Who's Underrepresented? Modeling Undercount in the U.S. Census

Maria Tackett
Duke University

JSM
August 2020

🔗 bit.ly/jsm2020-teach

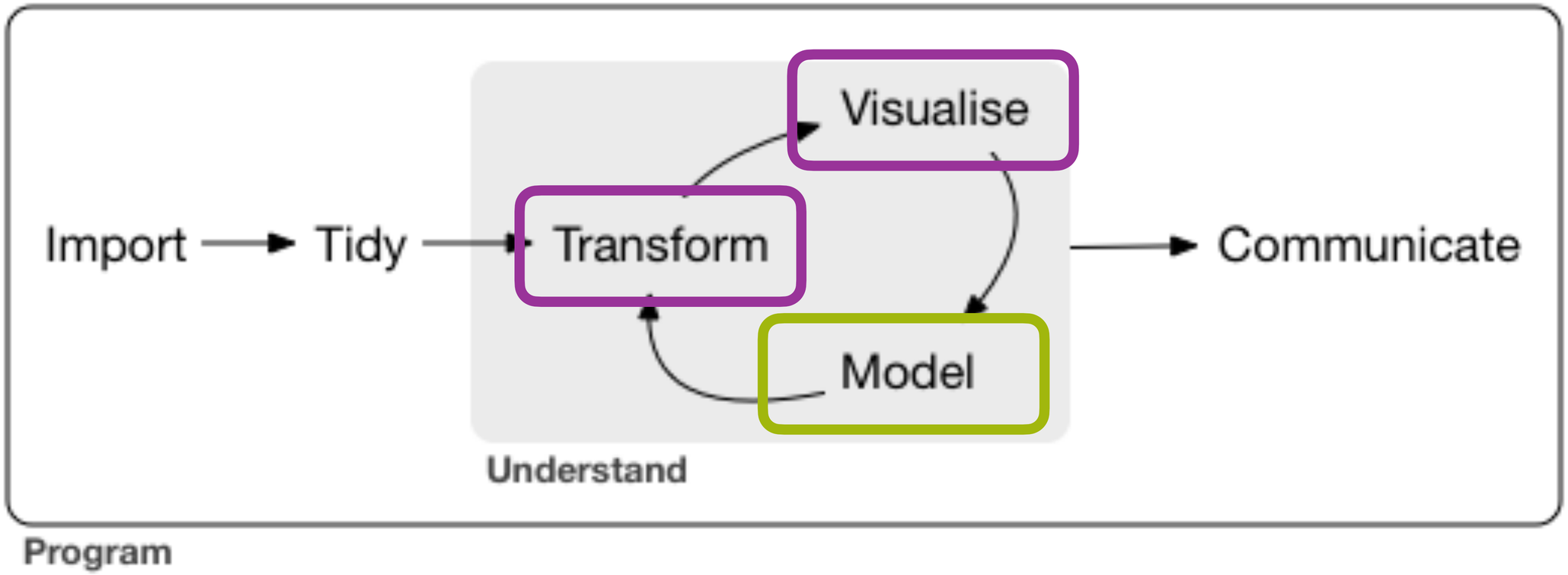
The course

Second semester undergraduate statistics course (~ 90 students)

