Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Overview of Multiple Imputation

Jonathan Lee Helm
Friday May 17th, 2019

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Single Imputation

Single Imputation

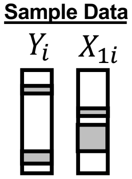
- Replace missing values with a 'guess'
  - Different approaches for choosing the guess

Single Imputation

- Replace missing values with a 'guess'
  - *Creates a complete data set*
    - *Different approaches for choosing the guess*
- Analyze complete data

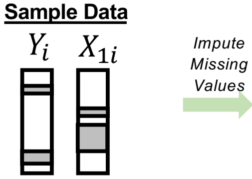
Conceptual Diagram of Single Imputation

Sample Data

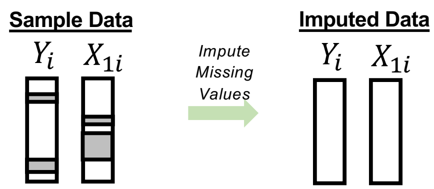


Conceptual Diagram of Single Imputation

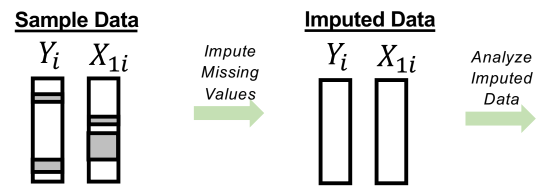
Sample Data



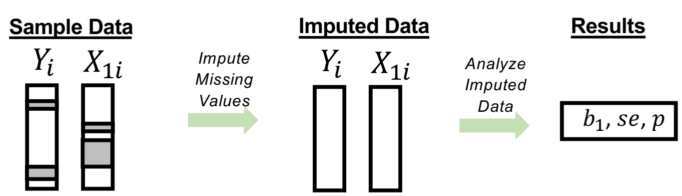
Conceptual Diagram of Single Imputation



Conceptual Diagram of Single Imputation



Conceptual Diagram of Single Imputation



Single Imputation

- How can we create a guess?

Single Imputation

- How can we create a guess?
- Conceptually, the simplest way is through regression

Observed Data

i	JS^{obs}	IQ^{obs}
1	--	78
2	--	84
3	--	84
4	--	85
5	--	87
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Observed Data

i	JS^{obs}	IQ^{obs}
1	--	78
2	--	84
3	--	84
4	--	85
5	--	87
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Regression:

$JS^{obs} = b_0 + b_1 IQ^{obs}$

Observed Data

i	JS^{obs}	IQ^{obs}
1	--	78
2	--	84
3	--	84
4	--	85
5	--	87
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Regression:

$JS^{obs} = b_0 + b_1 IQ^{obs}$

	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Observed Data

i	JS^{obs}	IQ^{obs}
1	--	78
2	--	84
3	--	84
4	--	85
5	--	87
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Regression:

$JS^{obs} = b_0 + b_1 IQ^{obs}$

	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imputed Data

i	JS^{imp}	IQ^{imp}
1	7.56	78
2	8.31	84
3	8.31	84
4	8.43	85
5	8.68	87
6	9.17	91
7	9.29	92
8	9.54	94
9	9.54	94
10	9.79	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Observed Data

i	JS^{obs}	IQ^{obs}
1	--	78
2	--	84
3	--	84
4	--	85
5	--	87
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imputed Data

i	JS^{imp}	IQ^{imp}
1	7.56	78
2	8.31	84
3	8.31	84
4	8.43	85
5	8.68	87
6	9.17	91
7	9.29	92
8	9.54	94
9	9.54	94
10	9.79	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Mean for JS^{obs} = 11.7
SD for JS^{obs} = 2.71

Mean for JS^{imp} = 10.28
SD for JS^{imp} = 2.42

Complete Data

i	JS^{com}	IQ^{com}
1	9	78
2	13	84
3	10	84
4	8	85
5	7	87
6	7	91
7	9	92
8	9	94
9	11	94
10	7	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imputed Data

i	JS^{imp}	IQ^{imp}
1	7.56	78
2	8.31	84
3	8.31	84
4	8.43	85
5	8.68	87
6	9.17	91
7	9.29	92
8	9.54	94
9	9.54	94
10	9.79	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Mean for JS^{obs} = 11.7
SD for JS^{obs} = 2.71

Mean for JS^{imp} = 10.28
SD for JS^{imp} = 2.42

Mean for JS^{com} = 10.35
SD for JS^{com} = 2.68

Single Imputation

- We impute values based on the observed data
- This will work well when data are MCAR or MAR, but not MNAR

