Precept 8: Missing Data

Soc 504: Advanced Social Statistics

Ian Lundberg

Princeton University

April 6, 2018

Outline

1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

Learning goals

By the end of precept, you should be able to:

1. Feel comfortable with three common **assumptions** about missing data
   - Missing completely at random
   - Missing at random
   - Non-ignorable
2. Be able to reason about the plausibility of these assumptions using **substantive knowledge** in real research settings.
3. Connect assumptions to concrete **strategies** to deal with missing data
   - Listwise deletion
   - Multiple imputation
   - Bounds

Outline

1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

[1]

Abraham Wald

- b. 1902, Austria-Hungary
- Jewish, persecuted in WWII
- Fled to U.S. in 1938
- Namesake of the Wald test
- Statistical consultant for U.S. Navy in WWII

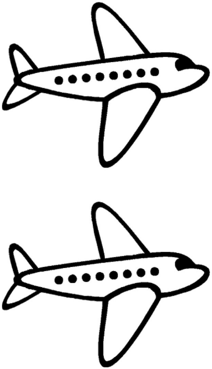
[1] PC: Wikimedia commons

Question: Where should armor be added to protect planes?

Data: Suppose we saw the following planes.[2]

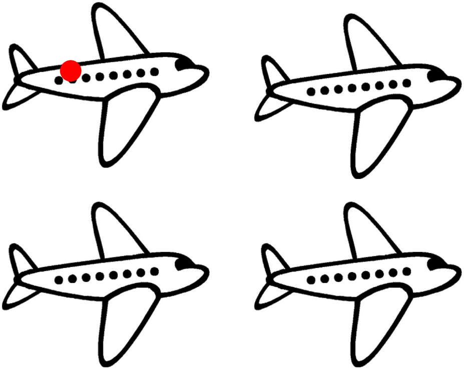
[2] Story told by Mangel and Samaniego 1984 [link].
Presentation style inspired by Joe Blitzstein. See the original here [link]

Observed data: Planes that came home with damage



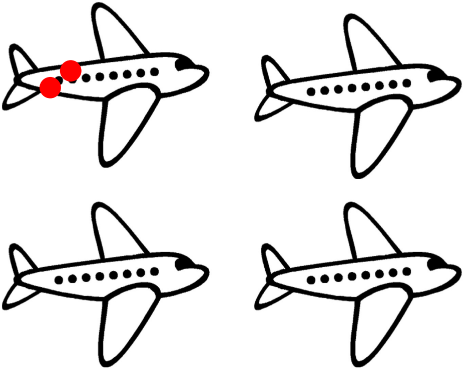
Q: Where would you add armor?

Observed data: Planes that came home with damage



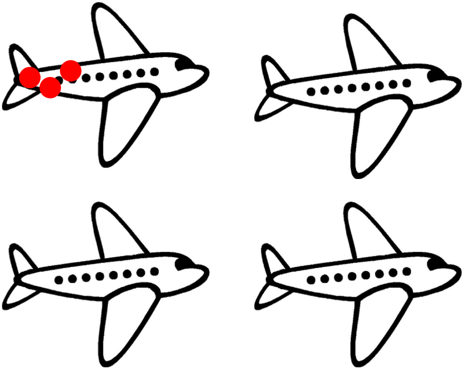
Q: Where would you add armor?

Observed data: Planes that came home with damage



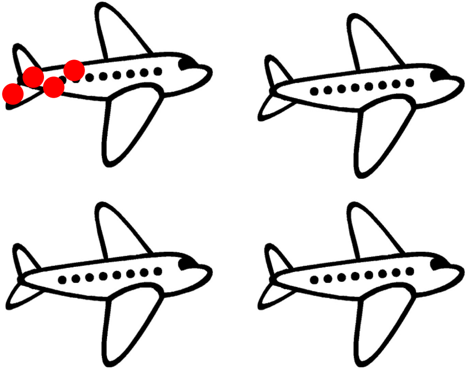
Q: Where would you add armor?

Observed data: Planes that came home with damage



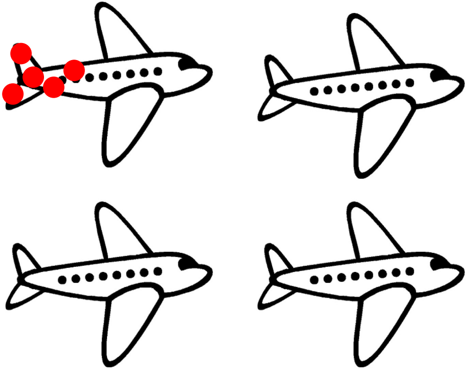
Q: Where would you add armor?

Observed data: Planes that came home with damage



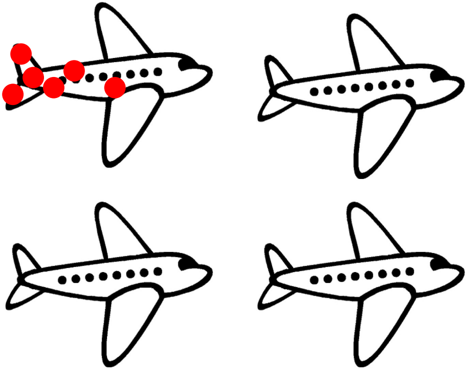
Q: Where would you add armor?

Observed data: Planes that came home with damage



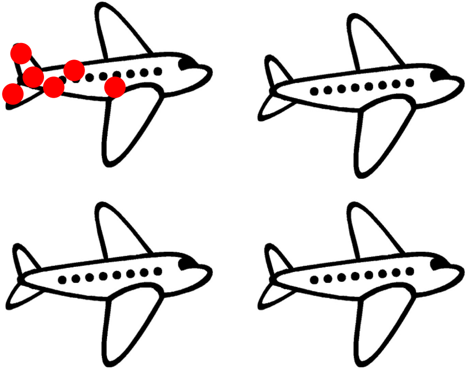
Q: Where would you add armor?

Observed data: Planes that came home with damage



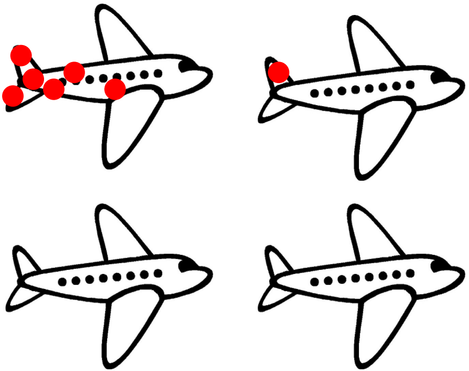
Q: Where would you add armor?

Observed data: Planes that came home with damage



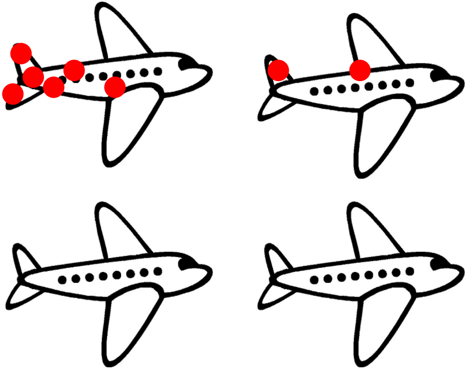
Q: Where would you add armor?

Observed data: Planes that came home with damage



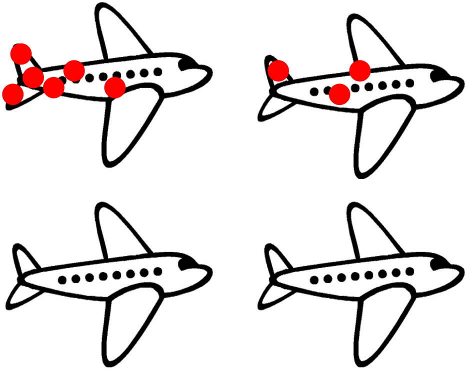
Q: Where would you add armor?

Observed data: Planes that came home with damage



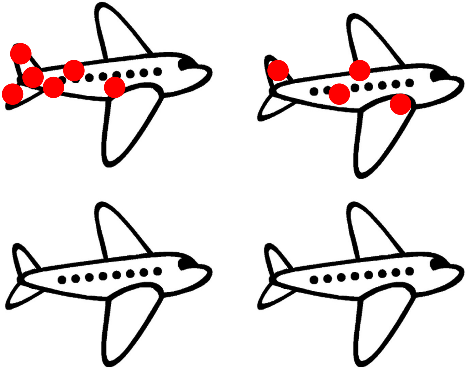
Q: Where would you add armor?

Observed data: Planes that came home with damage



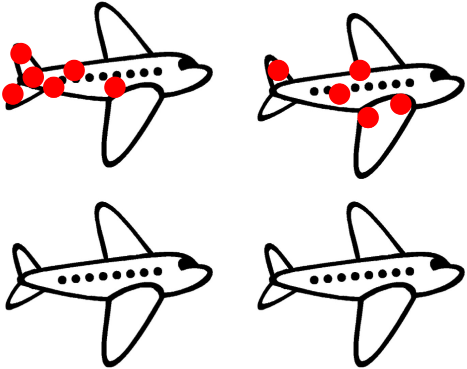
Q: Where would you add armor?

Observed data: Planes that came home with damage



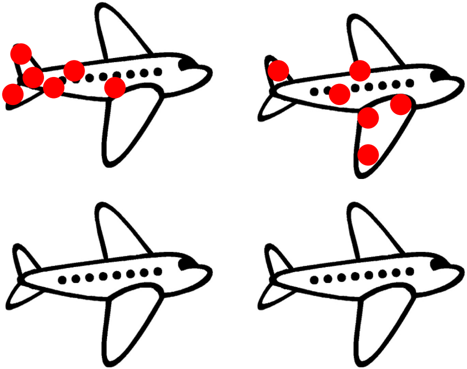
Q: Where would you add armor?

Observed data: Planes that came home with damage



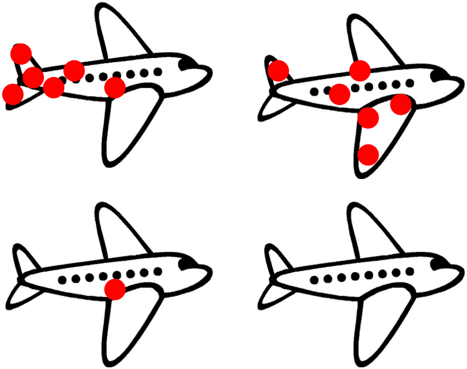
Q: Where would you add armor?

Observed data: Planes that came home with damage



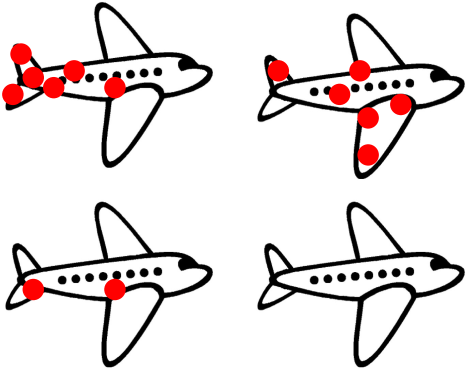
Q: Where would you add armor?

Observed data: Planes that came home with damage



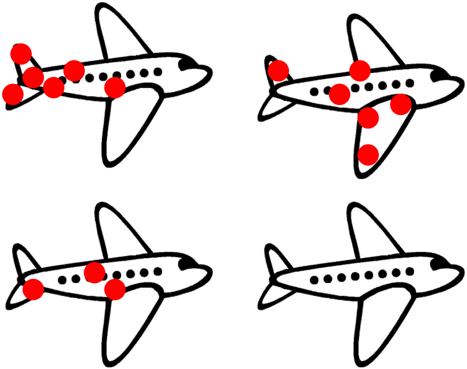
Q: Where would you add armor?

Observed data: Planes that came home with damage



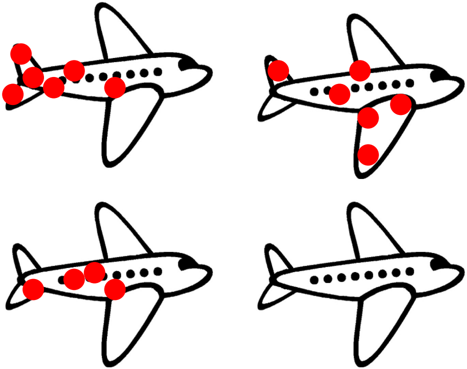
Q: Where would you add armor?

