Revisiting the Available Cases Method for Missing Values

Xiao (Max) Gu and Norm Matloff
University of California at Davis

JSM 2015
Taxonomy of Methods
Taxonomy of Methods

Major current methods:

- Use only complete cases (CC).
- Multiple imputation (MI).
- MLE.

Forgotten method:
- Available cases (AC). Use partially-intact cases when possible.
Taxonomy of Methods

Major current methods:

- Use only complete cases (CC).
- Multiple imputation (MI).
- MLE.
Taxonomy of Methods

Major current methods:

- Use only complete cases (CC).
- Multiple imputation (MI).
- MLE.

Forgotten method:
Taxonomy of Methods

Major current methods:

- Use only complete cases (CC).
- Multiple imputation (MI).
- MLE.

Forgotten method:

- Available cases (AC). Use partially-intact cases when possible.
Overview of AC Method
Overview of AC Method

E.g. linear regression (random-X).
Overview of AC Method

E.g. linear regression (random-X).

\[
\hat{\beta} = (X'X)^{-1}X'Y = \left[\frac{1}{n}(X'X)^{-1}\right] \left[\frac{1}{n}X'Y\right] = A^{-1}D \quad (1)
\]
Overview of AC Method

E.g. linear regression (random-X).

\[ \hat{\beta} = (X'X)^{-1}X'Y = \left[ \frac{1}{n}(X'X)^{-1} \right] \left[ \frac{1}{n}X'Y \right] = A^{-1}D \] (1)

A estimates quantities like
Overview of AC Method

E.g. linear regression (random-X).

\[ \hat{\beta} = (X'X)^{-1}X'Y = \left[ \frac{1}{n}(X'X)^{-1} \right] \left[ \frac{1}{n}X'Y \right] = A^{-1}D \]  \hspace{1cm} (1)

A estimates quantities like

\[ E[X^{(i)}X^{(j)}] \]  \hspace{1cm} (2)
Overview of AC Method

E.g. linear regression (random-X).

\[ \hat{\beta} = (X'X)^{-1}X'Y = \left[ \frac{1}{n}(X'X)^{-1} \right] \left[ \frac{1}{n}X'Y \right] = A^{-1}D \]  \hspace{1cm} (1)

\( A \) estimates quantities like

\[ E[X^{(i)}X^{(j)}] \]  \hspace{1cm} (2)

while \( D \) estimates quantities like
Overview of AC Method

E.g. linear regression (random-X).

\[ \hat{\beta} = (X'X)^{-1}X'Y = \left[ \frac{1}{n}(X'X)^{-1} \right] \left[ \frac{1}{n}X'Y \right] = A^{-1}D \quad (1) \]

A estimates quantities like

\[ E[X^{(i)}X^{(j)}] \quad (2) \]

while \( D \) estimates quantities like

\[ E[X^{(i)}Y] \quad (3) \]
AC Overview, cont’d.
AC Overview, cont’d.

CC seems wasteful.
AC Overview, cont’d.

CC seems wasteful.

• In estimating, say, $E[X^{(2)}Y]$, why throw out cases in which $X^{(2)}$ and $Y$ are intact but $X^{(5)}$ is missing?
AC Overview, cont’d.

**CC seems wasteful.**

- In estimating, say, $E[X^{(2)}Y]$, why throw out cases in which $X^{(2)}$ and $Y$ are intact but $X^{(5)}$ is missing?
- Instead, estimate by $E[X^{(i)}Y]$ by

$$
\frac{1}{M} \sum_{X^{(i)}, \ Y \ \text{intact}} X^{(i)} Y_k
$$

where $M = \# \text{ of cases with both } X^{(i)} \text{ and } Y \ \text{intact}$.  (4)
CC seems wasteful.

- In estimating, say, \( E[X^{(2)} Y] \), why throw out cases in which \( X^{(2)} \) and \( Y \) are intact but \( X^{(5)} \) is missing?
- Instead, estimate by \( E[X^{(i)} Y] \) by

\[
\frac{1}{M} \sum_{X^{(i)}, Y \text{ intact}} X^{(i)}_k Y_k
\] (4)

where \( M = \# \) of cases with both \( X^{(i)} \) and \( Y \) intact.
- Same for the quantities \( E[X^{(i)} X^{(j)}] \).
AC Sounds Good, But Not Popular

Lack of positive definiteness is unlikely to occur, and it's unclear whether it's important anyway.
AC Sounds Good, But Not Popular

- AC should be more accurate than CC — uses more data.
AC Sounds Good, But Not Popular

- AC should be more accurate than CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature,
AC Sounds Good, But Not Popular

