MULTIPLE IMPUTATION IN MPLUS

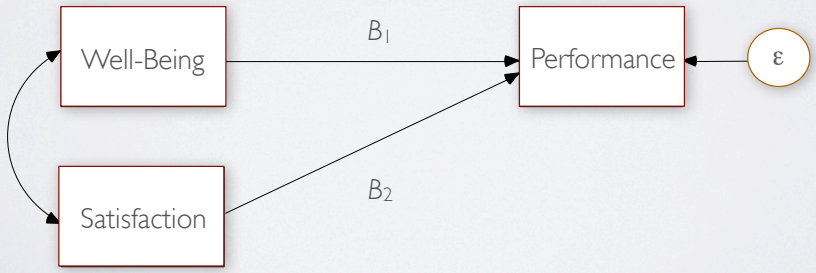
EMPLOYEE DATA

- Data set containing scores from 480 employees on eight work-related variables
- Variables:
- Age, gender, job tenure, IQ, psychological well-being, job satisfaction, job performance, and turnover intentions
- 33% of the cases have missing well-being scores, and 33% have missing satisfaction scores

ANALYSIS EXAMPLE

- Multiple regression model that predicts job performance from psychological well-being and job satisfaction

$$jobperf = B_0 + B_1(wbeing) + B_2(jobsat) + \varepsilon$$



MPLUS COMMANDS FOR DIAGNOSTIC ANALYSIS

- TITLE (optional)
- DATA (same as ML analysis)
- VARIABLE (same as ML analysis)
- ANALYSIS
- MODEL
- OUTPUT
- PLOT

DATA COMMAND

- Full file path (data in different directory than input file)

```
DATA:
file = 'c:\Data\employee.dat';
```

- Omitted file path (data in same directory as input file)

```
DATA:
file = employee.dat;
```

VARIABLE COMMAND

- The VARIABLE command serves the same function as in an ML analysis
- The USEVARIABLES list should include all variables that will be part of the imputation process

```
VARIABLE:
! Information about the contents of the data file;
names = id age tenure female wbeing jobsat jobperf turnover iq;
usevariables = age tenure wbeing jobsat jobperf iq;
missing = all (-99);
```

ANALYSIS COMMAND

- The following commands apply to the preliminary MCMC analysis that generates graphical diagnostics

```
ANALYSIS:
! Bayesian estimation;
estimator = bayes;
! Random number seed for MCMC algorithm;
bseed = 48932;
! Convergence criterion (.05 = PSR < 1.05);
bconvergence = .05;
```

BCONVERGENCE SUBCOMMAND

- Mplus generates imputed data sets only after the MCMC algorithm has converged (this feature prevents imputations from pathological chains)
- Setting BCONVERGENCE = .05 (the default) specifies that convergence is achieved once the PSR drops below 1.05
- Smaller values give a more conservative picture of convergence but require more computational time

MODEL COMMAND

- In the diagnostic run, MODEL should specify an analysis that estimates all means, variances, and covariances (i.e., a saturated model)

```
MODEL:
[jobperf tenure wbeing jobsat turnover iq];   ! Means;
jobperf tenure wbeing jobsat turnover iq;     ! Variances;
jobperf tenure wbeing jobsat turnover iq with
   jobperf tenure wbeing jobsat turnover iq;  ! Covariances;
```

OUTPUT COMMAND

- The TECH8 option prints the PSR statistic (updated after every 100th iteration) to the output file

```
OUTPUT:
! Tech8 gives the PSR statistic;
tech8;
```

PLOT COMMAND

- The PLOT command generates graphical diagnostics
- This command is only operational with the ESTIMATOR = BAYES option (typically used only for the diagnostic run)

```
PLOT:
! Graphical convergence diagnostics;
type = plot2;
```