Observed data: Planes that came home with damage



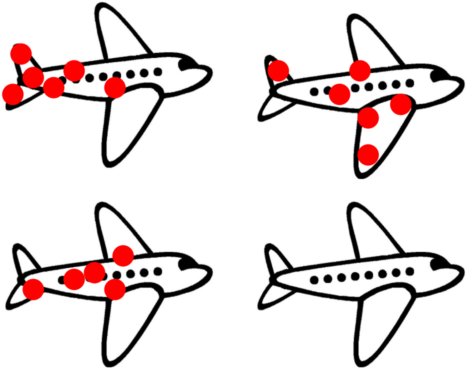
Q: Where would you add armor?

Observed data: Planes that came home with damage



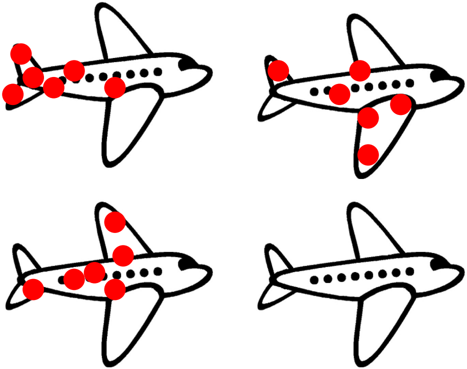
Q: Where would you add armor?

Observed data: Planes that came home with damage



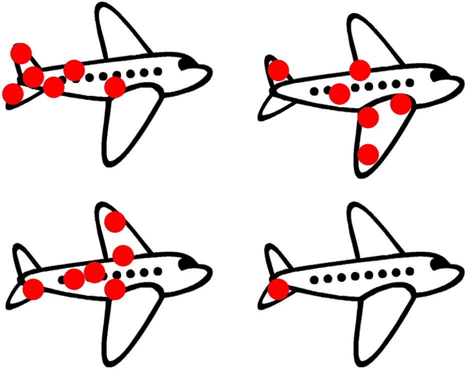
Q: Where would you add armor?

Observed data: Planes that came home with damage



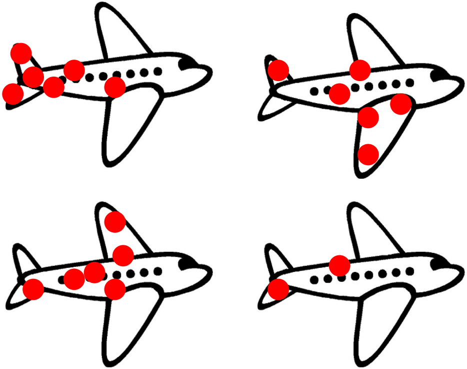
Q: Where would you add armor?

Observed data: Planes that came home with damage



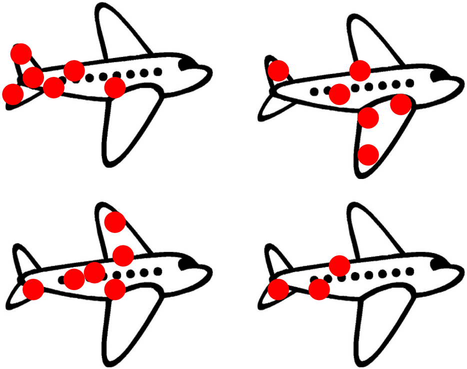
Q: Where would you add armor?

Observed data: Planes that came home with damage



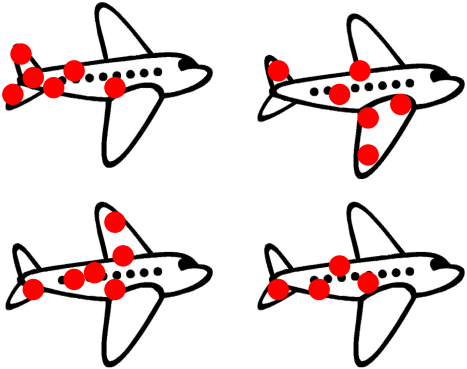
Q: Where would you add armor?

Observed data: Planes that came home with damage



Q: Where would you add armor?

Observed data: Planes that came home with damage



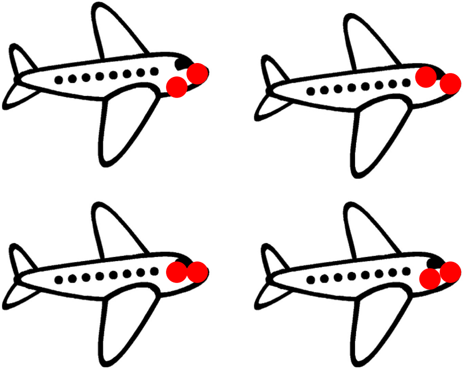
Q: Where would you add armor?

Discussion

Don't peek at the next slide if you downloaded these!

Missing data: Planes that never returned

Missing data: Planes that never returned



Now where should we add armor?

Now where should we add armor? To the nose!

Now where should we add armor? To the nose!

Results from the observed planes were misleading because data were not missing at random!

Now where should we add armor? To the nose!

Results from the observed planes were misleading because data were not missing at random!

Missing data requires careful thought.

Now where should we add armor? To the nose!

Results from the observed planes were misleading because data were not missing at random!

Missing data requires careful thought.

No algorithm solves it for you!

We will walk through the assumptions and implementation of multiple imputation.

We will walk through the assumptions and implementation of multiple imputation.

Our example will be the 2016 General Social Survey (GSS).

We will walk through the assumptions and implementation of multiple imputation.

Our example will be the 2016 General Social Survey (GSS).

The GSS measures Americans' attitudes toward lots of issues.

We will walk through the assumptions and implementation of multiple imputation.

Our example will be the 2016 General Social Survey (GSS).

The GSS measures Americans' attitudes toward lots of issues.

List of files (we use 2016): `http://gss.norc.org/get-the-data/spss`
Link directly to data download

We will walk through the assumptions and implementation of multiple imputation.

Our example will be the 2016 General Social Survey (GSS).

The GSS measures Americans' attitudes toward lots of issues.

List of files (we use 2016): `http://gss.norc.org/get-the-data/spss`
Link directly to data download

The GSS is a complex sample design. We will draw inferences about the sample. In two weeks we will learn how to generalize these to the population using weights.

Research question:

What is the relationship between the
respondent's education
and the respondent's
father's education?

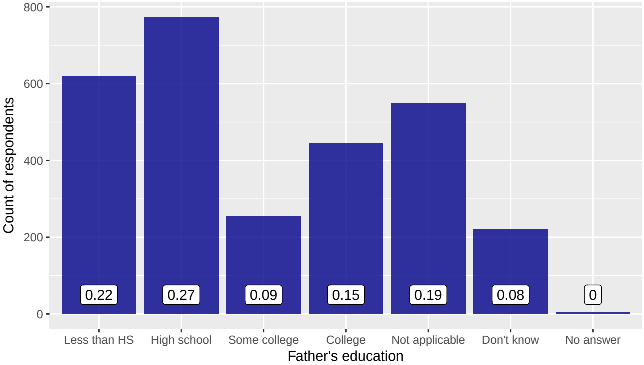
Research question:

What is the relationship between the
respondent's education
and the respondent's
father's education?

(Beyond attitudinal questions, the GSS asks several questions related to mobility) `paeduc` captures father's education in years.

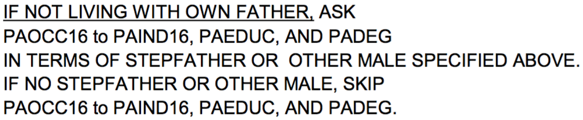
paeduc captures father's education in years.

But it's sometimes missing. We need to know why!



Why is father's education missing?

Check the codebook (p. 176) [link]

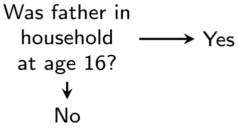


IF NOT LIVING WITH OWN FATHER, ASK
PAOCC16 to PAIND16, PAEDUC, AND PADEG
IN TERMS OF STEPFATHER OR OTHER MALE SPECIFIED ABOVE.
IF NO STEPFATHER OR OTHER MALE, SKIP
PAOCC16 to PAIND16, PAEDUC, AND PADEG.

These are the **"Not applicable"** cases.

You should always know what your variables are!

Questionnaire logic, graphically



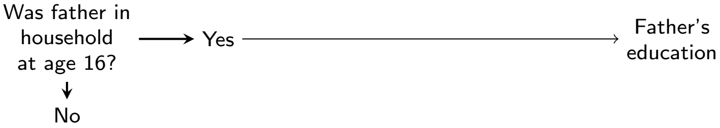
Was father in household at age 16? → Yes

↓

No

We must decide what to do with the not applicable, don't know, and no answer cases.

Questionnaire logic, graphically



Was father in household at age 16?

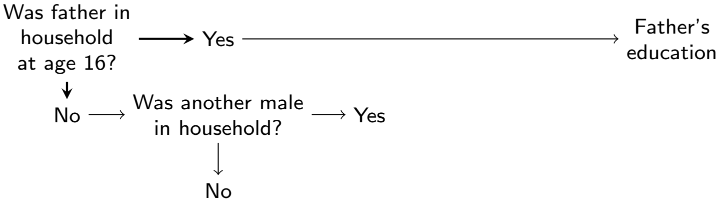
Yes

No

Father's education

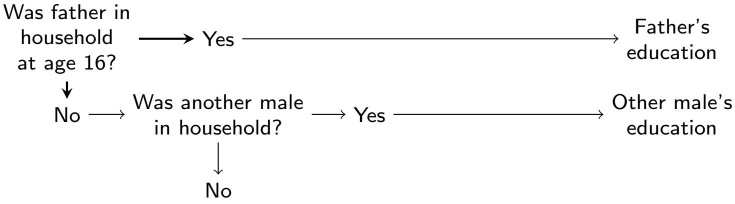
We must decide what to do with the not applicable, don't know, and no answer cases.

Questionnaire logic, graphically


Was father in household at age 16?
Yes
Father's education
No
Was another male in household?
Yes
No

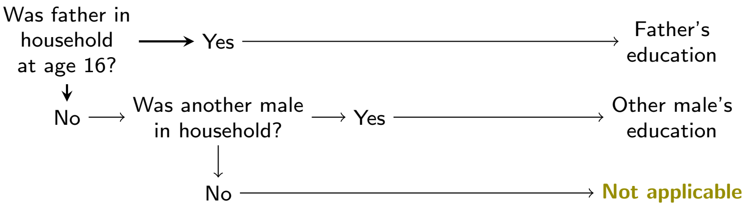
We must decide what to do with the not applicable, don't know, and no answer cases.

Questionnaire logic, graphically


Was father in household at age 16?
Yes
Father's education
No
Was another male in household?
Yes
Other male's education
No

We must decide what to do with the not applicable, don't know, and no answer cases.

Questionnaire logic, graphically



Was father in household at age 16?

Yes → Father's education

No → Was another male in household?

Yes → Other male's education

No → **Not applicable**

We must decide what to do with the not applicable, don't know, and no answer cases.

1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

1 Motivation

2 Listwise deletion

3 Bounds

4 Multiple imputation

5 Rubin's rules

6 Simulation

7 Review

8 Bonus slides to help with optional homework problem: EM

Listwise deletion

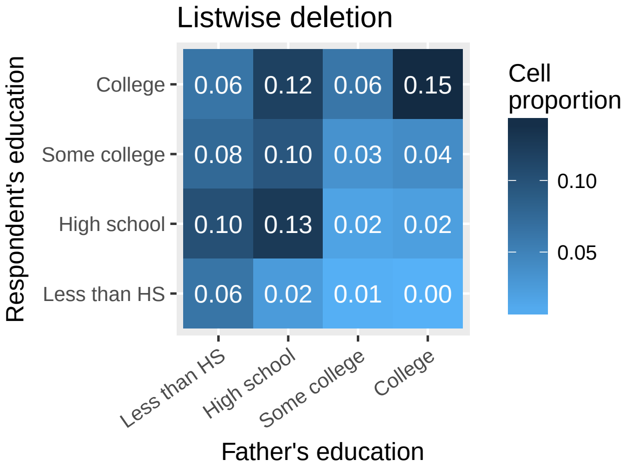
We could just drop anybody with missing values.

