Modern Methods for Missing Data

Paul D. Allison, Ph.D.
Statistical Horizons LLC
www.statisticalhorizons.com
Introduction

Missing data problems are nearly universal in statistical practice.

Last 25 years have seen revolutionary developments in missing data methods.

Two major methods: multiple imputation (MI) and maximum likelihood (ML).

MI more popular, but ML is superior.

This talk:
- Why ML is better for missing data.
- How to do ML with SAS and Stata.
Assumptions

Missing completely at random (MCAR)

Suppose some data are missing on Y. These data are said to be MCAR if the probability that Y is missing is unrelated to Y or other variables X (where X is a vector of observed variables).

\[ \Pr(Y \text{ is missing} | X, Y) = \Pr(Y \text{ is missing}) \]

- MCAR is the ideal situation.
- What variables must be in the X vector? Only variables in the model of interest.
Assumptions

Missing at random (MAR)

Data on $Y$ are missing at random if the probability that $Y$ is missing does not depend on the value of $Y$, after controlling for observed variables

$$Pr \left( Y \text{ is missing} \mid X, Y \right) = Pr(Y \text{ is missing} \mid X)$$

E.g., the probability of missing income depends on marital status, but within each marital status, the probability of missing income does not depend on income.

- Considerably weaker assumption than MCAR
- Only $X$’s in the model must be considered. *But*, including other $X$’s (correlated with $Y$) can make MAR more plausible.
- Can test whether missingness on $Y$ depends on $X$
- Cannot test whether missingness on $Y$ depends on $Y$
Ignorability

The missing data mechanism is said to be ignorable if
- The data are missing at random and
- Parameters that govern the missing data mechanism are distinct from parameters to be estimated (unlikely to be violated)

In practice, “MAR” and “ignorable” are used interchangeably.
- If MAR but not ignorable (parameters not distinct), methods assuming ignorability would still be good, just not optimal.
- If missing data are ignorable, no need to model the missing data mechanism.
- Any general purpose method for handling missing data must assume that the missing data mechanism is ignorable.
Assumptions

Not missing at random (NMAR)

If the MAR assumption is violated, the missing data mechanism must be modeled to get good parameter estimates.

Heckman’s regression model for sample selection bias is a good example.

Effective estimation for NMAR missing data requires very good prior knowledge about missing data mechanism.

- Data contain no information about what models would be appropriate
- No way to test goodness of fit of missing data model
- Results often very sensitive to choice of model
- Listwise deletion able to handle one important kind of NMAR
Conventional Imputation

Any method that substitutes estimated values for missing values

- Replacement with means
- Regression imputation (replace with conditional means)

Problems

- Often leads to biased parameter estimates (e.g., variances)
- Usually leads to standard error estimates that are biased downward
  - Treats imputed data as real data, ignores inherent uncertainty in imputed values.
Multiple Imputation

Three basic steps:

1. Introduce random variation into the imputation process and generate several data sets, each with different imputed values.
2. Perform the desired analysis on each data set.
3. Combine the results into a single set of parameter estimates, standard errors, and test statistics.
Multiple Imputation Properties

If assumptions are met and MI is done properly, resulting estimators are
1. Consistent (implies approximately unbiased).
2. Asymptotically efficient (almost) – minimum sampling variability
3. Asymptotically normal – justifies uses of normal table for $p$-values and confidence intervals.
MCMC Method in SAS and Stata

Assumptions: 1. Ignorability (-> MAR)
   2. Multivariate normality

Multivariate normality implies
- All variables are normally distributed
- All conditional expectation functions are linear
- All conditional variance functions are homoscedastic.

Multivariate normality also implies that optimal imputation can be done by linear regression.

Example: a data set has three variables, $X$, $Y$, and $Z$. Suppose $X$ and $Y$ are fully observed, but $Z$ has missing data for 20% of the cases.
Regression Imputation

Conventional imputation: a regression of $Z$ on $X$ and $Y$ for the complete cases yields the imputation equation

$$\hat{Z} = b_0 + b_1X + b_2Y$$

Problem: Variance of imputed values is too small.

Solution: $\hat{Z} = b_0 + b_1X + b_2Y + sE$

where $E$ is a random draw from a standard normal distribution and $s$ is the root mean squared error from the regression.
Multiple Random Imputation

Problems with single, random imputation
- Not fully efficient because of random variation.
- Reported standard errors are too low.