MPLUS DIAGNOSTIC PROGRAM

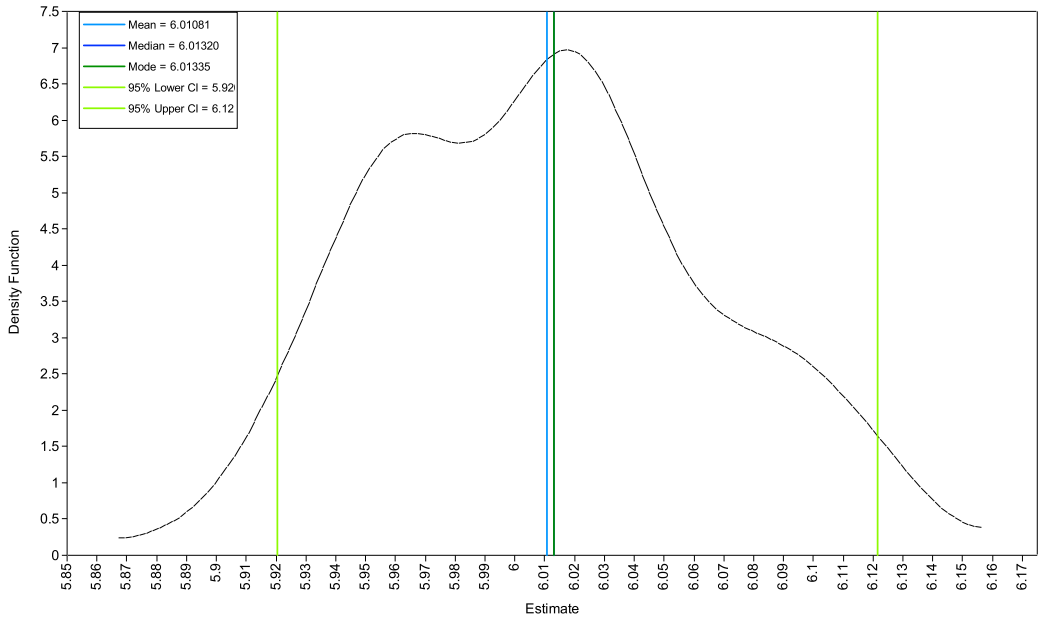
```
DATA:
file = employee.dat;
VARIABLE:
names = id age tenure female wbeing jobsat jobperf turnover iq;
usevariables = age tenure wbeing jobsat jobperf iq;
missing = all (-99);
ANALYSIS:
estimator = bayes;
bseed = 48932;
bconvergence = .05;
MODEL:
[age tenure wbeing jobsat jobperf iq];
age tenure wbeing jobsat jobperf iq;
age tenure wbeing jobsat jobperf iq with
   age tenure wbeing jobsat jobperf iq;
OUTPUT:
tech8;
PLOT:
type = plot2;
```

PSR OUTPUT (TECH8 OPTION)