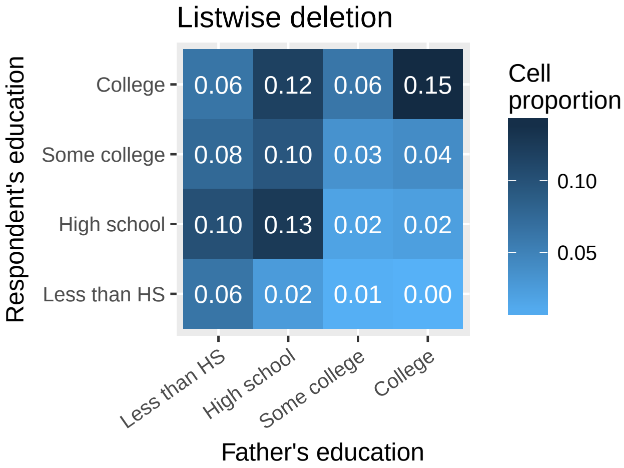
Listwise deletion

We could just drop anybody with missing values.

	Father's education (total $N = 2,630$)				
Respondent's education	Less than HS	High school	Some college	College	**MISSING**
College	122	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	7	168
MISSING	1	2	0	0	6

Table: GSS data on respondent's and father's education: Sample counts


Listwise deletion
Respondent's education
College
Some college
High school
Less than HS
0.06 0.12 0.06 0.15
0.08 0.10 0.03 0.04
0.10 0.13 0.02 0.02
0.06 0.02 0.01 0.00
Less than HS
High school
Some college
College
Father's education
Cell proportion
0.10
0.05



	Less than HS	High school	Some college	College
College	0.06	0.12	0.06	0.15
Some college	0.08	0.10	0.03	0.04
High school	0.10	0.13	0.02	0.02
Less than HS	0.06	0.02	0.01	0.00

Q: Under **what assumption** are these estimates unbiased?

Missing Completely at Random (MCAR)

Data are missing completely at random if $\mathrm{P}(M \mid X) = \mathrm{P}(M)$

Missing Completely at Random (MCAR)

Data are missing completely at random if $P(M \mid X) = P(M)$

	Father's education (total $N = 2,630$)				
Respondent's education	Less than HS	High school	Some college	College	**MISSING**
College	122	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	7	168
MISSING	1	2	0	0	6

Missing Completely at Random (MCAR)

Data are missing completely at random if $\text{P}(M \mid X) = \text{P}(M)$

	Father's education (total $N = 2,630$)				
Respondent's education	Less than HS	High school	Some college	College	**MISSING**
College	122	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	7	168
MISSING	1	2	0	0	6

In this case, MCAR requires that missingness is independent of the true values of father's and respondent's education.

Missing Completely at Random (MCAR)

Data are missing completely at random if $\mathrm{P}(M \mid X) = \mathrm{P}(M)$

	Father's education (total $N = 2,630$)				
Respondent's education	Less than HS	High school	Some college	College	**MISSING**
College	122	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	7	168
MISSING	1	2	0	0	6

In this case, MCAR requires that missingness is independent of the true values of father's and respondent's education.

If data are MCAR, then listwise deletion is unbiased.

1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

1. Motivation
2. Listwise deletion
3. **Bounds**
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

Bounds

MCAR is a strong assumption that requires substantive theory.

Bounds

MCAR is a strong assumption that requires substantive theory.

Could we bound some proportions with weaker assumptions?

Bounds

MCAR is a strong assumption that requires substantive theory.

Could we bound some proportions with weaker assumptions?

Place an upper bound on the proportion of the sample with extreme upward mobility:

father did not complete high school
but respondent attained a college degree

Bounds

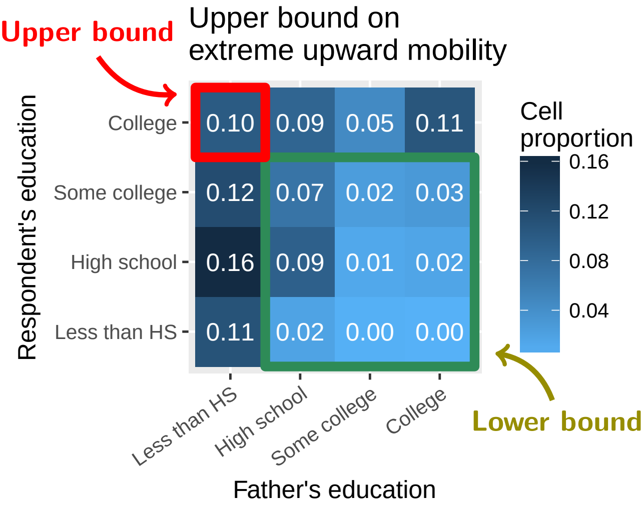
MCAR is a strong assumption that requires substantive theory.

Could we bound some proportions with weaker assumptions?

Place an upper bound on the proportion of the sample with extreme upward mobility:

father did not complete high school
but respondent attained a college degree

	Father's education (total $N = 2,630$)				
Respondent's education	Less than HS	High school	Some college	College	**MISSING**
College	**122**	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	7	168
MISSING	1	2	0	0	6



Upper bound

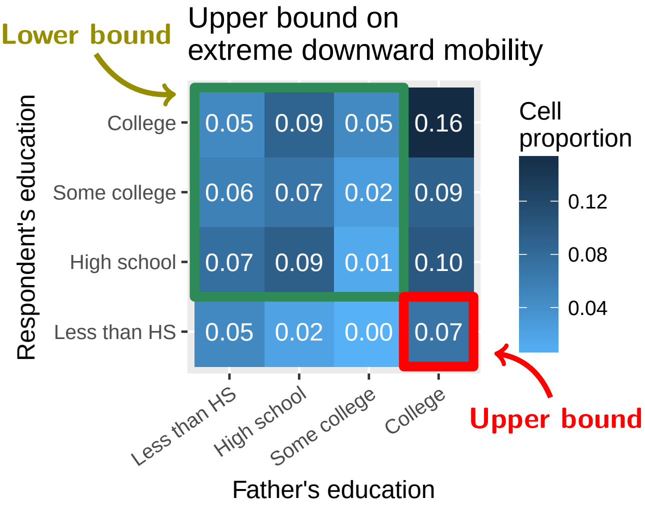
Lower bound

Respondent's education	Less than HS	High school	Some college	College
College	0.10	0.09	0.05	0.11
Some college	0.12	0.07	0.02	0.03
High school	0.16	0.09	0.01	0.02
Less than HS	0.11	0.02	0.00	0.00

Place an upper bound on the proportion of the sample with extreme downward mobility:

father completed college
but respondent did not complete high school

Respondent's education	Father's education (total $N = 2,630$)				
	Less than HS	High school	Some college	College	**MISSING**
College	122	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	**7**	168
MISSING	1	2	0	0	6



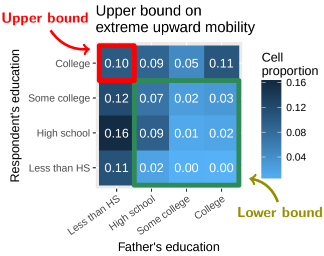
Lower bound

Upper bound on extreme downward mobility

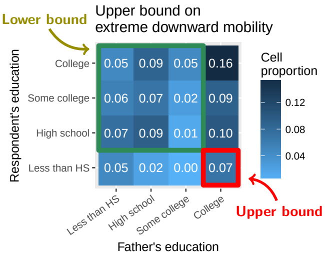
Upper bound

Respondent's education	Less than HS	High school	Some college	College
College	0.05	0.09	0.05	0.16
Some college	0.06	0.07	0.02	0.09
High school	0.07	0.09	0.01	0.10
Less than HS	0.05	0.02	0.00	0.07

Summarizing our bounds analysis



Respondent's education	Less than HS	High school	Some college	College
College	0.10	0.09	0.05	0.11
Some college	0.12	0.07	0.02	0.03
High school	0.16	0.09	0.01	0.02
Less than HS	0.11	0.02	0.00	0.00



Respondent's education	Less than HS	High school	Some college	College
College	0.05	0.09	0.05	0.16
Some college	0.06	0.07	0.02	0.09
High school	0.07	0.09	0.01	0.10
Less than HS	0.05	0.02	0.00	0.07

	Father's education (total $N = 2,630$)				
Respondent's education	Less than HS	High school	Some college	College	MISSING
College	122	230	119	280	135
Some college	146	184	65	76	164
High school	197	246	33	41	227
Less than HS	122	48	11	7	168
MISSING	1	2	0	0	6

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:
  - Y_i is bounded with support $[a, b]$

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:
  - Y_i is bounded with support $[a, b]$
  - We assume stable outcomes $Y_i^* = Y_i M_i + (\text{NA})(1 - M_i)$

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:
  - Y_i is bounded with support $[a, b]$
  - We assume stable outcomes $Y_i^* = Y_i M_i + (\text{NA})(1 - M_i)$
- We obtain sharp bounds for $E[Y]$ by first plugging in a for all missing values to get the lower bound, followed by plugging in b for all missing values to get the upper bound.

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:
  - Y_i is bounded with support $[a, b]$
  - We assume stable outcomes $Y_i^* = Y_i M_i + (\text{NA})(1 - M_i)$
- We obtain sharp bounds for $E[Y]$ by first plugging in a for all missing values to get the lower bound, followed by plugging in b for all missing values to get the upper bound.
- This leaves our quantity set identified as opposed to our usual point identified

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:
  - Y_i is bounded with support $[a, b]$
  - We assume stable outcomes $Y_i^* = Y_i M_i + (\text{NA})(1 - M_i)$
- We obtain sharp bounds for $E[Y]$ by first plugging in a for all missing values to get the lower bound, followed by plugging in b for all missing values to get the upper bound.
- This leaves our quantity set identified as opposed to our usual point identified
- Without further assumptions we can do no better.

Generalizing: When can you use sharp bounds? (Slide adapted from materials by Brandon Stewart)

The following gives us sharp Manski bounds (see e.g. Aronow and Miller Chapter 4)

- Assumptions:
  - Y_i is bounded with support $[a, b]$
  - We assume stable outcomes $Y_i^* = Y_i M_i + (\text{NA})(1 - M_i)$
- We obtain sharp bounds for $E[Y]$ by first plugging in a for all missing values to get the lower bound, followed by plugging in b for all missing values to get the upper bound.
- This leaves our quantity set identified as opposed to our usual point identified
- Without further assumptions we can do no better.
- This only works with bounded support and becomes much harder with missingness on many variables

1. Motivation

2. Listwise deletion

3. Bounds

4. Multiple imputation

5. Rubin's rules

6. Simulation

7. Review

8. Bonus slides to help with optional homework problem: EM

1 Motivation

2 Listwise deletion

3 Bounds

4 Multiple imputation

5 Rubin's rules

6 Simulation

7 Review

8 Bonus slides to help with optional homework problem: EM

Multiple imputation

Multiple imputation is a strategy to report a good point estimate with accurate uncertainty.

Multiple imputation

Multiple imputation is a strategy to report a good point estimate with accurate uncertainty.

If data are **missing at random** (MAR), then multiply imputed estimates are **unbiased**.

Multiple imputation

Multiple imputation is a strategy to report a good point estimate with accurate uncertainty.

If data are **missing at random** (MAR), then multiply imputed estimates are **unbiased**.

In many (but not all) settings, MAR is more plausible than MCAR.

Multiple imputation

Multiple imputation is a strategy to report a good point estimate with accurate uncertainty.

If data are **missing at random** (MAR), then multiply imputed estimates are **unbiased**.

In many (but not all) settings, MAR is more plausible than MCAR.

