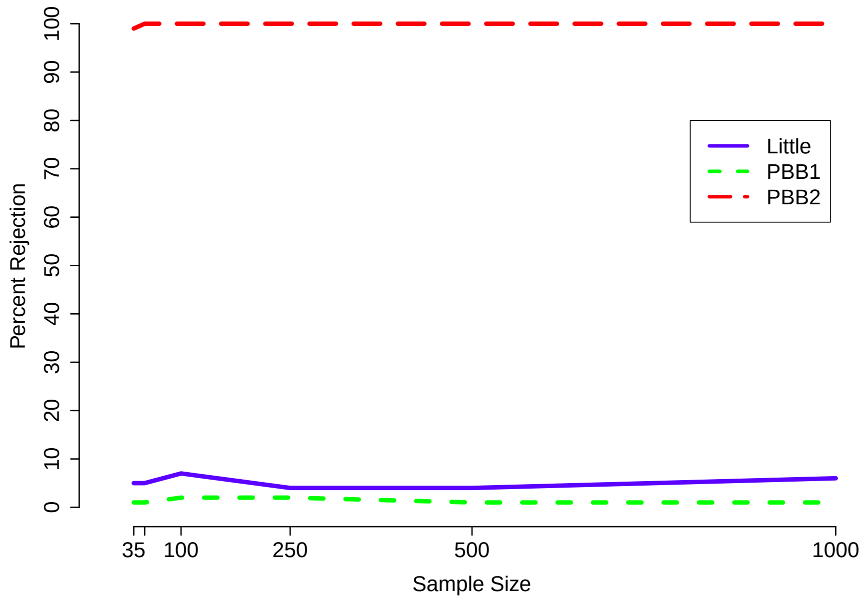


Figure 7.4: Full Simulation Study: MDM-1, Number of Variables = 2, Probability of missingness = 0.6, Correlation Among Variables = 0.9

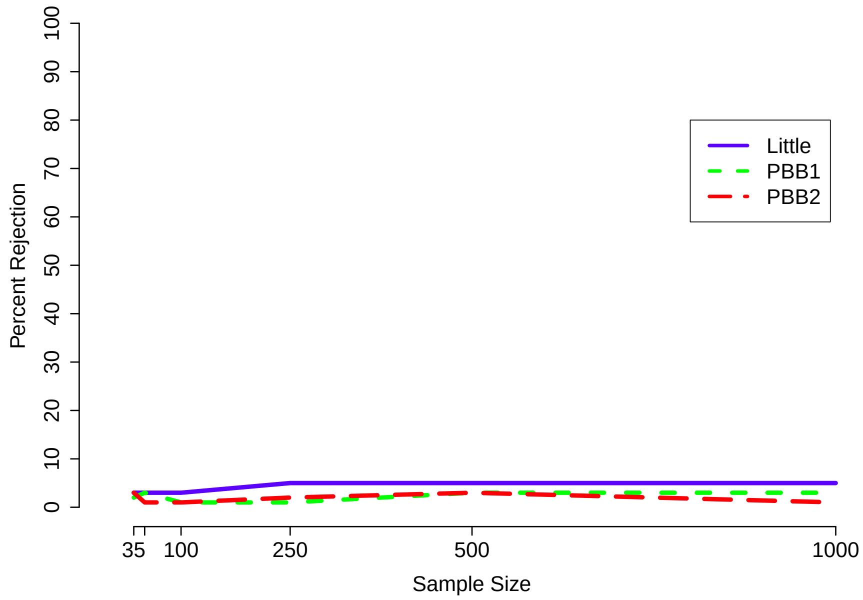


Figure 7.5: Full Simulation Study: MDM-1, Number of Variables = 3, Probability of Missingness = 0.2, No Correlation Among Variables

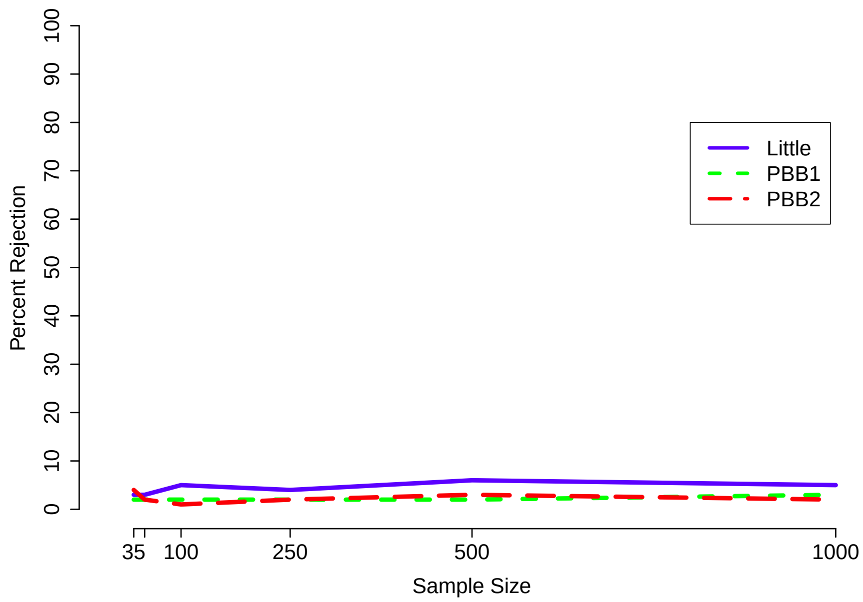


Figure 7.6: Full Simulation Study: MDM-1, Number of Variables = 3, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

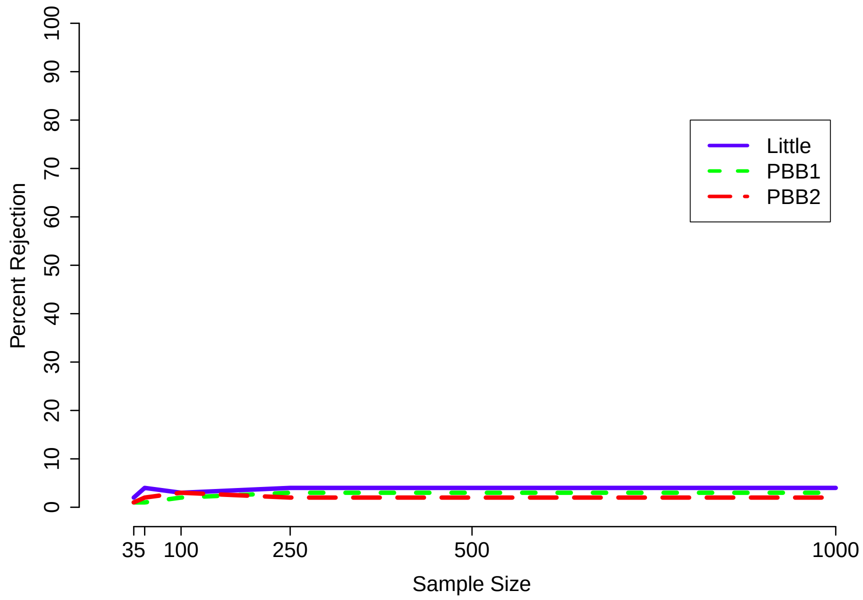


Figure 7.7: Full Simulation Study: MDM-1, Number of Variables = 3, Probability of Missingness = 0.6, No Correlation Among Variables

