Missing Data Workshop Exercises

Data For Workshop Exercises

The examples are inspired by an Institute of Educational Sciences-funded project featuring a cluster-randomized trial of a novel math problem-solving intervention

Variables include baseline and end-of-year problem-solving test scores and a number of student covariates and background variables

exercise.csv

Variable	Description	Missing	Metric
student	Student identifier		Nominal
abilitygrp	Ability grouping (3-group classification)	*	Nominal
female	Female dummy code		Nominal
stanmath	Standardized math test scores	*	Numeric
frlunch	Lunch assistance dummy code	*	Nominal
efficacy	Math self-efficacy rating scale	*	Ordinal
probsolve1	Math problem-solving score at baseline	*	Numeric
probsolve7	Math problem-solving score at final wave	*	Numeric

Analysis Model For Exercise 1

The analysis model is a multiple regression with baseline problem-solving scores, gender, and standardized math scores predicting end-of-year problem-solving

$$Probsolve7_i = \beta_0 + \beta_1 (Probsolve1_i) + \beta_2 (Female_i) + \beta_3 (Stanmath_i) + e_i$$

Exercise 1

1. Perform a diagnostic analysis to evaluate convergence of the MCMC algorithm

2. Using the information that you gleaned from the diagnostic run, generate 20 or more imputations for analysis in the software package of your choosing

3. Estimate the model and pool the estimates and standard errors in the software package of your choosing

Analysis Model For Exercise 2

The analysis model is a multiple regression with a mixture of categorical and continuous variables

Ability group is a three-group nominal variable (1 = learning disabled, 2 = low achieving, 3 = average achieving) represented as two dummy variables in the analysis model

$$Probsolve7_i = \beta_0 + \beta_1 (Probsolve1_i) + \beta_2 (LearnDis_i) + \beta_3 (LowAch_i) + \beta_4 (Efficacy_i) + e_i$$

Exercise 2

1. Perform a diagnostic analysis to evaluate convergence of the MCMC algorithm. Use standardized math scores and the lunch assistance dummy variable as auxiliary variables.

2. Using the information that you gleaned from the diagnostic run, generate 20 or more imputations for analysis in the software package of your choosing

3. Estimate the model and pool the estimates and standard errors in the software package of your choosing

Analysis Model For Exercise 3

The analysis model is a multiple regression with an interaction effect involving math self-efficacy and gender (dummy coded as 0 = male and 1 = female)

$$Probsolve7_i = \beta_0 + \beta_1 (Probsolve1_i) + \beta_2 (Efficacy_i) + \beta_3 (Female_i)$$
$$+\beta_4 (Efficacy_i)(Female_i) + e_i$$

Exercise 3

1. Perform a diagnostic analysis to evaluate convergence of the MCMC algorithm. Use standardized math scores and the lunch assistance dummy variable as auxiliary variables.

2. Using the information that you gleaned from the diagnostic run, generate 20 or more imputations for analysis in the software package of your choosing

3. Estimate the model and pool the estimates and standard errors in the software package of your choosing

Analysis Model For Exercise 4

The analysis model is a multiple regression with baseline problem-solving scores, gender, and standardized math scores predicting end-of-year problem-solving

$$Probsolve7_i = \beta_0 + \beta_1 (Probsolve1_i) + \beta_2 (Female_i) + \beta_3 (Stanmath_i) + e_i$$

Exercise 4

1. Write an Mplus program to fit the analysis model with full information maximum likelihood estimation. Specify a normal distribution for all predictor variables to avoid excluding observations.

Exercise 5

1. Write an Mplus program to fit the analysis model with full information maximum likelihood estimation. Specify a normal distribution for all predictor variables to avoid excluding observations.

2. Use either the saturated correlates or the extra dependent variable approaches to include the lunch assistance indicator and math self-efficacy as auxiliary variables.