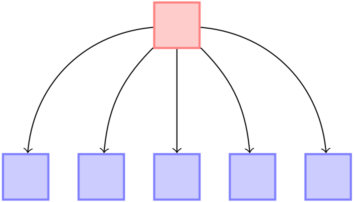
Missing at Random	Missing *Completely* at Random
$\text{P}(M \mid X_{\text{Obs}}, X_{\text{Miss}}) = \text{P}(M \mid X_{\text{Obs}})$	$\text{P}(M \mid X) = \text{P}(M)$
Missingness is independent of the true values given the observed values.	Missingness is independent of the true values.
Implies that multiple imputation is unbiased.	Implies that listwise deletion is unbiased.

The Multiple Imputation Scheme (from lecture)

The Multiple Imputation Scheme (from lecture)

incomplete data

The Multiple Imputation Scheme (from lecture)

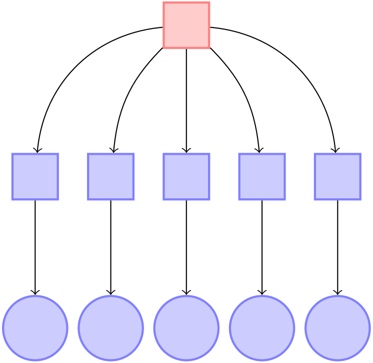


incomplete data

imputation

imputed datasets

The Multiple Imputation Scheme (from lecture)



incomplete data

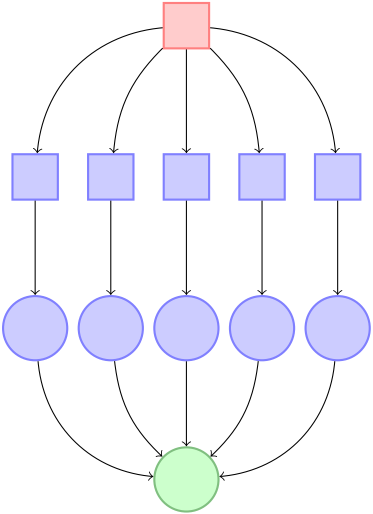
imputation

imputed datasets

analysis

separate results

The Multiple Imputation Scheme (from lecture)



incomplete data

imputation

imputed datasets

analysis

separate results

combination

final results

Choosing variables

We want to impute father's education with the set of variables $\vec{X}$ such that missingness is ignorable given the observed values of $\vec{X}$.

Choosing variables

We want to impute father's education with the set of variables $\vec{X}$ such that missingness is ignorable given the observed values of $\vec{X}$.

- Father's education
- Respondent's education
- Age
- Race
- Sex

Choosing variables

We want to impute father's education with the set of variables $\vec{X}$ such that missingness is ignorable given the observed values of $\vec{X}$.

- Father's education
- Respondent's education
- Age
- Race
- Sex

We have to argue for the MAR assumption: which observations are missing for each variable is independent of the true value.

Connection to causal inference

The **missing at random** assumption is analogous to the assumption of **selection on observables** in causal inference.

Connection to causal inference

The **missing at random** assumption is analogous to the assumption of **selection on observables** in causal inference.

In both cases, we assume the assignment of a binary variable (missingness or treatment assignment) is **conditionally ignorable** given $\vec{X}$.

Connection to causal inference

The **missing at random** assumption is analogous to the assumption of **selection on observables** in causal inference.

In both cases, we assume the assignment of a binary variable (missingness or treatment assignment) is **conditionally ignorable** given $\vec{X}$.

Ex. The two are very close when we have

- missingness on Y_i indicated by M_i
- but complete knowledge of $\vec{X}_i$

Connection to causal inference

	Multiple imputation	Causal inference (ATT)
Ultimate product	Infer Y_i for use in a model when $M_i = 1$.	Infer $\text{E}(Y_i(0) \mid D_i = 1)$ for use in estimation of the ATT: $\text{E}(Y_i(1) - Y_i(0) \mid D_i = 1)$
Goal	For those who are missing, estimate the outcome under non-missingness.	For those who are untreated, estimate the potential outcome under treatment.
Assumption	Missing at random $Y_i \perp\!\!\!\perp M_i \mid \vec{X}_i$	Selection on observables $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i \mid \vec{X}_i$

Choosing variables

```
d <- gss %>%
  filter(age >= 25) %>%
  select(id, age, race, sex,
         paeduc, educ) %>%
  mutate(sex = factor(sex, labels = c("Male","Female")),
         race = factor(race, labels = c("White","Black","Other")),
         paeduc = factor(ifelse(paeduc < 12, 1,
                                ifelse(paeduc == 12, 2,
                                       ifelse(paeduc < 16, 3,
                                              ifelse(paeduc >= 16 & paeduc <= 2
                                                     paeduc)))),
                         labels = c("Less than HS","High school",
                                    "Some college","College")),
         educ = factor(ifelse(educ < 12, 1,
                              ifelse(educ == 12, 2,
                                     ifelse(educ < 16, 3,
                                            ifelse(educ >= 16 & educ <= 20, 4,
                                                   educ)))),
                       labels = c("Less than HS","High school",
                                  "Some college","College")))
```

Choosing variables

```
> summary(d)
       id              age            race          sex      
 Min.   :   1.0   Min.   :25.00   White:1941   Male  :1164  
 1st Qu.: 706.2   1st Qu.:37.00   Black: 444   Female:1466  
 Median :1432.5   Median :52.00   Other: 245                
 Mean   :1429.3   Mean   :51.53                             
 3rd Qu.:2143.8   3rd Qu.:63.00                             
 Max.   :2867.0   Max.   :89.00                             
          paeduc               educ     
 Less than HS:588   Less than HS:356  
 High school :710   High school :744  
 Some college:228   Some college:635  
 College     :404   College     :886  
 NA's        :700   NA's        :  9  
```

Should we transform variables?

Quoted from Amelia documentation [link], p. 16:

> *As it turns out, much evidence in the literature (discussed in King et al. 2001) indicates that the multivariate normal model used in Amelia usually works well for the imputation stage even when discrete or non-normal variables are included and when the analysis stage involves these limited dependent variable models.*

In our example, we will still transform so we can use the nominal variables later in their nominal form.

Implementation in Amelia

Run Amelia

```
filled <- amelia(data.frame(d) %>%
                   mutate(educ = educ,
                          paeduc = paeduc,
                          race = race,
                          sex = sex),
                 ords = c("paeduc","educ"),
                 noms = c("race","sex"),
                 idvars = "id")
```

Run Amelia

```
> summary(filled)

Amelia output with 5 imputed datasets.
Return code:  1
Message:  Normal EM convergence.

Chain Lengths:
--------------
Imputation 1:  3
Imputation 2:  3
Imputation 3:  3
Imputation 4:  3
Imputation 5:  3
```

Run Amelia

```
Rows after Listwise Deletion:  1927
Rows after Imputation:  2630
Patterns of missingness in the data:  4

Fraction Missing for original variables:
-----------------------------------------

       Fraction Missing
id          0.000000000
age         0.000000000
race        0.000000000
sex         0.000000000
paeduc      0.266159696
educ        0.003422053
```

Patterns of missingness

Amelia told us there were 20 patterns of missignness. What were they?

Patterns of missingness

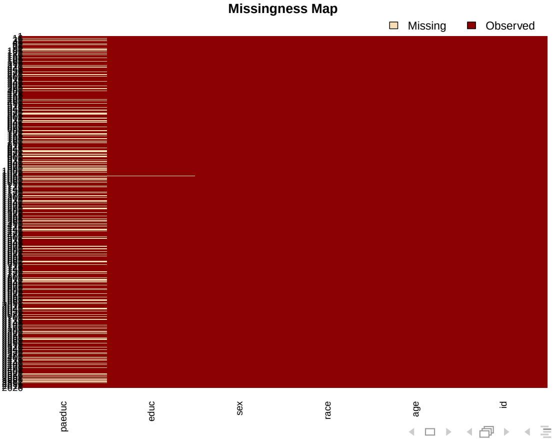
Amelia told us there were 20 patterns of missignness. What were they?

```
missmap(filled)
```

Patterns of missingness

Amelia told us there were 20 patterns of missignness. What were they?

```
missmap(filled)
```



Checking convergence

EM can sometimes end up in weird places.

Checking convergence

EM can sometimes end up in weird places.

We want to know our results converge the same place regardless of the starting values.

Checking convergence

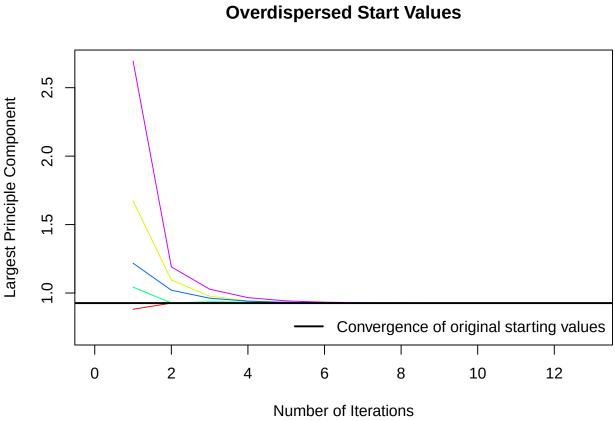
EM can sometimes end up in weird places.

We want to know our results converge the same place regardless of the starting values.

Amelia's `disperse()` command shows us that the first principle component (a unidimensional summary of the data) converges to the same value regardless of a few randomly chosen starting points.

Checking convergence

```
disperse(filled, dims = 1, m = 5)
```



Amelia objects

Your Amelia object holds lots of things, including 5 versions of the data.

Amelia objects

Your Amelia object holds lots of things, including 5 versions of the data.

```
> head(filled$imputations$imp1)
  id age  race    sex       paeduc         educ
1  1  47 White   Male      College      College
2  2  61 White   Male Less than HS  High school
3  3  72 White   Male  High school      College
4  4  43 White Female Less than HS  High school
5  5  55 White Female      College      College
6  6  53 White Female Less than HS Some college
```

`transform`: Operating on an Amelia object

What if we now want the respondent's education to be coded as college or not?

`transform`: Operating on an Amelia object

What if we now want the respondent's education to be coded as college or not? `transform` operates on all imputations at once.

transform: Operating on an Amelia object

What if we now want the respondent's education to be coded as college or not? transform operates on all imputations at once.

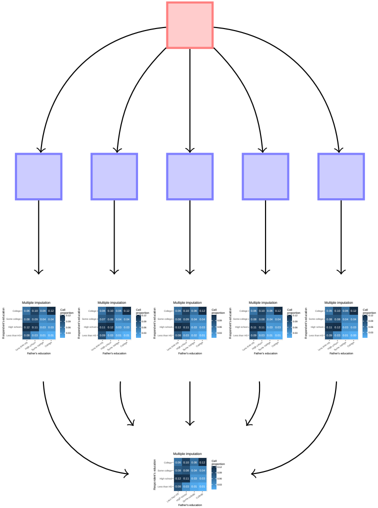
```
filled_transformed <- transform(
  filled,
  college = (educ == "College")
)
```

transform: Operating on an Amelia object

What if we now want the respondent's education to be coded as college or not? transform operates on all imputations at once.

```
filled_transformed <- transform(
  filled,
  college = (educ == "College")
)
> head(filled_transformed$imputations$imp1)
  id age  race    sex       paeduc          educ college
1  1  47 White   Male      College      College    TRUE
2  2  61 White   Male Less than HS  High school   FALSE
3  3  72 White   Male  High school      College    TRUE
4  4  43 White Female Less than HS  High school   FALSE
5  5  55 White Female      College      College    TRUE
6  6  53 White Female Less than HS Some college   FALSE
```

The Multiple Imputation Scheme (from lecture)



incomplete data

imputation

imputed datasets

analysis

separate results

combination

final results

1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

Combining results: Rubin's rules

Multiple imputation captures uncertainty by combining our uncertainty

- **within** each imputation and
- **across** imputations.

Combining results: Rubin's rules

Multiple imputation captures uncertainty by combining our uncertainty

- **within** each imputation and
- **across** imputations.

