Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Missing Data Workshop

Jonathan Lee Helm
Friday May 17th, 2019

@jonathanleehelm 1

Grand Overview

- Introduction to missing data
- Review of sampling
- Patterns, causes, and mechanisms of missing data
- The problem with missing data

@jonathanleehelm 2

Grand Overview

- Introduction to missing data
- Review of sampling
- Patterns, causes, and mechanisms of missing data
- The problem with missing data

@jonathanleehelm 3

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Introduction to Missing Data

@jonathanleehelm 4

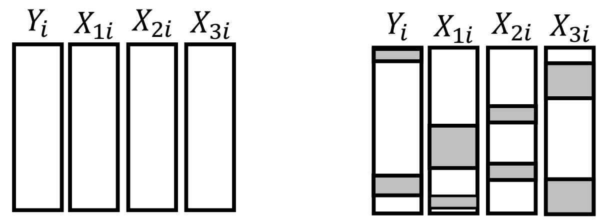
Missing Data

- Missing data occur when we do not have one or more observations for a given variable in our data

@jonathanleehelm 5

Missing Data

- Missing data occur when we do not have one or more observations for a given variable in our data



@jonathanleehelm 6

Missing Data

- Missing data are ubiquitous in psychological science

- Can anyone think of a real world example that has complete data in psychological science?

@jonathanleehelm 7

Missing Data

- Missing data pose potential problems

@jonathanleehelm 8

Missing Data

- Missing data pose potential problems

- Potential for biased estimates
- Potentially increase standard errors
  - Smaller sample size
  - Higher Type 2 error rate (i.e., harder to detect sig. effects)

Missing Data

- Missing data pose potential problems

- Potential for biased estimates
- Potentially increase standard errors
  - Smaller sample size
  - Higher Type 2 error rate (i.e., harder to detect sig. effects)

- These potential problems can be mitigated by different analyses
  - Multiple imputation

Missing Data

- The reasons that missing data may cause problems are closely linked to the importance of random sampling

- Let's review random sampling as a segue into missing data

Grand Overview

- Introduction to missing data
- Review of sampling
- Patterns, causes, and mechanisms of missing data
- The problem with missing data

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Review of Sampling

Sampling

- The problems associated with missing data can be conceptualized in terms of sampling

Sampling

- Scientific process:
  1. Random sample
  2. Measure
  3. Analyze
  4. Draw conclusions

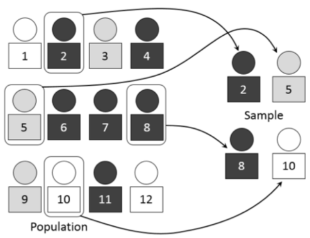
Sampling

- Scientific process:
  1. Random sample
  2. Measure
  3. Analyze
  4. Draw conclusions

This is an important feature!

Random sampling implies generalizability

Sampling



Random Sample

- All members of pop. have equal chance of being selected
- The results from the analysis should generalize to the population

Sampling

- Example of a random sample
- And a random sample of a random sample

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

The sample mean for JS = 10.3
The sample SD for JS = 2.8

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Generalizability is a function of sampling

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

The sample mean for JS = 10.3
The sample SD for JS = 2.8

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Generalizability is a function of sampling

Sample mean for JS = 10.3
Sample SD for JS = 2.8

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Sample mean for JS = 10.2
Sample SD for JS = 3.1

Random Sample of Random Sample

i	JS	IQ
1	10	72
⋮	⋮	⋮
49	9	118
50	14	134

Sampling

- Example of a non-random sample
- And a non-random sample of a random sample

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Non Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	4	80
100	6	73

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Non Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	4	80
100	6	73

Sample mean for JS = 7.1
Sample SD for JS = 1.1

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Generalizability does not hold!