Multiple linear regression, logistic regression, ANOVA

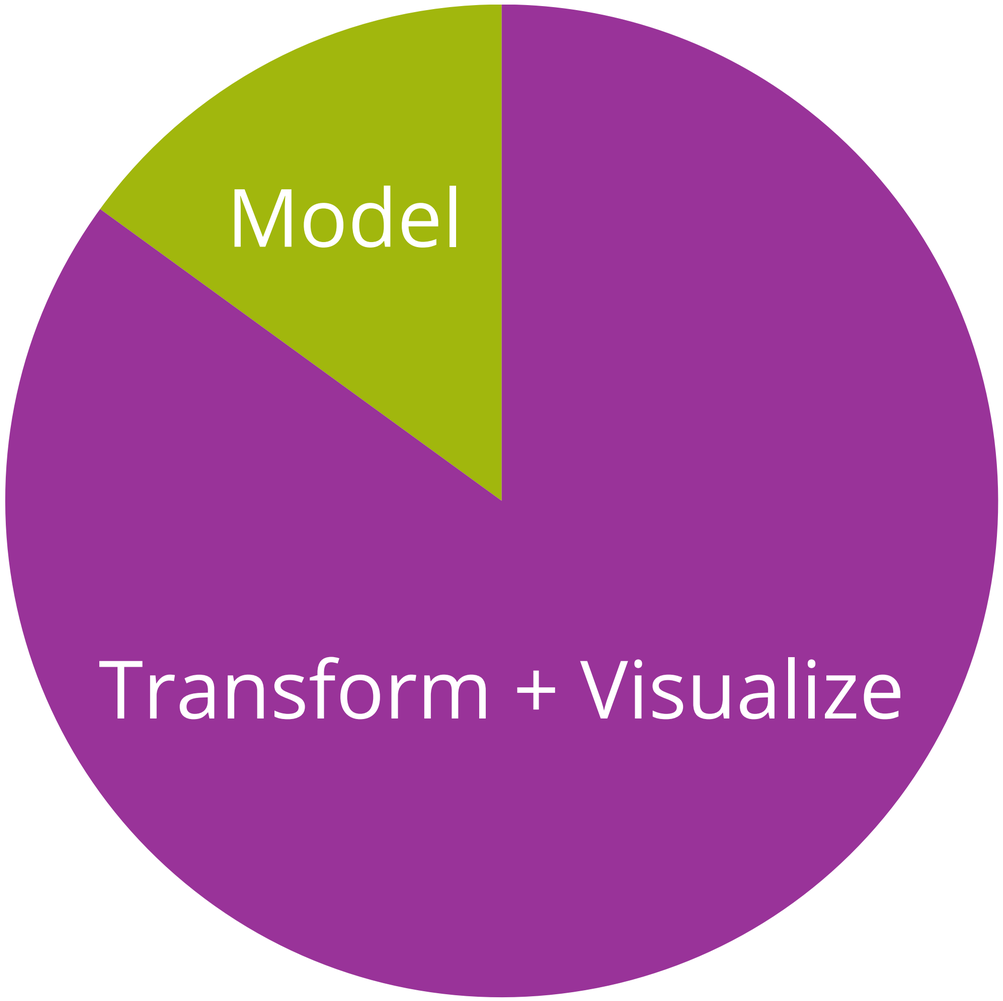
Computing using R and GitHub

Data science life cycle

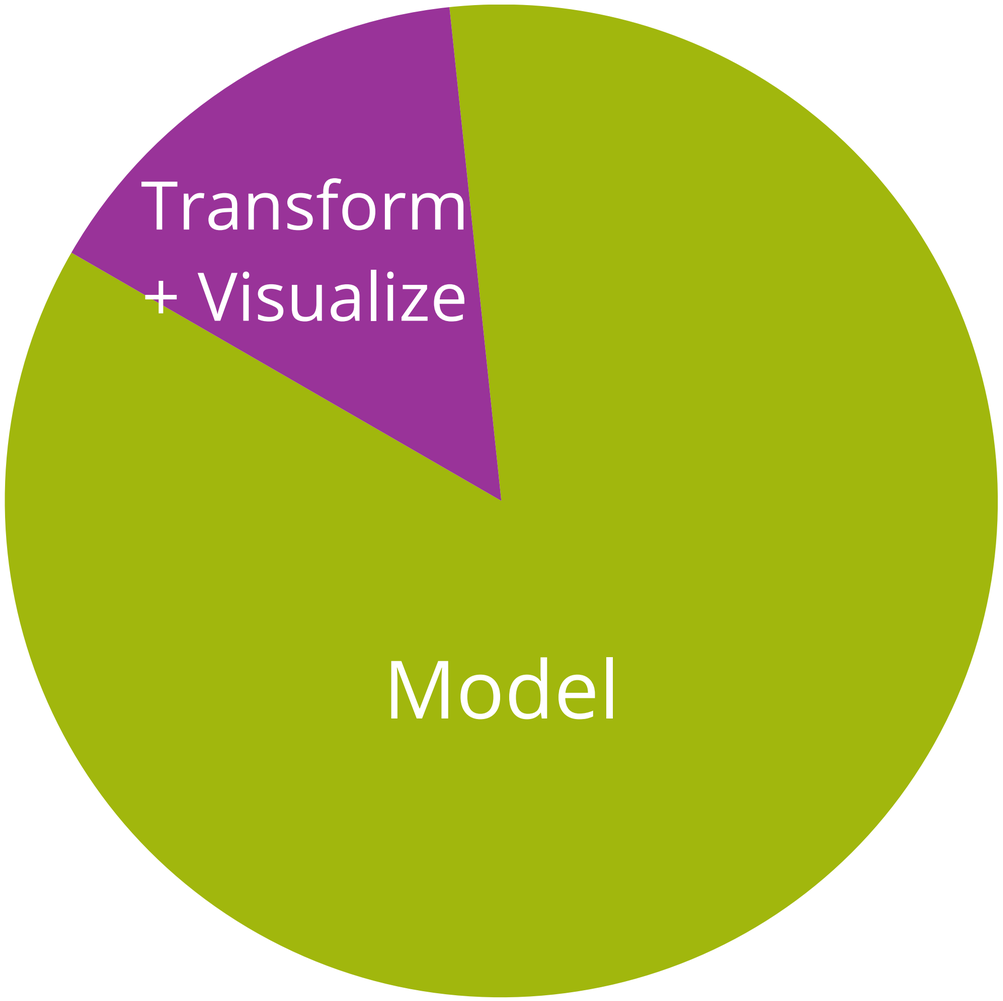


Data Science Life Cycle from R for Data Science

In practice



In class



Dealing with missing data

✓Identify different types of missingness

✓Use simple imputation methods to handle item nonresponse

✓***Think critically about unit nonresponse***

- ***Who is missing***
- ***Impact on analysis and conclusions***

Census 2020 (there's still time to fill it out!)

2020census.gov

- Headcount of every person living in the United States
- Occurs every 10 years
- Data is used to allocate…

✓ seats in U.S. House of Representatives

✓ federal funding for public programs

Why use census data in class?