- AC should be more accurate than CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
  - AC assumes MCAR, the strongest among the famous assumption sets.
  - Still, AC seems worth revisiting.
  - Lack of positive definiteness is unlikely to occur, and it's unclear whether it's important anyway.
  - The most common alternative assumption set, MAR, is also quite strong.
  - More on this later.
AC Sounds Good, But Not Popular

- AC should be more accurate than CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
AC Sounds Good, But Not Popular

- AC should be more accurate that CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
  - AC assumes MCAR, the strongest among the famous assumption sets.
AC Sounds Good, But Not Popular

- AC should be more accurate than CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
  - AC assumes MCAR, the strongest among the famous assumption sets.
- Still, AC seems worth revisiting.
AC Sounds Good, But Not Popular

- AC should be more accurate than CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
  - AC assumes MCAR, the strongest among the famous assumption sets.
- Still, AC seems worth revisiting.
  - Lack of positive definiteness is unlikely to occur, and it’s unclear whether it’s important anyway.
AC Sounds Good, But Not Popular

- AC should be more accurate that CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
  - AC assumes MCAR, the strongest among the famous assumption sets.
- Still, AC seems worth revisiting.
  - Lack of positive definiteness is unlikely to occur, and it’s unclear whether it’s important anyway.
  - The most common alternative assumption set, MAR, is also quite strong.
**AC Sounds Good, But Not Popular**

- AC should be more accurate than CC — uses more data.
- Yet, AC seems to have been dismissed early on in the Missing Value literature, apparently because:
  - The modified $X'X$ may not be positive definite.
  - AC assumes MCAR, the strongest among the famous assumption sets.
- Still, AC seems worth revisiting.
  - Lack of positive definiteness is unlikely to occur, and it’s unclear whether it’s important anyway.
  - The most common alternative assumption set, MAR, is also quite strong. (More on this later.)
Our Study: AC vs. CC, MI
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC,
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression,
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression, and 2 new ones: PCA and log-linear model.
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression, and 2 new ones: PCA and log-linear model.
- We look at these criteria:
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression, and 2 new ones: PCA and log-linear model.
- We look at these criteria:
  - Applicability.
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression, and 2 new ones: PCA and log-linear model.
- We look at these criteria:
  - Applicability.
  - Variance, bias.
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression, and 2 new ones: PCA and log-linear model.
- We look at these criteria:
  - Applicability.
  - Variance, bias.
  - Run time.
Our Study: AC vs. CC, MI

- Here we “reopen the case” regarding AC, comparing to CC and MI.
- We look at the old application, linear regression, and 2 new ones: PCA and log-linear model.
- We look at these criteria:
  - Applicability.
  - Variance, bias.
  - Run time.
- For MI, we use Amelia 2.
Linear Regression

Revisiting the Available Cases Method for Missing Values

Xiao (Max) Gu and Norm Matloff
University of California at Davis

Line 1: All 3 methods are applicable.

Line 2: Simulation results: n = 10000, p = 3, 10% missing, β₁ = 1

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Variance</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9996</td>
<td>0.0002</td>
<td>0.79</td>
</tr>
<tr>
<td>MI</td>
<td>0.9784</td>
<td>0.0002</td>
<td>142.02</td>
</tr>
<tr>
<td>AC</td>
<td>1.0027</td>
<td>0.0010</td>
<td>23.80</td>
</tr>
</tbody>
</table>

Note: Most time in AC spent in finding numeric derivs for standard errors.

Line 3: MI slightly biased.

Line 4: AC terrible MSE. (Some intuition....)

Line 5: MI terrible run time.

Verdict: Use CC.
Linear Regression

- All 3 methods are applicable.
Linear Regression

- All 3 methods are applicable.
- Simulation results: $n = 10000$, $p = 3$, 10% missing, $\beta_1 = 1$
Linear Regression

- All 3 methods are applicable.
- Simulation results: $n = 10000$, $p = 3$, 10% missing, $\beta_1 = 1$

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9996</td>
<td>0.0002</td>
<td>0.79</td>
</tr>
<tr>
<td>MI</td>
<td>0.9784</td>
<td>0.0002</td>
<td>142.02</td>
</tr>
<tr>
<td>AC</td>
<td>1.0027</td>
<td>0.0010</td>
<td>23.80</td>
</tr>
</tbody>
</table>

Note: Most time in AC spent in finding numeric derivs for standard errors.
Revisiting the Available Cases Method for Missing Values

Xiao (Max) Gu and Norm Matloff
University of California at Davis

Linear Regression

- All 3 methods are applicable.
- Simulation results: $n = 10000$, $p = 3$, 10% missing, $\beta_1 = 1$

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9996</td>
<td>0.0002</td>
<td>0.79</td>
</tr>
<tr>
<td>MI</td>
<td>0.9784</td>
<td>0.0002</td>
<td>142.02</td>
</tr>
<tr>
<td>AC</td>
<td>1.0027</td>
<td>0.0010</td>
<td>23.80</td>
</tr>
</tbody>
</table>

Note: Most time in AC spent in finding numeric derivs for standard errors.

