

Older Missing Data Handling Approaches

1

A Small Collection of Methods

Listwise and pairwise deletion

Mean imputation

Regression imputation

2

Motivating Example

Number of years smoking and
number of cigarettes smoked

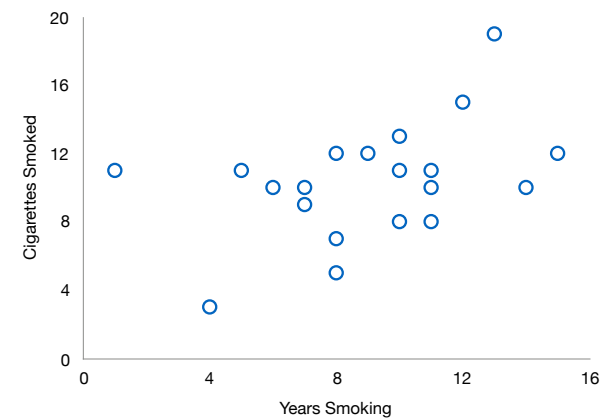
More years smoking is
associated with a higher
likelihood of not reporting
smoking behavior (MAR)

Self-efficacy is missing
completely at random (MCAR)

Years	Cigs	Efficacy
7	9	NA
8	NA	NA
1	11	16
4	3	21
6	10	17
8	5	10
8	7	13
10	NA	10
15	NA	11
5	11	13
9	12	11
11	11	16
14	NA	10
13	19	9
12	NA	5
11	8	7
10	13	10
10	8	NA
7	10	7
11	10	6

3

Hypothetical Complete-Data Scatterplot



4

Listwise Deletion

Listwise deletion (complete-case analysis) removes all cases with complete data

The resulting data set is complete

Estimates are valid with an MCAR mechanism, but eliminating data reduces power

5

Pairwise Deletion

Pairwise deletion (available-case analysis) removes data on an analysis-by-analysis basis

The effective sample size varies across analyses

Estimates are valid with an MCAR mechanism, but eliminating data reduces power and standard errors from multivariate analyses are ambiguous

6

Listwise and Pairwise Correlation Matrices

Listwise Deletion (Complete-Case)

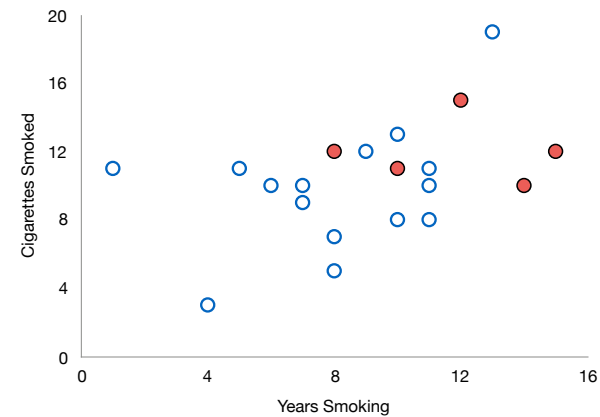
	Years	Cigs	Efficacy
Years	1.00 N = 13		
Cigs	.43 N = 13	1.00 N = 13	
Efficacy	-.62 N = 13	-.34 N = 13	1.00 N = 13

Pairwise Deletion (Available-Case)

	Years	Cigs	Efficacy
Years	1.00 N = 20		
Cigs	.40 N = 15	1.00 N = 17	
Efficacy	-.59 N = 17	-.34 N = 13	1.00 N = 15