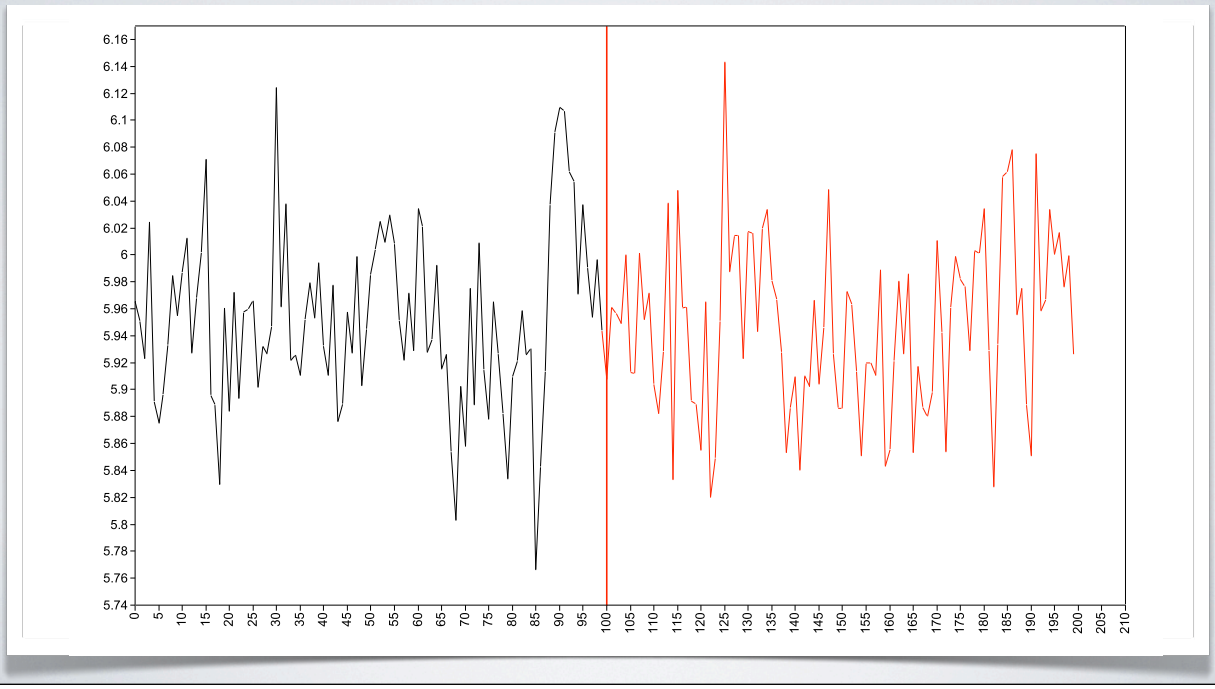
- The PSR dropped below 1.05 (i.e., the MCMC algorithm converged) by the 300th iteration
- A between-imputation interval of 200-300 is appropriate

```
TECHNICAL 8 OUTPUT

   TECHNICAL 8 OUTPUT FOR BAYES ESTIMATION

                    POTENTIAL          PARAMETER WITH
     ITERATION     SCALE REDUCTION       HIGHEST PSR
     100               1.134                13
     200               1.058                15
```