Single Imputation: MCAR

- Is single imputation reasonable?
- If data are MCAR, then the imputed values will just be random guesses
  - *This should not impact parameter estimates*
  - *We will use a larger sample size, so decreased standard errors*

Single Imputation: MAR

- Is single imputation reasonable?
- If the data are MAR, then the other variables in the model that relate to missingness will create good predicted values
  - *This should create less biased parameter estimates*
  - *Increase sample size*

Single Imputation: MNAR

- Is single imputation reasonable?
- If the data are MNAR, then the other variables in the analysis won't account for missingness
  - *This won't fully account for the bias*

Single Imputation: Limitation

- The limitation is that we are not accounting for the uncertainty of the regression

Single Imputation

- The limitation is that we are not accounting for the uncertainty of the regression

Regression:

$JS^{obs} = b_0 + b_1 IQ^{obs}$

$\sigma_\varepsilon^2 = 2.95$
$\sigma_\varepsilon = 1.72$

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

@jonathanleehelm 25

Single Imputation

- We can take certainty into accounted by creating multiple data sets

Regression:

$JS^{obs} = b_0 + b_1 IQ^{obs}$

$\sigma_\varepsilon^2 = 2.95$
$\sigma_\varepsilon = 1.72$

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

@jonathanleehelm 26

Single Imputation

- We can take certainty into accounted by creating multiple data sets (***Multiple imputation***)

Regression:

$JS^{obs} = b_0 + b_1 IQ^{obs}$

$\sigma_\varepsilon^2 = 2.95$
$\sigma_\varepsilon = 1.72$

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

@jonathanleehelm 27

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

@jonathanleehelm 28

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

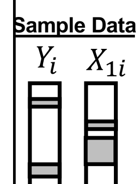
Multiple Imputation

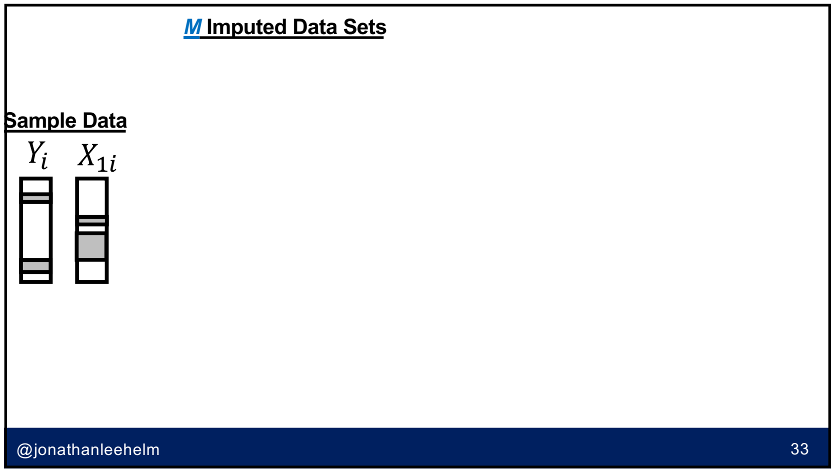
Multiple Imputation

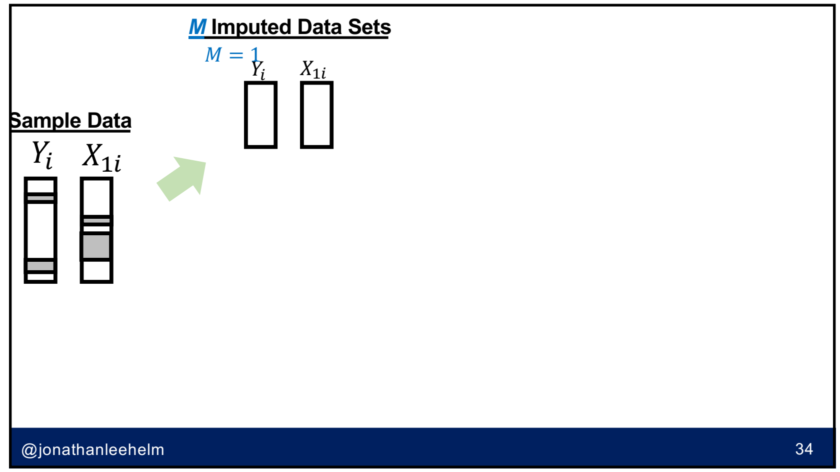
- Multiple imputation extends single imputation by creating/analyzing more than one imputed data set