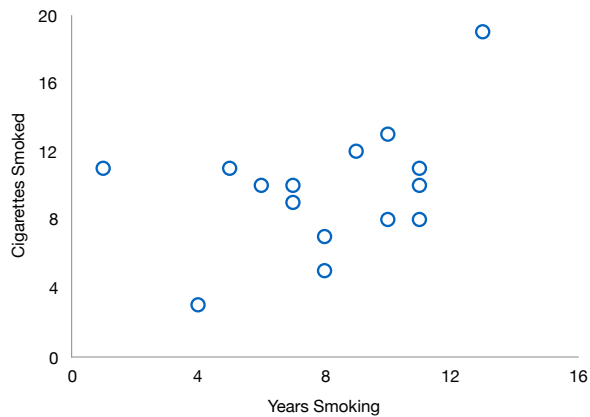
7

Deletion Scatterplot



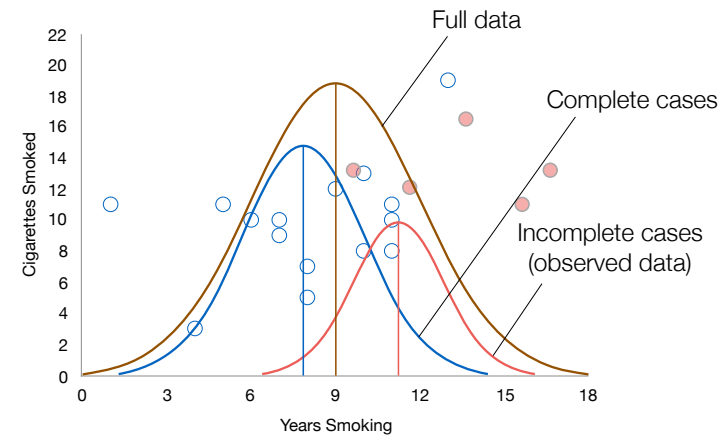
8

Deletion Scatterplot



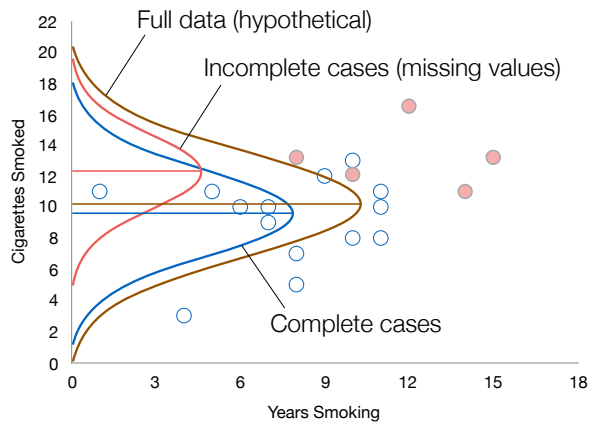
9

Distribution of Years Smoking



10

Distribution of Cigarettes Smoked



11

Mean Imputation

Mean imputation replaces (imputes) missing values with the average of the available scores

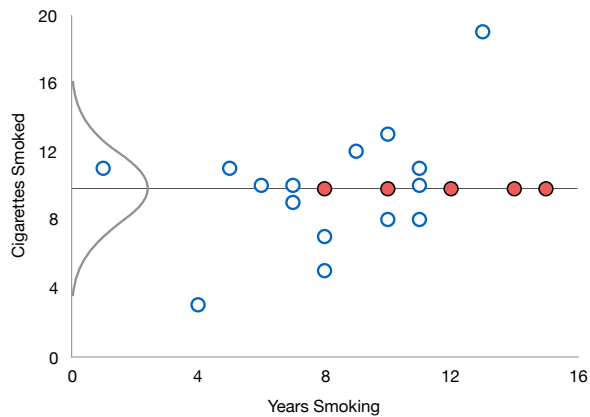
Variability and correlations are attenuated because imputations are constant

Estimates are biased under any mechanism

Probably the worst possible option ...