We do this using Rubin's rules (idea is important, formula is not).

$$\hat{\mathrm{V}}(\hat{\theta}) = \underbrace{\frac{1}{m}\sum_{i=1}^{m}\hat{\mathrm{V}}(\hat{\theta}_i)}_{\substack{\text{Mean} \\ \textbf{within-imputation}\text{ variance}}} + \underbrace{\left(1+\frac{1}{m}\right)}_{\substack{\text{Inflation} \\ \text{factor}}}\underbrace{\left(\frac{1}{m-1}\sum_{i=1}^{m}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]\right)}_{\textbf{Between-imputation}\text{ variance}}$$

Combining results: Rubin's rules

Multiple imputation captures uncertainty by combining our uncertainty

- **within** each imputation and
- **across** imputations.

We do this using Rubin's rules (idea is important, formula is not).

$$\hat{\text{V}}(\hat{\theta}) = \underbrace{\frac{1}{m}\sum_{i=1}^{m} \hat{\text{V}}(\hat{\theta}_i)}_{\substack{\text{Mean} \\ \textbf{within-imputation}\text{ variance}}} + \underbrace{\left(1+\frac{1}{m}\right)}_{\substack{\text{Inflation} \\ \text{factor}}} \underbrace{\left(\frac{1}{m-1}\sum_{i=1}^{m}\left[\hat{\theta}_i - \bar{\hat{\theta}}\right]\right)}_{\textbf{Between-imputation}\text{ variance}}$$

Ignored by single imputation

Example: Rubin's rules

Proportion of population with extreme upward mobility
(father $<$ HS, respondent college)

	Proportion	Confidence interval
Imputation 1	0.057	(0.049, 0.066)
Imputation 2	0.061	(0.052, 0.070)
Imputation 3	0.057	(0.048, 0.065)
Imputation 4	0.059	(0.050, 0.068)
Imputation 5	0.057	(0.049, 0.066)

Example: Rubin's rules

Proportion of population with extreme upward mobility
(father $<$ HS, respondent college)

	Proportion	Confidence interval
Imputation 1	0.057	(0.049, 0.066)
Imputation 2	0.061	(0.052, 0.070)
Imputation 3	0.057	(0.048, 0.065)
Imputation 4	0.059	(0.050, 0.068)
Imputation 5	0.057	(0.049, 0.066)

↑

Uncertainty
across
imputations

Example: Rubin's rules

Proportion of population with extreme upward mobility
(father $<$ HS, respondent college)

	Proportion	Confidence interval
Imputation 1	0.057	(0.049, 0.066)
Imputation 2	0.061	(0.052, 0.070)
Imputation 3	0.057	(0.048, 0.065)
Imputation 4	0.059	(0.050, 0.068)
Imputation 5	0.057	(0.049, 0.066)

↑ Uncertainty across imputations

↑ Uncertainty within each imputation

Example: Rubin's rules

Proportion of population with extreme upward mobility
(father $<$ HS, respondent college)

	Proportion	Confidence interval
Imputation 1	0.057	(0.049, 0.066)
Imputation 2	0.061	(0.052, 0.070)
Imputation 3	0.057	(0.048, 0.065)
Imputation 4	0.059	(0.050, 0.068)
Imputation 5	0.057	(0.049, 0.066)

↑ Uncertainty across imputations

↑ Uncertainty within each imputation

Overall estimate: 0.058 (0.049, 0.068)

The above focuses on intuition.

The next slides walk through doing this in code.

In code: Applying Rubin's rules. Step 1

Produce a list of estimates from each imputation

```
estimates <- lapply(filled$imputations, function(imp) {
  within_imp <- imp %>%
    summarize(proportion = mean(paeduc == 1 & educ == 4))
  return(within_imp$proportion)
})
```

In code: Applying Rubin's rules. Step 2

Produce a list of the variance of each imputation-specific estimate

```
estimate_variances <- lapply(filled$imputations, function(imp) {
  within_imp <- imp %>%
    summarize(proportion = mean(paeduc == 1 & educ == 4),
              num = n()) %>%
    mutate(var_proportion = proportion * (1 - proportion) / num)
  return(within_imp$var_proportion)
})
```

In code: Applying Rubin's rules. Step 3

The mitools package has a function MIcombine that applies Rubin's rules for you.

You give it a list of results and a list of variances estimated within each imputation.

```
combined <- MIcombine(
  results = estimates,
  variances = estimate_variances
)
```

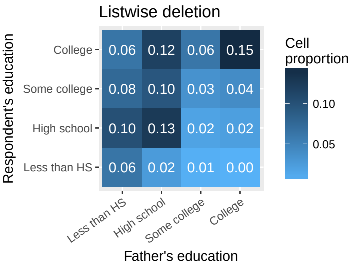
In code: Applying Rubin's rules. Step 4

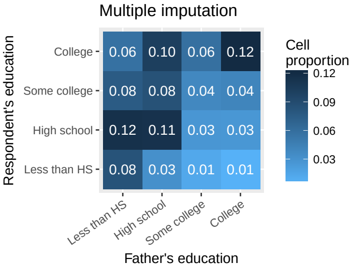
Report the estimate with a 95% confidence interval.

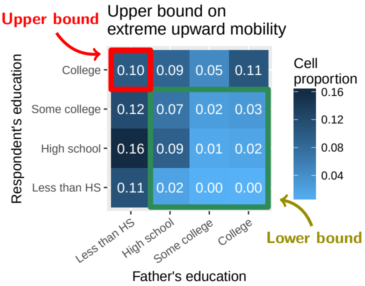
```
combined$coefficients
c(combined$coefficients - qnorm(.975) * sqrt(combined$variance),
  combined$coefficients + qnorm(.975) * sqrt(combined$variance))
```

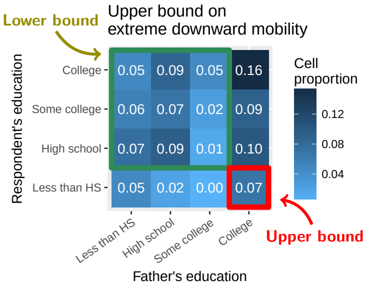
Overall estimate: 0.058 (0.049, 0.068)

All estimators together









1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

1 Motivation

2 Listwise deletion

3 Bounds

4 Multiple imputation

5 Rubin's rules

6 Simulation

7 Review

8 Bonus slides to help with optional homework problem: EM

Combining by simulation

We can do the same thing by **simulation**.

1. On each imputation $j = 1, \ldots, m$
    1. Fit a model
    2. Store your estimate $\hat{\tau}_j = h\left(\hat{\vec{\theta}}_j\right)$
    3. Draw $r = 1,000$ simulations
2. Report your **point estimate**:

$$\hat{\tau} = \frac{1}{m}\sum_{j=1}^{m}\hat{\tau}_j$$

3. Report **uncertainty**
    1. **Pool** all the simulations.
    2. Report a 95% quantile-based confidence interval

Combining by simulation: Step 1. Fit a model

College completion

$$
\begin{aligned}
Y_i &\sim \text{Bernoulli}(\pi_i) \\
\text{logit}(\pi_i) &= \vec{X}_i \vec{\beta}
\end{aligned}
$$

Father's education

Store our estimate $\hat{\tau}_j = \text{logit}^{-1}\left(x\hat{\vec{\beta}}_j\right)$

Draw 1,000 samples of our quantity of interest

$$\tilde{\vec{\beta}} \sim \text{Normal}\left(\hat{\vec{\beta}}, \hat{\text{V}}\left(\hat{\vec{\beta}}\right)\right))$$

$$\tilde{\tau}_{j[r]} = \text{logit}^{-1}\left(x\tilde{\vec{\beta}}_j\right)$$

Combining by simulation: Step 1. Fit a model

Combining by simulation: Step 1. Fit a model

```
r <- 1000
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
fits <- foreach(i = 1:length(filled_transformed$imputations)) %do% {
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
fits <- foreach(i = 1:length(filled_transformed$imputations)) %do% {

  fit <- glm(college ˜ paeduc,
             data = filled_transformed$imputations[[i]],
             family = binomial(link = "logit"))
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
fits <- foreach(i = 1:length(filled_transformed$imputations)) %do% {

  fit <- glm(college ~ paeduc,
             data = filled_transformed$imputations[[i]],
             family = binomial(link = "logit"))
  estimate <- plogis(setx %*% coef(fit))
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
fits <- foreach(i = 1:length(filled_transformed$imputations)) %do% {

  fit <- glm(college ~ paeduc,
             data = filled_transformed$imputations[[i]],
             family = binomial(link = "logit"))
  estimate <- plogis(setx %*% coef(fit))
  sim_beta <- rmvnorm(r, mean = coef(fit), sigma = vcov(fit))
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
fits <- foreach(i = 1:length(filled_transformed$imputations)) %do% {

  fit <- glm(college ˜ paeduc,
             data = filled_transformed$imputations[[i]],
             family = binomial(link = "logit"))
  estimate <- plogis(setx %*% coef(fit))
  sim_beta <- rmvnorm(r, mean = coef(fit), sigma = vcov(fit))
  simulations <- plogis(setx %*% t(sim_beta))
```

Combining by simulation: Step 1. Fit a model

```
r <- 1000
setx <- c(1,0,0,0)
fits <- foreach(i = 1:length(filled_transformed$imputations)) %do% {

  fit <- glm(college ~ paeduc,
             data = filled_transformed$imputations[[i]],
             family = binomial(link = "logit"))
  estimate <- plogis(setx %*% coef(fit))
  sim_beta <- rmvnorm(r, mean = coef(fit), sigma = vcov(fit))
  simulations <- plogis(setx %*% t(sim_beta))
  return(list(estimate = estimate,
              simulations = simulations))
}
```

Combining by simulation: Step 2. Report a point estimate

This is the same as with Rubin's rules:
average the imputation-specific estimates.

$$\hat{\tau} = \frac{1}{m}\sum_{j=1}^{m}\hat{\tau}_j$$

Combining by simulation: Step 2. Report a point estimate

This is the same as with Rubin's rules:
average the imputation-specific estimates.

$$\hat{\tau} = \frac{1}{m}\sum_{j=1}^{m}\hat{\tau}_j$$

$$\begin{aligned}\hat{\text{P}}(\text{College} \mid \text{Father} < \text{HS}) &= \frac{1}{m}\sum_{i=1}^{m}\hat{\tau}_j \\ &= \frac{1}{5}(0.18 + 0.17 + 0.18 + 0.17 + 0.16) \\ &= 0.17\end{aligned}$$

Combining by simulation: Step 2. Report a point estimate

This is the same as with Rubin's rules:
average the imputation-specific estimates.

$$\hat{\tau} = \frac{1}{m}\sum_{j=1}^{m}\hat{\tau}_j$$

$$\begin{aligned}\hat{\mathrm{P}}(\text{College} \mid \text{Father} < \text{HS}) &= \frac{1}{m}\sum_{i=1}^{m}\hat{\tau}_j \\ &= \frac{1}{5}(0.18 + 0.17 + 0.18 + 0.17 + 0.16) \\ &= 0.17\end{aligned}$$

```
estimate <- mean(sapply(fits, function(fit) fit$estimate))
```

Combining by simulation: Step 3. Report uncertainty

Pool all simulations

$$\tilde{\tau} = \begin{bmatrix} \tilde{\tau}_1 \\ \vdots \\ \tilde{\tau}_m \end{bmatrix}$$

Combining by simulation: Step 3. Report uncertainty

Pool all simulations

$$\tilde{\tau} = \begin{bmatrix} \tilde{\tau}_1 \\ \vdots \\ \tilde{\tau}_m \end{bmatrix}$$

Report a quantile-based confidence interval

$$\left(\tilde{\tau}_{(.025)}, \tilde{\tau}_{(.975)}\right) = (0.15, 0.20)$$

Combining by simulation: Step 3. Report uncertainty

Pool all simulations

$$\tilde{\tau} = \begin{bmatrix} \tilde{\tau}_1 \\ \vdots \\ \tilde{\tau}_m \end{bmatrix}$$

Report a quantile-based confidence interval

$$\left(\tilde{\tau}_{(.025)}, \tilde{\tau}_{(.975)}\right) = (0.15, 0.20)$$

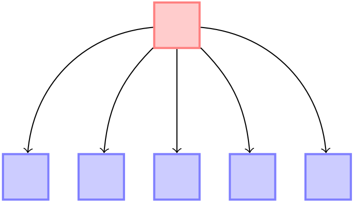
```
pooled_simulations <- do.call(
  c,
  lapply(fits, function(fit) fit$simulations)
)
ci <- quantile(pooled_simulations, c(.025, .975))
```