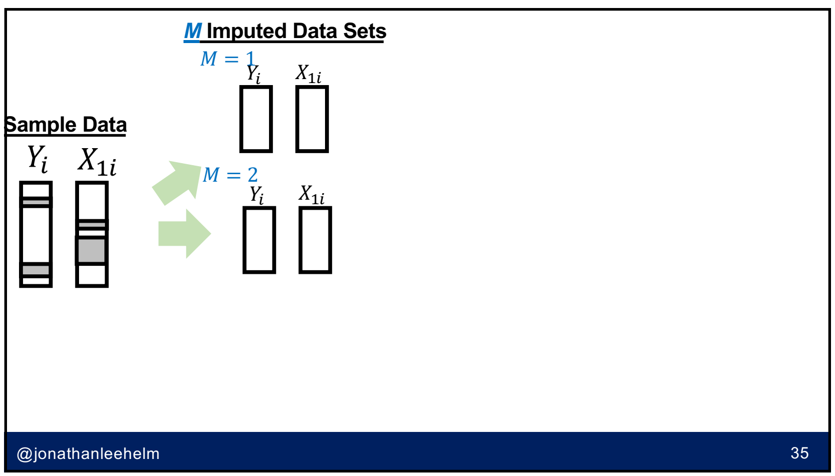
Multiple Imputation

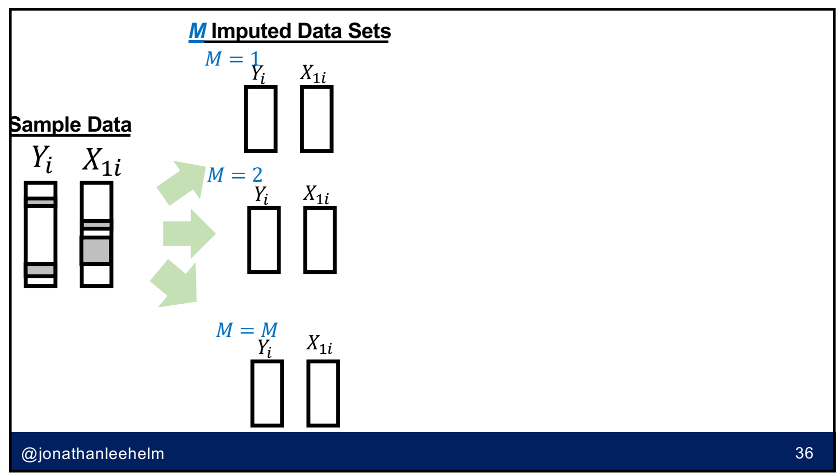
- Multiple imputation extends single imputation by creating/analyzing more than one imputed data set
- We create *M* imputed data sets
- Each data set includes some uncertainty for the imputed value

