Inferences on missing information under multiple imputation and two-stage multiple imputation

Ofer Harel

Department of Statistics
University of Connecticut

Outline

✓ Background
✓ Methods
  ★ Multiple imputation (MI)
  ★ Two-stage MI
  ★ Asymptotic results
✓ Number of imputations
✓ Discussion
✓ Current/future work
Background

The missing data problem – Most statistical analysis and estimation procedures were not designed to handle missing values.

✓ Even a small amount of missing values can cause great difficulty.

✓ The missing-data aspect is nuisance, not of primary interest.

The Goal – To make statistically valid inferences about population parameters from an incomplete data set.

✓ Not to estimate, predict, or recover the missing values themselves.

✓ Untestable assumptions are inevitable.

✓ Sensitivity analysis is helpful.
The information concept was introduced by Fisher in the 30’s. The use of information, information matrices and other functions of the information concepts have become very common. For example

- The inverse of the covariance matrix.
- Information based tests such as AIC and BIC.

The information concept became important for both frequentists and bayesians.
Background

Information

✓ I am interested in using the information concept in an incomplete-data situation.

✓ Rubin introduced the rate of the missing information, after partitioning the complete information into observed and missing parts.

✓ Both Rubin and Schafer argue for the importance of this measure.

✓ I am going to develop this concept and show its importance.

✓ I am going to touch on possible extensions.

✓ How do we define the rates of missing information?
Why do we care?

✓ Can we use MI (or any other missing data procedure) if we have 1%, 5%, 10%, 25%, 50%, 75%, 90% missing values?

✓ what is the cut point?
Why do we care?

- Can we use MI (or any other missing data procedure) if we have 1%, 5%, 10%, 25%, 50%, 75%, 90% missing values?
- What is the cut point?
- How confident are we in the results from incomplete data?
- When using MI, how do we determine the number of imputations?
- Can we evaluate the effect of a missing value or a group of missing values?
- Can we evaluate our missing data model, and/or its effect on our inference?
Multiple imputation (Rubin, 1987): a simulation-based approach to missing data.

**Imputation:** Create $m$ imputations of the missing data, $Y_{mis}^{(1)}, \ldots, Y_{mis}^{(m)}$, under a suitable model.

**Analysis:** Analyze each of the $m$ completed data sets in the same way.

**Combination:** Combine the $m$ sets of estimates and variances using Rubin’s (1987) rules.
Multiple imputation

MI is separated into 3 stages: Imputation stage, Analysis stage, and Combining Results stage.

![Diagram of multiple imputation process]

- Observed Data
- Imputation
  1 2 ... m
  ?  ?  ?  ?
Multiple imputation

Imputation: Under an ignorability assumption we draw \( m \) sets of imputations from

\[
P(Y_{mis}|Y_{obs}),
\]

while under non-ignorable models we draw the imputations from

\[
P(Y_{mis}|Y_{obs}, M).
\]

Analysis: Given the complete data sets we can use any complete-data procedures to analyze the data.
Multiple imputation

Combining rules: Calculate and store the estimates \( \hat{Q}(j) \) and variances \( U(j) \) for \( j = 1, \ldots, m \) and combine:

\[
\bar{Q} = m^{-1} \sum_{j=1}^{m} \hat{Q}(j)
\]

\[
\bar{U} = m^{-1} \sum_{j=1}^{m} U(j)
\]

\[
B = (m - 1)^{-1} \sum_{j=1}^{m} \left( \hat{Q}(j) - \bar{Q} \right)^2
\]

\[
T = \bar{U} + (1 + m^{-1})B
\]
Which leads to an approximate 95% interval for $Q$

$$\bar{Q} \pm t_\nu \sqrt{T},$$

where the degrees of freedom are

$$\nu = (m - 1) \left[ \frac{(1 + m^{-1})B}{T} \right]^{-2}.$$ 

The rate of missing information is

$$\lambda = \frac{\bar{U}^{-1} - (\nu + 1)(\nu + 3)^{-1}T^{-1}}{\bar{U}^{-1}}.$$ 

This is approximately $\lambda = \frac{B}{U+B}$ when $\nu$ is large.
Rates of missing information

✓ The rate of missing information is the proportion of missing information to the complete information.