Non Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	4	80
100	6	73

Sample mean for JS = 7.1
Sample SD for JS = 1.1

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Pop. mean for JS = 10
Pop. SD for JS = 3

Generalizability is a function of sampling

Sample mean for JS = 10.3
Sample SD for JS = 2.8

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Sample mean for JS = 7.1
Sample SD for JS = 1.1

Non-Random Sample of Random Sample

i	JS	IQ
1	10	72
⋮	⋮	⋮
49	9	80
50	14	73

Sampling

- Missing data, in some cases can be conceptualized as a sample of a sample
- This occurs when we use listwise deletion
  - *Delete any row that has a missing value*

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Sample Remaining after Listwise Deletion

i	JS	IQ
1	10	72
⋮	⋮	⋮
49	9	80
50	14	73

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Listwise deletion will produce unbiased results if the remaining sample is still a random sample

Sample Remaining after Listwise Deletion

i	JS	IQ
1	10	72
⋮	⋮	⋮
49	9	80
50	14	73

Sampling

- Take aways:
  1. We have to perform sampling
  2. Our results will generalize if we have a random sample
  3. Missing data with listwise deletion can be conceptualized as a sample of a sample

Grand Overview

- Introduction to missing data
- Review of sampling
- Patterns, causes, and mechanisms of missing data
- The problem with missing data

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

Patterns, Causes, and Mechanisms of Missing Data

Patterns vs Causes vs Mechanisms

- Patterns of missingness
- Causes of missingness
- Missing data mechanisms
  - *Missing data assumptions*

Patterns vs Causes vs Mechanisms

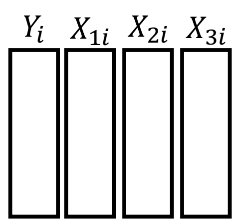
- Patterns of missingness
- Causes of missingness
- Missing data mechanisms
  - *Missing data assumptions*

Patterns of Missingness

- A pattern of missingness is a pattern that occurs in a specific data set

Patterns of Missingness

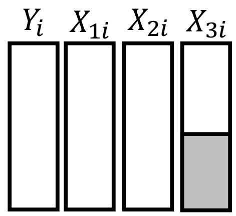
A pattern of missingness is a pattern that occurs in a specific data set



Complete data
No missingness
No missingness pattern

Patterns of Missingness

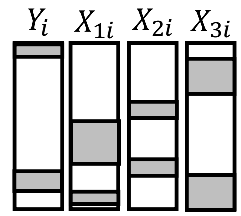
A pattern of missingness is a pattern that occurs in a specific data set



Missing on X_{3i}
Univariate pattern

Patterns of Missingness

A pattern of missingness is a pattern that occurs in a specific data set

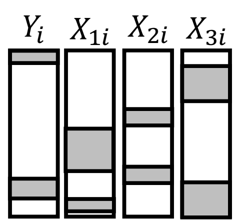


Missing on more than one variable
General pattern

Patterns of Missingness

In psychology, we usually have general patterns

We would like methods that can account for general patterns



Patterns vs Causes vs Mechanisms

- Patterns of missingness
- Causes of missingness
- Missing data mechanisms
  - *Missing data assumptions*

Cause of Missingness

- The ***true*** (not assumed) reason why the data are missing

Cause of Missingness

- Example 1
- A researcher asks individuals to report their income
- Individuals with lower income tend to not report their income

- Cause of missingness: Those with lower income do not report their income

Cause of Missingness

- Example 2
- A researcher performs an intervention on smoking
- Individuals that find the intervention challenging drop out
- There is missing data on smoking behavior at follow up

- Cause of missingness: Those that find the intervention to be challenging drop out

Cause of Missingness

- Example 3
- A researcher performs an intervention on smoking
- Individuals that find the intervention challenging drop out
- There is missing data on smoking behavior at follow up

- Cause of missingness: Those that find the intervention to be challenging drop out

Cause of Missingness

- Example 4
- A researcher performs measures intelligence longitudinally
- The researcher randomly assigns half of the sample to be measured at ages 5, and 7; and half at 6 and 8
- Cause of missingness: Planned missingness: The researcher creates the missingness by design

Cause of Missingness

- The ***true*** (not assumed) reason why the data are missing
- Typically not known to the researcher
  - Counter: Planned missingness

Patterns vs Causes vs Mechanisms

- Patterns of missingness
- Causes of missingness
- Missing data mechanisms
  - *Missing data assumptions*