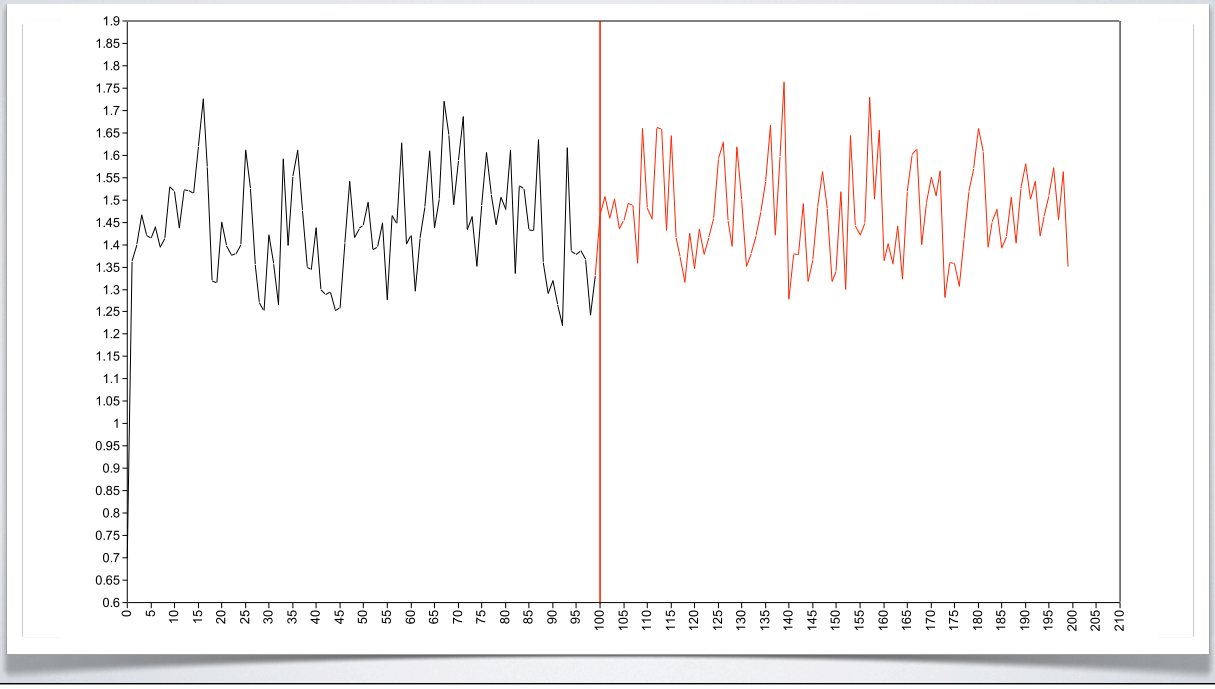
POSTERIOR DISTRIBUTION OF THE JOB SATISFACTION MEAN



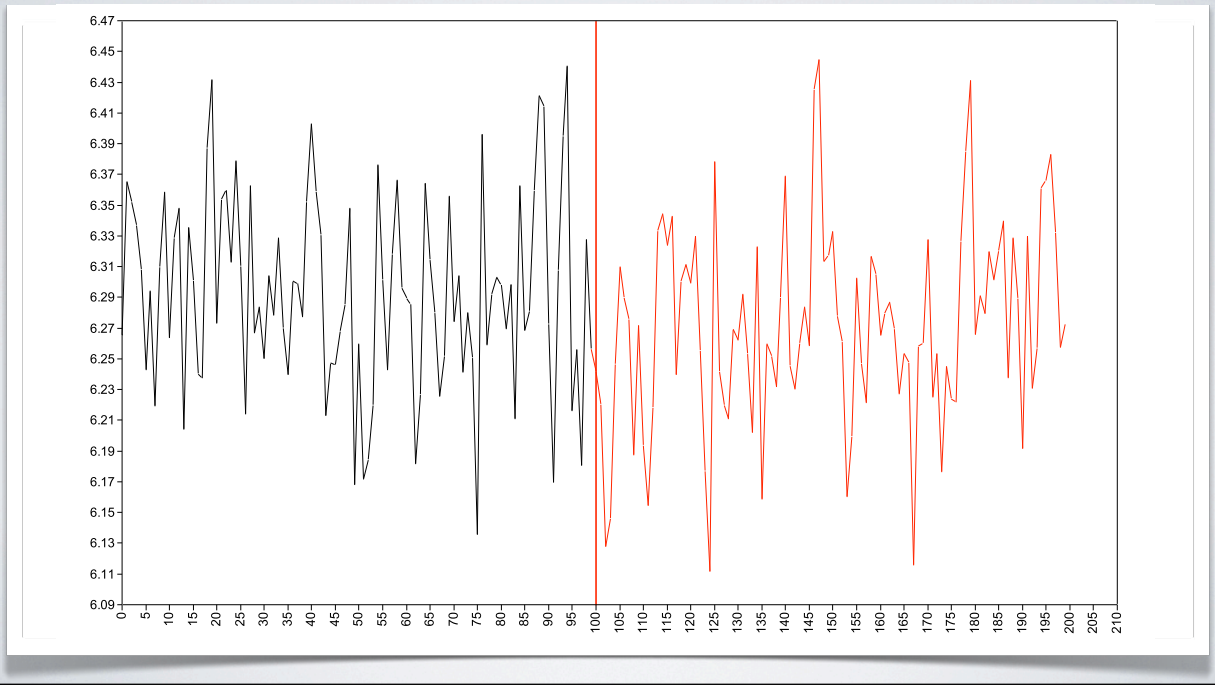