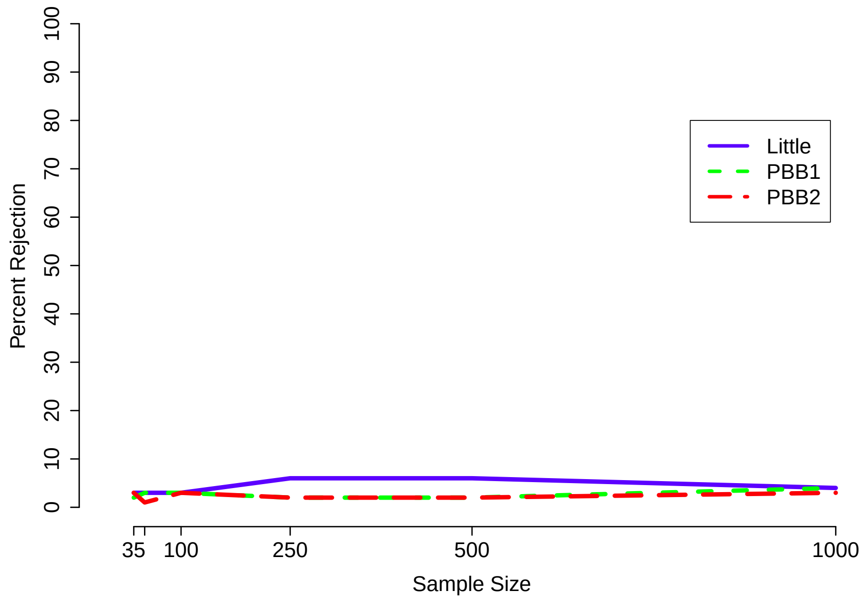


Figure 7.8: Full Simulation Study: MDM-1, Number of Variables = 3, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

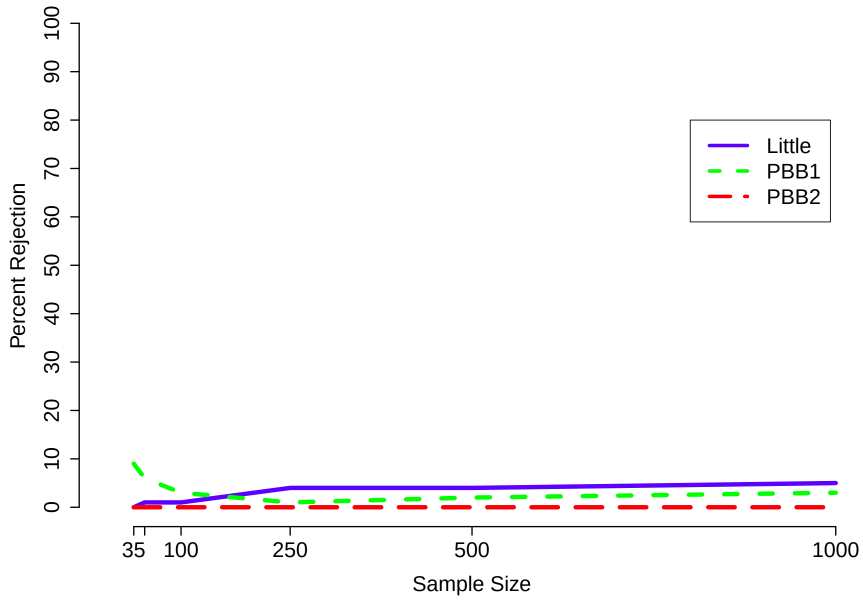


Figure 7.9: Full Simulation Study: MDM-1, Number of Variables = 5, Probability of Missingness = 0.2, No Correlation Among Variables

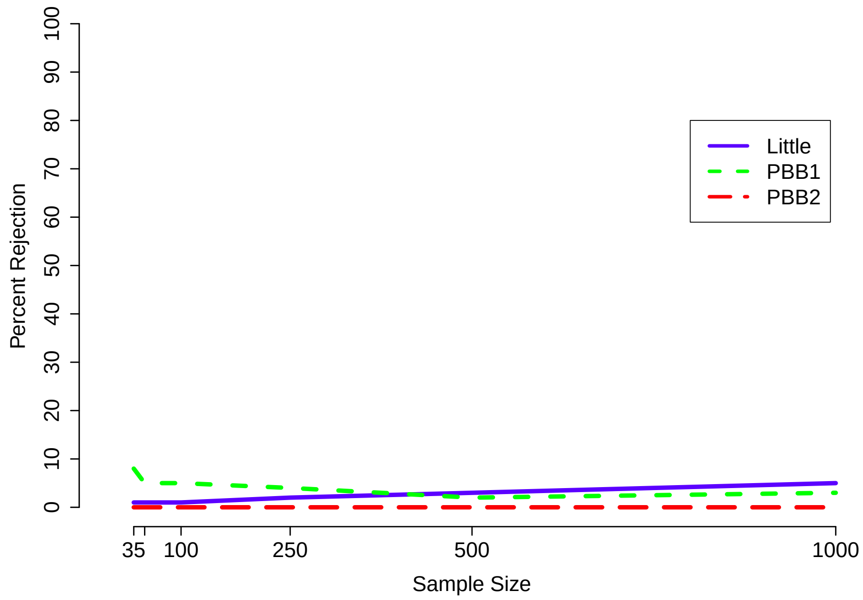


Figure 7.10: Full Simulation Study: MDM-1, Number of Variables = 5, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

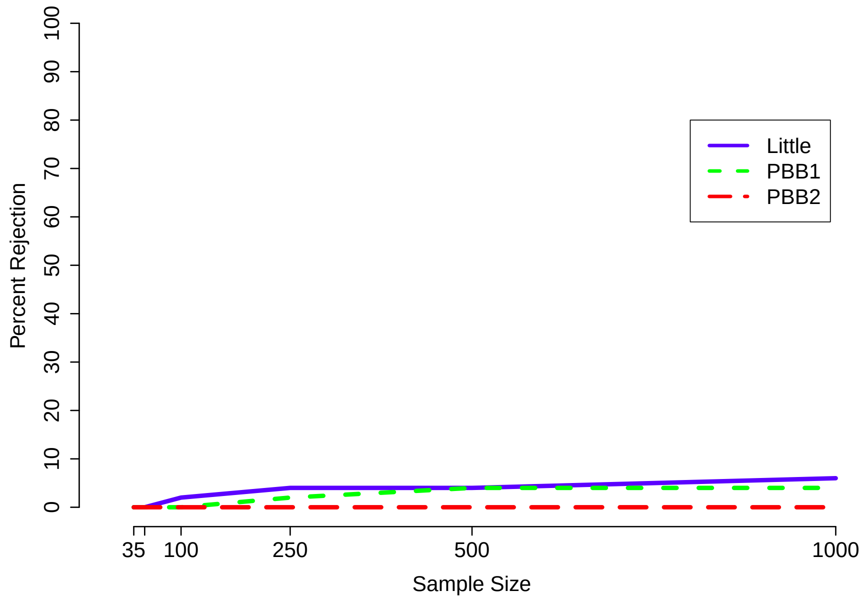


Figure 7.11: Full Simulation Study: MDM-1, Number of Variables = 5, Probability of Missingness = 0.6, No Correlation Among Variables

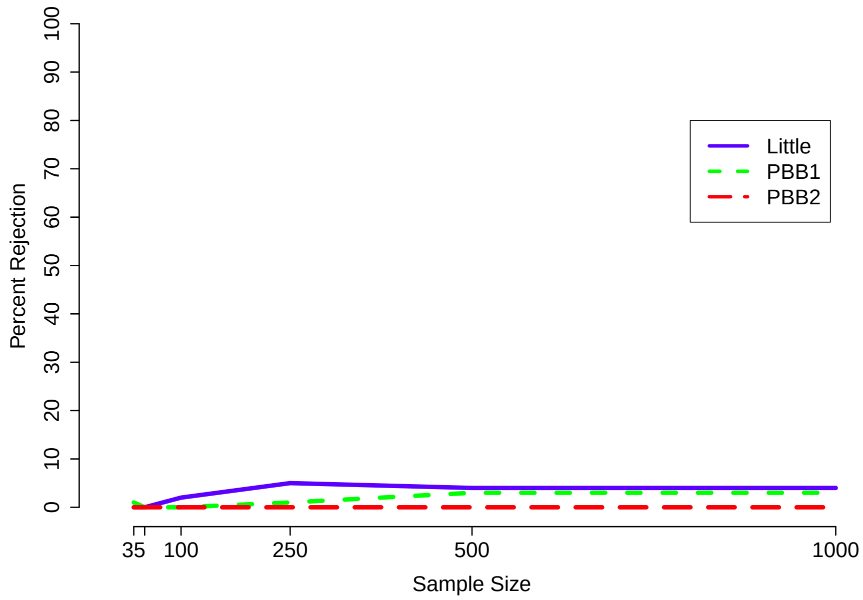


Figure 7.12: Full Simulation Study: MDM-1, Number of Variables = 5, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