- MI slightly biased.
Linear Regression

- All 3 methods are applicable.
- Simulation results: $n = 10000$, $p = 3$, 10% missing, $\beta_1 = 1$

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9996</td>
<td>0.0002</td>
<td>0.79</td>
</tr>
<tr>
<td>MI</td>
<td>0.9784</td>
<td>0.0002</td>
<td>142.02</td>
</tr>
<tr>
<td>AC</td>
<td>1.0027</td>
<td>0.0010</td>
<td>23.80</td>
</tr>
</tbody>
</table>

Note: Most time in AC spent in finding numeric derivs for standard errors.

- MI slightly biased.
- AC terrible MSE. (Some intuition....)
Linear Regression

- All 3 methods are applicable.
- Simulation results: \( n = 10000, p = 3, 10\% \) missing, \( \beta_1 = 1 \)

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9996</td>
<td>0.0002</td>
<td>0.79</td>
</tr>
<tr>
<td>MI</td>
<td>0.9784</td>
<td>0.0002</td>
<td>142.02</td>
</tr>
<tr>
<td>AC</td>
<td>1.0027</td>
<td>0.0010</td>
<td>23.80</td>
</tr>
</tbody>
</table>

Note: Most time in AC spent in finding numeric derivs for standard errors.
- MI slightly biased.
- AC terrible MSE. (Some intuition....)
- MI terrible run time.
Linear Regression

- All 3 methods are applicable.
- Simulation results: \( n = 10000, p = 3, 10\% \) missing, \( \beta_1 = 1 \)

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.9996</td>
<td>0.0002</td>
<td>0.79</td>
</tr>
<tr>
<td>MI</td>
<td>0.9784</td>
<td>0.0002</td>
<td>142.02</td>
</tr>
<tr>
<td>AC</td>
<td>1.0027</td>
<td>0.0010</td>
<td>23.80</td>
</tr>
</tbody>
</table>

Note: Most time in AC spent in finding numeric derivs for standard errors.

- MI slightly biased.
- AC terrible MSE. (Some intuition....)
- MI terrible run time.
- Verdict: Use CC.
PCA
• CC, AC methods applicable.
PCA

- CC, AC methods applicable.
- MI sometimes gave error message ("perfectly collinear...").
PCA

- CC, AC methods applicable.
- MI sometimes gave error message (“perfectly collinear…”).
- Simulation results: $n = 100$, $p = 10$, 10% missing; largest eigenvalue; $\rho$ matrix
• CC, AC methods applicable.
• MI sometimes gave error message ("perfectly collinear...").
• Simulation results: \( n = 100, p = 10, 10\% \) missing; largest eigenvalue; \( \rho \) matrix

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>2.3328</td>
<td>0.0517</td>
</tr>
<tr>
<td>AC</td>
<td>2.1012</td>
<td>0.0218</td>
</tr>
</tbody>
</table>
• CC, AC methods applicable.
• MI sometimes gave error message ("perfectly collinear...").
• Simulation results: $n = 100$, $p = 10$, 10% missing; largest eigenvalue; $\rho$ matrix

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>2.3328</td>
<td>0.0517</td>
</tr>
<tr>
<td>AC</td>
<td>2.1012</td>
<td>0.0218</td>
</tr>
</tbody>
</table>
A Note on PCA

• PCA is upward biased anyway (even with no NAs), since PCA naturally overfits.

• The means of 2.1 and 2.3 we got for $n = 100$ become about 1.97 for $n = 1000$.