Missing Data Mechanisms

- A.K.A:
  1. Categories of missingness
  2. Types of missingness
- These are ***assumptions*** regarding the missing data
- Not known causes of the missingness

Missing Data Mechanisms

- Three Missing Data Mechanisms

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Missing not at Random (MNAR)

Missing Data Mechanisms

- Three Missing Data Mechanisms

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Missing not at Random (MNAR)

Missing Data Mechanisms

- A given analysis (e.g., *t*-test) ***inherently assumes*** a missing data mechanism
- If the assumption is incorrect, the parameter estimates may be biased

Missing Data Mechanisms: MCAR

- Missingness on a variable Y_i is MCAR if the *probability* of missingness is *unrelated* to
  1. The values Y_i (including the missing values!)
  2. Or to any other variable in the analysis*

* You can have variables in your data set that are not in the analysis

Missing Data Mechanisms: MCAR

- Missingness on a variable Y_i is MCAR if the *probability* of missingness is *unrelated* to
  1. The values Y_i (including the missing values!)
  2. Or to any other variable in the analysis*

- ***The observed are a random sub-sample of the complete sample***

* You can have variables in your data set that are not in the analysis

@jonathanleehelm 57

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

@jonathanleehelm 58

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

Observed Data

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

@jonathanleehelm 59

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	13	84	0
3	--	84	1
4	8	85	0
5	7	87	0 0
6	7	91	0
7	9	92	0
8	9	94	0
9	11	94	0
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	--	134	1

@jonathanleehelm 60

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	13	84	0
3	--	84	1
4	8	85	0
5	7	87	0 0
6	7	91	0
7	9	92	0
8	9	94	0
9	11	94	0
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	--	134	1

MCAR:
Y^{com} is not related to Y^{ind} or any other variable in the analysis

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	13	84	0
3	--	84	1
4	8	85	0
5	7	87	0 0
6	7	91	0
7	9	92	0
8	9	94	0
9	11	94	0
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	--	134	1

MCAR:
Y^{com} is not related to Y^{ind}

***t*-test:**
Mean JS^{com} for $JS^{ind} = 0$
10.6

Mean JS^{com} for $JS^{ind} = 1$
9.6

p-value for difference
.3857

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	13	84	0
3	--	84	1
4	8	85	0
5	7	87	0 0
6	7	91	0
7	9	92	0
8	9	94	0
9	11	94	0
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	--	134	1

MCAR:
Y^{com} is not related to Y^{ind}
Or any other variable

***t*-test:**
Mean IQ^{com} for $JS^{ind} = 0$
99.73

Mean IQ^{com} for $JS^{ind} = 1$
100.8

p-value for difference
.9235

Missing Data Mechanisms: MCAR

- MCAR: missingness is not related to any variable in the data set
  - The observed are a random sample of your sample

Missing Data Mechanisms: MCAR

- MCAR: missingness is not related to any variable in the data set
  - The observed are a random sample of your sample
- Can we ever know that data are MCAR in practice?

Missing Data Mechanisms: MCAR

- MCAR: missingness is not related to any variable in the data set
  - The observed are a random sample of your sample
- Can we ever know that data are MCAR in practice?
  - *No, we would need the complete data*

Missing Data Mechanisms

- Three Missing Data Mechanisms

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Missing not at Random (MNAR)

Missing Data Mechanisms: MAR

- Missingness on a variable Y_i is ***Missing at Random*** if the probability of missingness is unrelated to Y_i ***after controlling for other variables*** in the analysis

Missing Data Mechanisms: MAR

- Missingness on a variable Y_i is ***Missing at Random*** if the probability of missingness is unrelated to Y_i ***after controlling for other variables*** in the analysis
- Missingness on Y_i is related to another variable in the analysis
- You have other measures of the other variable

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

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

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

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	--	84	1
3	--	84	1
4	--	85	1
5	--	87	1 1
6	--	91	1
7	--	92	1
8	--	94	1
9	--	94	1
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

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

Observed Data			
i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	--	84	1
3	--	84	1
4	--	85	1
5	--	87	1 1
6	--	91	1
7	--	92	1
8	--	94	1
9	--	94	1
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MAR:
Y^{com} is not related to Y^{ind} after controlling for other variables in the analysis