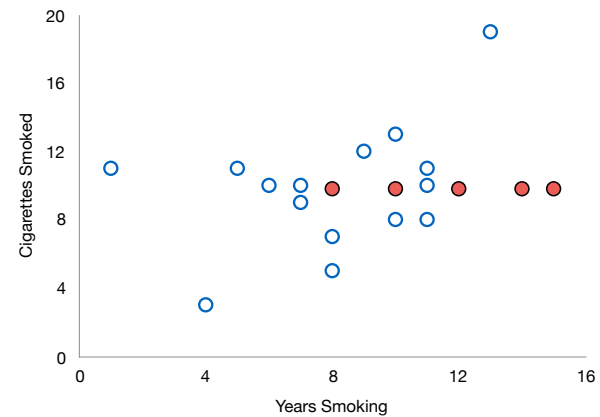
12

Mean Imputation



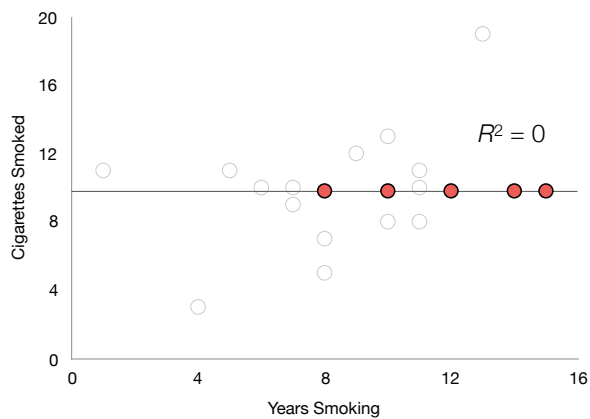
13

Mean Imputation Scatterplot



14

Imputed Values



15

Regression (Conditional Mean) Imputation

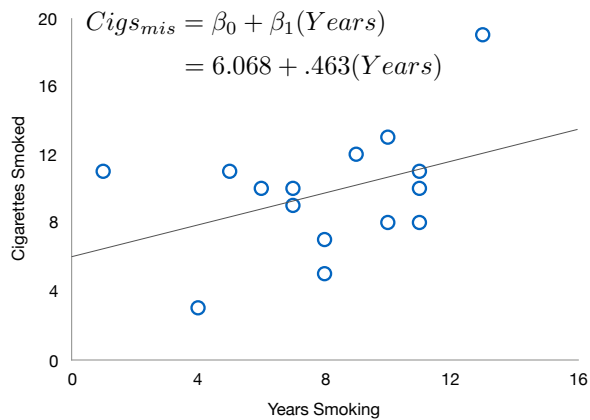
Regression imputation replaces missing values with predicted scores from a regression equation where the complete variables predict the incomplete variables

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

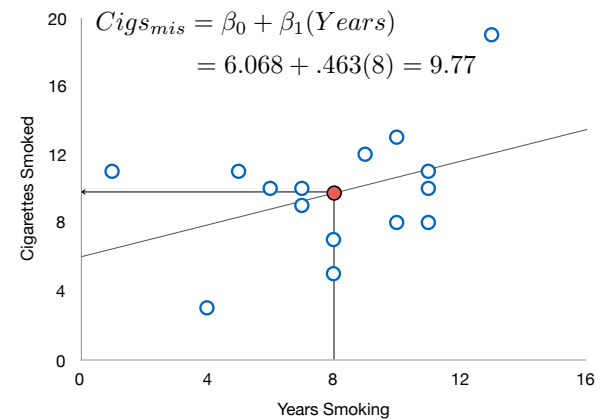
16

Regression Imputation



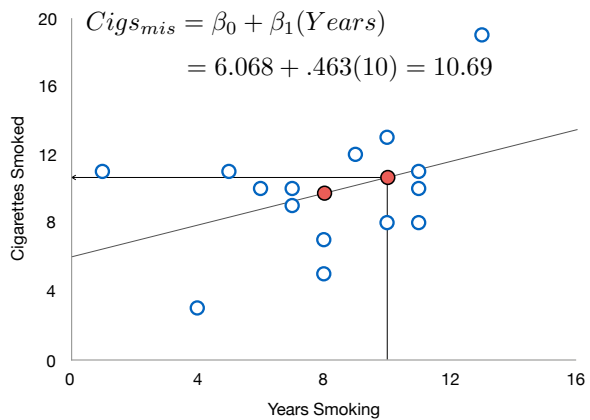
17

Imputation Example: Years = 8



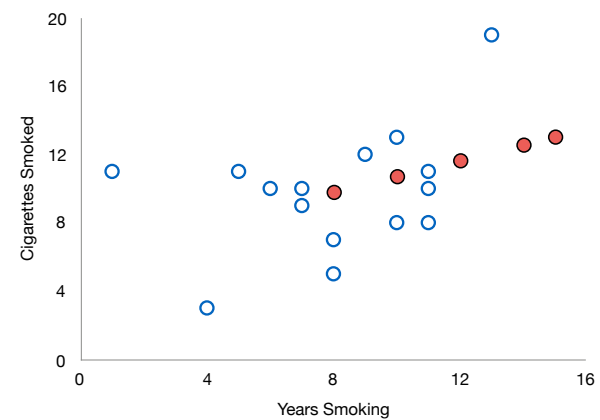
18

Imputation Example: Years = 10



19

Regression Imputation Scatterplot



20

Imputed Values