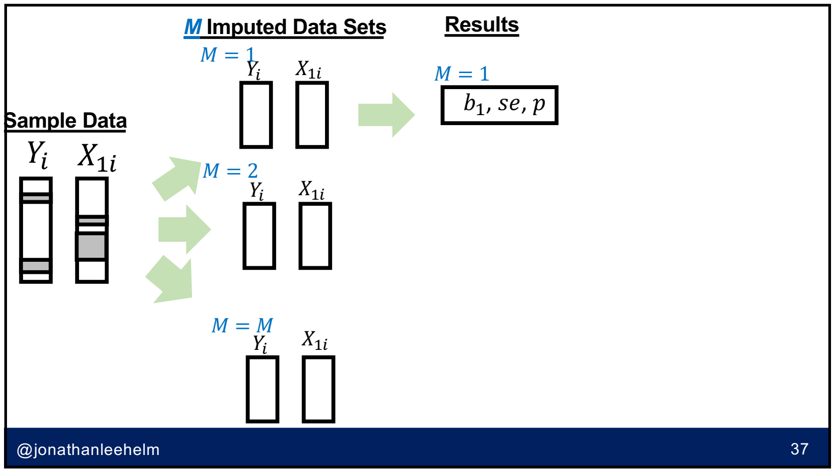


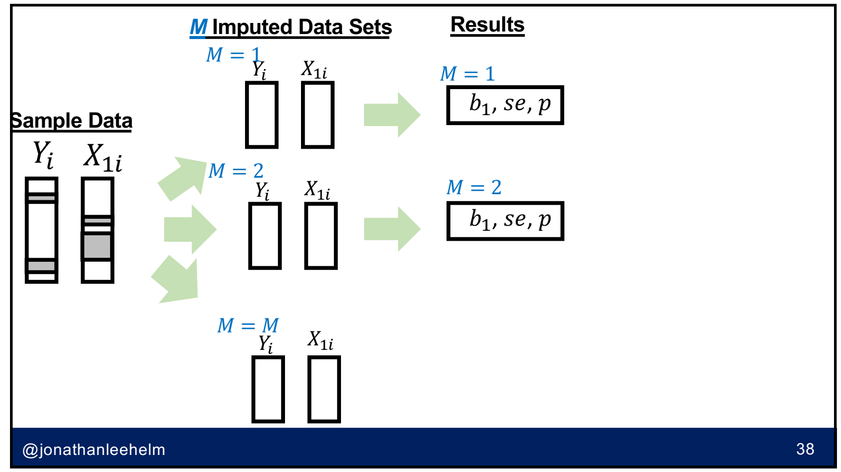

M Imputed Data Sets
Sample Data
Y_i X_{1i}
@jonathanleehelm
33

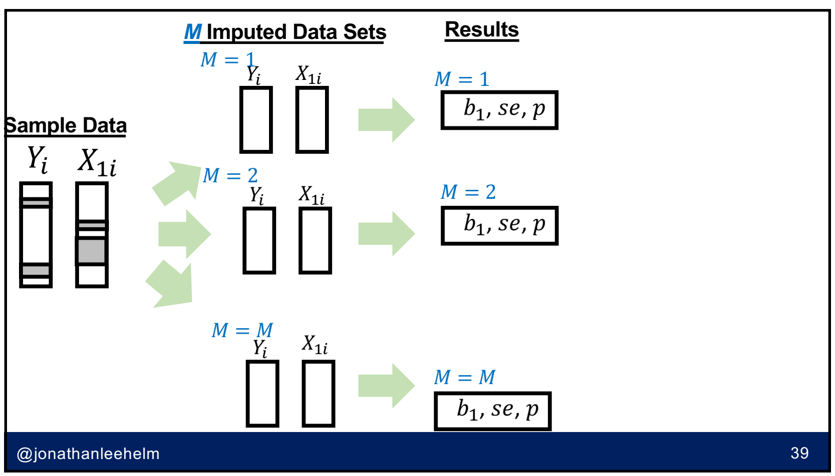

M Imputed Data Sets
$M = 1$
Y_i X_{1i}
Sample Data
Y_i X_{1i}
@jonathanleehelm
34

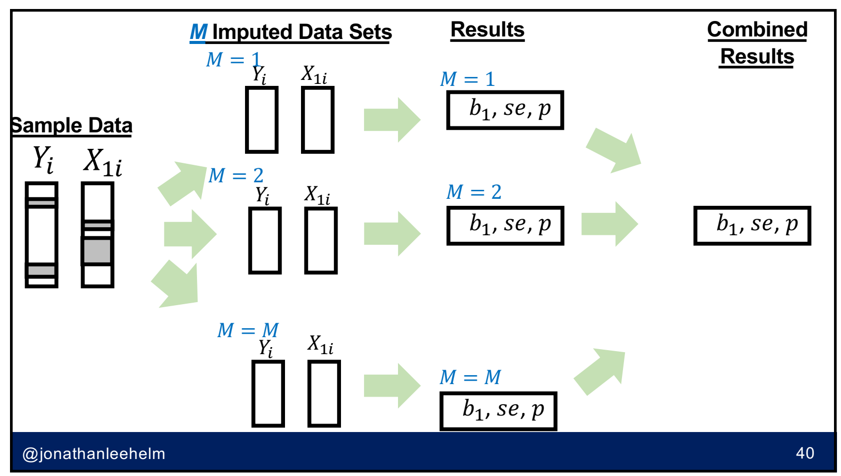

M Imputed Data Sets
$M = 1$
Y_i X_{1i}
Sample Data
Y_i X_{1i}
$M = 2$
Y_i X_{1i}
@jonathanleehelm
35


M Imputed Data Sets
$M = 1$
Y_i X_{1i}
Sample Data
Y_i X_{1i}
$M = 2$
Y_i X_{1i}
$M = M$
Y_i X_{1i}
@jonathanleehelm
36


M Imputed Data Sets
Results
Sample Data
Y_i X_{1i}
$M = 1$
$M = 2$
$M = M$
b_1, se, p
@jonathanleehelm
37