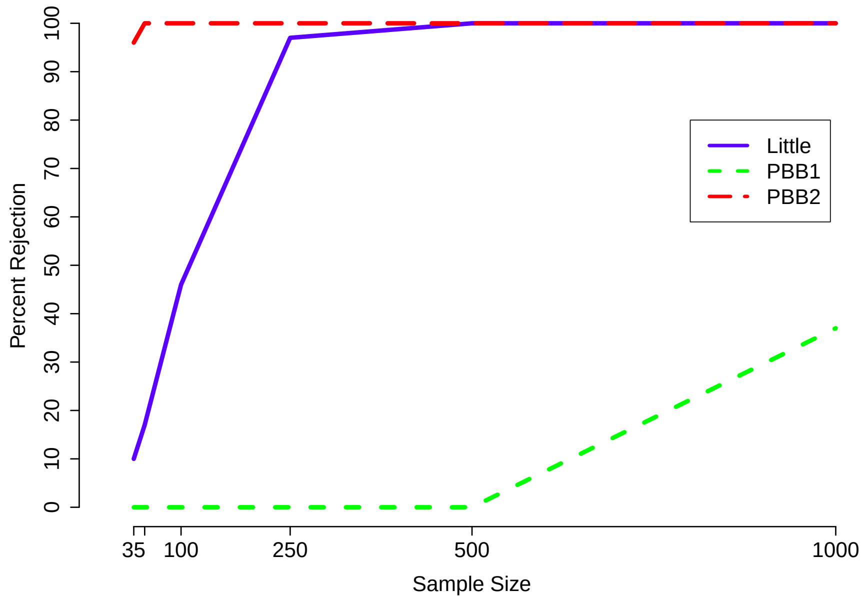


Figure 7.13: Full Simulation Study: MDM-2, Number of Variables = 2, Probability of Missingness = 0.2, No Correlation Among Variables

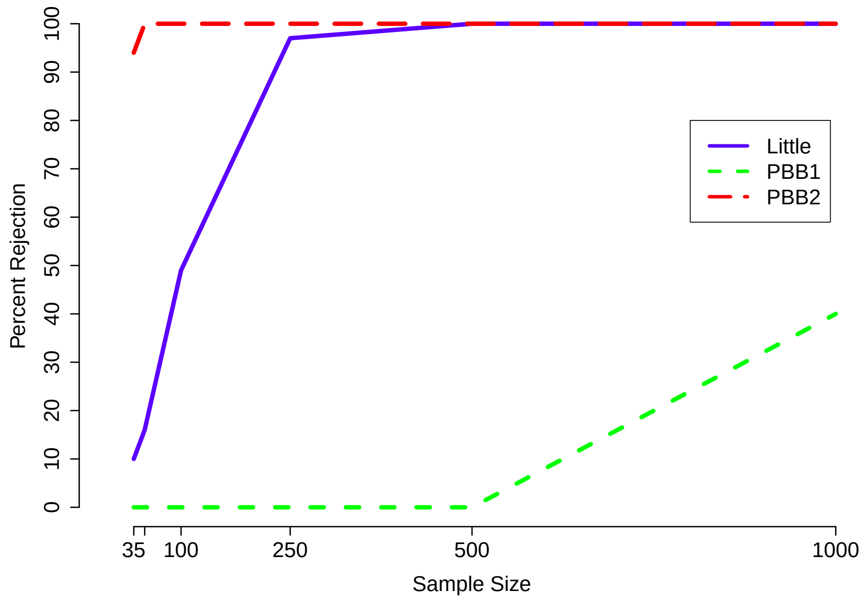


Figure 7.14: Full Simulation Study: MDM-2, Number of Variables = 2, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

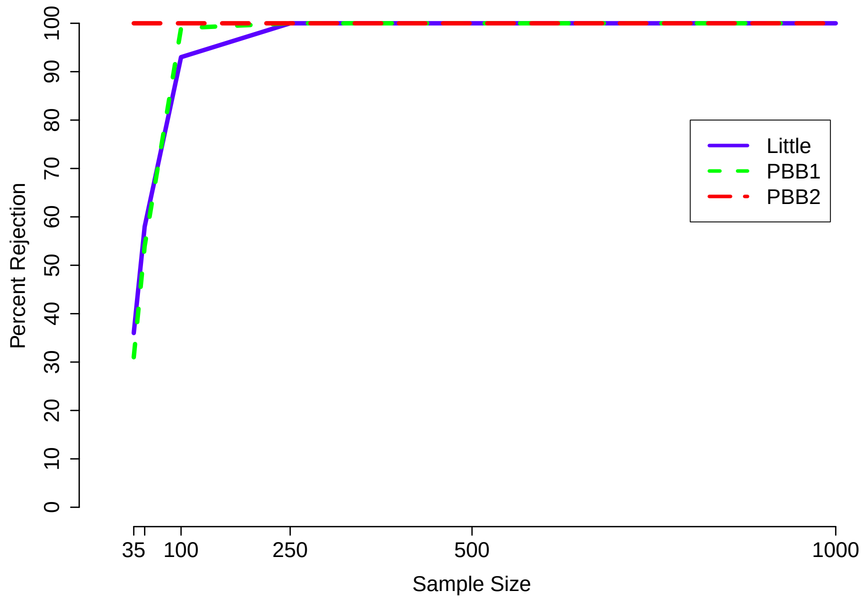


Figure 7.15: Full Simulation Study: MDM-2, Number of Variables = 2, Probability of Missingness = 0.6, No Correlation Among Variables

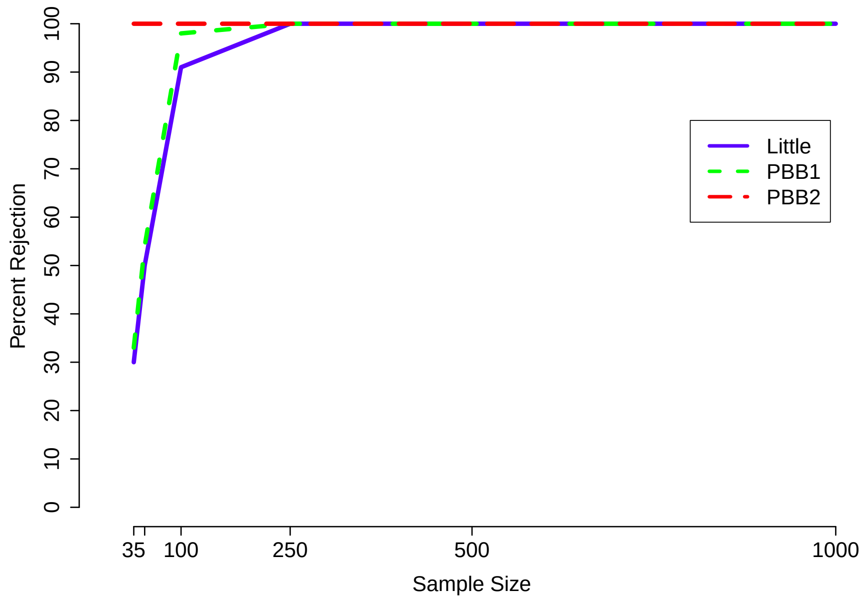


Figure 7.16: Full Simulation Study: MDM-2, Number of Variables = 2, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