TRACE PLOT OF THE JOB SATISFACTION MEAN



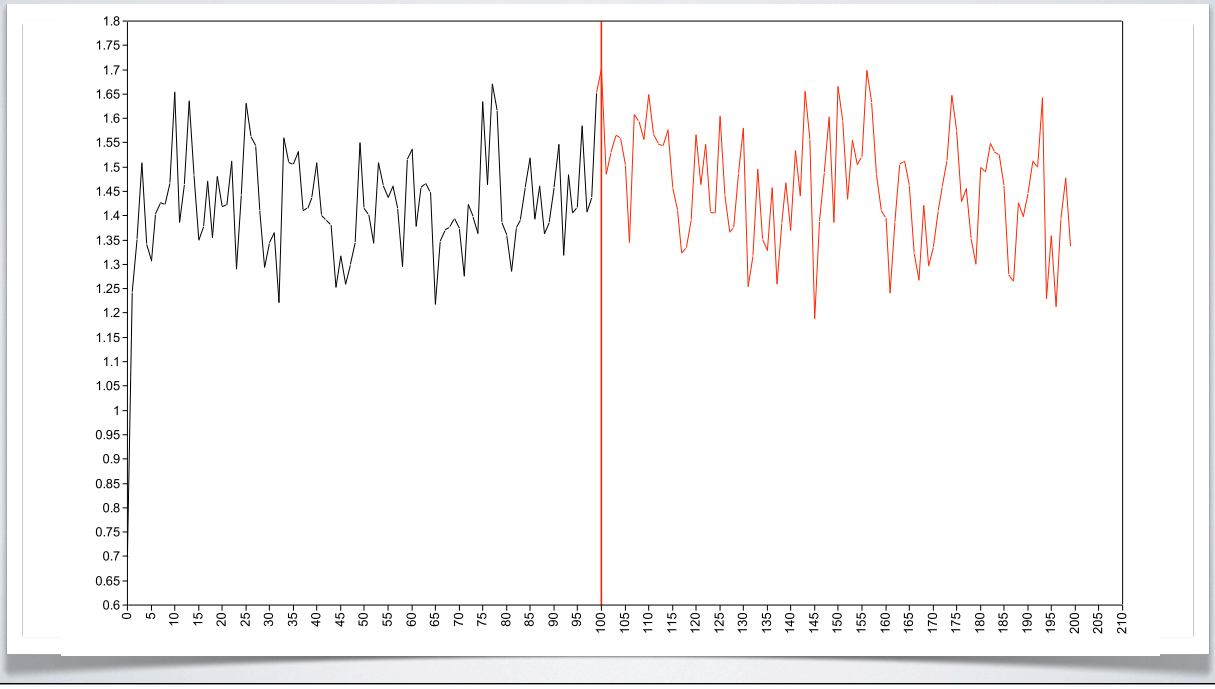
TRACE PLOT OF THE JOB SATISFACTION VARIANCE