@jonathanleehelm 73

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

Observed Data			
i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	--	84	1
3	--	84	1
4	--	85	1
5	--	87	1 1
6	--	91	1
7	--	92	1
8	--	94	1
9	--	94	1
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MAR:
Y^{com} is not related to Y^{ind} after controlling for other variables in the analysis

$Y^{com} = b_0 + b_1 Y^{ind}$

	Est.	s.e.	p
b_0	11.7	.75	<.01
b_1	-2.7	1.05	.02

@jonathanleehelm 74

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

Observed Data			
i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	--	84	1
3	--	84	1
4	--	85	1
5	--	87	1 1
6	--	91	1
7	--	92	1
8	--	94	1
9	--	94	1
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MAR:
Y^{com} is not related to Y^{ind} after controlling for other variables in the analysis

$Y^{com} = b_0 + b_1 IQ^{obs}$

	Est.	s.e.	p
b_0	0.07	3.79	.98
b_1	.10	.04	.01

@jonathanleehelm 75

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

Observed Data			
i	JS^{obs}	IQ^{obs}	JS^{ind}
1	--	78	1
2	--	84	1
3	--	84	1
4	--	85	1
5	--	87	1 1
6	--	91	1
7	--	92	1
8	--	94	1
9	--	94	1
10	--	96	1
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MAR:
Y^{com} is not related to Y^{ind} after controlling for other variables in the analysis

$Y^{com} = b_0 + b_1 IQ^{obs} + b_2 Y^{ind}$

	Est.	s.e.	p
b_0	3.97	7.80	.62
b_1	.07	.07	.33
b_2	-1.11	1.92	.57

@jonathanleehelm 76

Missing Data Mechanisms: MAR

- Missingness on random indicates that there is some other variable in the analyses that accounting for the missingness

Missing Data Mechanisms: MAR

- Missingness on random indicates that there is some other variable in the analyses that accounting for the missingness

- *i.e., that variable can be a proxy for the cause of the missingness*
- *Once you account for that variable, you have a random sample again*

Missing Data Mechanisms: MAR

- Can we ever know for certain that data are MAR?

- *No, we would need the complete data to be certain*
- *Even then, it would still be an assumption*

Missing Data Mechanisms

- Three Missing Data Mechanisms

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Missing not at Random (MNAR)

Missing Data Mechanisms: MNAR

- Missingness on a variable Y_i is ***Missing Not at Random*** if the probability of missingness is ***still related*** to Y_i ***after controlling for other variables*** in the analysis

Missing Data Mechanisms: MAR

- Missingness on a variable Y_i is ***Missing Not at Random*** if the probability of missingness is ***still related*** to Y_i ***after controlling for other variables*** in the analysis
- Missingness on Y_i is still related to Y_i after controlling for other variables

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

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

Observed Data

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

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	9	78	0
2	13	84	0
3	10	84	0
4	8	85	0
5	7	87	0 0
6	--	91	1
7	--	92	1
8	--	94	1
9	11	94	0
10	7	96	0
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	9	78	0
2	13	84	0
3	10	84	0
4	8	85	0
5	7	87	0 0
6	--	91	1
7	--	92	1
8	--	94	1
9	11	94	0
10	7	96	0
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MNAR:
Y^{com} is still related to Y^{ind} after controlling for other variables in the analysis

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	9	78	0
2	13	84	0
3	10	84	0
4	8	85	0
5	7	87	0 0
6	--	91	1
7	--	92	1
8	--	94	1
9	11	94	0
10	7	96	0
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MNAR:
Y^{com} is still related to Y^{ind} after controlling for other variables in the analysis

$Y^{com} = b_0 + b_1 Y^{ind}$

	Est.	s.e.	p
b_0	11.4	.51	<.01
b_1	-4.2	1.02	<.01

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	9	78	0
2	13	84	0
3	10	84	0
4	8	85	0
5	7	87	0 0
6	--	91	1
7	--	92	1
8	--	94	1
9	11	94	0
10	7	96	0
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MNAR:
Y^{com} is still related to Y^{ind} after controlling for other variables in the analysis