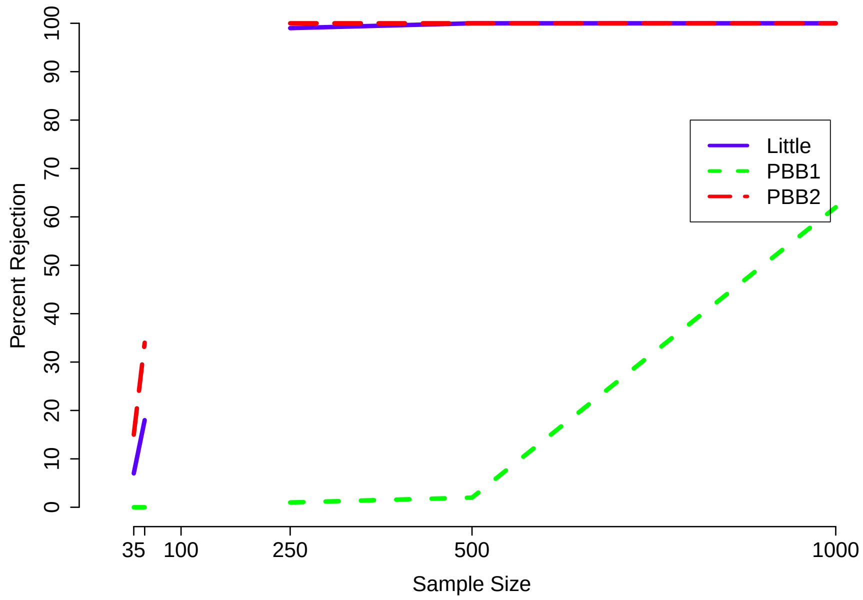


Figure 7.17: Full Simulation Study: MDM-2, Number of Variables = 3, Probability of Missingness = 0.2, No Correlation Among Variables

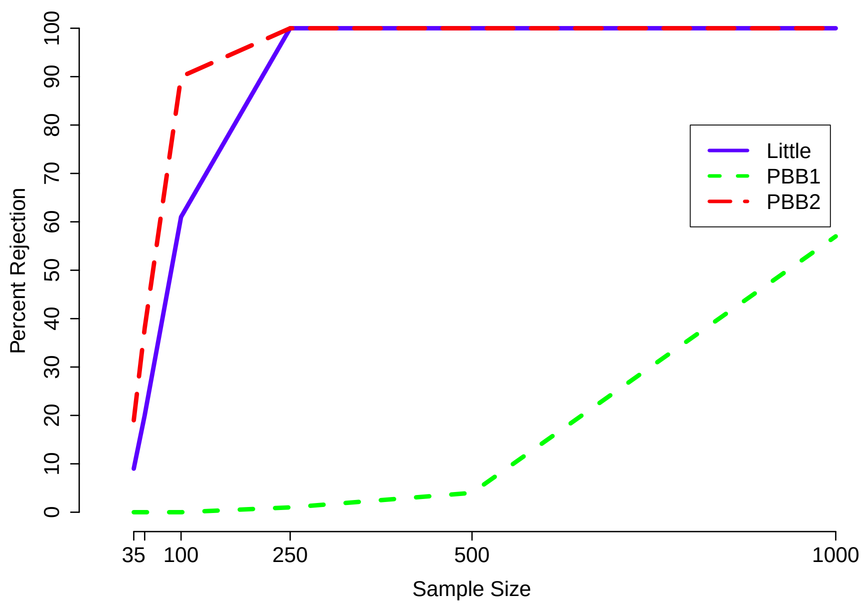


Figure 7.18: Full Simulation Study: MDM-2, Number of Variables = 3, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

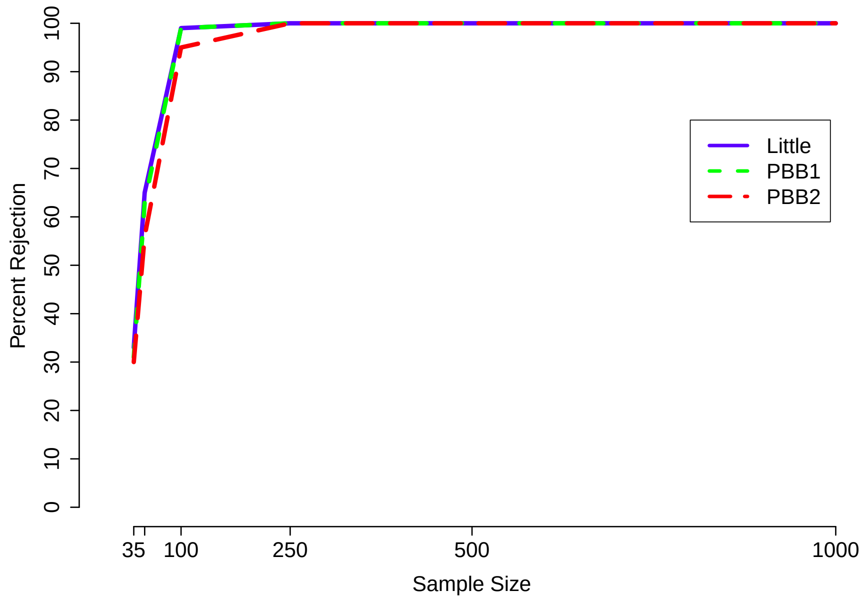


Figure 7.19: Full Simulation Study: MDM-2, Number of Variables = 3, Probability of Missingness = 0.6, No Correlation Among Variables

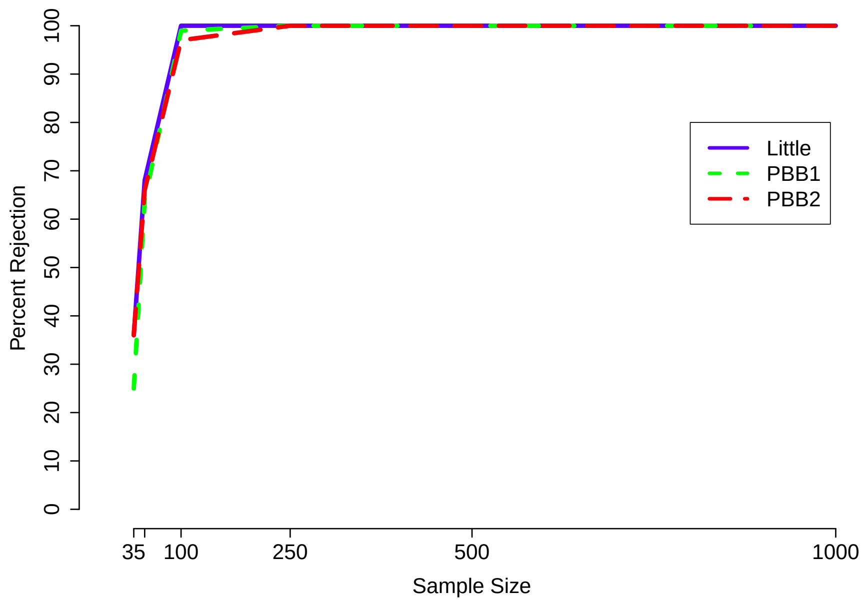


Figure 7.20: Full Simulation Study: MDM-2, Number of Variables = 3, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

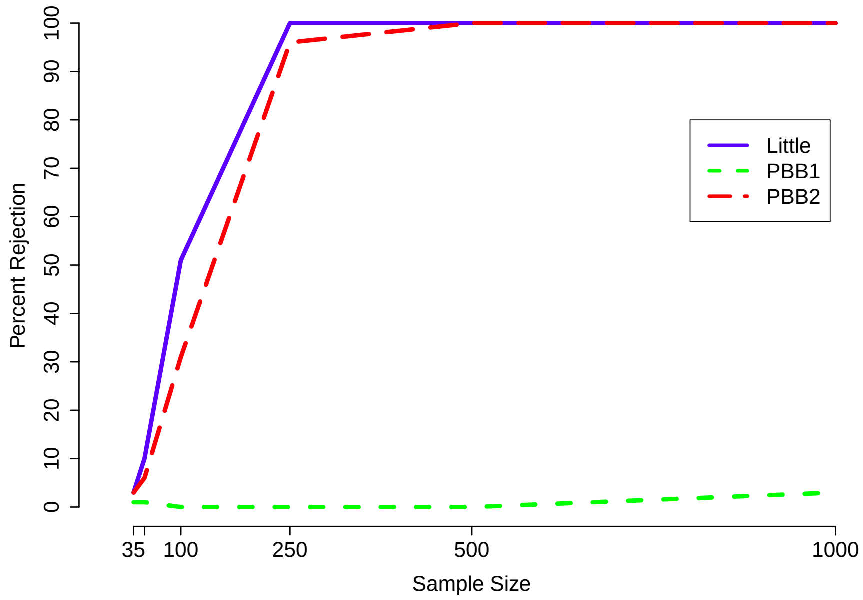


Figure 7.21: Full Simulation Study: MDM-2, Number of Variables = 5, Probability of Missingness = 0.2, No Correlation Among Variables

