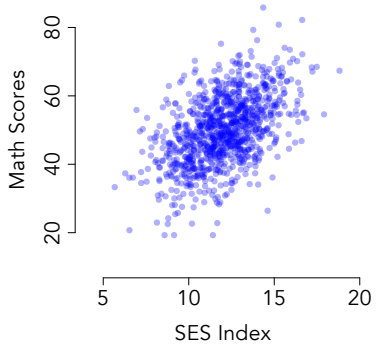
2: Older Missing Data Handling Methods

Motivating Example

Bivariate scenario involving SES and math achievement

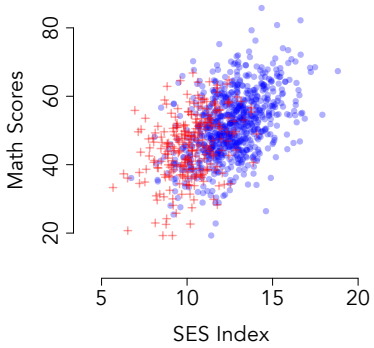
Use simulated data to illustrate biases associated with different methods



Missing At Random Mechanism

25% of the math scores are missing due to low SES (e.g., high mobility)

MAR mechanism where missingness is related to observed data



Listwise Deletion

Listwise deletion (i.e., complete-case analysis) removes all cases with complete data

Convenient, the resulting data set is complete

Valid estimates require an MCAR mechanism, and even then discarding data decreases power

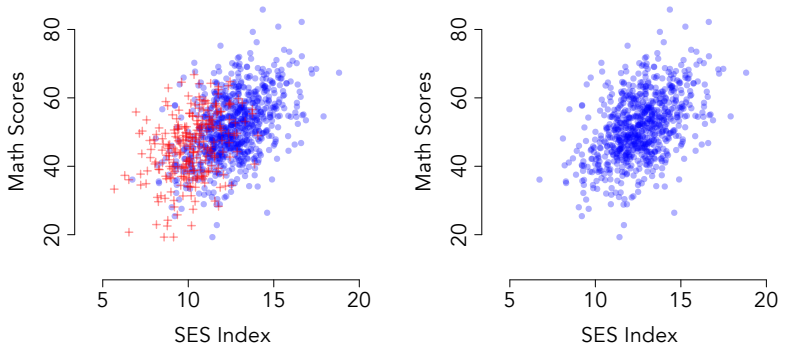
Pairwise Deletion

Pairwise deletion (i.e., available-case analysis) removes data on an analysis-by-analysis basis

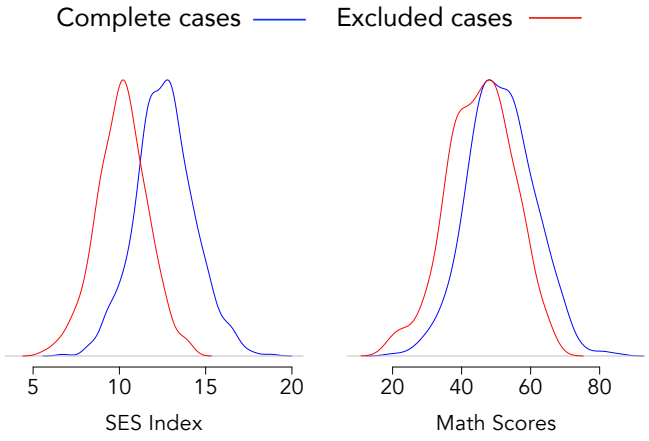
Valid estimates require an MCAR mechanism, but discarding data decreases power

Sample size varies across pairs of variables, so it isn't clear which N to use for standard errors