Solution: Do it multiple times
- Produce multiple data sets, each with different imputed values. Estimate parameters on each data set.
- Averaging the parameter estimates dampens the variation. A few data sets yields close to full efficiency.
- Variability among the estimates provides information for correcting the standard errors. Yields accurate standard errors that fully account for the uncertainty about the missing values.
Complications

- To get the right amount of variability in the imputed values, the imputation parameters must be random draws from their posterior distribution.
- When more than one variable has missing data, imputation typically requires an iterative method of repeated imputations.
- Both of these issues can be solved by the MCMC algorithm available in many software packages.
- An alternative algorithm is the fully conditional specification (FCS), also known as multiple imputation by chained equations (MICE).
Maximum Likelihood

Choose as parameter estimates those values which, if true, would maximize the probability of observing what has, in fact, been observed.

Likelihood function: Expresses the probability of the data as a function of the data and the unknown parameter values.

Example: Let $f(y|\theta)$ be the probability density for $y$, given $\theta$ (a vector of parameters). For a sample of $n$ independent observations, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(y_i | \theta)$$
Properties of Maximum Likelihood

To get ML estimates, we find the value of $\theta$ that maximizes the likelihood function.

Under usual conditions, ML estimates have the following properties:

- Consistent (implies approximately unbiased in large samples)
- Asymptotically efficient
- Asymptotically normal
ML with Ignorable Missing Data

Suppose we have $n$ independent observations on $k$ variables $(y_1, y_2, ..., y_k)$. With no missing data, the likelihood function is

$$L = \prod_{i=1}^{n} f_i (y_{i1}, y_{i2}, ..., y_{ik}; \theta)$$

Now suppose that data are missing for individual $i$ for $y_1$ and $y_2$. The likelihood for that individual is just the probability (or density) of observing the remaining variables. If $y_1$ and $y_2$ are discrete that’s found by

$$f_i^* (y_{i3}, ..., y_{ik}; \theta) = \sum_{y_1} \sum_{y_2} f_i (y_{i1}, ..., y_{ik}; \theta)$$
ML with Ignorable Missing Data

If the missing variables are continuous, the likelihood for \( i \) is

\[
f_i^* (y_{i3}, \ldots, y_{ik}; \theta) = \int \int f_i (y_{i1}, y_{i2}, \ldots, y_{ik}) \, dy_2 \, dy_1
\]

If there are \( m \) observations with complete data and \( n-m \) observations with data missing on \( y_1 \) and \( y_2 \), the likelihood function for the full data set becomes

\[
L = \prod_{i=1}^{m} f_i (y_{i1}, y_{i2}, \ldots, y_{ik}; \theta) \prod_{i=m+1}^{n} f_i^* (y_{i3}, \ldots, y_{ik}; \theta)
\]

This can be maximized by standard methods to get ML estimates of \( \theta \).
Why ML is Better Than MI

1. ML is more efficient than MI
   - Smaller true standard errors.
   - Full efficiency for MI requires an infinite number of data sets. How many is enough?
   - This may be important for small samples.

2. For a given data set, ML always gives the same results. MI gives a different result each time you use it.
   - Can force MI to give the same results by setting the “seed”. (Computers only do pseudo-random numbers.)
   - No reason to prefer one seed to any other seed.
   - Reduce variability by using more data sets. How many?
Why ML is Better

3. MI requires many uncertain decisions. ML needs far fewer.
   - MCMC vs. FCS?
   - If FCS, what model for each variable?
   - How many data sets, is it enough?
   - How many iterations between data sets?
   - Prior distributions?
   - How to incorporate interactions and non-linearities?
   - Which method for multivariate testing?
Why ML is Better

4. With MI, there’s always potential conflict between imputation model and analysis model. No conflict with ML because only one model.

- Imputation model involves choice of variables, specification of relationships, probability distributions.
- For results to be correct, analysis model must be compatible with imputation model.
- Some common sources of incompatibility:
  - Analysis model contains variables that were not in the imputation model.
  - Analysis model contains interactions and non-linearities, but imputation model was strictly linear.
- Most problems occur when imputation model is more restrictive than analysis model.
Why Ever Do Multiple Imputation?