• But in all simulation runs, AC was less upward biased, and had small variance, compared to CC. This was severe for larger values of $p$. 
A Note on PCA

• PCA is upward biased anyway (even with no NAs), since PCA naturally overfits.
A Note on PCA

- PCA is upward biased anyway (even with no NAs), since PCA naturally overfits. (First comp. maxes var. of lin. combs. of length 1.)
A Note on PCA

- PCA is upward biased anyway (even with no NAs), since PCA naturally overfits. (First comp. maxes var. of lin. combs. of length 1.)
- The means of 2.1 and 2.3 we got for $n = 100$ become about 1.97 for $n = 1000$. 
A Note on PCA

- PCA is upward biased anyway (even with no NAs), since PCA naturally overfits. (First comp. maxes var. of lin. combs. of length 1.)
- The means of 2.1 and 2.3 we got for $n = 100$ become about 1.97 for $n = 1000$.
- But in all simulation runs, AC was less upward biased, and had small variance, compared to CC.
A Note on PCA

- PCA is upward biased anyway (even with no NAs), since PCA naturally overfits. (First comp. maxes var. of lin. combs. of length 1.)
- The means of 2.1 and 2.3 we got for $n = 100$ become about 1.97 for $n = 1000$.
- But in all simulation runs, AC was less upward biased, and had small variance, compared to CC. This was severe for larger values of $p$. 
Contingency Table Models

MI not appropriate, since assumes MV normal data. (Though MI methods do exist for this setting.)

Example: Factors $X, Y, Z$; (12)(13) model — $Y$ and $Z$ independent, given $X$.

In terms of marginal distributions:

$$p_{ijk} = p_{i.} p_{.j} p_{..k}$$

E.g. set $\hat{p}_{i..k}$ to the proportion of cases in which $X = i$, $Z = k$, among cases in which $X$ and $Z$ are intact.

Simulation example: (1)(23) model, $n = 100$, est. $p_{111}$.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.1246591</td>
<td>0.0009020450</td>
</tr>
<tr>
<td>AC</td>
<td>0.1249168</td>
<td>0.0007548656</td>
</tr>
</tbody>
</table>

AC advantage more if have more factors or higher NA %.
Contingency Table Models

- MI not appropriate, since assumes MV normal data.
Contingency Table Models

- MI not appropriate, since assumes MV normal data. (Though MI methods do exist for this setting.)
Contingency Table Models

- MI not appropriate, since assumes MV normal data. (Though MI methods do exist for this setting.)
- Example: Factors $X$, $Y$, $Z$;
Contingency Table Models

- MI not appropriate, since assumes MV normal data. (Though MI methods do exist for this setting.)
- Example: Factors $X, Y, Z$; (12)(13) model — $Y$ and $Z$ independent, given $X$. 

---

Simulation example: (1)(23) model, $n = 100$, est. $p_{111}$ method mean var

<table>
<thead>
<tr>
<th>CC</th>
<th>AC</th>
<th>AC advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1246591</td>
<td>0.1249168</td>
<td>higher NA %</td>
</tr>
</tbody>
</table>

E.g. set $\hat{p}_{i.k}$ to the proportion of cases in which $X = i, Z = k$, among cases in which $X$ and $Z$ are intact.
Contingency Table Models

- MI not appropriate, since assumes MV normal data. (Though MI methods do exist for this setting.)
- Example: Factors $X, Y, Z$; (12)(13) model — $Y$ and $Z$ independent, given $X$.
- In terms of marginal distributions:

$$p_{ijk} = p_{i..} \frac{p_{i.j} p_{i..}}{p_{i..} p_{i..}} = \frac{p_{i.j} p_{i..}}{p_{i..}}$$

(5)

- E.g. set $\hat{p}_{i..k}$ to the proportion of cases in which $X = i$, $Z = k$, among cases in which $X$ and $Z$ are intact.
- Simulation example: (1)(23) model, $n = 100$, est. $p_{111}$.

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>var</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.1246591</td>
<td>0.0009020450</td>
</tr>
<tr>
<td>AC</td>
<td>0.1249168</td>
<td>0.0007548656</td>
</tr>
</tbody>
</table>
Contingency Table Models

- MI not appropriate, since assumes MV normal data. (Though MI methods do exist for this setting.)
- Example: Factors $X$, $Y$, $Z$; (12)(13) model — $Y$ and $Z$ independent, given $X$.
- In terms of marginal distributions:

$$\frac{p_{ijk}}{p_{i..}p_{i..}} = \frac{p_{i.j}p_{i..}}{p_{i..}p_{i..}}$$ (5)

- E.g. set $\hat{p}_{i..}$ to the proportion of cases in which $X = i$, $Z = k$, among cases in which $X$ and $Z$ are intact.
- Simulation example: (1)(23) model, $n = 100$, est. $p_{111}$.

<table>
<thead>
<tr>
<th>method</th>
<th>mean</th>
<th>var</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.1246591</td>
<td>0.0009020450</td>
</tr>
<tr>
<td>AC</td>
<td>0.1249168</td>
<td>0.0007548656</td>
</tr>
</tbody>
</table>

AC advantage more if have more factors or higher NA %.
On Assumptions

CC, AC assume MCAR, stronger than MI's MAR.

Arguably, \( \text{MAR} \cap \text{MCAR} \) rare in practice.

