# Introduction to Proc MI and Proc MIAnalyze for Multiple Imputation of Missing Data in SAS

Presented at
Michigan SAS Users Group

Kathy Welch
CSCAR, The University of Michigan
May 12, 2009

# Missing Data is a Problem in Many Studies

- Attrition in longitudinal studies
  - Drug doesn’t work/ side effects
  - Drug works too well
- Item non-response in surveys
  - Don’t answer questions on income, but answer other questions
- Partial non-response
  - Participate in interview, but not physical exam

# Missing Data Mechanisms

- MCAR: missing completely at random, missingness (M) has no relationship to data (Y)

  $p(M|Y) = p(M)$, for all Y
- MAR: Missing At Random, missingness only related to observed data values

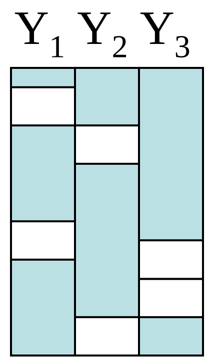
  $p(M|Y)=p(M|Yobs)$, for all Y
- NMAR: Not Missing at random, missingness depends on the ***unobserved*** values of Y
  - AKA, informative missing, non-ignorable missing
- Rubin, 1976, Little and Rubin, 2002

# What to do if NMAR

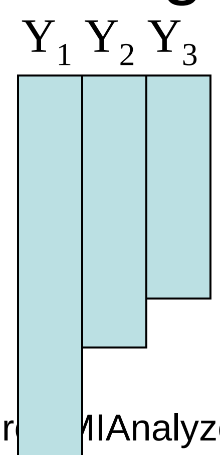
- If data are not missing at random, the missing data *mechanism* needs to be modeled
- There are methods for this, but not implemented in standard software packages

# Missing Data Patterns

- General pattern:




- Monotone pattern:
  - When one value is missing, all subsequent values are missing



# Some methods for handling missing data

- CC: Complete Cases analysis, weighted complete case analysis
- Single Imputation methods
  - Mean imputation
  - Conditional mean imputation
- Multiple Imputation
- Other methods, e.g., bootstrap imputations
- All methods have limitations

# Complete Case Analysis

- Throw away cases with any missing data
- Default method in most SAS procedures (e.g., Proc Reg, Proc GLM)
- Possible problems
  - Can result in serious bias, if missing cases are not MCAR
  - Variances of estimates will be greater than if all cases were used
- May be fine, esp. if amount of missing data is small
- Include in analysis variables that predict missingness (Allison)

# Single Imputation Methods

- Use incomplete data to get a plausible predicted value for the missing data
- Create a “complete” data set for analysis
- Advantage: Once through the data
- But can have problems
  – May be seriously biased
  – May Understate the variability of estimates

# Mean Imputation

- Replace missing data for continuous variables by Mean of non-missing values
- Replace missing data for categorical variables by the mode
- Because marginal distributions are used, associations are distorted
- Standard deviations of completed data are too small (variability is artificially reduced)
- Sample size is overestimated
- Can be less efficient than CC analysis
- Can result in serious bias

# Conditional Mean Imputation

- Conditional on observed values
- Use a regression model to predict the missing values
- Better than unconditional mean, more plausible values
- Standard errors of estimates are still too small
- CC analysis may be better

# Multiple Imputation

- Create m sets of imputations (complete data)
- Analyze the m sets of imputations with usual statistical methods
- Combine estimates to get imputation inference
- Very useful for large public datasets, so users can analyze complete data

# Number of Imputations

- Relative efficiency of a small number (m) of imputations compared to the theoretical infinite number of imputations is high
- 5 imputations the default for Proc MI

Relative Efficiencies*

| | $\lambda$ | | | | |
|---|---|---|---|---|---|
| m | 10% | 20% | 30% | 50% | 70% |
| 3 | 0.9677 | 0.9375 | 0.9091 | 0.8571 | 0.8108 |
| 5 | 0.9804 | 0.9615 | 0.9434 | 0.9091 | 0.8772 |
| 10 | 0.9901 | 0.9804 | 0.9709 | 0.9524 | 0.9346 |
| 20 | 0.9950 | 0.9901 | 0.9852 | 0.9756 | 0.9662 |