- Once you’ve imputed the missing values, you can use the data set with any other statistical software.
  - No need to learn new software.
  - Lots of possibilities

- ML requires specialized procedures.
  - Nothing in SAS, Stata, SPSS or R when predictors are missing for logistic regression, Cox regression, Poisson regression, etc.
How to Do ML

Scenario 1: Regression model with missing data on dependent variable only.

- Any kind of regression (linear, logistic, etc.)
- Assume MAR.
- Maximum likelihood reduces to listwise deletion (complete case analysis).

Could possibly do better if there is a good “auxiliary variable” (one highly correlated with dependent variable but not in the model).

If data are NMAR, could possibly do better with ML using Heckman’s method for selection bias. But that’s problematic.
Repeated Measures

Scenario 2: Repeated measures regression with data missing on dependent variable only.

Solution: Use ML (or REML) to estimate mixed model.

- Under MAR, estimates have the usual optimal properties.
- No need to do anything special.

Example: 595 people surveyed annually for 7 years. So 4165 records in data set with these variables:

- LWAGE = log of hourly wage (dependent variable)
- FEM = 1 if female, 0 if male
- T = year of the survey, 1 to 7
- ID = ID number for each person
Repeated Measures Example

No missing data in original sample.
- Estimated a linear model with main effects of FEM and T, and their interaction.
- Used robust standard errors to correct for dependence

**SAS**

PROC SURVEYREG DATA=my.wages;
  MODEL lwage = fem t fem*t;
  CLUSTER id;

**Stata**

reg lwage fem t fem#c.t, cluster(id)
### Estimated Regression Coefficients

| Parameter | Estimate | Standard Error | t Value | Pr > |t| |
|-----------|----------|----------------|---------|------|---|
| Intercept | 6.3399174| 0.01569031     | 404.07  | <.0001 |
| FEM       | -0.4555891| 0.05227259     | -8.72   | <.0001 |
| T         | 0.0974641 | 0.00188545     | 51.69   | <.0001 |
| FEM*T     | -0.0047193| 0.00553191     | -0.85   | 0.3940 |
Repeated Measures With Dropout

I deliberately made some data MAR. The probability of dropping out in year $t+1$, given response at time $t$ was

$$p_{t+1} = \frac{1}{1 + \exp(-8.5 - lwage_t)}$$

After 7 years, over half the people had dropped out. Re-estimating the model with PROC SURVEYREG yielded:

| Parameter | Estimate  | Standard Error | t Value | Pr > |t| |
|-----------|-----------|----------------|---------|-------|---|
| Intercept | 6.3815417 | 0.01722919     | 370.39  | <.0001| |
| FEM       | -0.4817529| 0.05357062     | -8.99   | <.0001| |
| t         | 0.0599360 | 0.00373687     | 16.04   | <.0001| |
| FEM*t     | 0.0205201 | 0.00851239     | 2.41    | 0.0162| |
Mixed Model

Let’s estimate a random intercepts model:

\[ y_{it} = \mu + \beta x_{it} + \alpha_i + \varepsilon_{it} \]

**SAS:**

PROC MIXED DATA=my.wagemiss;
    MODEL lwage = fem t fem*t / SOLUTION;
    RANDOM INTERCEPT / SUBJECT=id;

**Stata:**

    xtreg lwage fem t fem#c.t, i(id) mle
### PROC MIXED Output

| Effect    | Estimate | Standard Error | DF  | t Value | Pr > |t| |
|-----------|----------|----------------|-----|---------|-------|---|
| Intercept | 6.3368   | 0.01607        | 593 | 394.39  | <.0001|
| FEM       | -0.4623  | 0.04749        | 2237| -9.73   | <.0001|
| t         | 0.09573  | 0.001674       | 2237| 57.18   | <.0001|
| FEM*t     | 0.000759 | 0.004135       | 2237| 0.18    | 0.8545|

#### Number of Observations

<table>
<thead>
<tr>
<th>Number of Observations Read</th>
<th>4165</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations Used</td>
<td>2834</td>
</tr>
<tr>
<td>Number of Observations Not Used</td>
<td>1331</td>
</tr>
</tbody>
</table>
Binary Example

Data source: National Drug Abuse Treatment Clinical Trials Network sponsored by National Institute on Drug Abuse

Sample: 134 opioid-addicted youths, half randomly assigned to a new drug treatment over a 12-week period. Other half got standard detox therapy. Dependent variable = 1 if subject tested positive for opiates in each of 12 weeks, otherwise 0. A lot of missing data: Percentage missing at each week ranged from 10% to 63%.