M Imputed Data Sets
Results
Sample Data
Y_i X_{1i}
$M = 1$
$M = 2$
$M = M$
b_1, se, p
b_1, se, p
@jonathanleehelm
38


M Imputed Data Sets
Results
Sample Data
Y_i X_{1i}
$M = 1$
$M = 2$
$M = M$
b_1, se, p
b_1, se, p
b_1, se, p
@jonathanleehelm
39


M Imputed Data Sets
Results
Combined Results
Sample Data
Y_i X_{1i}
$M = 1$
$M = 2$
$M = M$
b_1, se, p
b_1, se, p
b_1, se, p
b_1, se, p
@jonathanleehelm
40

Observed Data		
i	JS^{obs}	IQ^{obs}
1	--	78
2	--	84
3	--	84
4	--	85
5	--	87
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Observed Data			Imp Data M = 1		
i	JS^{obs}	IQ^{obs}	*i*	JS^{imp}	IQ^{imp}
1	--	78	1	15	78
2	--	84	2	7	84
3	--	84	3	10	84
4	--	85	4	10	85
5	--	87	5	15	87
6	--	91	6	11	91
7	--	92	7	10	92
8	--	94	8	15	94
9	--	94	9	10	94
10	--	96	10	10	96
11	7	99	11	7	99
⋮	⋮	⋮	⋮	⋮	⋮
19	16	118	19	16	118
20	12	134	20	12	134

Observed Data			Imp Data M = 1			Imp Data M = 2		
i	JS^{obs}	IQ^{obs}	*i*	JS^{imp}	IQ^{imp}	*i*	JS^{imp}	IQ^{imp}
1	--	78	1	15	78	1	11	78
2	--	84	2	7	84	2	7	84
3	--	84	3	10	84	3	10	84
4	--	85	4	10	85	4	10	85
5	--	87	5	15	87	5	10	87
6	--	91	6	11	91	6	10	91
7	--	92	7	10	92	7	7	92
8	--	94	8	15	94	8	15	94
9	--	94	9	10	94	9	11	94
10	--	96	10	10	96	10	15	96
11	7	99	11	7	99	11	7	99
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	16	118	19	16	118	19	16	118
20	12	134	20	12	134	20	12	134

Observed Data			Imp Data M = 1			Imp Data M = 2			Imp Data M = 3		
i	JS^{obs}	IQ^{obs}	*i*	JS^{imp}	IQ^{imp}	*i*	JS^{imp}	IQ^{imp}	*i*	JS^{imp}	IQ^{imp}
1	--	78	1	15	78	1	11	78	1	7	78
2	--	84	2	7	84	2	7	84	2	7	84
3	--	84	3	10	84	3	10	84	3	15	84
4	--	85	4	10	85	4	10	85	4	10	85
5	--	87	5	15	87	5	10	87	5	10	87
6	--	91	6	11	91	6	10	91	6	10	91
7	--	92	7	10	92	7	7	92	7	10	92
8	--	94	8	15	94	8	15	94	8	7	94
9	--	94	9	10	94	9	11	94	9	10	94
10	--	96	10	10	96	10	15	96	10	7	96
11	7	99	11	7	99	11	7	99	11	7	99
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
19	16	118	19	16	118	19	16	118	19	16	118
20	12	134	20	12	134	20	12	134	20	12	134

Combining Results

- Perform the analysis on each data set separately
- Collect the statistics of interest
  - *q*
    - *e.g., Mean, SD, regression coefficient*
- Calculated the average across the statistics

Observed Data

i	JS^obs	IQ^obs
1	--	78

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

i	JS^obs	IQ^obs
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 1

i	JS^imp	IQ^imp
1	15	78
2	7	84
3	10	84
4	10	85
5	15	87
6	11	91
7	10	92
8	15	94
9	10	94
10	10	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 2

i	JS^imp	IQ^imp
1	11	78
2	7	84
3	10	84
4	10	85
5	10	87
6	10	91
7	7	92
8	15	94
9	11	94
10	15	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 3

i	JS^imp	IQ^imp
1	7	78
2	7	84
3	15	84
4	10	85
5	10	87
6	10	91
7	10	92
8	7	94
9	10	94
10	7	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Observed Data

i	JS^obs	IQ^obs
1	--	78

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

i	JS^obs	IQ^obs
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 1

i	JS^imp	IQ^imp
1	15	78
2	7	84
3	10	84

	Est.	s.e.	*p*
b_0	5.09	4.22	.24
b_1	.064	.04	.14

i	JS^imp	IQ^imp
8	15	94
9	10	94
10	10	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 2