Listwise Deletion Scatterplot



Comparison Of Distributions



Comparison Of Estimates

Complete-Data Estimates

	M	SD	r
SES	12.00	1.96	0.50
Math	49.93	9.90	

Listwise Deletion Estimates

	M	SD	r
SES	12.63	1.70	0.46
Math	51.47	9.59	

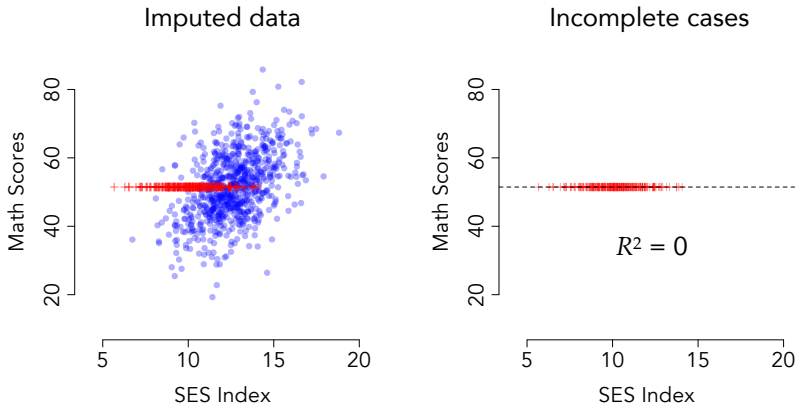
Mean Imputation

Mean imputation replaces missing values with a variable's average (i.e., the column mean), computed from the complete cases

Variability and correlations are attenuated because imputations are constant

Estimates are biased under any mechanism, so one of the worst options available

Mean Imputation Scatterplot



Comparison Of Estimates

Complete-Data Estimates

	M	*SD*	*r*
SES	12.00	1.96	0.50
Math	49.93	9.90	

Mean Imputation Estimates

	M	*SD*	*r*
SES	12.00	1.96	0.35
Math	51.47	8.29	

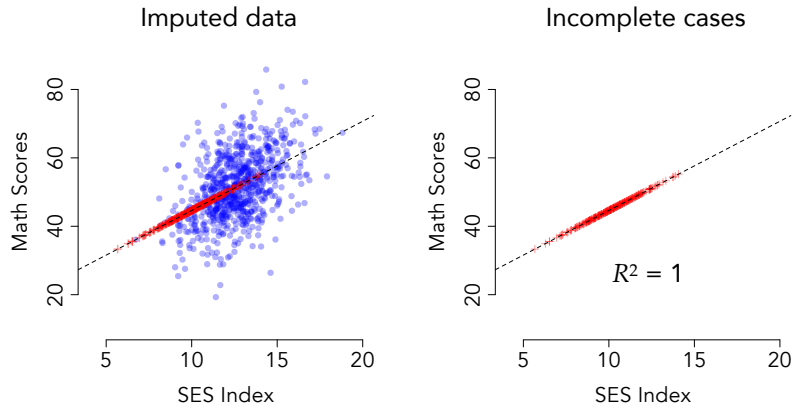
Regression Imputation

An incomplete variable is regressed on complete variables, and the regression equation generates predicted values that replace missing scores

Imputations lack variability because they fall directly on a regression line (or plane)

Means are valid under an MCAR mechanism, but measures of variation and association are biased

Regression Imputation Scatterplot



Comparison Of Estimates

Complete-Data Estimates

	M	*SD*	*r*
SES	12.00	1.96	0.50
Math	49.93	9.90	

Mean Imputation Estimates

	M	*SD*	*r*
SES	12.00	1.96	0.57
Math	49.84	8.96	

Last Observation Carried Forward (LOCF)

LOCF replaces missing values in a longitudinal design with the last available observed score

Person	Wave 1	Wave 2	Wave 3
1	1	2	1
2	5	NA	4
3	3	2	4
4	3	NA	NA

Person	Wave 1	Wave 2	Wave 3
1	1	2	1
2	5	5	4
3	3	2	4
4	3	3	3

LOCF Continued

Conventional wisdom says LOCF is conservative because it assumes no post-attrition change

Methodological work tells us that LOCF can attenuate or exaggerate developmental trajectories and is generally prone to bias

LOCF is a poor strategy for longitudinal data …