*Table 44.5 from SAS documentation for Proc MI, SAS 9.2

# Proper Multiple Imputation Methods

- Each dataset is a random draw from the joint posterior predictive distribution of missing values and parameters of the distribution
- The uncertainty in estimating $\theta$ is taken into account
  – Propagate the error in estimating $\theta$ in the imputations (this is good)
- Estimates have high efficiency
- The accuracy increases with m, and the percent of complete data

# Imputation Model vs. Analysis Model

- Imputation model may use more variables than analysis model
- This is OK
  - Imputation may use many more variables, even though not of substantive interest
  - Different analysts may want to use different predictors in their models
- Improved efficiency and less bias by using more variables for imputation model
- But try to use variables in imputation that are at least helpful for imputation of missing values

# Combining Estimates from Multiple Imputation

$$\overline{\Theta} = \frac{1}{m}\sum_{m=1}^{M}\hat{\Theta}_m$$ Combined Parameter Estimate

$$T = \overline{W} + (1 + \frac{1}{m})B$$ Total variance

$$\overline{W} = \frac{1}{m}\sum_{m=1}^{M}W_m$$ Within-imputation variance

$$B = \frac{1}{m-1}\sum_{m=1}^{M}(\hat{\Theta}_m - \overline{\Theta})^2$$ Between-imputation variance

# Combining Estimates (Cont)

Fraction of Missing Information

$$r = \frac{(1+1/m)B}{\bar{W} + (1+1/m)B}$$

Degrees of Freedom

$$\nu = (m-1)/r^2$$

Degrees of Freedom are increased by having more imputations, or less missing data

# SAS Proc MI Methods for Multiple Imputation

- Methods available in Proc MI depend on missing data pattern
- For arbitrary missing data pattern use Markov-chain Monte Carlo (MCMC) method
  - Based on Bayes methods
  - Proper Imputation
  - Assumes a multivariate normal distribution
  - This may be problematic, esp. for missing categorical variables, but not too bad
- Can be used to get to a monotone missing data pattern, then another method, such as logistic regression for categorical variables, may be used

# Proc MI Methods Available for Different Missing Data Patterns

| Pattern of Missingness | Type of Imputed Variable | Recommended Methods |
|---|---|---|
| Monotone | Continuous | Regression<br>Predicted Mean Matching<br>Propensity Score |
| Monotone | Classification (Ordinal) | Logistic Regression |
| Monotone | Classification (Nominal) | Discriminant Function Method |
| Arbitrary | Continuous | MCMC Full-Data Imputation<br>MCMC Monotone-Data Imputation |

Table 44.3 from Proc MI documentation, SAS 9.1.3

# Example Using Proc MI and MIAnalyze

- LSOA study (Longitudinal Study of Aging, NCHS)

http://www.cdc.gov/nchs/about/otheract/aging/lsoa1.htm

- Baseline data collected in 1984, follow-up in 1986.
- 5151 participants, aged 70 to 99 in 1984
- 64% female
- Want to model changes in ADL (restrictions in Activities of Daily Living) from 1984 to 1986
- Missing data on over 21% for ADL in 1986

# Variables in LSOA Dataset

Variables in Creation Order

| # | Variable | Type | Len | Label |
|---|---|---|---|---|
| 1 | SEX | Num | 3 | 1=MALE 2=FEMALE |
| 2 | AGE84 | Num | 3 | Age during 1984 |
| 3 | RACER | Num | 3 | 1=WHITE 2=BLACK 3=OTHER |
| 4 | EDUC | Num | 3 | Education of individual-completed years |
| 5 | POVERTY | Num | 3 | NHIS poverty index |
| 6 | OWNBUYR | Num | 3 | Own/buying recode |
| 7 | MORTGAGE | Num | 3 | Fully paid for or mortgage being paid |
| 8 | AMTOWED | Num | 4 | Amount principal still owed |
| 9 | PRESVAL | Num | 4 | Present value of place |
| 10 | NUMADL | Num | 3 | Number of ADLs, 1984 |
| 11 | NUMADL2R | Num | 3 | Number of ADLs, 1986 |
| 12 | FNLWGT2 | Num | 3 | Final 1986 LSOA weight |
| 13 | STRATUM | Num | 8 | |
| 14 | PSU | Num | 8 | |