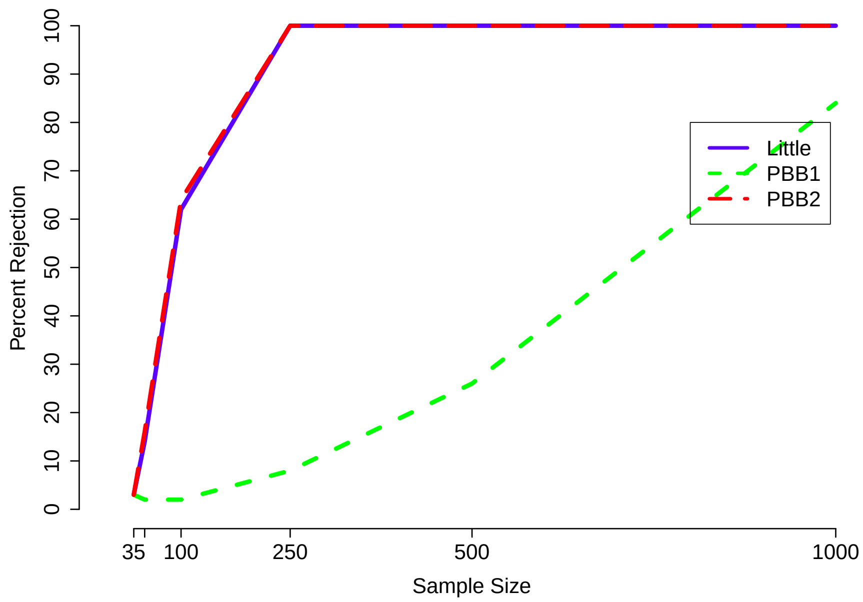


Figure 7.22: Full Simulation Study: MDM-2, Number of Variables = 5, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

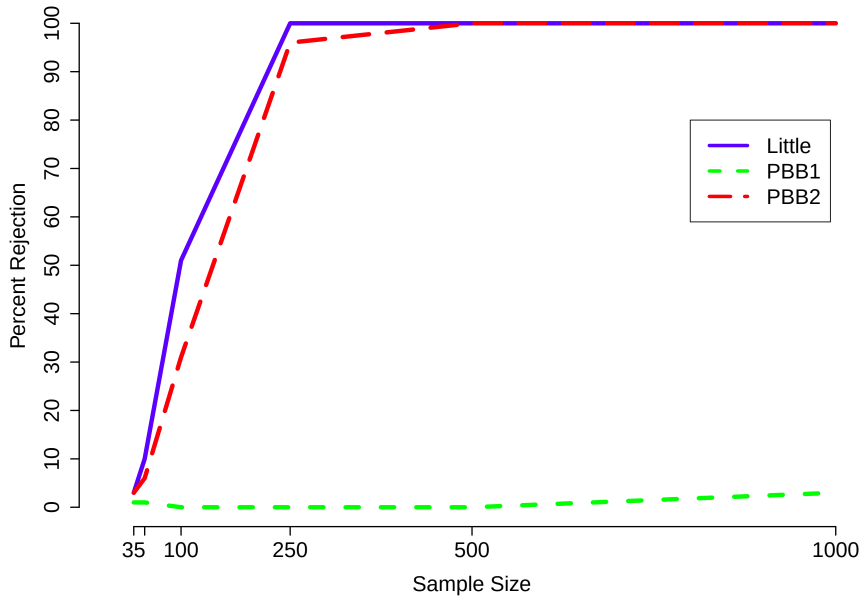


Figure 7.23: Full Simulation Study: MDM-2, Number of Variables = 5, Probability of Missingness = 0.2, No Correlation Among Variables

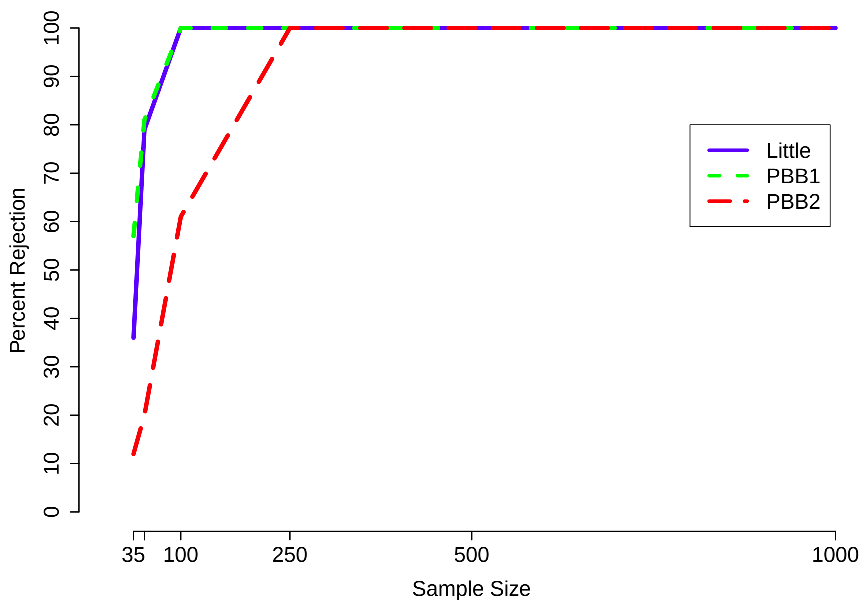


Figure 7.24: Full Simulation Study: MDM-2, Number of Variables = 5, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

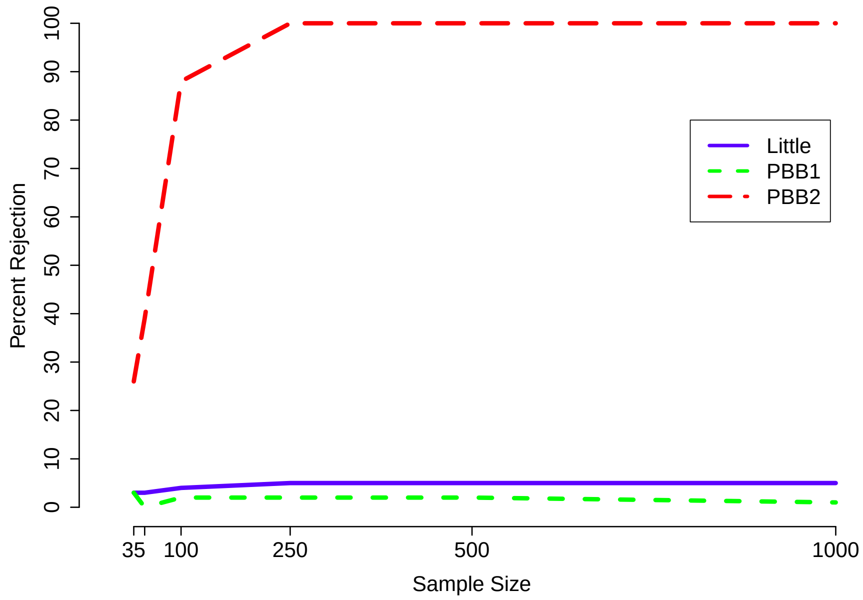


Figure 7.25: Full Simulation Study: MDM-3, Number of Variables = 2, Probability of Missingness = 0.2, No Correlation Among Variables