*"Using **real data in context** is crucial in teaching and learning statistics, both to give students experience with analyzing genuine data and to illustrate the usefulness and fascination of our discipline."*

2016 Guidelines for Assessment and Instruction in Statistics Education (GAISE)

Problem: Undercount!

APR. 23, 2019, AT 9:38 AM

How The Citizen
Could Break The

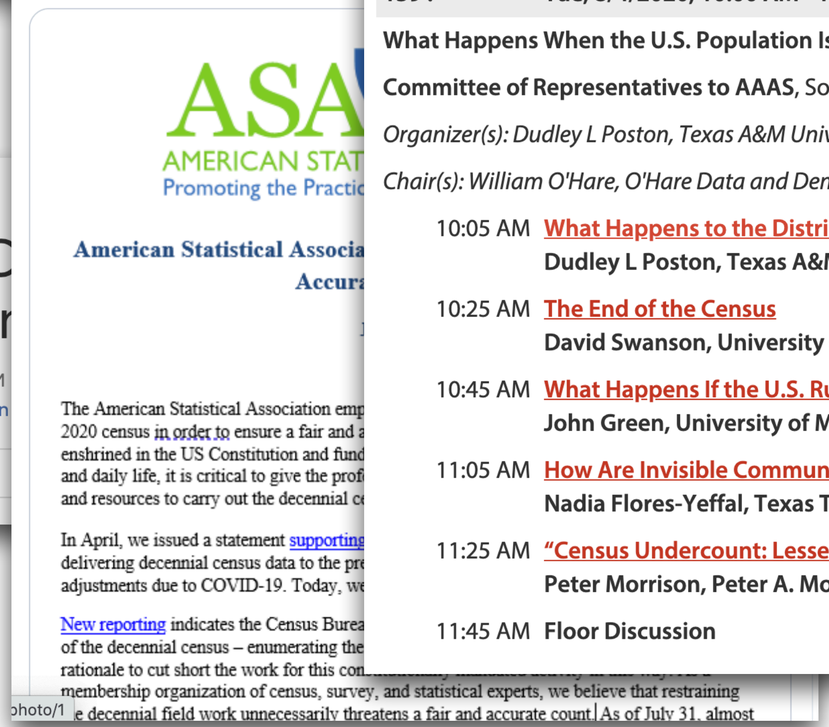
By Amelia Thomson-DeVeaux

Filed under Supreme Court

Get the data on GitHub

ASA Science Policy
@ASA_SciPol

American Statistical Association Board issues Statement on Ensuring Fair and Accurate 2020 Census, saying "the Census Bureau should be allowed to continue the timeline they proposed this spring" for carrying out non-response follow up. @Am
#DataIntegrity