# LSOA Analysis Plan

- Check pattern of missing data
- Create five multiply imputed datasets using Proc MI
- Carry out a regression analysis of the five datasets using Proc Reg
- Combine the estimates from the five datasets to get the MI estimates, using Proc MIAnalyze
- Carry out a CC analysis for comparison

# LSOA Descriptives for Original Dataset

The MEANS Procedure

| Variable | N | N Miss | Mean | Std Dev |
|---|---|---|---|---|
| diffadl | 4048 | 1103 | 0.2197777 | 0.5834531 |
| POVERTY | 4229 | 922 | 1.1941357 | 0.3955806 |
| own | 5015 | 136 | 0.6638086 | 0.4724524 |
| EDUC | 5053 | 98 | 9.7573719 | 3.7452672 |
| AGE84 | 5151 | 0 | 78.2071442 | 6.0010675 |
| SEX | 5151 | 0 | 1.6396816 | 0.4801394 |
| black | 5151 | 0 | 0.1087168 | 0.3113137 |
| other | 5151 | 0 | 0.0108717 | 0.1037091 |

Much missing data on the outcome variable

# Proc MI SAS Code

```
proc mi data=lsoa nimpute=5 out=OUTMI
seed=3355;

var poverty own sex black other educ
age84 logadl logadl2r ;

run;
```

OUTMI contains the five multiply-imputed datasets

_IMPUTATION_ variable indexes imputations

# Proc MI Model Information

```
           The MI Procedure

                       Model Information

Data Set                              WORK.LSOA
Method                                MCMC
Multiple Imputation Chain             Single Chain
Initial Estimates for MCMC            EM Posterior Mode
Start                                 Starting Value
Prior                                 Jeffreys
Number of Imputations                 5
Number of Burn-in Iterations          200
Number of Iterations                  100
Seed for random number generator      3355
```

# Proc MI Missing Data Patterns

Missing Data Patterns

| Group | POVERTY | own | SEX | black | other | EDUC | AGE84 | logadl | logadl2r | Freq |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | X | X | X | X | X | X | X | X | X | 3268 |
| 2 | X | X | X | X | X | X | X | X | . | 791 |
| 3 | X | X | X | X | X | X | X | . | X | 7 |
| 4 | X | X | X | X | X | X | X | . | . | 7 |
| 5 | X | X | X | X | X | . | X | X | X | 28 |
| 6 | X | X | X | X | X | . | X | X | . | 27 |
| 7 | X | X | X | X | X | . | X | . | X | 1 |
| 8 | X | . | X | X | X | X | X | X | X | 77 |
| 9 | X | . | X | X | X | X | X | X | . | 21 |
| 10 | X | . | X | X | X | . | X | X | X | 2 |
| 11 | . | X | X | X | X | X | X | X | X | 624 |