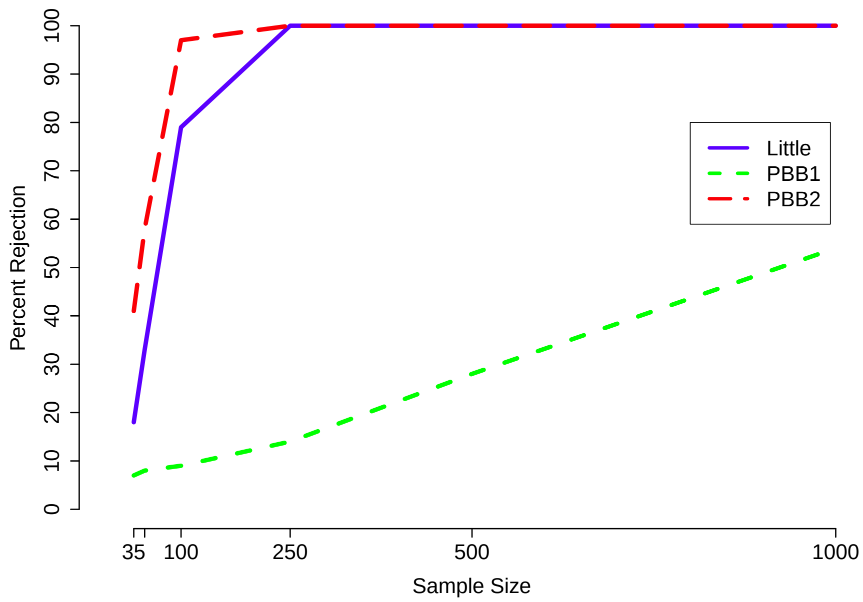


Figure 7.26: Full Simulation Study: MDM-3, Number of Variables = 2, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

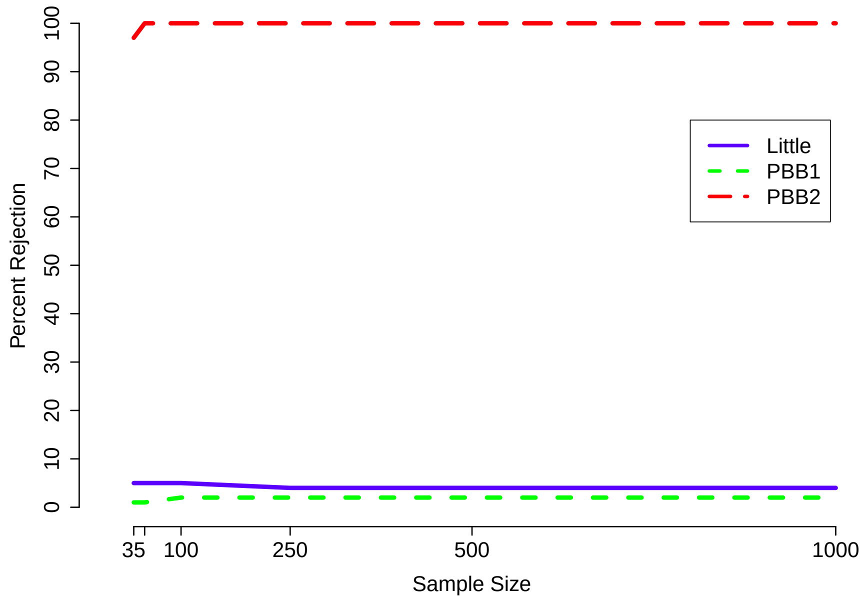


Figure 7.27: Full Simulation Study: MDM-3, Number of Variables = 2, Probability of Missingness = 0.6, No Correlation Among Variables

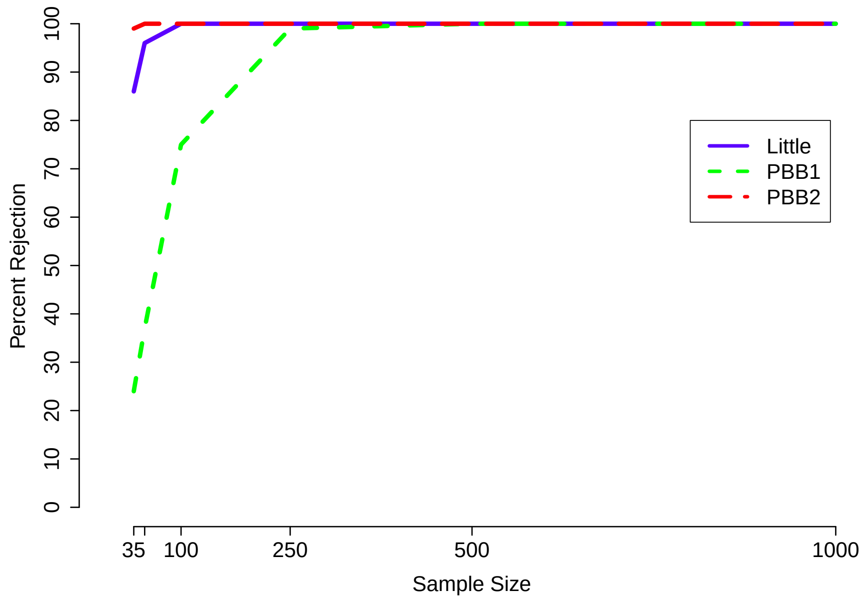


Figure 7.28: Full Simulation Study: MDM-3, Number of Variables = 2, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

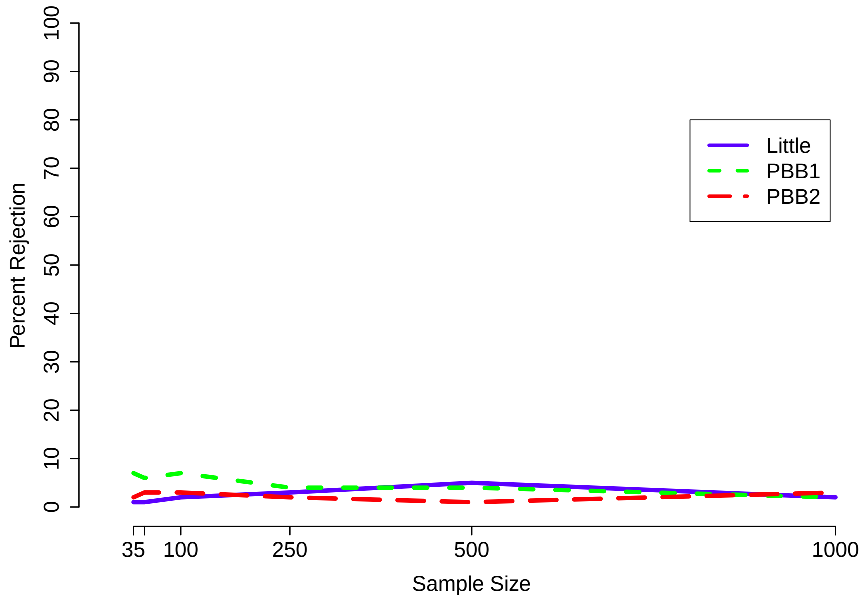


Figure 7.29: Full Simulation Study: MDM-3, Number of Variables = 3, Probability of Missingness = 0.2, No Correlation Among Variables

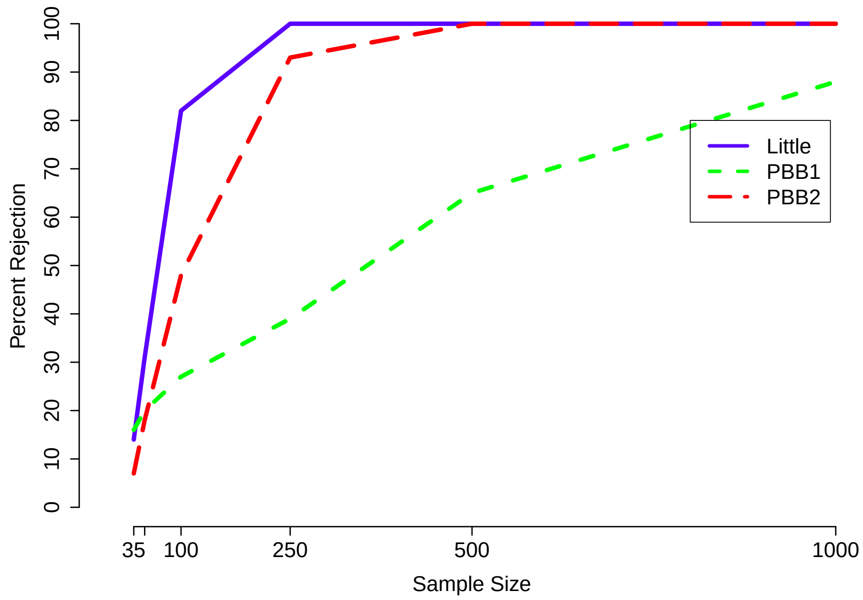


Figure 7.30: Full Simulation Study: MDM-3, Number of Variables = 3, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