ASA
AMERICAN STAT
Promoting the Practic

American Statistical Associa
Accura

The American Statistical Association emp
2020 census in order to ensure a fair and a
enshrined in the US Constitution and fund
and daily life, it is critical to give the profi
and resources to carry out the decennial c

In April, we issued a statement supporting
delivering decennial census data to the pre
adjustments due to COVID-19. Today, we

New reporting indicates the Census Burea
of the decennial census – enumerating the
rationale to cut short the work for this con
membership organization of census, survey, and statistical experts, we believe that restraining
e decennial field work unnecessarily threatens a fair and accurate count As of July 31, almost

chart shows how badly the census could
undercount people of color

hout the citizenship question, things don't look great.

NATIONAL

Census C
By A Mor

August 3, 2020 · 9:07 PM
Heard on Morning Edition


HANSI LO WANG

+ Add to My Program

Tweet

159 ! Tue, 8/4/2020, 10:00 AM - 11:50 AM Virtual

What Happens When the U.S. Population Is Undercounted in the Decennial Census? — Topic Contributed Papers

Committee of Representatives to AAAS, Social Statistics Section, Government Statistics Section

Organizer(s): Dudley L Poston, Texas A&M University

Chair(s): William O'Hare, O'Hare Data and Demographic Services LLC

10:05 AM **What Happens to the Distribution of Seats in the U.S. House of Representatives with a Census Undercount?**
Dudley L Poston, Texas A&M University

10:25 AM **The End of the Census**
David Swanson, University of California, Riverside

10:45 AM **What Happens If the U.S. Rural Population Is Undercounted?: Challenges and Community-Level Responses**
John Green, University of Mississippi Center for Population Studies

11:05 AM **How Are Invisible Communities of Immigrants in the United States Counted? What Happens If They're Undercounted?**
Nadia Flores-Yeffal, Texas Tech University

11:25 AM **"Census Undercount: Lessening a Community's Financial Loss"**
Peter Morrison, Peter A. Morrison & Associates, Inc.