$Y^{com} = b_0 + b_1 IQ^{obs}$

	Est.	s.e.	p
b_0	0.07	3.79	.98
b_1	.10	.04	.01

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

Observed Data

i	JS^{obs}	IQ^{obs}	JS^{ind}
1	9	78	0
2	13	84	0
3	10	84	0
4	8	85	0
5	7	87	0 0 0
6	--	91	1
7	--	92	1
8	--	94	1
9	11	94	0
10	7	96	0
11	7	99	0
⋮	⋮	⋮	⋮
19	16	118	0
20	12	134	0

MNAR:
Y^{com} is still related to Y^{ind} after controlling for other variables in the analysis

$Y^{com} = b_0 + b_1 IQ^{obs} + b_2 Y^{ind}$

	Est.	s.e.	*p*
b_0	4.88	3.30	.16
b_1	.06	.03	.06
b_2	-3.48	1.01	<.01

Missing Data Mechanisms: MNAR

- Data are missing not at random when the missingness on Y_i is related to the values of Y_i, even after controlling for other variables in the analysis

Missing Data Mechanisms: MAR

- Missingness not at random occurs when the missingness on Y_i is related to the values of Y_i, even after controlling for other variables in the analysis
- *i.e., the sample is still not a random sample from the population, even after accounting for other variables in the analysis*

Missing Data Mechanisms: MAR

- Can we ever know for certain that data are MNAR?

Missing Data Mechanisms: MAR

- Can we ever know for certain that data are MNAR?

- *No, we would need the complete data to be certain*
- *Even then, it would still be an assumption*

Missing Data Mechanisms

- Three Missing Data Mechanisms

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Missing not at Random (MNAR)

Missing Data Mechanisms

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Missing not at Random (MNAR)

- These are ***assumptions*** regarding missingness
- Different statistical analysis will be valid under different assumptions

Grand Overview

- Introduction to missing data
- Review of sampling
- Patterns, causes, and mechanisms of missing data
- The problem with missing data

Missing Data Workshop
Joint Doctoral Program in Clinical Psyc

The Problem with Missing Data

Problems with Missing Data

- Most analytic techniques require complete data
  - Descriptive statistics (e.g., means, SDs, correlations)
  - *t*-tests
  - Significance testing of correlation
  - ANOVA
  - Regression

Problems with Missing Data

- If our data contain missingness, then we typically perform listwise deletion

Problems with Missing Data

- If our data contain missingness, then we typically perform listwise deletion
- If data are missing completely at random, then we should obtain unbiased estimates

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

- Missingness is MCAR
- Sub-sample is a random sample
- Parameter estimates are unbiased ☺
- Standard errors are still larger ☹

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Sample Remaining after Listwise Deletion

i	JS	IQ
1	10	72
⋮	⋮	⋮
49	9	80
50	14	73

Problems with Missing Data

- If our data contain missingness, then we typically perform listwise deletion
- If data are missing completely at random, then we should obtain unbiased estimated
- If data are missing at random or missing not at random then we will obtain biased estimates

Population Data

i	JS	IQ
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
⋮	⋮	⋮
N-1	16	118
N	12	134

- Missingness is MAR or MNAR
- Sub-sample is not random a sample
- Parameter estimates can be biased ☹
- Standard errors are larger ☹

Random Sample

i	JS	IQ
1	10	72
2	12	74
3	7	90
⋮	⋮	⋮
99	9	118
100	14	134

Sample Remaining after Listwise Deletion

i	JS	IQ
1	10	72
⋮	⋮	⋮
49	9	80
50	14	73

Problems with Missing Data

- Therefore, many analytic techniques assume data are missing completely at random
  - *And they didn't even tell you!*

Problems with Missing Data

- So the major question is, is there anything we can do to at least assume MAR instead of the more restrictive MCAR?
  - *Multiple Imputation*

Grand Overview

- Introduction to missing data
- Review of sampling
- Patterns, causes, and mechanisms of missing data
- The problem with missing data