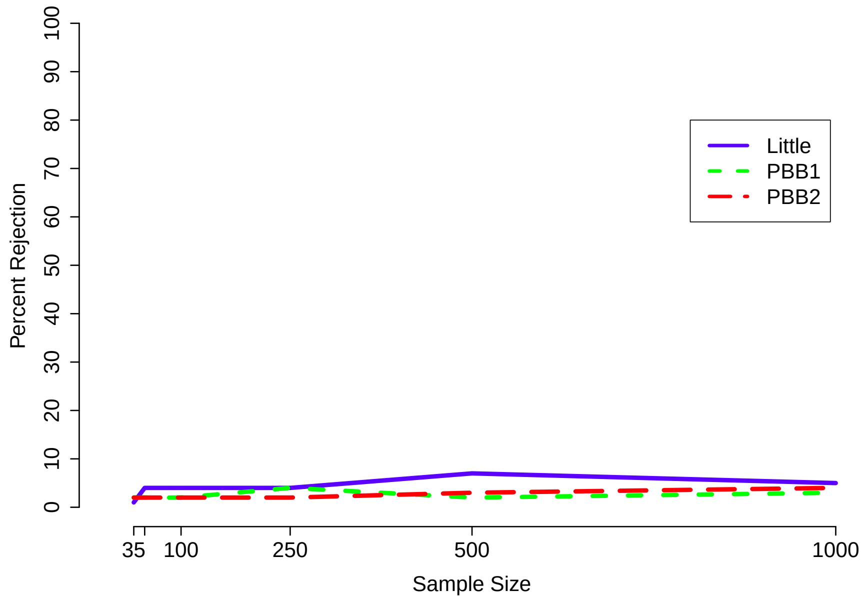


Figure 7.31: Full Simulation Study: MDM-3, Number of Variables = 3, Probability of Missingness = 0.6, No Correlation Among Variables

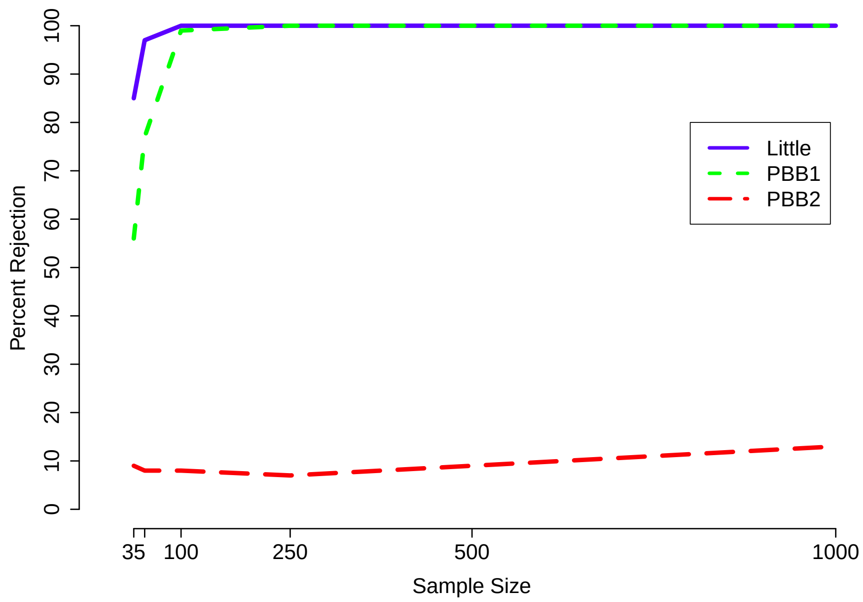


Figure 7.32: Full Simulation Study: MDM-3, Number of Variables = 3, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

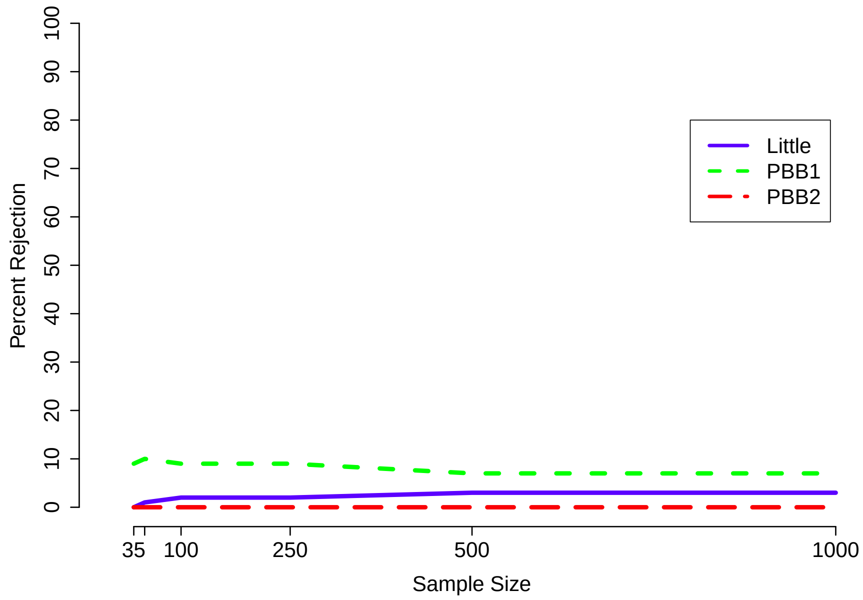


Figure 7.33: Full Simulation Study: MDM-3, Number of Variables = 5, Probability of Missingness = 0.2, No Correlation Among Variables

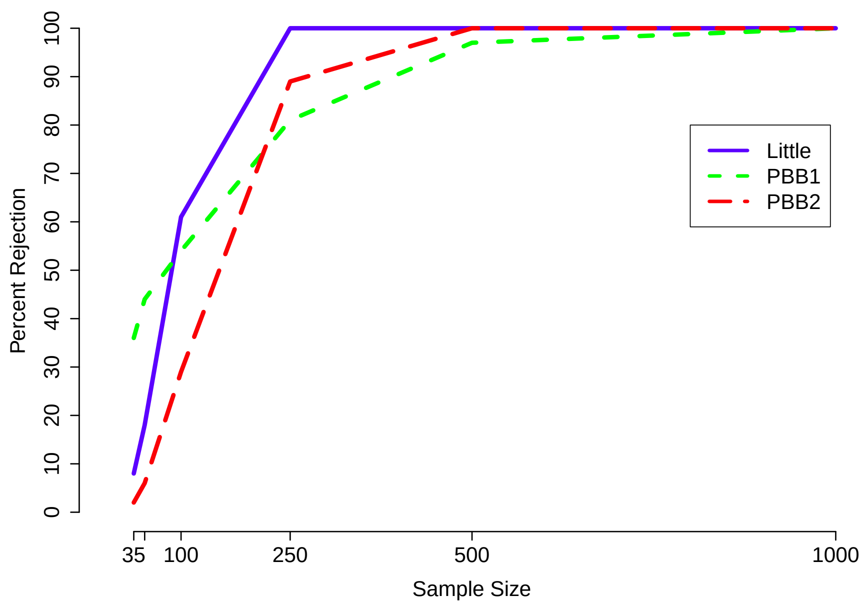


Figure 7.34: Full Simulation Study: MDM-3, Number of Variables = 5, Probability of Missingness = 0.2, Correlation Among Variables = 0.9