# Proc MI Missing Data Patterns: Group Means

Missing Data Patterns

| | | Group Means | | | |
|---|---|---|---|---|---|
| Group | Percent | POVERTY | own | SEX | black |
| 1 | 63.46 | 1.186047 | 0.686965 | 1.646879 | 0.109241 |
| 2 | 15.36 | 1.231353 | 0.596713 | 1.573957 | 0.113780 |
| 3 | 0.14 | 1.285714 | 0.857143 | 1.571429 | 0 |
| 4 | 0.14 | 1.142857 | 0.571429 | 1.428571 | 0.142857 |
| 5 | 0.54 | 1.178571 | 0.571429 | 1.678571 | 0.178571 |
| 6 | 0.52 | 1.259259 | 0.370370 | 1.666667 | 0.222222 |
| 7 | 0.02 | 1.000000 | 1.000000 | 2.000000 | 0 |
| 8 | 1.50 | 1.155844 | . | 1.532468 | 0.077922 |
| 9 | 0.41 | 1.142857 | . | 1.476190 | 0.095238 |
| 10 | 0.04 | 1.000000 | . | 1.500000 | 0.500000 |
| 11 | 12.12 | . | 0.692308 | 1.684295 | 0.105769 |

# Proc MI Variance Information

Multiple Imputation Variance Information

| Variable | Variance Between | Variance Within | Variance Total | DF |
|---|---|---|---|---|
| POVERTY | 0.000001824 | 0.000030417 | 0.000032606 | 749.29 |
| own | 0.000000864 | 0.000043349 | 0.000044386 | 2981.6 |
| EDUC | 0.000022363 | 0.002728 | 0.002755 | 4547.1 |
| logadl | 0.000000234 | 0.000070284 | 0.000070565 | 5024.8 |
| logadl2r | 0.000024269 | 0.000098270 | 0.000127 | 75.09 |

Multiple Imputation Variance Information

| Variable | Relative Increase in Variance | Fraction Missing Information | Relative Efficiency |
|---|---|---|---|
| POVERTY | 0.071951 | 0.069216 | 0.986346 |
| own | 0.023922 | 0.023630 | 0.995296 |
| EDUC | 0.009836 | 0.009787 | 0.998046 |
| logadl | 0.003991 | 0.003983 | 0.999204 |
| logadl2r | 0.296360 | 0.248006 | 0.952743 |

# Proc MI Parameter Estimates

Multiple Imputation Parameter Estimates

| Variable | Mean | Std Error | 95% Confidence Limits | | DF |
|---|---|---|---|---|---|
| POVERTY | 1.194470 | 0.005710 | 1.183260 | 1.205680 | 749.29 |
| own | 0.664994 | 0.006662 | 0.651931 | 0.678057 | 2981.6 |
| EDUC | 9.743557 | 0.052489 | 9.640652 | 9.846461 | 4547.1 |
| logadl | 0.349539 | 0.008400 | 0.333070 | 0.366007 | 5024.8 |
| logadl2r | 0.559964 | 0.011287 | 0.537480 | 0.582448 | 75.09 |

Multiple Imputation Parameter Estimates

| Variable | Minimum | Maximum | Mu0 | t for H0: Mean=Mu0 | Pr > \|t\| |
|---|---|---|---|---|---|
| POVERTY | 1.193273 | 1.196416 | 0 | 209.18 | <.0001 |
| own | 0.663869 | 0.665959 | 0 | 99.81 | <.0001 |
| EDUC | 9.738145 | 9.749231 | 0 | 185.63 | <.0001 |
| logadl | 0.348811 | 0.350080 | 0 | 41.61 | <.0001 |
| logadl2r | 0.554439 | 0.567925 | 0 | 49.61 | <.0001 |

# Modify Data and Check Means

```
/*Create Difference on Log Scale AFTER Imputation*/
data OUTMI;
  set OUTMI;
  diffadl = logadl2r - logadl;
run;
proc means data=outmi;
by _imputation_;
run;
```

# Descriptives for Imputation 1

```
------------------------------- _Imputation_=1 --------------------------------

                              The MEANS Procedure

                             N
Variable          N       Miss         Minimum         Maximum            Mean
------------------------------------------------------------------------------
SEX            5150          0       1.0000000       2.0000000       1.6398058
AGE84          5150          0      70.0000000      99.0000000      78.2077670
RACER          5150          0       1.0000000       3.0000000       1.1304854
EDUC           5150          0      -0.6051802      18.0000000       9.7473659
POVERTY        5150          0      -0.3943656       2.7625847       1.1964162
own            5150          0      -0.7833835       1.8704920       0.6651582
logadl2r       5150          0      -1.1814288       2.8560990       0.5544391
logadl         5150          0      -0.2376750       2.0794415       0.3498080
black          5150          0               0       1.0000000       0.1087379
other          5150          0               0       1.0000000       0.0108738
ownmiss        5150          0               0       1.0000000       0.0264078
povmiss        5150          0               0       1.0000000       0.1788350
diffadl        5150          0      -2.0794415       2.3073451       0.2046311
------------------------------------------------------------------------------
```