The Multiple Imputation Scheme (again)

The Multiple Imputation Scheme (again)

incomplete data

The Multiple Imputation Scheme (again)

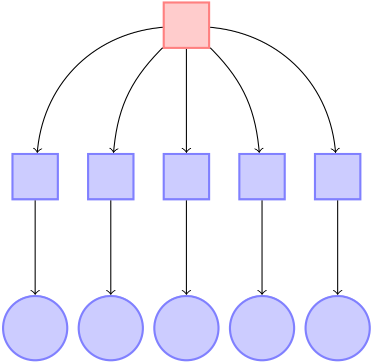


incomplete data

imputation

imputed datasets

The Multiple Imputation Scheme (again)



incomplete data

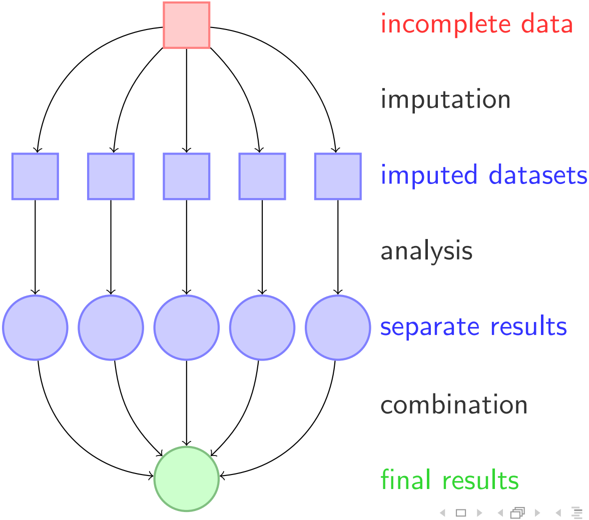
imputation

imputed datasets

analysis

separate results

The Multiple Imputation Scheme (again)



Missingness Assumptions (adapted from lecture)

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

$$P(M|X) = P(M)$$

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

$$P(M|X) = P(M)$$

Missingness (M) is unrelated to father's education (X)

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

   $$P(M|X) = P(M)$$

   Missingness (M) is unrelated to father's education (X)

2. **MAR**: Missing At Random (empirical)

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

   $$P(M|X) = P(M)$$

   Missingness (M) is unrelated to father's education (X)

2. **MAR**: Missing At Random (empirical)

   $$P(M|X,Z) = P(M|Z)$$

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

   $$P(M|X) = P(M)$$

   Missingness (M) is unrelated to father's education (X)

2. **MAR**: Missing At Random (empirical)

   $$P(M|X,Z) = P(M|Z)$$

   Missingness is not a function of the missing variable (X = Father's education), conditional on measured variables (Z = Mother's education)
   e.g., Children with lesser-educated mothers are more likely to have missing fathers

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

   $$P(M|X) = P(M)$$

   Missingness (M) is unrelated to father's education (X)

2. **MAR**: Missing At Random (empirical)

   $$P(M|X, Z) = P(M|Z)$$

   Missingness is not a function of the missing variable ($X =$ Father's education), conditional on measured variables ($Z =$ Mother's education)
   e.g., Children with lesser-educated mothers are more likely to have missing fathers

3. **NI**: Non-ignorable (fatalistic)
   $P(M|X)$ doesn't simplify

Missingness Assumptions (adapted from lecture)

1. **MCAR**: Missing Completely At Random (naive)

   $$P(M|X) = P(M)$$

   Missingness (M) is unrelated to father's education (X)

2. **MAR**: Missing At Random (empirical)

   $$P(M|X,Z) = P(M|Z)$$

   Missingness is not a function of the missing variable ($X =$ Father's education), conditional on measured variables ($Z =$ Mother's education)
   e.g., Children with lesser-educated mothers are more likely to have missing fathers

3. **NI**: Non-ignorable (fatalistic)
   $P(M|X)$ doesn't simplify
   e.g., within cells of mother's education, missingness is still related to father's education

   Adding variables to predict father's education can change NI to MAR

Assumptions in actual sociology research

Next, we will discuss these assumptions in the context of real publications.

I will walk through an example from AJS.

Then in groups you will do the same thing for 4 papers from the current issue of ASR.

We will highlight **what they do well** as well as things we might do differently. Remember these are very good papers!

Assumptions in actual sociology research

In groups, take 10 minutes to discuss the ways missing data were addressed in actual papers.

Then summarize for the class:

- Broadly what the paper argues (very brief)
- The missing data problem in the paper
- How the authors addressed it
- Whether this was appropriate
- Any other information you wish the author provided
- What you would do

Lasting Consequences of the Summer Learning Gap

Karl L. Alexander
Johns Hopkins University

Doris R. Entwisle
Johns Hopkins University

Linda Steffel Olson
Johns Hopkins University

Prior research has demonstrated that summer learning rooted in family and community influences widens the achievement gap across social lines, while schooling offsets those family and community influences. In this article, we examine the long-term educational consequences of summer learning differences by family socioeconomic level. Using data from the Baltimore Beginning School Study youth panel, we decompose achievement scores at the start of high school into their developmental precursors, back to the time of school entry in 1st grade. We find that cumulative achievement gains over the first nine years of children's schooling mainly reflect school-year learning, whereas the high SES–low SES achievement gap at 9th grade mainly traces to differential summer learning over the elementary years. These early out-of-school summer learning differences, in turn, substantially account for achievement-related differences by family SES in high school track placements (college preparatory or not), high school noncompletion, and four-year college attendance. We discuss implications for understanding the bases of educational stratification, as well as educational policy and practice.

Comparisons of school-year and summer learning inform fundamental questions of educational stratification and help parse school, family, and community influences on children's academic development. With children "in" their homes, schools, and communities during the school year, but just "in" their homes and communities over the summer months, the academic

gaps by family SES (socioeconomic status) and race/ethnicity widen more during the summer months than during the school year.

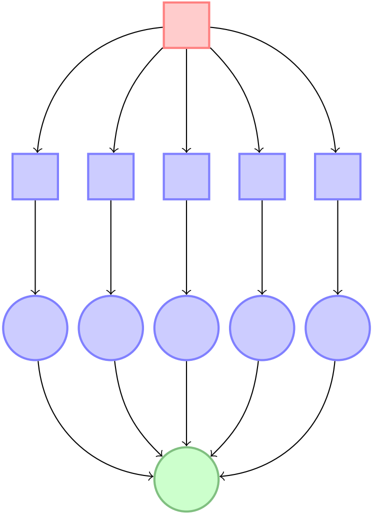
Although the detailed results of subsequent research on the seasonality of learning do not line up perfectly (see Cooper and colleagues' [1996] meta-analysis for an overview), the patterns documented by Heyns in the 1970s for

SAMPLE AND METHODS

The BSS panel consists of a representative random sample of Baltimore school children whose educational progress has been monitored from 1st grade through age 22. The project began in the fall of 1982, when the study participants (N = 790), randomly selected from 20 public elementary schools within strata defined by school spring of year 9). This is an uncommonly rich set of testing data, but owing to absences, transfers outside the city school system, and other complications, not all children were tested on every occasion. Case coverage when screened on complete testing data is 326 (from 790 originally). Additionally, some positive selection is evident—while fall of 1st grade scores are close (281.7 for the listwise sample and 280.6 for the full sample), by year 9 the listwise group's spring average is .18 SD above the full sample average. However, the 464 excluded cases include many with nearly complete testing records, and some useful testing data are available for just about everyone. For example, 81 have data for at least four test scores.[2] To take advantage of this circumstance, we generated an imputed version of the raw data (based on 10 imputations) using multiple imputation methods (e.g., Allison 2002). These methods predict missing scores from the available data (including spring scores over years 6 through 8, which are not used in the substantive analyses), plus race, sex, and family SES background (the continuous version), which are known for all but three cases, and high school track placement. We

The imputed achievement data were derived as the average of the 10 versions generated by the imputation process. We then used these scores (fall and spring over the elementary years and spring of year 9) to calculate the four achievement components used in the analyses:

The Multiple Imputation Scheme (again)



incomplete data

imputation

imputed datasets

analysis

separate results

combination

final results

Assumptions in actual sociology research

In groups, take 10 minutes to discuss the ways missing data were addressed in actual papers.

Then summarize for the class:

- Broadly what the paper argues (very brief)
- The missing data problem in the paper
- How the authors addressed it
- Whether this was appropriate
- Any other information you wish the author provided
- What you would do

American Sociological Review
2018, Vol. 83(2) 278–304
© American Sociological
Association 2018
DOI: 10.1177/0003122418759651
journals.sagepub.com/home/asr

Social and Genetic Pathways in Multigenerational Transmission of Educational Attainment

Hexuan Liu[a]

Abstract

This study investigates the complex roles of the social environment and genes in the multigenerational transmission of educational attainment. Drawing on genome-wide data and educational attainment measures from the Framingham Heart Study (FHS) and the Health and Retirement Study (HRS), I conduct polygenic score analyses to examine genetic confounding in the estimation of parents' and grandparents' influences on their children's and grandchildren's educational attainment. I also examine social genetic effects (i.e., genetic effects that operate through the social environment) in the transmission of educational attainment across three generations. Two-generation analyses produce three important findings. First, about one-fifth of the parent-child association in education reflects genetic inheritance. Second, up to half of the association between parents' polygenic scores and children's education is mediated by parents' education. Third, about one-third of the association between children's polygenic scores and their educational attainment is attributable to parents' genotypes and education. Three-generation analyses suggest that genetic confounding on the estimate of the direct effect of grandparents' education on grandchildren's education (net of parents' education) may be inconsequential, and I find no evidence that grandparents' genotypes significantly influence grandchildren's education through non-biological pathways. The three-generation results are suggestive, and the results may change when different samples are used.

Sensitivity Analyses

The three-generation results may suffer from sample attrition. In FHS, genotypes are only available for 21 percent of the participants in the original cohort (G1) (versus 71 percent in G2 and 95 percent in G3). Compared to participants whose genotypes are missing, those who provided genotypes are younger and better educated. This may lead to biased model estimates (Domingue et al. 2016; Liu and Guo 2015).

parents and children. To conduct a sensitivity test, I imputed missing PGSs in G1 based on G2's PGSs and G1's educational attainment using the multiple imputation technique (Rubin 1987). As a result, I imputed PGSs of an additional 1,897 G1 participants.

American Sociological Review
2018, Vol. 83(2) 305–330
© American Sociological Association 2018
DOI: 10.1177/0003122418759544
journals.sagepub.com/home/asr

Precarious Sexuality: How Men and Women Are Differentially Categorized for Similar Sexual Behavior

Trenton D. Mize[a] and Bianca Manago[b,c]

Abstract

Are men and women categorized differently for similar sexual behavior? Building on theories of gender, sexuality, and status, we introduce the concept of *precarious sexuality* to suggest that men's—but not women's—heterosexuality is an especially privileged identity that is easily lost. We test our hypotheses in a series of survey experiments describing a person who has a sexual experience conflicting with their sexual history. We find that a single same-sex sexual encounter leads an observer to question a heterosexual man's sexual orientation to a greater extent than that of a heterosexual woman in a similar situation. We also find that a different-sex sexual encounter is more likely to change others' perceptions of a lesbian woman's sexual orientation—compared to perceptions of a gay man's sexual orientation. In two conceptual replications, we vary the level of intimacy of the sexual encounter and find consistent evidence for our idea of precarious sexuality for heterosexual men. We close with a general discussion of how status beliefs influence categorization processes and with suggestions for extending our theoretical propositions to other categories beyond those of sexual orientation.

STUDY 1

Study 1 was conducted as an online survey experiment on a nationally representative panel. Data were collected by GfK with support from Time-Sharing Experiments for the Social Sciences (TESS) (Freese and Druckman 2015). A total of 2,035 participants were randomly selected from the GfK panel, which recruits participants via a combination of random-digit dialing and address-based sampling methods to ensure representativeness of the adult U.S. population.