i	JS^imp	IQ^imp
1	11	78
2	7	84
3	10	84
4	10	85
5	10	87
6	10	91
7	7	92
8	15	94
9	11	94
10	15	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 3

i	JS^imp	IQ^imp
1	7	78
2	7	84
3	15	84
4	10	85
5	10	87
6	10	91
7	10	92
8	7	94
9	10	94
10	7	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Observed Data

i	JS^obs	IQ^obs
1	--	78

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

i	JS^obs	IQ^obs
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 1

i	JS^imp	IQ^imp
1	15	78
2	7	84
3	10	84

	Est.	s.e.	*p*
b_0	5.09	4.22	.24
b_1	.064	.04	.14

i	JS^imp	IQ^imp
8	15	94
9	10	94
10	10	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 2

i	JS^imp	IQ^imp
1	11	78
2	7	84
3	10	84
4	10	85

	Est.	s.e.	*p*
b_0	4.64	3.03	.14
b_1	.064	.03	.05

i	JS^imp	IQ^imp
9	11	94
10	15	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Imp Data M = 3

i	JS^imp	IQ^imp
1	7	78
2	7	84
3	15	84
4	10	85
5	10	87
6	10	91
7	10	92
8	7	94
9	10	94
10	7	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

Observed Data

i	JS^{obs}	IQ^{obs}
1	--	78
6	--	91
7	--	92
8	--	94
9	--	94
10	--	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1

i	JS^{imp}	IQ^{imp}
1	15	78
2	7	84
3	10	84
7	10	92
8	15	94
9	10	94
10	10	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

	Est.	s.e.	*p*
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2

i	JS^{imp}	IQ^{imp}
1	11	78
2	7	84
3	10	84
4	10	85
9	11	94
10	15	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

	Est.	s.e.	*p*
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3

i	JS^{imp}	IQ^{imp}
1	7	78
2	7	84
3	15	84
4	10	85
5	10	87
10	7	96
11	7	99
⋮	⋮	⋮
19	16	118
20	12	134

	Est.	s.e.	*p*
b_0	4.09	2.99	.19
b_1	.069	.03	.03

Observed Data

	Est.	s.e.	*p*
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1

	Est.	s.e.	*p*
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2

	Est.	s.e.	*p*
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3

	Est.	s.e.	*p*
b_0	4.09	2.99	.19
b_1	.069	.03	.03

$$\frac{5.09 + 4.64 + 4.09}{3} = 4.60$$

$$\frac{.064 + .064 + .069}{3} = .066$$

Pooled Estimates across Imputations

	Est.	s.e.	*p*
b_0	4.60	3.51	.21
b_1	.066	.034	.07

Combining Results

- What about standard errors?

Combining Results

- What about standard errors?
  - Within imputation variability (W):
    - Average the squares of the s.e. (variances within each imputation)
  - Between imputation variability (B):
    - Include the variance across imputation estimates

Observed Data	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1	Est.	s.e.	p
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2	Est.	s.e.	p
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3	Est.	s.e.	p
b_0	4.09	2.99	.19
b_1	.069	.03	.03

$$W = \frac{4.22^2 + 3.03^2 + 2.99^2}{3}$$

Pooled Estimates across Imputations

	Est.	s.e.	p
b_0	4.60	3.51	.21
b_1	.066	.034	.07

@jonathanleehelm 53

Observed Data	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1	Est.	s.e.	p
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2	Est.	s.e.	p
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3	Est.	s.e.	p
b_0	4.09	2.99	.19
b_1	.069	.03	.03

$$W = \frac{4.22^2 + 3.03^2 + 2.99^2}{3}$$

$$B = var(5.09, 4.64, 4.09)$$

Pooled Estimates across Imputations

	Est.	s.e.	p
b_0	4.60	3.51	.21
b_1	.066	.034	.07

@jonathanleehelm 54

Observed Data	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1	Est.	s.e.	p
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2	Est.	s.e.	p
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3	Est.	s.e.	p
b_0	4.09	2.99	.19
b_1	.069	.03	.03

$$W = \frac{4.22^2 + 3.03^2 + 2.99^2}{3}$$

$$B = var(5.09, 4.64, 4.09)$$

$$V_{b_0} = W + \left(1 + \frac{1}{M}\right)B$$

Pooled Estimates across Imputations

	Est.	s.e.	p
b_0	4.60	3.51	.21
b_1	.066	.034	.07

@jonathanleehelm 55

Observed Data	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1	Est.	s.e.	p
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2	Est.	s.e.	p
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3	Est.	s.e.	p
b_0	4.09	2.99	.19
b_1	.069	.03	.03