11:45 AM **Floor Discussion**

Disclaimer: Spring 2020 didn’t go as planned

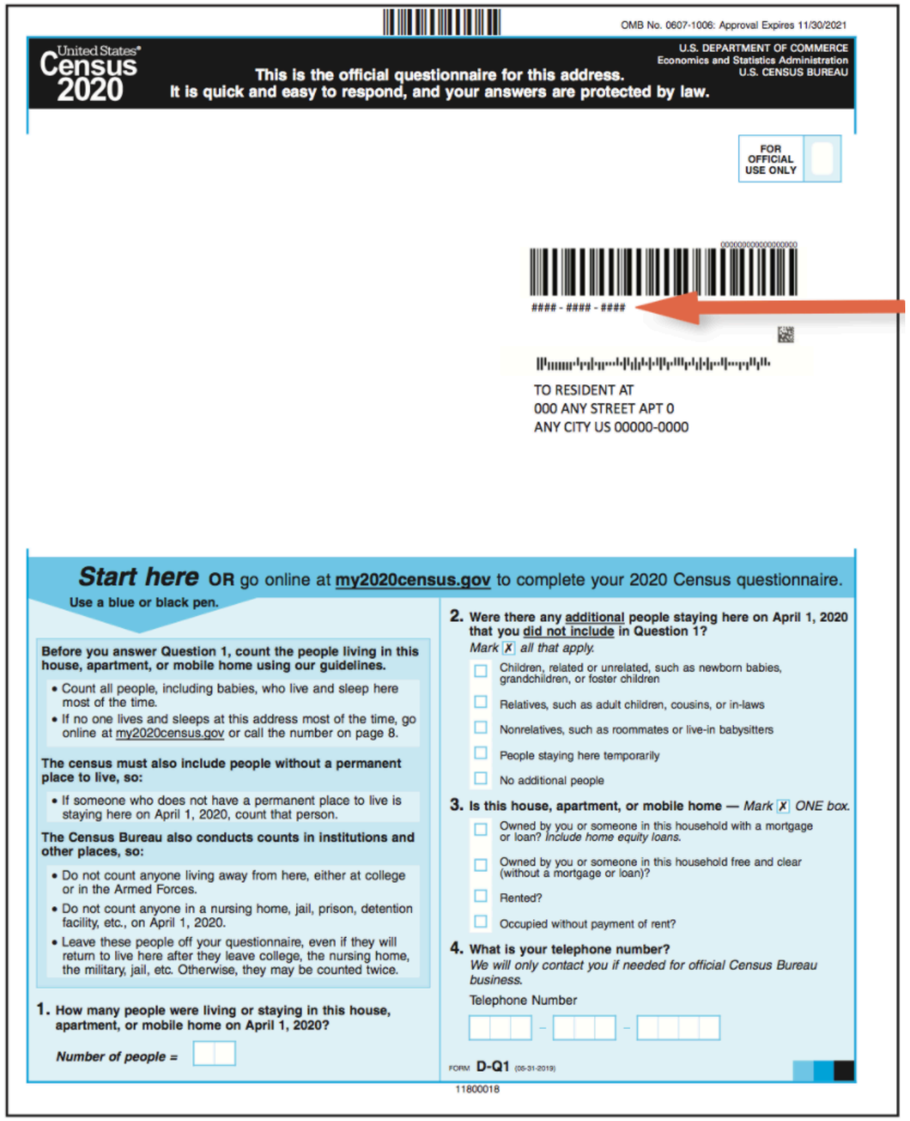
Larry the Cat Retweeted

Larry the Cat
@Number10cat

My plans: 2020:

5:16 PM · May 19, 2020 · Twitter for iPhone

Collecting data for the census



OMB No. 0607-1006: Approval Expires 11/30/2021

United States® Census 2020

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

This is the official questionnaire for this address.
It is quick and easy to respond, and your answers are protected by law.

FOR OFFICIAL USE ONLY

- #### -

TO RESIDENT AT
000 ANY STREET APT 0
ANY CITY US 00000-0000

Start here **OR** go online at **my2020census.gov** to complete your 2020 Census questionnaire.

Use a blue or black pen.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.
- If no one lives and sleeps at this address most of the time, go online at my2020census.gov or call the number on page 8.

The census must also include people without a permanent place to live, so:

- If someone who does not have a permanent place to live is staying here on April 1, 2020, count that person.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away from here, either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2020.
- Leave these people off your questionnaire, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2020?

Number of people =

2. Were there any additional people staying here on April 1, 2020 that you did not include in Question 1?
Mark [X] all that apply.

- ☐ Children, related or unrelated, such as newborn babies, grandchildren, or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in babysitters
- ☐ People staying here temporarily
- ☐ No additional people

3. Is this house, apartment, or mobile home — *Mark [X] ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

4. What is your telephone number?
We will only contact you if needed for official Census Bureau business.

Telephone Number

FORM **D-Q1** (05-31-2019)

11800018

- Invitation sent to households in March
- Respond by mail, phone, or online
- Door knocking effort in August to interview those who haven't responded

Image source: https://my2020census.gov/static/letter.html?lang=en

Data collection discussion

- What populations are most likely to be *hard to count* and therefore underrepresented in the U.S. Census? Why?