✓ The rate of missing information governs the convergence rates of the EM algorithm and the data augmentation.

✓ The lower the rate of missing information, the faster the convergence will be.

✓ The rate of missing information is important in assessing how missing information contributes to inferential uncertainty about the population quantity of interest.
Rates of missing information

✓ The naive estimate of the rate of missing information is the rate of missing values.

✓ The naive estimate will be accurate in only very limited situations.

✓ The rate of missing information is a function not only of the amount of missing values but also of their importance.

✓ When using MI with a small number of imputations, the estimate of the rates of missing information can be very noisy.
Two-types of missing values

In many situations, missing values may be of **two qualitatively different types**. Examples include:

- planned missingness versus unplanned.
- unit non-response versus item non-response.
- refusal versus “don’t know”.
- latent variables versus missing manifest items.
- deaths versus other types of dropout.
- dropouts versus subjects who return.
- dropout for reasons that might be closely related to outcomes, versus dropout for other reasons.
Two-stage multiple imputation

Extends Rubin’s (1987) framework to two types of missing values.

**Imputation:** Create $m$ imputations of $Y_{mis}^A$ the first level of the missing data, $Y_{mis}^{A(1)}, \ldots, Y_{mis}^{A(m)}$. Then, for each $Y_{mis}^A$, generate $n$ imputations of $Y_{mis}^B$ from a conditional predictive distribution given $Y_{mis}^A$.

**Analysis:** Analyze each of the $mn$ completed data sets in the same way.

**Combination:** Combine the $mn$ sets of estimates and variances by Shen’s (2000) rules.
Two-stage MI
Shen rules

From Shen (2000, unpublished Ph.D. thesis) calculate and store:

\[ \hat{Q}(j,k) = \text{estimate of } Q \quad U(j,k) = \text{standard error}^2 \]

for \( j = 1, \ldots, m \) and \( k = 1, \ldots, n \). Then:

\[
\bar{Q}_{..} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} \hat{Q}(j,k) \\
\bar{Q}_j = \frac{1}{n} \sum_{k=1}^{n} \hat{Q}(j,k)
\]

\[
\bar{U} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{k=1}^{n} U(j,k)
\]

\[
B = \frac{1}{m-1} \sum_{j=1}^{m} \left( \bar{Q}_j - \bar{Q}_{..} \right)^2
\]

\[
W = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n-1} \sum_{k=1}^{n} \left( \hat{Q}(j,k) - \bar{Q}_j \right)^2
\]
Shen rules - Cont.

The total variance is

\[ T = \bar{U}_. + (1 + \frac{1}{m})B + (1 - \frac{1}{n})W. \]

An approximate 95% interval for \( Q \) is

\[ \bar{Q} \pm t_\nu \sqrt{T}, \]

where the degrees of freedom are

\[ \nu^{-1} = \frac{1}{m(n-1)} \left( \frac{(1 - \frac{1}{n})W}{T} \right)^{-2} + \frac{1}{(m-1)} \left( \frac{(1 + \frac{1}{m})B}{T} \right)^{-2}. \]
The overall estimated rate of missing information is

\[ \hat{\lambda} = \frac{B + (1 - n^{-1})W}{\bar{U} + B + (1 - n^{-1})W}. \]

The estimated rate of missing information due to \( Y_{mis}^B \) if \( Y_{mis}^A \) was known is

\[ \hat{\lambda}^{B|A} = \frac{W}{\bar{U} + W}, \]

and the difference \( \hat{\lambda}^A = \hat{\lambda} - \hat{\lambda}^{B|A} \) represents the decrease in the rate of missing information if \( Y_{mis}^A \) becomes known.
Rates of missing information

✓ The overall estimated rate of missing information in two-stage MI is equivalent to the rate of missing information in conventional MI.

✓ The partition of the rates of missing information is based on the partition of the missing values.

✓ There are many ways to partition the missing data.

✓ Similar to conventional MI, a small number of imputations may lead to a noisy estimate of the rates of missing information.
Asymptotic Results - Conventional MI

✓ The random imputations, $Y_{mis}^{(j)} \sim P(Y_{mis}|Y_{obs})$, induces a probability distribution to $B$.