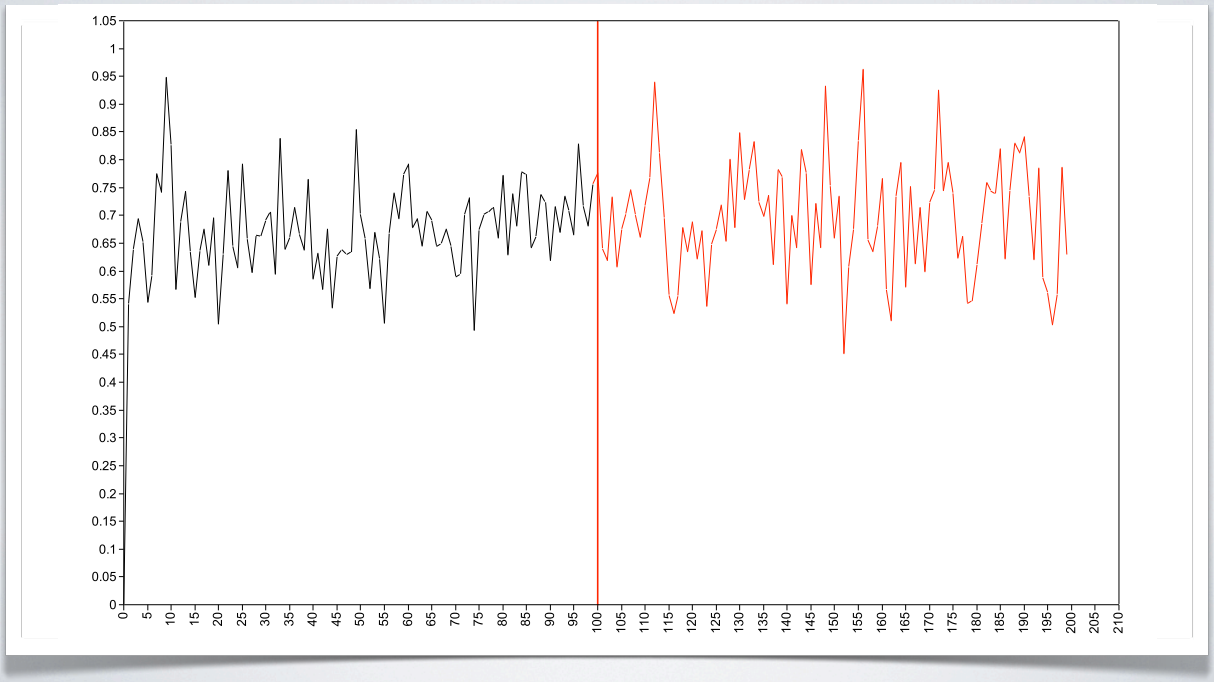
TRACE PLOT OF THE WELL-BEING MEAN



TRACE PLOT OF THE WELL-BEING VARIANCE



TRACE PLOT OF THE SATISFACTION - WELL-BEING COVARIANCE



MPLUS COMMANDS FOR IMPUTATION ANALYSIS

- TITLE (optional)
- DATA (same as ML analysis)
- VARIABLE (same as ML analysis)
- ANALYSIS
- DATA IMPUTATION
- OUTPUT

ANALYSIS COMMAND

- The following commands apply to the final MCMC run that generates the imputed data sets

```
ANALYSIS:
! Saturated imputation model;
type = basic;
! Random number seed for MCMC algorithm;
bseed = 48932;
! Convergence criterion (.05 = PSR < 1.05);
bconvergence = .05;
```

DATA IMPUTATION COMMAND

- The following commands apply to the final MCMC run that generates the imputed data sets

```
DATA IMPUTATION:
! Incomplete variables to be imputed;
! The USEVARIABLES list specifies other variables in the model;
impute = wbeing jobsat;
! Number of imputed data sets;
ndatasets = 50;
! File name prefix for imputed data sets;
save = employeeimp*.dat;
! Between-imputation interval;
thin = 300;
```

MPLUS IMPUTATION PROGRAM

```
DATA:
file = employee.dat;
VARIABLE:
names = id age tenure female wbeing jobsat jobperf turnover iq;
usevariables = age tenure wbeing jobsat jobperf iq;
missing = all (-99);
ANALYSIS:
type = basic;
bseed = 48932;
bconvergence = .05;
DATA IMPUTATION:
impute = wbeing jobsat;
ndatasets = 50;
save = employeeimp*.dat;
thin = 300;
OUTPUT:
tech8;
```

IMPUTATION OUTPUT FILE

- Mplus lists the variable order in the imputed data near the bottom of the output file
- Use this variable list for all subsequent analyses

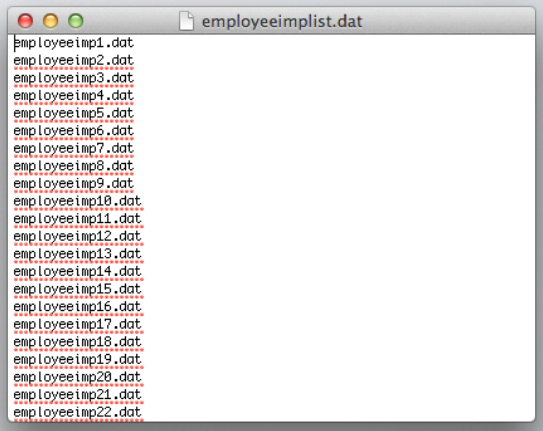
```
SAVEDATA INFORMATION

  Order of variables

    AGE
    TENURE
    WBEING
    JOBSAT
    JOBPERF
    IQ
```