$$W = \frac{4.22^2 + 3.03^2 + 2.99^2}{3}$$

$$B = var(5.09, 4.64, 4.09)$$

$$V_{b_0} = W + \left(1 + \frac{1}{M}\right)B$$

$$se_{b_0} = \sqrt{V_{b_0}} = 3.51$$

Pooled Estimates across Imputations

	Est.	s.e.	p
b_0	4.60	3.51	.21
b_1	.066	.034	.07

@jonathanleehelm 56

Observed Data

	Est.	s.e.	p
b_0	-2.06	9.92	.84
b_1	.123	.09	.20

Imp Data M = 1

	Est.	s.e.	p
b_0	5.09	4.22	.24
b_1	.064	.04	.14

Imp Data M = 2

	Est.	s.e.	p
b_0	4.64	3.03	.14
b_1	.064	.03	.05

Imp Data M = 3

	Est.	s.e.	p
b_0	4.09	2.99	.19
b_1	.069	.03	.03

$$W = \frac{.04^2 + .03^2 + .03^2}{3}$$

$$B = var(.064, .064, .069)$$

$$V_{b_0} = W + \left(1 + \frac{1}{M}\right) B$$

$$se_{b_0} = \sqrt{V_{b_0}} = 0.34$$

Pooled Estimates across Imputations

	Est.	s.e.	p
b_0	4.60	3.51	.21
b_1	.066	.034	.07

Multiple Imputation: Summary

- Take aways:
  - Multiple imputation extends single imputation by creating/analyzing more than one imputed data set
  - Each data set includes some uncertainty for the imputed value

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

The Imputation Process

Technical Aspects of Imputation

- Multiple imputation software rarely uses multiple regression to impute values for missingness
- The actual process is a bit more technical, but it can be conceptually related to regression
  - *Hence the way I teach it*

Technical Aspects of Imputation

- Multiple imputation software rarely uses multiple regression to impute values for missingness
- The actual process is a bit more technical, but it can be conceptually related to regression
  - *Hence the way I teach it*
- http://www.stat.columbia.edu/~gelman/arm/missing.pdf
- https://www.jstatsoft.org/article/view/v045i03/v45i03.pdf

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Multiple Imputation
When does it work?

When Does MI Perform Well?

- Multiple imputation will perform well if one of the variables in the imputation accounts for the missingness
  - *The data are missing at random (MAR)*

When Does MI Perform Well?

- Multiple imputation will perform well if one of the variables in the imputation accounts for the missingness
  - *The data are missing at random (MAR)*
- Multiple imputation will also perform well if the missingness is not related to any variable in the data set
  - *The data are missing completely at random (MCAR)*

When Does MI Perform Well?

- Multiple imputation will not perform well if the missingness cannot be accounted for by the data
  - *The data are missing not at random (MNAR)*

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Assumptions in
Scientific Inference

Statistics in Science

- Statistics are foundation underlying scientific evidence
- Statistics are the language scientists use to make arguments

Statistics and Assumptions

- Virtually all inferential statistics rely on assumptions

Statistics and Assumptions

- Virtually all inferential statistics rely on assumptions

1. Random sample from the population
2. Certain variables follow a normal distribution
3. No measurement error
4. Independent observations
5. Equal variance across groups (or across the regression line)
6. The model is correct in the population

Statistics and Assumptions

- Regression: $Y_i = b_0 + b_1 X_i + \varepsilon_i$
- $b_1 = 5,\ p = .001$

- Common interpretation:
  - X_i significantly affects Y_i in the population

Statistics and Assumptions

- Regression: $Y_i = b_0 + b_1 X_i + \varepsilon_i$; $b_1 = 5,\ p = .001$

- Correct interpretation:
  - If I have a random sample from the population
  - and X_i is measured without error
  - and $Y_i = b_0 + b_1 X_i$ is the true model (i.e., nothing else affects Y_i)
  - and ε_i actually follows a normal distribution
  - and the variance of ε_i is constant around Y_i
  - and all of the observations are independent
- Then if the null hypothesis is true (if b_1 actually equals 0 in the pop.), the probability of getting 5 (or a value more extreme) is equals .001