✓ Rubin argues that $\frac{(m-1)B}{B_{\infty}} \sim \chi^2_{m-1}$, where $B_{\infty} = V(\hat{Q}(Y_{obs}, Y_{mis})|Y_{obs})$.

✓ Assuming that the above is reasonable, the Central Limit Theorem indicates that

$$\frac{\sqrt{m/2(B - B_{\infty})}}{B_{\infty}} \rightarrow N(0, 1)$$

as $m \rightarrow \infty$, where $E(B) = B_{\infty}$ and $V(B) = 2B_{\infty}^2/(m - 1)$.
Because \( \hat{\lambda} = B/(B+U) \), we can use the delta method to derive

\[
\frac{\sqrt{m}(\hat{\lambda} - \lambda)}{\sqrt{\frac{2B^2 \tilde{U}^2}{(B+U)^4}}} \to N(0, 1),
\]

which may also be written as

\[
\frac{\sqrt{m}(\hat{\lambda} - \lambda)}{\sqrt{2\lambda(1 - \lambda)}} \to N(0, 1).
\]

Therefore, an approximate 95% confidence interval for the population rate of missing information is

\[
\hat{\lambda} \pm \frac{1.96\Hat{\lambda}(1 - \hat{\lambda})}{\sqrt{m/2}}.
\]
The rate of missing information is a proportion.

In situations where the rate of missing information is close to zero or one, it may be useful to apply a normal approximation on the logit scale,

\[ \text{Logit}(p) = \log\left[\frac{p}{1 - p}\right]; \] this result in

\[ \sqrt{m}(\text{Logit}(\hat{\lambda}) - \text{Logit}(\lambda)) \rightarrow N(0, 2). \]

The logit transformation is a first order variance-stabilizing transformation for \( \hat{\lambda} \)
Asymptotic Results - two-stage MI

✓ In two-stage MI, the sources of the variance can be separated into a between and within nest.

\[
\hat{Q}_{jk} - \bar{Q}_{..} = (\bar{Q}_j - \bar{Q}_{..}) + (\hat{Q}_{jk} - \bar{Q}_j).
\]

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>(m - 1)</td>
<td>(n \sum_{j=1}^{m} (\bar{Q}<em>j - \bar{Q}</em>{..})^2)</td>
<td>nB</td>
<td>(\sigma_W^2 + n\sigma_B^2)</td>
</tr>
<tr>
<td>Within</td>
<td>(m(n - 1))</td>
<td>(\sum_{j=1}^{m} \sum_{k=1}^{n} (\hat{Q}_{jk} - \bar{Q}_j)^2)</td>
<td>W</td>
<td>(\sigma_W^2)</td>
</tr>
</tbody>
</table>
Asymptotic Results - two-stage MI

From the ANOVA literature it is known that the mean square errors are independent chi-squares random variables such that,

\[
\left( \frac{\sigma_w^2 + n\sigma_B^2}{m - 1} \right)^{-1} nB \sim \chi^2_{m-1} \\
\left( \frac{\sigma_w^2}{m(n - 1)} \right)^{-1} W \sim \chi^2_{m(n-1)}.
\]
A first-order approximation to the joint distribution of $B$ and $W$ is

$$\sqrt{m} \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}^{-1} \left( \begin{pmatrix} B \\ W \end{pmatrix} - E \begin{pmatrix} B \\ W \end{pmatrix} \right) \sim N(0, I),$$

where $V_1 = 2(\sigma_W^2 + n\sigma_B^2)^2/n^2$ and $V_2 = 2(\sigma_W^2)^2/(n - 1)$.