# Descriptives for Imputation 2

------------------------------- _Imputation_=2 -------------------------------

| Variable | N | N Miss | Minimum | Maximum | Mean |
|---|---|---|---|---|---|
| SEX | 5150 | 0 | 1.0000000 | 2.0000000 | 1.6398058 |
| AGE84 | 5150 | 0 | 70.0000000 | 99.0000000 | 78.2077670 |
| RACER | 5150 | 0 | 1.0000000 | 3.0000000 | 1.1304854 |
| EDUC | 5150 | 0 | 0 | 18.0000000 | 9.7492314 |
| POVERTY | 5150 | 0 | -0.0334959 | 2.4663084 | 1.1953344 |
| own | 5150 | 0 | -0.8350119 | 2.0972619 | 0.6657782 |
| logadl2r | 5150 | 0 | -1.6669717 | 3.1120230 | 0.5679253 |
| logadl | 5150 | 0 | -0.6301672 | 2.0794415 | 0.3500805 |
| black | 5150 | 0 | 0 | 1.0000000 | 0.1087379 |
| other | 5150 | 0 | 0 | 1.0000000 | 0.0108738 |
| ownmiss | 5150 | 0 | 0 | 1.0000000 | 0.0264078 |
| povmiss | 5150 | 0 | 0 | 1.0000000 | 0.1788350 |
| diffadl | 5150 | 0 | -2.0794415 | 2.1250189 | 0.2178448 |

# SAS Code for Proc Reg

```
proc reg data=OUTMI outest=OUTREG covout ;
   by _Imputation_;
   model diffadl=poverty own educ logadl age84 sex black other;
run;
```

# Output from Regression for Imputation 1

```
                       The REG Procedure
                         Model: MODEL1
                 Dependent Variable: diffadl

         Number of Observations Read        5150
         Number of Observations Used        5150

                      Parameter Estimates

                       Parameter      Standard
Variable     DF         Estimate         Error    t Value    Pr > |t|

Intercept     1         -1.06429       0.11564      -9.20      <.0001
POVERTY       1          0.00564       0.02052       0.27      0.7836
own           1         -0.04346       0.01666      -2.61      0.0091
EDUC          1         -0.00714       0.00220      -3.25      0.0012
logadl        1         -0.33447       0.01335     -25.06      <.0001
AGE84         1          0.01892       0.00135      14.05      <.0001
SEX           1         -0.00591       0.01627      -0.36      0.7163
black         1          0.08336       0.02585       3.22      0.0013
other         1         -0.13478       0.07383      -1.83      0.0680
```

# Output from Regression for Imputation 2

The REG Procedure
Model: MODEL1
Dependent Variable: diffadl

| | |
|---|---|
| Number of Observations Read | 5150 |
| Number of Observations Used | 5150 |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | -1.11195 | 0.11696 | -9.51 | <.0001 |
| POVERTY | 1 | 0.04296 | 0.02096 | 2.05 | 0.0405 |
| own | 1 | -0.04476 | 0.01688 | -2.65 | 0.0080 |
| EDUC | 1 | -0.00755 | 0.00223 | -3.39 | 0.0007 |
| logadl | 1 | -0.33505 | 0.01350 | -24.83 | <.0001 |
| AGE84 | 1 | 0.01895 | 0.00136 | 13.89 | <.0001 |
| SEX | 1 | 0.00683 | 0.01648 | 0.41 | 0.6784 |
| black | 1 | 0.07137 | 0.02622 | 2.72 | 0.0065 |
| other | 1 | -0.16516 | 0.07474 | -2.21 | 0.0272 |