Model: Mixed logistic regression where $p$ is probability of positive test, $z$ is treatment indicator and $t$ is time:

$$
\log \left( \frac{p_{it}}{1 - p_{it}} \right) = \mu + \beta z_i + \delta t + \gamma t z_i + \theta t^2 + \phi t^2 z_i + \alpha_i
$$
PROC GLIMMIX

SAS:

PROC GLIMMIX DATA=my.nidalong METHOD=QUAD(QPOINTS=5);
  CLASS usubjid;
  WHERE week NE 0;
  MODEL opiates = week|week|treat / D=B SOLUTION ;
  RANDOM INTERCEPT / SUBJECT=usubjid;

Stata:

xtlogit opiates c.week##c.week##treat if week~=0, i(usubjid)
  i(usubjid)
### GLIMMIX Output

#### Solutions for Fixed Effects

| Effect          | Estimate | Standard Error | DF  | t Value | Pr > |t| |
|-----------------|----------|----------------|-----|---------|------|-----|
| Intercept       | -0.5545  | 0.4928         | 132 | -1.13   | 0.2625|
| treat           | 1.1327   | 0.6766         | 786 | 1.67    | 0.0945|
| week            | 0.4198   | 0.1714         | 786 | 2.45    | 0.0145|
| treat*week      | -1.2558  | 0.2336         | 786 | -5.38   | <.0001|
| week*week       | -0.02804 | 0.01297        | 786 | -2.16   | 0.0309|
| treat*week*week | 0.09299  | 0.01776        | 786 | 5.24    | <.0001|
Figure 1. Predicted Probability of a Positive Test in Weeks 1 Through 12, for Treatment and Control Groups.
Missing Data on Predictors

Scenario 3: Linear models with missing data on predictors.

Mixed modeling software does nothing for cases with observed $y$ but missing $x$’s. Those cases are deleted, losing potentially useful information.

Solution: Full information ML using structural equation modeling (SEM) software.
Full Information ML

Also known as “raw” ML or “direct” ML

Directly maximize the likelihood for the specified model

Several computer packages can do this for any “LISREL” model

- Amos (www.spss.com/amos)
- Mplus (www.statmodel.com)
- LISREL (www.ssicentral.com/lisrel)
- MX (freeware) (views.vcu.edu/mx)
- EQS (www.mvsoft.com)
- PROC CALIS 9.22 (support.sas.com)
- Stata 12 (www.stata.com)
NLSY Example

581 children surveyed in 1990 as part of the National Longitudinal Survey of Youth.

Variables:

- **ANTI** antisocial behavior, a scale ranging from 0 to 6.
- **SELF** self-esteem, a scale ranging from 6 to 24.
- **POV** poverty status of family, 1 if in poverty, else 0.
- **BLACK** 1 if child is black, otherwise 0
- **HISPANIC** 1 if child is Hispanic, otherwise 0
- **CHILDAGE** child’s age in 1990
- **DIVORCE** 1 if mother was divorced in 1990, otherwise 0
- **GENDER** 1 if female, 0 if male
- **MOMAGE** mother’s age at birth of child
- **MOMWORK** 1 if mother was employed in 1990, otherwise 0

Data missing on **SELF, POV, BLACK, HISPANIC, MOMWORK**
Listwise Deletion

Goal is to estimate a linear regression with ANTI as the dependent variable. Listwise deletion removes 386 of the 581 cases:

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|--------------------|----------------|---------|-------|-----|
| Intercept| 1  | 2.86533            | 1.99117        | 1.44    | 0.1516|
| self     | 1  | -0.04531           | 0.03135        | -1.45   | 0.1498|
| pov      | 1  | 0.71946            | 0.23739        | 3.03    | 0.0027|
| black    | 1  | 0.05069            | 0.24918        | 0.20    | 0.8390|
| hispanic | 1  | -0.35696           | 0.25537        | -1.40   | 0.1636|
|childage | 1  | 0.00197            | 0.17072        | 0.01    | 0.9908|
| divorce  | 1  | 0.08703            | 0.24499        | 0.36    | 0.7228|
| gender   | 1  | -0.33470           | 0.19844        | -1.69   | 0.0931|
| momage   | 1  | -0.01198           | 0.04611        | -0.26   | 0.7953|
| momwork  | 1  | 0.25440            | 0.21751        | 1.17    | 0.2435|
FIML in SAS and Stata

Like MCMC (in SAS and Stata), FIML assumes

- Missing at random.
- Multivariate normality.