Using a multivariate delta method expansion, similar to the procedure we have done for conventional MI we get
Asymptotic Results - two-stage MI

\[ \sqrt{m} \Sigma_1^{-1} \left( \begin{pmatrix} \lambda \\ \lambda_{B|A} \end{pmatrix} - \begin{pmatrix} \hat{\lambda} \\ \hat{\lambda}_{B|A} \end{pmatrix} \right) \sim N(0, I), \]

where

\[ \Sigma_1 = \begin{pmatrix} \frac{V_1 U^2 + V_2 U^2(1-n^{-1})}{(U+B+(1-n^{-1})W)^4} & \frac{V_2 U^2(1-n^{-1})}{(U+B+(1-n^{-1})W)^2(U+W)^2} \\ \frac{V_2 U^2(1-n^{-1})}{(U+B+(1-n^{-1})W)^2(U+W)^2} & \frac{V_2 U^2}{(U+W)^4} \end{pmatrix} \].
Asymptotic Results - two-stage MI

Rewriting the covariance matrix $\Sigma_1$ as a function of $\lambda$ and $\lambda^{B|A}$ gives

$$
\Sigma_1 = \begin{pmatrix}
2(1 - \lambda)^4 & \frac{2(1-\lambda)^2(\lambda^{B|A})^2}{n} \\
\frac{2(1-\lambda)^2(\lambda^{B|A})^2}{n} & \frac{2(1-\lambda)^2(\lambda^{B|A})^2}{(n-1)}
\end{pmatrix}.
$$

When the rates of missing information are close to zero or one, we can calculate an approximation on a logit scale

$$
\sqrt{m\Sigma_2^{-1}} \left( \begin{pmatrix}
\text{Logit}(\lambda) \\
\text{Logit}(\lambda^{B|A})
\end{pmatrix} - \begin{pmatrix}
\text{Logit}(\hat{\lambda}) \\
\text{Logit}(\lambda^{B|A})
\end{pmatrix} \right) \sim N(0, I),
$$

where

$$
\Sigma_2 = \begin{pmatrix}
\frac{2(1-\lambda)^2}{\lambda^2} & \frac{2(1-\lambda)^2}{\lambda^2} \\
\frac{2(1-\lambda)(\lambda^{B|A})^2}{\lambda(1-\lambda^{B|A})n} & \frac{2(1-\lambda)(\lambda^{B|A})^2}{\lambda(1-\lambda^{B|A})n}
\end{pmatrix}.
$$
Asymptotic Results - simulations

✓ Simulation studies were designed to evaluate the theoretical distributions.

✓ Simulation studies were performed for both conventional MI and two-stage MI.

✓ Different levels of rates of missing information and number of imputations were simulated for conventional MI and the estimates, variances and coverage were tested.

✓ For two-stage MI, the rates of missing information and the number of imputations in each level were simulated, and again the estimates variances (covariances) and coverage were tested.

✓ The simulated and theoretical results were very close.
Simulations results

Comparison of asymptotic and simulated behavior of the estimated rate of missing information. Plot of $SE(\hat{\lambda}) \times 100$
Simulations results

Comparison of asymptotic and simulated behavior of the estimated rate of missing information. Plot of $SE(\hat{\lambda}) \times 100$
Rubin (1987) demonstrated that a limited number of imputations (5-20) is enough.

This is true since the relative efficiency on the variance scale of a point estimate based on \( m \) imputations, to one based on an infinite number of imputations, is \( (1 + \lambda/m)^{-1} \).

<table>
<thead>
<tr>
<th>( m \backslash \lambda )</th>
<th>0.1</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91</td>
<td>80</td>
<td>67</td>
<td>57</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>97</td>
<td>92</td>
<td>86</td>
<td>80</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>98</td>
<td>95</td>
<td>91</td>
<td>87</td>
<td>85</td>
</tr>
<tr>
<td>10</td>
<td>99</td>
<td>98</td>
<td>95</td>
<td>93</td>
<td>92</td>
</tr>
<tr>
<td>( \infty )</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
✓ From the standpoint of point-estimation alone, there is little to gain from using more than a few imputations unless the rates of missing information are unusually large.

✓ When we are interested in the rates of missing information, estimates of these rates can be very unstable (Schafer, 1997).