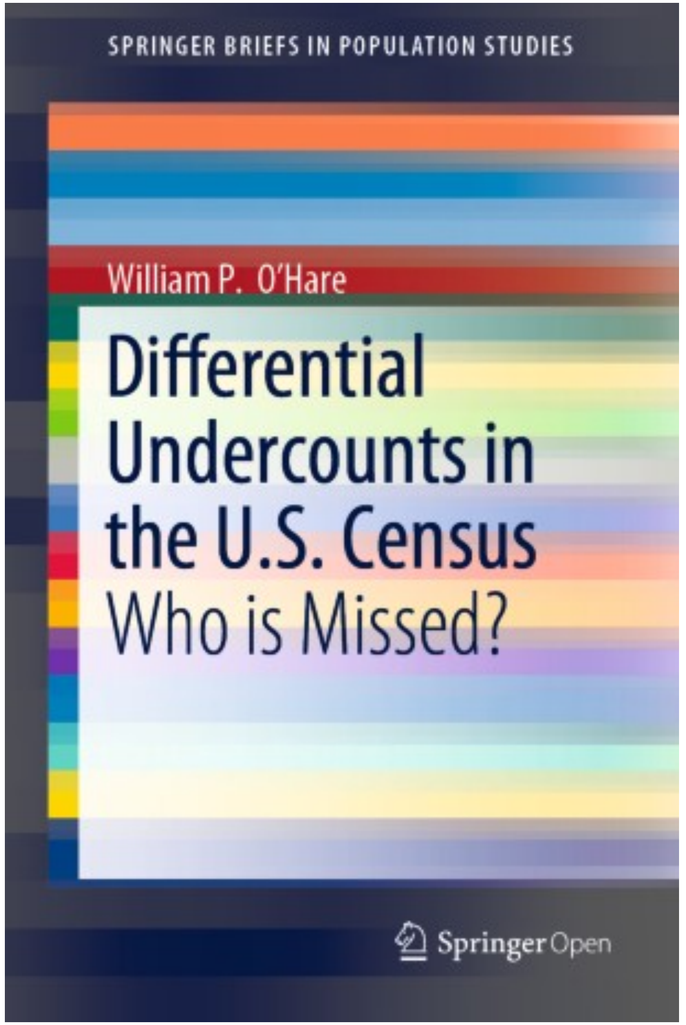
- What are the potential impacts of having underrepresented subgroups in the data when using census data to…
  - allocate funds or make other societal decisions?
  - conduct statistical analysis?

Measuring undercount

Demographic Analysis (DA): Compare census counts to a independent population estimate

$$Pop_{0-74} = Births - Deaths + NetMig$$

Duel System Estimates (DSE): Compare census counts to results from a Post-Enumeration Survey (PES)



Differential Undercounts in the U.S. Census

Activity set up

Suppose you're part of an organization whose goal is to reach people in hard-to-reach populations and encourage them to fill out the Census.

The organization has limited resources, so you will use data to help determine how to prioritize your time and effort.

Fit a regression model that you can use to better understand the characteristics of hard-to-reach populations.

`tidycensus` R package

```
52  census2010 <- get_decennial(geography = "state",
53                              variables = c("P003001", "P003002")
54                              year = 2010,
55                              output = "wide",
56                              cache = TRUE)
```

```
92  avg_hh_size <- get_acs(geography = "state",
93                         year = 2010,
94                         table = "B25010",
95                         output = "wide",
96                         moe_level = 95,
97                         survey = "acs5",
98                         cache  = TRUE)
```

walker-data.com/tidycensus

Response variable

What we think the total population is in 2010

$$Pop_{2010} = Pop_{2009} + Births_{2010} - Deaths_{2010} + NetMigration_{2010}$$

What is the response variable for your model?

Use Pop_{2010} and the population from the 2010 Census to define a response variable

Explanatory variables

censusreporter.org

Topics

Learn more about the concepts and tables covered by the Census and American Community Survey. We'll be adding more of these pages in the next few months, so let us know if there are topics you'd like to see us explain.

- Getting Started
- About the Census
- Age and Sex
- Children
- Commute
- Employment
- Families
- Geography
- Health Insurance
- Housing
- Income
- Migration
- Poverty
- Public Assistance
- Race and Hispanic Origin
- Same-Sex Couples
- Seniors
- Table Codes
- Veterans and Military

💡 ***Provide data to students for short-term assignments.***

Model + conclusion

```
model <- lm(pct_diff ~ medinc + pct_public_asst + pct_0_4,
            data = state_char)
tidy(model, conf.int = TRUE) %>%
  kable(format = "html", digits = 3)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-0.002	0.027	-0.093	0.926	-0.056	0.051
pct_white	-0.018	0.017	-1.058	0.296	-0.052	0.016
pct_public_asst	-0.615	0.169	-3.647	0.001	-0.954	-0.276
pct_0_4	0.975	0.315	3.095	0.003	0.342	1.608

Based on your model, describe how you will prioritize your efforts to encourage people to respond to the U.S. Census.

Reflection questions

- What is one observation from your model about undercount in the Census? How does it compare to the results from using DA and DSE to measure undercount?
- Briefly explain why it is important to consider which subgroups are underrepresented in data used to build statistical models.
- What is one remaining question you have about data collection for the U.S. Census?

Thank You!

maria.tackett@duke.edu

@MT_statistics

bit.ly/jsm2020-teach