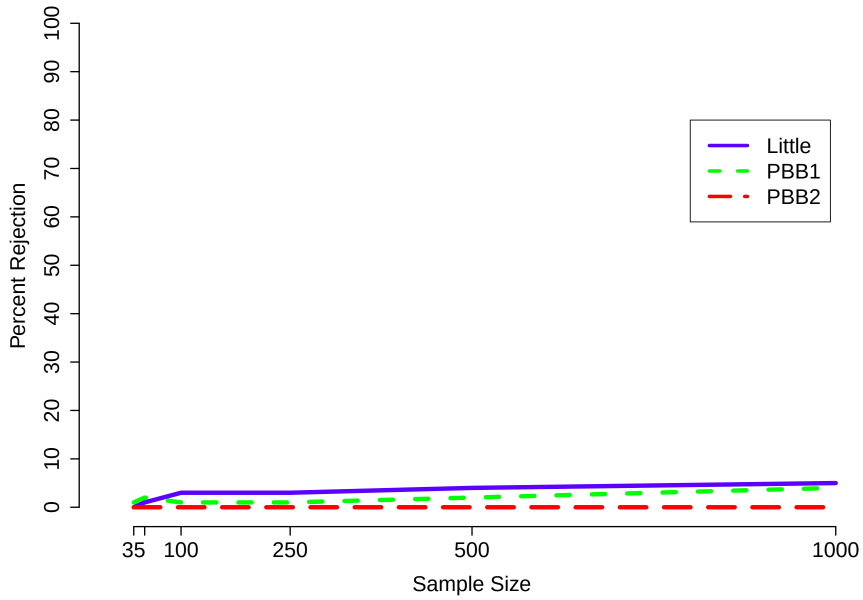


Figure 7.35: Full Simulation Study: MDM-3, Number of Variables = 5, Probability of Missingness = 0.6, No Correlation Among Variables

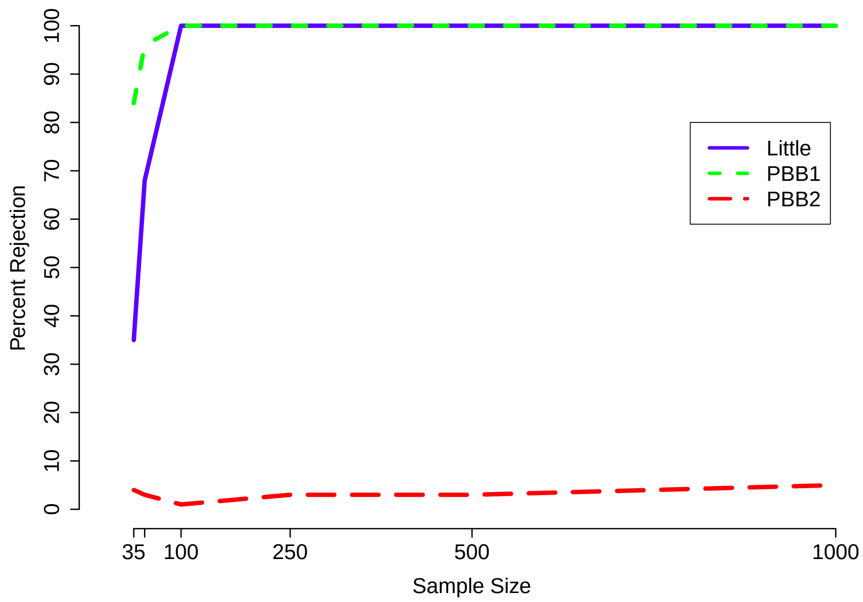


Figure 7.36: Full Simulation Study: MDM-3, Number of Variables = 5, Probability of Missingness = 0.6, Correlation Among Variables = 0.9

References

Allison, P. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545–557.

Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education policy analysis archives*, *18*(14), 1.

Anderson, T. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, *52*(278), 200–203.

Bentler, P. M. (2006). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bock, R., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, *12*(3), 261.

Braddock, J., et al. (1981). Race, athletics, and educational attainment: Dispelling the myths. *Youth and Society*, *12*(3), 335–50.

Braddock, J., et al. (1991). Bouncing back: Sports and academic resilience among African-American males. *Education and Urban Society*, *24*(1), 113–31.

Chen, H., & Little, R. (1999). A test of missing completely at random for generalised estimating equations with missing data. *Biometrika*, *86*(1), 1.

Crain, R. L. (1981). Making desegregation work: Extracurricular activities. *The Urban Review*, *13*(2), 121–127.

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 (Methodological)*, 1–38.

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(1), 128–141.

Enders, C. K. (2010). *Applied missing data analysis.* The Guilford Press.

Enders, C. K., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457.

Enders, C. K., et al. (2011). Missing not at random models for latent growth curve analyses. *Psychological methods*, *16*(1).

Greenlees, J., Reece, W., & Zieschang, K. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, *77*(378), 251–261.

Hanks, M. P., & Eckland, B. K. (1976). Athletics and social participation in the educational attainment process. *Sociology of Education*, 271–294.

IBM Corporation. (2011). *IBM SPSS Missing Values 20.* IBM Corporation.

Ingels, S., Pratt, D., Rogers, J., Siegel, P., & Stutts, E. (2004). Education longitudinal study of 2002: Base year data file user's manual. NCES 2004-405. *National Center for Education Statistics*.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, *118*(3), 843–877.

Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, 1–26.

Jamshidian, M., & Mata, M. (2008). Postmodeling sensitivity analysis to detect the effect of missing data mechanisms. *Multivariate Behavioral Research*, *43*(3), 432–452.

Jamshidian, M., & Schott, J. (2007). Testing equality of covariance matrices when data are incomplete. *Computational statistics & data analysis*, *51*(9), 4227–4239.

Jamshidian, M., Yuan, K., & Le, P. (in press). Using condence intervals to test for non-ignorable non-response.

Johnson, E. G. (1992). The design of the National Assessment of Educational Progress. *Journal of Educational Measurement*, *29*(2), 95–110.

Jordan, W. (1999). Black high school students' participation in school-sponsored sports activities: Effects on school engagement and achievement. *Journal of Negro Education*, 54–71.

Kim, K., & Bentler, P. (2002). Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika*, *67*(4), 609–623.

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 1198–1202.

Little, R., & Rubin, D. (2002). *Statistical analysis with missing data.* Wiley, New York.

Maydeu-Olivares, A., & Joe, H. (2005). Limited-and full-information estimation and goodness-of-fit testing in 2 n contingency tables. *Journal of the American Statistical Association*, *100*(471), 1009–1020.

McKnight, P. (2007). *Missing data: A gentle introduction.* The Guilford Press.

Melnick, M., Sabo, D., Vanfossen, B., et al. (1992). Educational effects of interscholastic athletic participation on African-American and Hispanic youth. *Adolescence*, *27*(106), 295.

Molenberghs, G., Beunckens, C., Sotto, C., & Kenward, M. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(2), 371–388.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431–462.

Park, T., & Davis, C. (1993). A test of the missing data mechanism for repeated categorical data. *Biometrics*, *49*(2), 631–638.

Peugh, J., & Enders, C. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525.

Rubin, D. (1976). Inference and missing data. *Biometrika*, *63*(3), 581.

Rubin, D. (2004). *Multiple imputation for nonresponse in surveys*. Wiley-IEEE.

Schafer, J., & Graham, J. (2002). Missing data: Our view of the state of the art. *Psychological methods*, *7*(2), 147–177.

Snyder, E., & Spreitzer, E. (1990). High school athletic participation as related to college attendance among Black, Hispanic, and White males: A research note. *Youth & Society*, *21*(3), 390–398.

Trent, W., & Braddock, J. (1992). Extracurricular activities in secondary schools. *Encyclopedia of Educational Research*, *2*, 476–480.

Trudeau, F., & Shephard, R. (2008). Physical education, school physical activity, school sports and academic performance. *International Journal of Behavioral Nutrition and Physical Activity*, *5*(1), 10.

Upton, J., Amos, D., & Ryman, A. (2011, May 3). For teachers, many ways and reasons to cheat on tests. Retrieved from `http://usatoday30.usatoday.com/news/education/2011-03-10-1aschooltesting10_cv_n.htm`