✓ I will determine the number of imputations necessary for obtaining reliable estimates for the missing information rate.
Table 1: Recommendations for choosing $m$ in Conventional MI to estimate $\lambda$ with given accuracy

<table>
<thead>
<tr>
<th>$\hat{\lambda}$</th>
<th>0.10</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>0.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SE(\hat{\lambda}) = 0.03$</td>
<td>19</td>
<td>79</td>
<td>139</td>
<td>79</td>
<td>18</td>
</tr>
<tr>
<td>$SE(\hat{\lambda}) = 0.02$</td>
<td>41</td>
<td>176</td>
<td>313</td>
<td>176</td>
<td>41</td>
</tr>
<tr>
<td>$SE(\hat{\lambda}) = 0.01$</td>
<td>163</td>
<td>704</td>
<td>1250</td>
<td>704</td>
<td>162</td>
</tr>
</tbody>
</table>
For two-stage MI, one can invoke similar arguments for the relative efficiency

\[(1 + \lambda^*/m)^{-1},\]

where

\[\lambda^* = \lambda - \frac{\lambda^B|A(1 - \lambda)(1 - n^{-1})}{(1 - \lambda^B|A)}.
\]

Because 0 ≤ \(\lambda^B|A\) ≤ \(\lambda\), the relative efficiency lies between \((1 + \lambda/m)^{-1}\) and \((1 + \lambda/(mn))^{-1}\).
Recommendations for choosing $m$ and $n$ in two-stage MI where 
$SE(\hat{\lambda}^{B|A}) \leq SE(\hat{\lambda}) \leq 0.03$

| $\hat{\lambda}^{B|A}$ | $n$  | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 |
|-----------------------|------|------|------|------|------|------|
| 0.05                  | 2    | 11   | 66   | 132  | 77   | 18   |
|                       | 5    | 7    | 60   | 128  | 76   | 18   |
| 0.10                  | 2    | 5    | 55   | 124  | 76   | 18   |
|                       | 5    | 5    | 42   | 116  | 74   | 18   |
| 0.25                  | 2    | 20   | 97   | 70   | 17   |
|                       | 5    | 4    | 75   | 65   | 17   |
| 0.50                  | 2    | 35   | 55   | 16   |
|                       | 5    | 6    | 42   | 15   |
| 0.75                  | 2    | 20   | 13   |
|                       | 5    | 4    | 10   |
| 0.90                  | 2    | 5    |      |
Recommendations for choosing $m$ and $n$ in two-stage MI where $SE(\hat{\lambda}^B|A) \leq SE(\hat{\lambda}) \leq 0.02$

| $\hat{\lambda}^B|A$ | $\lambda$ | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 |
|---------------------|-----------|------|------|------|------|------|
| 0.05                | 2         | 23   | 149  | 296  | 173  | 40   |
|                     | 5         | 16   | 134  | 287  | 171  | 40   |
| 0.10                | 2         | 10   | 122  | 279  | 170  | 40   |
|                     | 5         | 5    | 95   | 260  | 166  | 40   |
| 0.25                | 2         | 44   | 217  | 157  | 39   |
|                     | 5         | 7    | 168  | 146  | 39   |
| 0.50                | 2         | 79   | 122  | 36   |
|                     | 5         | 13   | 95   | 34   |
| 0.75                | 2         | 43   | 29   |
|                     | 5         | 7    | 22   |
| 0.90                | 2         |      |      |      |      |
Recommendations for choosing $m$ and $n$ in two-stage MI where $SE(\hat{\lambda}^B|A) \leq SE(\hat{\lambda}) \leq 0.01$

| $\hat{\lambda}^B|A$ | n  | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 |
|---------------------|----|------|------|------|------|------|
| 0.05                | 2  | 95   | 597  | 1185 | 691  | 161  |
|                     | 5  | 63   | 537  | 1147 | 684  | 161  |
| 0.10                | 2  | 41   | 489  | 1115 | 678  | 160  |
|                     | 5  | 7    | 378  | 1038 | 662  | 159  |
| 0.25                | 2  | 174  | 868  | 627  | 156  |
|                     | 5  | 28   | 673  | 583  | 153  |
| 0.50                | 2  | 313  | 489  | 145  |
|                     | 5  | 50   | 378  | 135  |
| 0.75                | 2  |      |      | 176  | 113  |
|                     | 5  |      |      | 28   | 87   |
| 0.90                | 2  |      |      |      | 41   |
Discussion

✓ When the researcher’s main interest is point estimates (and their variances), it is sufficient to use a modest number of imputations.