Participants. A total of 46 participants did not provide usable answers for any of the dependent measures, and an additional 24 participants had missing data on at least one demographic variable and are thus not included in the analysis (final $N = 1{,}965$).[7] Descriptive statistics of the Study 1 sample are in Table S4 in Part B of the online supplement. The Study 1 sample had 932 men and 1,033 women; their

[7] Specifically, 22 respondents refused all the questions on the survey and 24 answered "0 percent" to all the dependent measures.

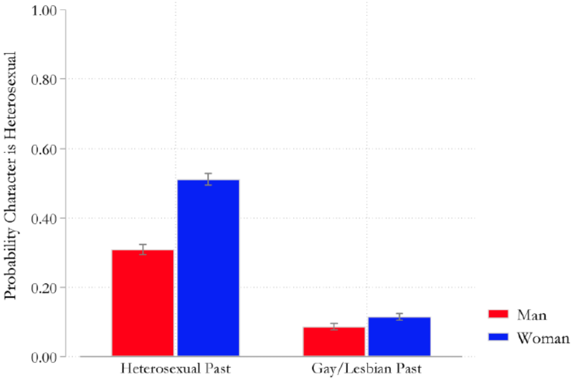


Figure 2. Probability Rating That Vignette Character Is Heterosexual, Study 1

American Sociological Review
2018, Vol. 83(2) 331–360
© American Sociological
Association 2018
DOI: 10.1177/0003122418762291
journals.sagepub.com/home/asr

The Mark of a Woman's Record: Gender and Academic Performance in Hiring

Natasha Quadlin[a]

Abstract

Women earn better grades than men across levels of education—but to what end? This article assesses whether men and women receive equal returns to academic performance in hiring. I conducted an audit study by submitting 2,106 job applications that experimentally manipulated applicants' GPA, gender, and college major. Although GPA matters little for men, women benefit from moderate achievement but *not* high achievement. As a result, high-achieving men are called back significantly more often than high-achieving women—at a rate of nearly 2-to-1. I further find that high-achieving women are most readily penalized when they major in math: high-achieving men math majors are called back three times as often as their women counterparts. A survey experiment conducted with 261 hiring decision-makers suggests that these patterns are due to employers' gendered standards for applicants. Employers value competence and commitment among men applicants, but instead privilege women applicants who are perceived as likeable. This standard helps moderate-achieving women, who are often described as sociable and outgoing, but hurts high-achieving women, whose personalities are viewed with more skepticism. These findings suggest that achievement invokes gendered stereotypes that penalize women for having good grades, creating unequal returns to academic performance at labor market entry.

Table 5 shows the effects of achievement and gender on applicants' chances of receiving a callback for jobs sorted by salary. The first two columns separate jobs into salary ranges: less than \$42,500 (the median for the sample), and \$42,500 or more. The third column uses the full sample and includes interactions between applicant characteristics and salary.[20]

[20] About 40 percent of job advertisements listed a starting salary or salary range. For jobs that posted a salary range, I assigned a salary equal to the median of that range. For jobs with missing salary data, I imputed the median salary for the job type (see Appendix Table A1) and region associated with that job. Ten salaries could not be imputed using the job type and region, so I imputed the median salary for the job type across all five regions.

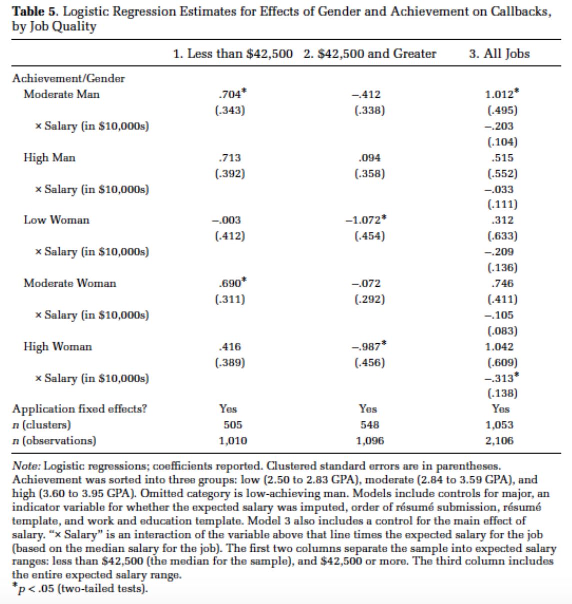


Table 5. Logistic Regression Estimates for Effects of Gender and Achievement on Callbacks, by Job Quality

	1. Less than $42,500	2. $42,500 and Greater	3. All Jobs
Achievement/Gender			
Moderate Man	.704*	–.412	1.012*
	(.343)	(.338)	(.495)
× Salary (in $10,000s)			–.203
			(.104)
High Man	.713	.094	.515
	(.392)	(.358)	(.552)
× Salary (in $10,000s)			–.033
			(.111)
Low Woman	–.003	–1.072*	.312
	(.412)	(.454)	(.633)
× Salary (in $10,000s)			–.209
			(.136)
Moderate Woman	.690*	–.072	.746
	(.311)	(.292)	(.411)
× Salary (in $10,000s)			–.105
			(.083)
High Woman	.416	–.987*	1.042
	(.389)	(.456)	(.609)
× Salary (in $10,000s)			–.313*
			(.138)
Application fixed effects?	Yes	Yes	Yes
n (clusters)	505	548	1,053
n (observations)	1,010	1,096	2,106

Note: Logistic regressions; coefficients reported. Clustered standard errors are in parentheses. Achievement was sorted into three groups: low (2.50 to 2.83 GPA), moderate (2.84 to 3.59 GPA), and high (3.60 to 3.95 GPA). Omitted category is low-achieving man. Models include controls for major, an indicator variable for whether the expected salary was imputed, order of résumé submission, résumé template, and work and education template. Model 3 also includes a control for the main effect of salary. "× Salary" is an interaction of the variable above that line times the expected salary for the job (based on the median salary for the job). The first two columns separate the sample into expected salary ranges: less than $42,500 (the median for the sample), and $42,500 or more. The third column includes the entire expected salary range.

$^*p < .05$ (two-tailed tests).

American Sociological Review
2018, Vol. 83(2) 390–417
© American Sociological Association 2018
DOI: 10.1177/0003122418759546
journals.sagepub.com/home/asr
SAGE

Mass Mobilization and the Durability of New Democracies

Mohammad Ali Kadivar[a]

Abstract

The "elitist approach" to democratization contends that "democratic regimes that last have seldom, if ever, been instituted by mass popular actors" (Huntington 1984:212). This article subjects this observation to empirical scrutiny using statistical analyses of new democracies over the past half-century and a case study. Contrary to the elitist approach, I argue that new democracies growing out of mass mobilization are more likely to survive than are new democracies that were born amid quiescence. Survival analysis of 112 young democracies in 80 different countries based on original data shows that the longer the mobilization, the more likely the ensuing democracy is to survive. I use a case study of South Africa to investigate the mechanisms. I argue that sustained unarmed uprisings have generated the longest-lasting new democracies—largely because they are forced to develop an organizational structure, which provides a leadership cadre for the new regime, forges links between the government and society, and strengthens checks on the power of the post-transition government.

Keywords

social movements, democratization, democracy, mobilization, organization, leadership, civil society

Using this measure, there are 115 new electoral democracies from 1960 to 2010. Because of missing socioeconomic data, I drop three regimes from the analysis.[4] Of the remaining

[4] Ghana 1956 to 1960, Tanzania 1960 to 1964, and Laos 1960 to 1962.

average age is 17.6, with a median of 18. This analysis only includes countries when they are democratic. Countries are not included in the sample before democratic transition or after democratic breakdown. Note that a given

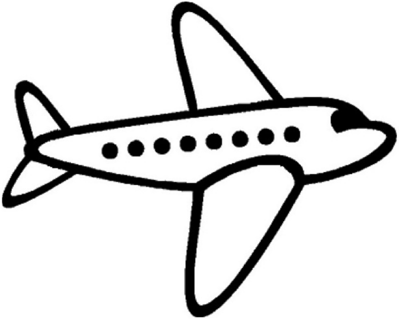
Revisiting our goals

By the end of precept, you should be able to:

1. Feel comfortable with three common **assumptions** about missing data
   - Missing completely at random
   - Missing at random
   - Non-ignorable
2. Be able to reason about the plausibility of these assumptions using **substantive knowledge** in real research settings.
3. Connect assumptions to concrete **strategies** to deal with missing data
   - Listwise deletion
   - Multiple imputation
   - Bounds

As you leave: Handout on poster and paper writing.
Next week: Design-based sampling.

Cards! Questions?



1. Motivation
2. Listwise deletion
3. Bounds
4. Multiple imputation
5. Rubin's rules
6. Simulation
7. Review
8. Bonus slides to help with optional homework problem: EM

1 Motivation

2 Listwise deletion

3 Bounds

4 Multiple imputation

5 Rubin's rules

6 Simulation

7 Review

8 Bonus slides to help with optional homework problem: EM

Mixture of exponentials

$$X_{0i} \sim \text{Exponential}(\lambda_0)$$

Mixture of exponentials

$$X_{0i} \sim \text{Exponential}(\lambda_0)$$
$$X_{1i} \sim \text{Exponential}(\lambda_1)$$

Mixture of exponentials

$$
\begin{aligned}
X_{0i} &\sim \text{Exponential}(\lambda_0) \\
X_{1i} &\sim \text{Exponential}(\lambda_1) \\
Z_i &\sim \text{Bernoulli}(p)
\end{aligned}
$$

Mixture of exponentials

$$
\begin{aligned}
X_{0i} &\sim \text{Exponential}(\lambda_0) \\
X_{1i} &\sim \text{Exponential}(\lambda_1) \\
Z_i &\sim \text{Bernoulli}(p) \\
Y_i &\equiv (1 - Z_i)X_{0i} + Z_i X_{1i}
\end{aligned}
$$

We observe Y_i but Z_i is **latent**: unobserved.
When there is a latent variable, you should think **EM**!

Simulate the data

We'll simulate some fake data and try to recover the parameters.

```
set.seed(08544)
```

Simulate the data

We'll simulate some fake data and try to recover the parameters.

```
set.seed(08544)
x0 <- rexp(100, rate = 0.5)
```

Simulate the data

We'll simulate some fake data and try to recover the parameters.

```
set.seed(08544)
x0 <- rexp(100, rate = 0.5)
x1 <- rexp(100, rate = 2)
```

Simulate the data

We'll simulate some fake data and try to recover the parameters.

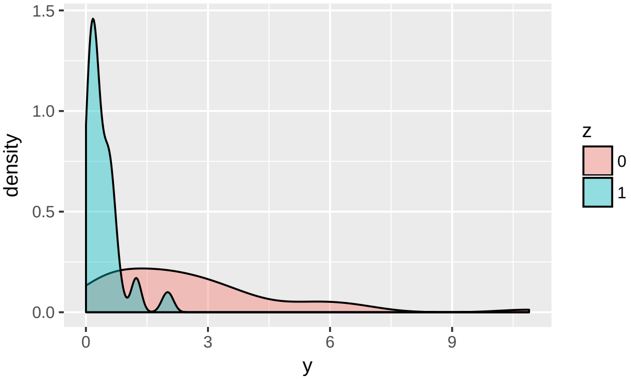
```
set.seed(08544)
x0 <- rexp(100, rate = 0.5)
x1 <- rexp(100, rate = 2)
z <- rbinom(100, size = 1, prob = .6)
```

Simulate the data

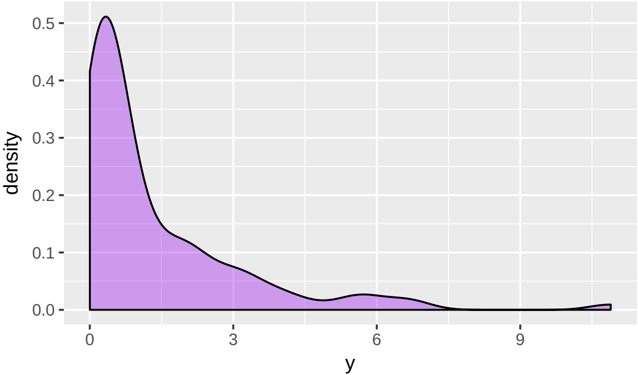
We'll simulate some fake data and try to recover the parameters.

```
set.seed(08544)
x0 <- rexp(100, rate = 0.5)
x1 <- rexp(100, rate = 2)
z <- rbinom(100, size = 1, prob = .6)
y <- (1 - z)*x0 + z*x1
```

Distribution within each (unknown) class Z



We observe this marginal distribution of Y



E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

We sometimes call these the responsibilities.