**SAS**

PROC CALIS DATA=nlsymiss METHOD=FIML;
PATH anti <- self pov black hispanic childage divorce gender momage momwork;

**Stata**

sem anti <- self pov black hispanic childage divorce gender momage momwork, method(mlmv)
method(mlmv)
## PATH List

<table>
<thead>
<tr>
<th>Path</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti &lt;--- self</td>
<td>_Parm1</td>
<td>-0.06718</td>
<td>0.02193</td>
<td>-3.06412</td>
</tr>
<tr>
<td>anti &lt;--- pov</td>
<td>_Parm2</td>
<td>0.64627</td>
<td>0.16366</td>
<td>3.94874</td>
</tr>
<tr>
<td>anti &lt;--- black</td>
<td>_Parm3</td>
<td>0.08500</td>
<td>0.16081</td>
<td>0.52857</td>
</tr>
<tr>
<td>anti &lt;--- hispanic</td>
<td>_Parm4</td>
<td>-0.32439</td>
<td>0.17011</td>
<td>-1.90694</td>
</tr>
<tr>
<td>anti &lt;--- childage</td>
<td>_Parm5</td>
<td>-0.00387</td>
<td>0.10055</td>
<td>-0.03844</td>
</tr>
<tr>
<td>anti &lt;--- divorce</td>
<td>_Parm6</td>
<td>-0.10599</td>
<td>0.14583</td>
<td>-0.72680</td>
</tr>
<tr>
<td>anti &lt;--- gender</td>
<td>_Parm7</td>
<td>-0.56116</td>
<td>0.11691</td>
<td>-4.79985</td>
</tr>
<tr>
<td>anti &lt;--- momage</td>
<td>_Parm8</td>
<td>0.02076</td>
<td>0.02817</td>
<td>0.73702</td>
</tr>
<tr>
<td>anti &lt;--- momwork</td>
<td>_Parm9</td>
<td>0.21895</td>
<td>0.14169</td>
<td>1.54528</td>
</tr>
</tbody>
</table>
Beyond Linear Regression.

CALIS and **sem** do FIML for a wide class of linear models

- Non-recursive simultaneous equations
- Latent variables
- Fixed and random effects for longitudinal data.

**Mplus (www.statmodel.com)** does the same, but it also does FIML for

- Logistic regression (binary, ordinal, multinomial).
- Poisson and negative binomial regression.
- Cox regression and piecewise exponential regression for survival data.
- Models in which data are not missing at random.
Data Not Missing at Random

Sometimes we suspect that data are not missing at random:
- people with high incomes may be less likely to report their income
- people who’ve been arrested may be less likely to report arrest status.

If data are NMAR, correct inference requires that the missing data mechanism be modeled as part of the inference. **IF** the model is correct, both ML and MI can produce optimal inferences in NMAR situations.

Problems:
1. Data contain no information to determine the correct model for the missing data mechanism.
2. Inference may be very sensitive to model choice.
Selection Models

Heckman’s model for selection bias

Designed for NMAR data on the dependent variable in a linear regression.

Linear model for $Y$:

\[ Y_i = \beta X_i + \epsilon_i \text{ where } \epsilon_i \text{ is normal with mean 0, constant variance, and uncorrelated with } X. \]

Probit model for $R$:

\[ \Pr(R_i=1|Y_i,X_i) = \Phi(a_0 + a_1 Y_i + a_2 X_i) \]

where $\Phi(.)$ is the cumulative distribution function for a standard normal variable.

If $a_1=0$, data are MAR. If $a_1=a_2=0$, data are MCAR.
Heckman’s Model (cont.)

This model implies a likelihood function whose parameters are all identified. It can be maximized in SAS with PROC QLIM or in Stata with `heckman` command.