# Proc MIAnalyze Code to Combine Estimates Across Imputations

```
proc mianalyze data=OUTREG;
   modeleffects Intercept poverty own educ logadl
age84 sex black other;
run;
```

# Proc MIAnalyze Output

The MIANALYZE Procedure

Model Information

| | |
|---|---|
| Data Set | WORK.OUTREG |
| Number of Imputations | 5 |

Multiple Imputation Variance Information

| Parameter | Variance Between | Variance Within | Variance Total | DF |
|---|---|---|---|---|
| Intercept | 0.001782 | 0.013592 | 0.015730 | 216.5 |
| poverty | 0.000324 | 0.000432 | 0.000821 | 17.816 |
| own | 0.000310 | 0.000282 | 0.000655 | 12.358 |
| educ | 0.000000724 | 0.000004891 | 0.000005760 | 175.62 |
| logadl | 0.000112 | 0.000181 | 0.000315 | 21.953 |
| age84 | 0.000000485 | 0.000001842 | 0.000002424 | 69.323 |
| sex | 0.000073830 | 0.000269 | 0.000357 | 65.112 |
| black | 0.000175 | 0.000680 | 0.000890 | 71.717 |
| other | 0.002603 | 0.005532 | 0.008655 | 30.722 |

# Proc MIAnalyze Output (Cont)

Multiple Imputation Parameter Estimates

| Parameter | Estimate | Std Error | 95% Confidence Limits | | DF |
|---|---|---|---|---|---|
| Intercept | -1.081213 | 0.125421 | -1.32842 | -0.83401 | 216.5 |
| poverty | 0.035685 | 0.028650 | -0.02455 | 0.09592 | 17.816 |
| own | -0.031520 | 0.025584 | -0.08708 | 0.02404 | 12.358 |

Multiple Imputation Parameter Estimates

| Parameter | Minimum | Maximum |
|---|---|---|
| Intercept | -1.132226 | -1.024379 |
| poverty | 0.005635 | 0.053830 |
| own | -0.044841 | -0.010716 |

Multiple Imputation Parameter Estimates

| Parameter | Theta0 | t for H0: Parameter=Theta0 | Pr > \|t\| |
|---|---|---|---|
| Intercept | 0 | -8.62 | <.0001 |
| poverty | 0 | 1.25 | 0.2291 |
| own | 0 | -1.23 | 0.2409 |

# Proc MIAnalyze Output (Cont)

Multiple Imputation Parameter Estimates

| Parameter | Estimate | Std Error | 95% Confidence Limits | | DF |
|---|---|---|---|---|---|
| educ | -0.008150 | 0.002400 | -0.01289 | -0.00341 | 175.62 |
| logadl | -0.341935 | 0.017754 | -0.37876 | -0.30511 | 21.953 |
| age84 | 0.018647 | 0.001557 | 0.01554 | 0.02175 | 69.323 |
| sex | 0.002601 | 0.018906 | -0.03516 | 0.04036 | 65.112 |
| black | 0.070117 | 0.029827 | 0.01065 | 0.12958 | 71.717 |
| other | -0.123315 | 0.093033 | -0.31313 | 0.06650 | 30.722 |

| Parameter | Minimum | Maximum |
|---|---|---|
| educ | -0.009317 | -0.007142 |
| logadl | -0.360206 | -0.334471 |
| age84 | 0.017851 | 0.019504 |
| sex | -0.005912 | 0.014089 |
| black | 0.049046 | 0.083361 |
| other | -0.165156 | -0.034706 |

| Parameter | Theta0 | t for H0: Parameter=Theta0 | Pr > \|t\| |
|---|---|---|---|
| educ | 0 | -3.40 | 0.0008 |
| logadl | 0 | -19.26 | <.0001 |
| age84 | 0 | 11.98 | <.0001 |
| sex | 0 | 0.14 | 0.8910 |
| black | 0 | 2.35 | 0.0215 |
| other | 0 | -1.33 | 0.1948 |