E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

We sometimes call these the responsibilities.

$$E(Z_i \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) =$$

E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

We sometimes call these the responsibilities.

$$E(Z_i \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) = P(Z_i = 1 \mid p^t, \lambda_0^t, \lambda_1^t, Y_i)$$

E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

We sometimes call these the responsibilities.

$$\begin{aligned} E(Z_i \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) &= P(Z_i = 1 \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) \\ &= \frac{P(Y_i \mid Z_i = 1) P(Z_i = 1)}{P(Y_i)} \end{aligned}$$

E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

We sometimes call these the responsibilities.

$$\begin{aligned} E(Z_i \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) &= P(Z_i = 1 \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) \\ &= \frac{P(Y_i \mid Z_i = 1)P(Z_i = 1)}{P(Y_i)} \\ &= \frac{P(Y_i \mid Z_i = 1)P(Z_i = 1)}{P(Y_i \mid Z_i = 1)P(Z_i = 1) + P(Y_i \mid Z_i = 0)P(Z_i = 0)} \end{aligned}$$

E-step

Find the expected value of the latent variable Z_i, given the parameters $\{p^t, \lambda_0^t, \lambda_1^t\}$ and the data Y_i.

We sometimes call these the responsibilities.

$$\begin{aligned}
E(Z_i \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) &= P(Z_i = 1 \mid p^t, \lambda_0^t, \lambda_1^t, Y_i) \\
&= \frac{P(Y_i \mid Z_i = 1)P(Z_i = 1)}{P(Y_i)} \\
&= \frac{P(Y_i \mid Z_i = 1)P(Z_i = 1)}{P(Y_i \mid Z_i = 1)P(Z_i = 1) + P(Y_i \mid Z_i = 0)P(Z_i = 0)} \\
&= \frac{\lambda_1 e^{-y_i \lambda_1} p}{\lambda_1 e^{-y_i \lambda_1} p + \lambda_0 e^{-y_i \lambda_0}(1 - p)}
\end{aligned}$$

Note: Conditioning on the parameters is not written explicitly after the first step to simplify the presentation. But all quantities throughout are conditional on p^t, λ_0^t, and $\lambda_1^t\}$. Likewise, P refers to both probability and probability densities for simplicity.

E-step

```
e.step <- function(p, lambda0, lambda1, y) {
  e.z <- lambda1 * exp(-y * lambda1) * p /
    lambda1 * exp(-y * lambda1) * p +
    lambda0 * exp(-y * lambda0) * (1 - p)
  return(e.z)
}
```

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) = f(y, z \mid p^t, \lambda_0^t, \lambda_1^t)$$

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$
\begin{aligned}
L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= f(y, z \mid p^t, \lambda_0^t, \lambda_1^t) \\
&= f(y \mid z, p^t, \lambda_0^t, \lambda_1^t) f(z)
\end{aligned}
$$

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$
\begin{aligned}
L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= f(y, z \mid p^t, \lambda_0^t, \lambda_1^t) \\
&= f(y \mid z, p^t, \lambda_0^t, \lambda_1^t) f(z) \\
&= \prod_{i=1}^{n} (\lambda_1 e^{-y_i \lambda_1})^{z_i} (\lambda_0 e^{-y_i \lambda_0})^{1-z_i} p^{z_i} (1-p)^{1-z_i}
\end{aligned}
$$

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$\begin{aligned} L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= f(y, z \mid p^t, \lambda_0^t, \lambda_1^t) \\ &= f(y \mid z, p^t, \lambda_0^t, \lambda_1^t) f(z) \\ &= \prod_{i=1}^{n} (\lambda_1 e^{-y_i \lambda_1})^{z_i} (\lambda_0 e^{-y_i \lambda_0})^{1-z_i} p^{z_i} (1-p)^{1-z_i} \\ \ell(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= \sum_{i=1}^{n} \Bigg(z_i (\log \lambda_1 - y_i \lambda_1) \end{aligned}$$

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$\begin{aligned}
L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= f(y, z \mid p^t, \lambda_0^t, \lambda_1^t) \\
&= f(y \mid z, p^t, \lambda_0^t, \lambda_1^t) f(z) \\
&= \prod_{i=1}^{n} (\lambda_1 e^{-y_i \lambda_1})^{z_i} (\lambda_0 e^{-y_i \lambda_0})^{1-z_i} p^{z_i} (1-p)^{1-z_i} \\
\ell(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= \sum_{i=1}^{n} \Bigg(z_i (\log \lambda_1 - y_i \lambda_1) \\
&\qquad + (1 - z_i)(\log \lambda_0 - y_i \lambda_0)
\end{aligned}$$

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$\begin{aligned}
L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= f(y, z \mid p^t, \lambda_0^t, \lambda_1^t) \\
&= f(y \mid z, p^t, \lambda_0^t, \lambda_1^t) f(z) \\
&= \prod_{i=1}^{n} (\lambda_1 e^{-y_i \lambda_1})^{z_i} (\lambda_0 e^{-y_i \lambda_0})^{1-z_i} p^{z_i} (1-p)^{1-z_i} \\
\ell(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= \sum_{i=1}^{n} \Big(z_i (\log \lambda_1 - y_i \lambda_1) \\
&\qquad + (1 - z_i)(\log \lambda_0 - y_i \lambda_0) \\
&\qquad + z_i \log p_i + (1 - z_i) \log(1 - p_i) \Big)
\end{aligned}$$

M-step

Find updated MLE estimates of $\{p^t, \lambda_0^t, \lambda_1^t\}$ using the data z^t created in the E-step.

First, write the complete data log likelihood, which includes both observed and latent variables.

$$
\begin{aligned}
L(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= f(y, z \mid p^t, \lambda_0^t, \lambda_1^t) \\
&= f(y \mid z, p^t, \lambda_0^t, \lambda_1^t) f(z) \\
&= \prod_{i=1}^{n} (\lambda_1 e^{-y_i \lambda_1})^{z_i} (\lambda_0 e^{-y_i \lambda_0})^{1-z_i} p^{z_i} (1-p)^{1-z_i} \\
\ell(p^t, \lambda_0^t, \lambda_1^t \mid y, z) &= \sum_{i=1}^{n} \Bigg(z_i(\log \lambda_1 - y_i \lambda_1) \\
&\qquad + (1 - z_i)(\log \lambda_0 - y_i \lambda_0) \\
&\qquad + z_i \log p_i + (1 - z_i) \log(1 - p_i) \Bigg)
\end{aligned}
$$

M-step

```
comp.data.log.lik <- function(par,z,y) {
```

M-step

```
comp.data.log.lik <- function(par,z,y) {
  p <- plogis(par[1])
  lambda0 <- exp(par[2])
  lambda1 <- exp(par[3])
```

M-step

```
comp.data.log.lik <- function(par,z,y) {
  p <- plogis(par[1])
  lambda0 <- exp(par[2])
  lambda1 <- exp(par[3])
  log.lik <- sum(z*(log(lambda1) - y*lambda1) +
                   (1 - z)*(log(lambda0) - y*lambda0) +
                   z*log(p) + (1 - z)*log(1 - p))
  return(log.lik)
}
```

M-step

Write a function to maximize that log likelihood

```
m.step <- function(z,y) {
```

M-step

Write a function to maximize that log likelihood

```
m.step <- function(z,y) {
  opt.out <- optim(
    par = c(0,0,0),
    z = z,
    y = y,
    fn = comp.data.log.lik,
    method = "BFGS",
    control = list(fnscale = -1)
  )
```

M-step

Write a function to maximize that log likelihood

```
m.step <- function(z,y) {
  opt.out <- optim(
    par = c(0,0,0),
    z = z,
    y = y,
    fn = comp.data.log.lik,
    method = "BFGS",
    control = list(fnscale = -1)
  )
  p <- plogis(opt.out$par[1])
  lambda0 <- exp(opt.out$par[2])
  lambda1 <- exp(opt.out$par[3])
  return(list(p = p, lambda0 = lambda0,
              lambda1 = lambda1))
}
```

Put E and M together!

Initialize the matrix to store parameters

```
par.estimates <- matrix(nrow = 11, ncol = 3)
colnames(par.estimates) <- c("p.t","lambda0.t","lambda1.t")
```

Put E and M together!

Initialize the matrix to store parameters

```
par.estimates <- matrix(nrow = 11, ncol = 3)
colnames(par.estimates) <- c("p.t","lambda0.t","lambda1.t"
```

Choose starting values

```
p.t <- 0.5
lambda0.t <- 1
lambda1.t <- 1
set.seed(12345)
z.t <- rbinom(n = length(y),
              size = 1,
              prob = .5)
```

Put E and M together!

Initialize the matrix to store parameters

```
par.estimates <- matrix(nrow = 11, ncol = 3)
colnames(par.estimates) <- c("p.t","lambda0.t","lambda1.t")
```

Choose starting values

```
p.t <- 0.5
lambda0.t <- 1
lambda1.t <- 1
set.seed(12345)
z.t <- rbinom(n = length(y),
              size = 1,
              prob = .5)
```

Store our starting parameters in the matrix

```
par.estimates[1,] <- c(p.t, lambda0.t, lambda1.t)
```

Put E and M together!

Iterate

```
for (i in 2:11) {
```

Put E and M together!

Iterate

```
for (i in 2:11) {
  z.t <- e.step(p = p.t,
                lambda0 = lambda0.t,
                lambda1 = lambda1.t,
                y = y)
```

Put E and M together!

Iterate

```
for (i in 2:11) {
  z.t <- e.step(p = p.t,
                lambda0 = lambda0.t,
                lambda1 = lambda1.t,
                y = y)

  m.out <- m.step(z = z.t, y = y)
```

Put E and M together!

Iterate

```
for (i in 2:11) {
  z.t <- e.step(p = p.t,
                lambda0 = lambda0.t,
                lambda1 = lambda1.t,
                y = y)

  m.out <- m.step(z = z.t, y = y)

  p.t <- m.out$p
  lambda0.t <- m.out$lambda0
  lambda1.t <- m.out$lambda1
```

Put E and M together!

Iterate

```
for (i in 2:11) {
  z.t <- e.step(p = p.t,
                lambda0 = lambda0.t,
                lambda1 = lambda1.t,
                y = y)

  m.out <- m.step(z = z.t, y = y)

  p.t <- m.out$p
  lambda0.t <- m.out$lambda0
  lambda1.t <- m.out$lambda1

  par.estimates[i,] <- c(p.t, lambda0.t, lambda1.t)
}
```

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548
5	0.2874	0.6019	1.9603

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548
5	0.2874	0.6019	1.9603
6	0.2835	0.6042	1.9580

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548
5	0.2874	0.6019	1.9603
6	0.2835	0.6042	1.9580
7	0.2851	0.6033	1.9589

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548
5	0.2874	0.6019	1.9603
6	0.2835	0.6042	1.9580
7	0.2851	0.6033	1.9589
8	0.2844	0.6037	1.9585

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548
5	0.2874	0.6019	1.9603
6	0.2835	0.6042	1.9580
7	0.2851	0.6033	1.9589
8	0.2844	0.6037	1.9585
9	0.2847	0.6035	1.9587

EM convergence

Iteration	p^t	λ_0^t	λ_1^t
0	0.5000	1.0000	1.0000
1	0.3482	0.5409	2.7712
2	0.2474	0.6271	1.8944
3	0.3010	0.5934	1.9726
4	0.2779	0.6076	1.9548
5	0.2874	0.6019	1.9603
6	0.2835	0.6042	1.9580
7	0.2851	0.6033	1.9589
8	0.2844	0.6037	1.9585
9	0.2847	0.6035	1.9587
10	0.2846	0.6036	1.9586