Statistics and Assumptions

- Regression: $Y_i = b_0 + b_1 X_i + \varepsilon_i$; $b_1 = 5,\ p = .001$

- Correct interpretation:
  - If I have a random sample from the
  - and X_i is measured without error
  - and $Y_i = b_0 + b_1 X_i$ is the true model (i.e., nothing else affects Y_i)
  - and ε_i actually follows a normal distribution
  - and the variance of ε_i is constant around Y_i
  - and all of the observations are independent
- Then if the null hypothesis is true (if b_1 actually equals 0 in the pop.), the probability of getting 5 (or a value more extreme) is equals .001

If we have missing data, then we need to add:
and the data are missing completely at random

Statistics and Assumptions

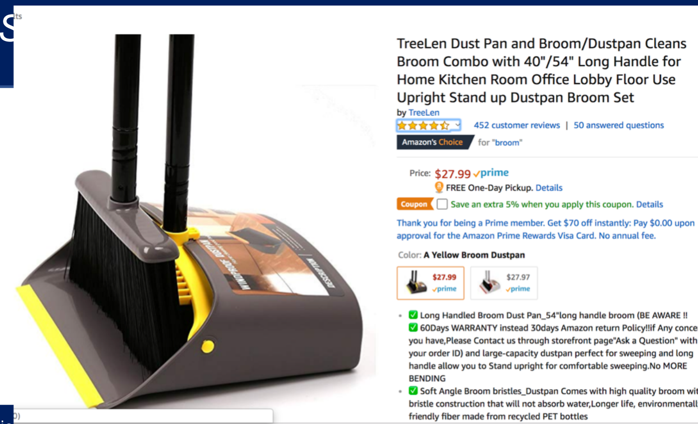
- Regression: $Y_i = b_0 + b_1 X_i + \varepsilon_i$; $b_1 = 5,\ p = .001$

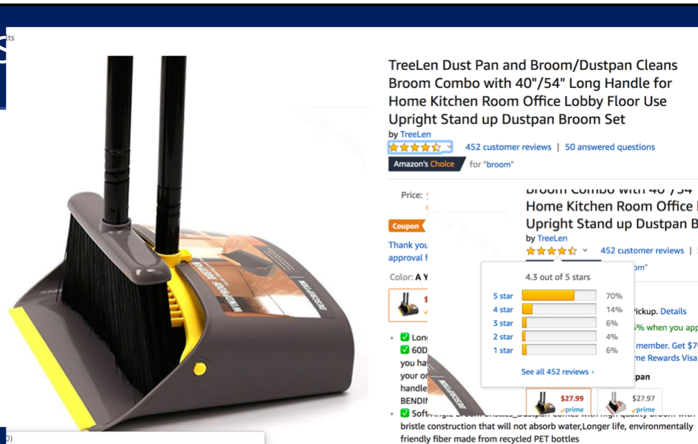
- Correct interpretation:
  - If I have a random sample from the
  - and X_i is measured without error
  - and $Y_i = b_0 + b_1 X_i$ is the true model (i.e., nothing else affects Y_i)
  - and ε_i actually follows a normal distribution
  - and the variance of ε_i is constant around Y_i
  - and all of the observations are independent
- Then if the null hypothesis is true (if b_1 actually equals 0 in the pop.), the probability of getting 5 (or a value more extreme) is equals .001

If we have missing data, and we performed multiple imputation:
and the data are missing at random

Statistics and Assumptions

- Don't panic





Statistics and Assumptions

- Interpretation:
  - If I buy the broom, I will be happy with it
- Correct interpretation:
  - If I have a random sample from the population
  - and ratings are measured without error
  - and ratings of the broom a direct reflection of broom satisfaction
    - *is the true model (i.e., nothing else affects Y_i)*
  - and all of the observations are independent
- Then I will be happy with the broom

Statistics and Assumptions

- Interpretation:
  - If I buy the broom, I will be happy with it
- Correct interpretation:
  - If I have a random sample from the population
  - and ratings are measured without error
  - and ratings of the broom a direct reflection of broom satisfaction
    - *is the true model (i.e., nothing else affects Y_i)*
  - and all of the observations are independent
- Then I will be happy with the broom

These assumptions are likely not true

And I would still buy the broom

Statistics and Assumptions

- Don't panic
- The most important part is to recognize that assumptions that you're making when you're drawing conclusions
- Missing data mechanisms are a part of those assumptions
- So include it, and draw conclusions accordingly

@jonathanleehelm 82

Grand Overview

- Single Imputation
- Multiple Imputation
- The Imputation Process
- When does Multiple Imputation work?
- A note about Assumptions

@jonathanleehelm 83