# Impute Values for Poverty (Binary) Using Proc Logistic on Imputed Data

```
data SIX;
  set OUTMI;
  if povmiss=1 then poverty = .;
run;
proc mi data=SIX nimpute=1 out=OUTMI2 seed=3355;
   by _Imputation_;
   class POVERTY;
   monotone logistic (POVERTY = own educ logadl logadl2r age84
sex black other / details);
   var  own educ logadl logadl2r age84 sex black other  poverty;
run;
proc reg data=OUTMI2  outest=OUTREG2 covout ;
  by _imputation_;
  model diffadl=poverty own educ logadl age84 sex black other;
run;
```

# Output from Complete Case Analysis

The REG Procedure
Model: MODEL1
Dependent Variable: diffadl

| | |
|---|---|
| Number of Observations Read | 5151 |
| Number of Observations Used | 3268 |
| Number of Observations with Missing Values | 1883 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 8 | 121.35294 | 15.16912 | 50.43 | <.0001 |
| Error | 3259 | 980.21561 | 0.30077 | | |
| Corrected Total | 3267 | 1101.56855 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 0.54843 | R-Square | 0.1102 |
| Dependent Mean | 0.20687 | Adj R-Sq | 0.1080 |
| Coeff Var | 265.10481 | | |

# Compare CC vs. MI Results

| | CC Analysis | | | | Proc MI | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Estimate | SE | df | p-value | Estimate | SE | df | p-value |
| Int | -0.84897 | 0.14929 | 3259 | <.0001 | -1.081213 | 0.125421 | 216.5 | <.0001 |
| Poverty | 0.03078 | 0.02626 | 3259 | 0.2412 | 0.035685 | 0.028650 | 17.816 | 0.2291 |
| Own | -0.01959 | 0.02131 | 3259 | 0.3580 | -0.031520 | 0.025584 | 12.358 | 0.2409 |
| Educ | -0.00922 | 0.00278 | 3259 | 0.0009 | -0.008150 | 0.002400 | 175.62 | 0.0008 |
| Logadl | -0.34402 | 0.01836 | 3259 | <.0001 | -0.341935 | 0.017754 | 21.953 | <.0001 |
| Age84 | 0.01631 | 0.00175 | 3259 | <.0001 | 0.018647 | 0.001557 | 69.323 | <.0001 |
| Sex | -0.02703 | 0.02056 | 3259 | 0.1886 | 0.002601 | 0.018906 | 65.112 | 0.8910 |
| Black | 0.03811 | 0.03223 | 3259 | 0.2372 | 0.070117 | 0.029827 | 71.717 | 0.0215 |
| Other | -0.08840 | 0.09632 | 3259 | 0.3588 | -0.123315 | 0.093033 | 30.722 | 0.1948 |

# Summing Up

- There are many methods for analyzing missing data with missing values
- No method is perfect
- CC analysis is better than some mean imputation methods
- Multiple imputation is better than single imputation
- Uncertainty in parameter estimates is correctly maintained in MI methods

# References


- Rubin, D.B. (1976) Inference and missing data (with discussion), Biometrika 63, 581-592
- Little, R.J., and Rubin, D. B. (2002), Statistical Analysis with Missing Data, 2nd Edition, New York: Wiley.
- Allison, Paul D.(2001), Missing Data, Series: Quantitative Applications in the Social Sciences, Volume 136, a Sage University Paper
- Little, R.J., and Ragunathan, T., Statistical Analysis with Missing Data, a CSCAR Workshop, May 17 and 18, 2005
- SAS Documentation for release 9.1.3 for Windows and for release 9.2 for Windows

# References (Cont)

- The data analysis and software code for this presentation was generated using SAS/STAT software, Version 9.1.3 of the SAS System for Windows. Copyright © 2002-2003 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.