\( \hat{\beta} \) still unbiased for \( \beta \) under CC, AC even under \( \text{MAR} \cap \text{MCAR} \).

In \( \text{MAR} \cap \text{MCAR} \) case, bias does arise if use CC or AC to estimate \( E_Y \) or \( E_X(\text{i}) \).

In such case, use Matloff, Biometrika, 1982.
On Assumptions

- CC, AC assume MCAR, stronger than MI’s MAR.
On Assumptions

- CC, AC assume MCAR, stronger than MI’s MAR.
- However:
On Assumptions

- CC, AC assume MCAR, stronger than MI’s MAR.
- However:
  - Arguably, \( \text{MAR} \cap \text{MCAR}^c \) rare in practice.
On Assumptions

- CC, AC assume MCAR, stronger than MI’s MAR.
- However:
  - Arguably, \( \text{MAR} \cap \text{MCAR}^c \) rare in practice.
  - \( \hat{\beta} \) still unbiased for \( \beta \) under CC, AC even under \( \text{MAR} \cap \text{MCAR}^c \).
On Assumptions

- CC, AC assume MCAR, stronger than MI’s MAR.
- However:
  - Arguably, $\text{MAR} \cap \text{MCAR}^c$ rare in practice.
  - $\hat{\beta}$ still unbiased for $\beta$ under CC, AC even under $\text{MAR} \cap \text{MCAR}^c$.
  - In $\text{MAR} \cap \text{MCAR}^c$ case, bias does arise if use CC or AC to estimate $EY$ or $EX^{(i)}$. 

In such case, use Matloff, *Biometrika*, 1982.
On Assumptions

- CC, AC assume MCAR, stronger than MI’s MAR.
- However:
  - Arguably, $\text{MAR} \cap \text{MCAR}^c$ rare in practice.
  - $\hat{\beta}$ still unbiased for $\beta$ under CC, AC even under $\text{MAR} \cap \text{MCAR}^c$.
  - In $\text{MAR} \cap \text{MCAR}^c$ case, bias does arise if use CC or AC to estimate $EY$ or $EX^{(i)}$. In such case, use Matloff, *Biometrika*, 1982.
Software
Revisiting the Available Cases Method for Missing Values

Xiao (Max) Gu and Norm Matloff
University of California at Davis

Software

- Code available at
  [https://github.com/maxguxiao/Available-Cases.git](https://github.com/maxguxiao/Available-Cases.git).
  Currently under development; check current status.
Software

- R’s `cov()`, `cor()` functions include the option `use = 'pairwise.complete.obs'`, which is the AC method.
Revisiting the Available Cases Method for Missing Values

Xiao (Max) Gu and Norm Matloff
University of California at Davis

Software

- Code available at https://github.com/maxguxiao/Available-Cases.git. Currently under development; check current status.
- R’s `cov()`, `cor()` functions include the option `use = 'pairwise.complete.obs'`, which is the AC method. This could be used to implement AC in two applications:
Software

- Code available at https://github.com/maxguxiao/Available-Cases.git. Currently under development; check current status.
- R’s `cov()`, `cor()` functions include the option `use = 'pairwise.complete.obs'`, which is the AC method. This could be used to implement AC in two applications:
  - For PCA, just run `eigen()` on either a covariance or correlation matrix computed for AC as above.
Software

- Code available at https://github.com/maxguxiao/Available-Cases.git. Currently under development; check current status.
- R’s `cov()`, `cor()` functions include the option `use = 'pairwise.complete.obs'`, which is the AC method. This could be used to implement AC in two applications:
  - For PCA, just run `eigen()` on either a covariance or correlation matrix computed for AC as above.
  - For linear regression, the matrices $A$ and $D$ both can be computed using `cov()`, after adjusting via a centering operation.
Software

- R’s `cov()`, `cor()` functions include the option `use = 'pairwise.complete.obs'`, which is the AC method. This could be used to implement AC in two applications:
  - For PCA, just run `eigen()` on either a covariance or correlation matrix computed for AC as above.
  - For linear regression, the matrices $A$ and $D$ both can be computed using `cov()`, after adjusting via a centering operation.
Conclusions

- Final score: AC had 2 wins, 1 loss.
- MI quite time-consuming, not recommended unless MCAR an issue.

Conclusions

- Final score: AC had 2 wins, 1 loss.
Conclusions

- Final score: AC had 2 wins, 1 loss.
- MI quite time-consuming, not recommended unless MCAR an issue.
Conclusions

- Final score: AC had 2 wins, 1 loss.
- MI quite time-consuming, not recommended unless MCAR an issue.

These slides available at