✓ When the rates of missing information are a concern, more imputations are recommended.

✓ Simulations have been made to test the accuracy of the asymptotic distribution.

✓ In most cases, producing more imputations is not a problem.

✓ The rates of missing information are an important model diagnostic tool.
Extensions

✓ There are numerous measures that assess the effect of an observation, group of observations, a variable, or variables and observations on the regression estimation (Outlier, influence point etc.).

✓ Can we assess the effect of a missing observation, a group of missing observations, an incomplete variable or any combination of these, on the overall estimation?

✓ We want this measure to be applicable under any missing data mechanism.

✓ And applicable to any data analysis problem.
Extensions

✓ I call this measure "Outfluence".

✓ The Outfluence is a relative measure comparing the effect of a missing value relative to the effect of the rest of the missing values.

✓ For example, consider a survey study in which we are interested in the effect of missing income on the estimation. The outfluence will measure the effect of the missing income compared to all other missing values.

✓ We already defined the measure (Harel, 2008) and introduce its limiting distribution (Harel & Stratton, 2009).

✓ More work needs to be devoted to the development of methods that will evaluate incomplete-data methodology.
Consider the next setup:

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Marijuana Example

✓ A pilot study presented by Weil et al (1968) and analyzed by Schafer (1997).

✓ Three treatments (low dose, high dose and placebo) were given to each subject.

✓ The change in heart beat was recorded.

✓ We will look at the mean of placebo in 90 minute since Schafer (1997) showed missing data have a large effect on this estimate.
## Example – Data

<table>
<thead>
<tr>
<th>Subject</th>
<th>Placebo</th>
<th>Low</th>
<th>High</th>
<th>Placebo</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>20</td>
<td>16</td>
<td>20</td>
<td>-6</td>
<td>-4</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>24</td>
<td>12</td>
<td>-6</td>
<td>-4</td>
<td>-8</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>8</td>
<td>26</td>
<td>-4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>8</td>
<td>NA</td>
<td>NA</td>
<td>20</td>
<td>-4</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>4</td>
<td>-8</td>
<td>NA</td>
<td>22</td>
<td>-8</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>20</td>
<td>28</td>
<td>-20</td>
<td>-4</td>
<td>-4</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>28</td>
<td>24</td>
<td>12</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>-8</td>
<td>20</td>
<td>24</td>
<td>-3</td>
<td>8</td>
<td>-24</td>
</tr>
<tr>
<td>9</td>
<td>NA</td>
<td>20</td>
<td>24</td>
<td>8</td>
<td>12</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Means</strong></td>
<td><strong>8.8</strong></td>
<td><strong>16.9</strong></td>
<td><strong>18.2</strong></td>
<td><strong>1.0</strong></td>
<td><strong>7.6</strong></td>
<td><strong>-3.2</strong></td>
</tr>
</tbody>
</table>

Example – Results

<table>
<thead>
<tr>
<th>Missing values</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\lambda}^{B,A}$</th>
<th>$\hat{\phi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4,+)</td>
<td>0.103</td>
<td>0.013</td>
<td>0.126</td>
</tr>
<tr>
<td>(4,4)</td>
<td>0.102</td>
<td>0.007</td>
<td>0.069</td>
</tr>
<tr>
<td>(5,4)</td>
<td>0.101</td>
<td>0.009</td>
<td>0.089</td>
</tr>
<tr>
<td>(+,4)</td>
<td>0.108</td>
<td>0.103</td>
<td>0.954</td>
</tr>
<tr>
<td>(9,+)</td>
<td>0.101</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Extensions and Future work

- Model diagnosis is an importance concept in statistics.
- The rates of missing information can be used as a diagnostic tool (e.g. Harel & Miglioretti 2007).
- Many measures were developed to compare the performance of competing models in complete data.
- To date there is almost no research about the performance of methodology with incomplete data.
- How can we use measures such as AIC, BIC, DIC to evaluate our missing data modeling?
References

✓ Harel, O. (2007) "Inferences on missing information under multiple imputation and two-stage multiple imputation" *Statistical Methodology*, 4, 75-89.