IMPUTED DATA FORMAT

- Mplus saves each imputed data set to a separate file
- The file names use the prefix specified in the SAVE command (e.g., employeeimp1.dat, employeeimp2.dat, etc.)
- The program also generates a list file that contains the file names of the imputed data sets (e.g., employeeimplist.dat)
- The list file serves as input data for all subsequent analyses

THE IMPLIST FILE

- The imputation program generates a text file (e.g., employeimplist.dat) that contains the imputed data set names
- The implist file serves as input data for all subsequent analyses



ANALYZING IMPUTED DATA

- Mplus fully automates the analysis and pooling phases
- Analyzing imputed data sets requires a small change to the DATA command, but the remaining commands are identical to a complete-data analysis
- The analyses simplify a bit (e.g., no need to list incomplete predictors, no need to use the auxiliary command)

DATA COMMAND

- Omitted file path (implist file in the same directory as input file)
- TYPE = imputation specifies multiply imputed data sets and invokes pooling rules

```
DATA:
file = employeeimplist.dat;    ! List of imputation file names;
type = imputation;             ! Imputation data;
```

MPLUS ANALYSIS PROGRAM

```
DATA:
file = employeeimplist.dat;
type = imputation;
VARIABLE:
names = age tenure wbeing jobsat jobperf iq;
usevariables = wbeing jobsat jobperf;
ANALYSIS:
estimator = ml;
MODEL:
jobperf on wbeing (b1);
jobperf on jobsat (b2);
MODEL TEST:
b1 = 0;
b2 = 0;
OUTPUT:
standardized;
```

DESCRIPTIVES

```
SAMPLE STATISTICS

NOTE:  These are average results over 50 data sets.

     SAMPLE STATISTICS


           Means
              JOBPERF        WBEING         JOBSAT
              ________       ________       ________
     1          6.021          6.296          5.949


           Covariances
              JOBPERF        WBEING         JOBSAT
              ________       ________       ________
JOBPERF        1.570
WBEING         0.676          1.380
JOBSAT         0.263          0.448          1.404
```

DESCRIPTIVES, CONTINUED

Correlations	JOBPERF	WBEING	JOBSAT
JOBPERF	1.000		
WBEING	0.459	1.000	
JOBSAT	0.177	0.322	1.000

WALD TEST (MODEL TEST COMMAND)

- The Wald statistic (a chi-square with 2 degrees of freedom) is akin to the omnibus *F* test in OLS regression

Wald Test of Parameter Constraints	
Value	136.976
Degrees of Freedom	2
P-Value	0.0000

- The significant chi-square, $\chi^2(2) = 136.976$, indicates that the set of predictors explain significant variation in the dependent variable

UNSTANDARDIZED ESTIMATES

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
JOBPERF ON				
WBEING	0.479	0.055	8.783	0.000
JOBSAT	0.035	0.056	0.622	0.534
Intercepts				
JOBPERF	2.797	0.402	6.959	0.000
Residual Variances				
JOBPERF	1.236	0.088	14.052	0.000

INTERPRETATIONS

- Interpret and report MI estimates in the same way as a complete-data analysis
- Controlling for job satisfaction, a one-point increase in psychological well-being results in a .479 increase in job performance, on average
- Controlling for psychological well-being, a one-point increase job satisfaction in results in a .035 increase in job performance, on average

STANDARDIZED ESTIMATES (STANDARDIZED OPTION)

```
STANDARDIZED MODEL RESULTS

STDYX Standardization

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 JOBPERF  ON
    WBEING           0.449      0.048      9.400      0.000
    JOBSAT           0.033      0.053      0.622      0.534
```

STANDARDIZED ESTIMATES, CONTINUED

```
R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    JOBPERF            0.213      0.041      5.250      0.000
```

INTERPRETATIONS

- The STDYX standardization gives beta weights
  - Controlling for job satisfaction, a one standard deviation increase in psychological well-being results in a .449 standard deviation increase in job performance, on average
  - Controlling for psychological well-being, a one standard deviation increase job satisfaction in results in a .033 standard deviation increase in job performance, on average
- Together, the two predictors explain 21.3% of the variance in job performance ratings