Example: Women’s labor force participation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>N</th>
<th>Mean</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>INLF</td>
<td>=1 if in lab force, 1975</td>
<td>751</td>
<td>0.5672437</td>
<td>0</td>
<td>1.0000000</td>
</tr>
<tr>
<td>NWIFEINC</td>
<td>(faminc - wage*hours)/1000</td>
<td>751</td>
<td>20.1405723</td>
<td>-0.0290575</td>
<td>96.0000000</td>
</tr>
<tr>
<td>EDUC</td>
<td>years of schooling</td>
<td>751</td>
<td>12.2876165</td>
<td>5.0000000</td>
<td>17.0000000</td>
</tr>
<tr>
<td>EXPER</td>
<td>actual labor mkt exper</td>
<td>751</td>
<td>10.6324900</td>
<td>0</td>
<td>45.0000000</td>
</tr>
<tr>
<td>EXPERSQ</td>
<td>exper**2</td>
<td>751</td>
<td>178.1797603</td>
<td>0</td>
<td>2025.00</td>
</tr>
<tr>
<td>AGE</td>
<td>woman's age in yrs</td>
<td>751</td>
<td>42.5645806</td>
<td>30.0000000</td>
<td>60.0000000</td>
</tr>
<tr>
<td>KIDSLT6</td>
<td># kids &lt; 6 years</td>
<td>751</td>
<td>0.2370173</td>
<td>0</td>
<td>3.0000000</td>
</tr>
<tr>
<td>KIDSGE6</td>
<td># kids 6-18</td>
<td>751</td>
<td>1.3501997</td>
<td>0</td>
<td>8.0000000</td>
</tr>
<tr>
<td>LWAGE</td>
<td>log(wage)</td>
<td>426</td>
<td>1.1914355</td>
<td>-2.0541637</td>
<td>3.2188759</td>
</tr>
</tbody>
</table>

Heckman's Model (cont.)
PROC QLIM DATA=my.mroz;
  MODEL inlf = nwifeinc educ exper expersq age kidslt6 kidsge6 /DISCRETE;
  MODEL lwage = educ exper expersq / SELECT(inlf=1);

| Parameter          | DF | Estimate | Standard Error | t Value | Approx Pr > |t| |
|--------------------|----|----------|----------------|---------|--------------|---|
| LWAGE.Intercept    | 1  | -0.534485| 0.262808       | -2.03   | 0.0420       |
| LWAGE.EDUC         | 1  | 0.107986 | 0.014908       | 7.24    | <.0001       |
| LWAGE.EXPER        | 1  | 0.041618 | 0.014974       | 2.78    | 0.0054       |
| LWAGE.EXPERSQ      | 1  | -0.000810| 0.000419       | -1.93   | 0.0532       |
| _Sigma.LWAGE       | 1  | 0.664033 | 0.022763       | 29.17   | <.0001       |
| INLF.Intercept     | 1  | 0.251903 | 0.509797       | 0.49    | 0.6212       |
| INLF.NWIFEINC      | 1  | -0.012049| 0.004879       | -2.47   | 0.0135       |
| INLF.EDUC          | 1  | 0.131315 | 0.025374       | 5.18    | <.0001       |
| INLF.EXPER         | 1  | 0.122833 | 0.018738       | 6.56    | <.0001       |
| INLF.EXPERSQ       | 1  | -0.001878| 0.000601       | -3.13   | 0.0018       |
| INLF.AGE           | 1  | -0.052466| 0.008492       | -6.18   | <.0001       |
| INLF.KIDSLT6       | 1  | -0.868107| 0.118966       | -7.30   | <.0001       |
| INLF.KIDSGE6       | 1  | 0.034870 | 0.043511       | 0.80    | 0.4229       |
| _Rho               | 1  | 0.017170 | 0.149062       | 0.12    | 0.9083       |
Conclusion

Whenever possible, ML is preferable to MI for handling missing data.

- Deterministic result.
- No conflict between imputation and analysis model.
- Fewer uncertain decisions.
- Usually, the same assumptions.

For many scenarios, ML is readily available in standard packages

- Listwise deletion for missing only on response variable.
- Mixed models for repeated measures with missing only on response variable.
- SEM software for missing predictors.
- Heckman models for NMAR on response variable.
- For other scenarios, use Mplus.
More information

To get a pdf of the paper on which this talk is based, go to

support.sas.com/resources/papers/proceedings12/312-2012.pdf

For information on short courses on a variety of statistical topics go to

www.statisticalhorizons.com