STAT 113
Introduction to Statistics

Colin Reimer Dawson

Oberlin College

February 18th, 2022

Outline

Small Group Warmup: Statistics in the News

Themes of This Course

Overview/Syllabus

Syllabus Nuts and Bolts

Statistics in the News

- Go to the URL `http://colindawson.net/stat113/misc/misleading-graphs.pdf`
- There are three examples of statistical results that are all misleading in some way
- (5 minutes) With your small group, identify the surface "takehome message" in each one as well as potentially missing context

Statistics in the News, Part 2

- Go to the URL `http://colindawson.net/stat113/misc/misleading-graphs-part2.pdf`
- Each of the three examples is depicted in a different way with some additional context
- (5 minutes) With your small group, identify how the "takehome" message is different, and what change led to such a difference
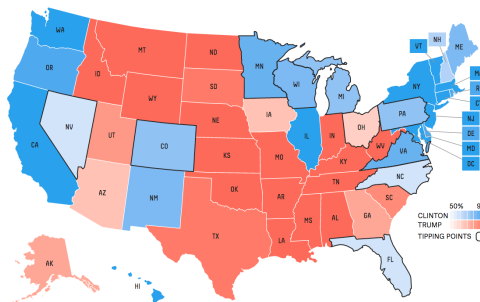
The Importance of the Baseline

- A theme of this course: "If there were nothing interesting going on, what data would we expect to see?"
- Interpret actual results relative to this "baseline"

Election Prediction: A Failure of Statistics?

Sampling Error

Noisy Polls Are to Be Expected

If Hillary Clinton were up by a modest margin, there would be plenty of polls showing a very close race — or even a Trump lead.

A simulation of 100 surveys, if Mrs. Clinton were really up 4 points nationally.
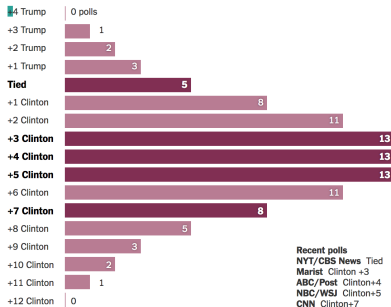
+4 Trump	0 polls
+3 Trump	1
+2 Trump	2
+1 Trump	3
Tied	5
+1 Clinton	8
+2 Clinton	11
+3 Clinton	13
+4 Clinton	13
+5 Clinton	13
+6 Clinton	11
+7 Clinton	8
+8 Clinton	5
+9 Clinton	3
+10 Clinton	2
+11 Clinton	1
+12 Clinton	0

Recent polls
NYT/CBS News Tied
Marist Clinton +3
ABC/Post Clinton+4
NBC/WSJ Clinton+5
CNN Clinton+7

Figure: Polls come from samples, not the population, and are subject to sampling error. Source: The New York Times "The Upshot" 7/18/16.

Sampling and Non-response Bias?

STATE	WHITE NON-COLLEGE SHARE	TRUMP MARGIN OF VICTORY		
		ADJ. POLLING AVERAGE	ACTUAL	ACTUAL VS. POLLS
West Virginia	65.7%	+27.9	+42.2	+14.3
Kentucky	62.2	+18.4	+29.8	+11.4
Iowa	62.0	+3.4	+9.4	+6.0
Maine	61.6	-6.9	-2.7	+4.2
Idaho	60.4	+19.7	+31.8	+12.1
North Dakota	59.6	+24.7	+35.7	+11.0
Wisconsin	57.2	-5.4	+0.7	+6.1
Montana	56.7	+16.4	+20.5	+4.1
New Hampshire	56.5	-3.5	-0.4	+3.1
Wyoming	55.0	+36.3	+46.3	+10.0

Virginia	36.7	-5.4	-5.3	+0.1
Georgia	34.2	+4.0	+5.2	+1.2
New Jersey	32.9	-11.2	-14.1	-2.9
Texas	31.4	+8.5	+9.1	+0.6
New York	29.8	-18.7	-21.2	-2.5
Maryland	29.2	-26.3	-26.6	-0.3
New Mexico	27.5	-5.3	-8.2	-2.9
California	26.4	-23.0	-30.0	-7.0
Hawaii	15.2	-20.8	-32.2	-11.4
D.C.	2.2	-69.3	-86.8	-17.5

Adjusted polling average is from the FiveThirtyEight polls-only forecast as of Nov. 8, 2016. Actual results are accurate as of Nov. 30.

Figure: Is the sample *representative* of the population? Polling errors strongly correlated with state share of whites w/o college degrees. Source: fivethirtyeight.com
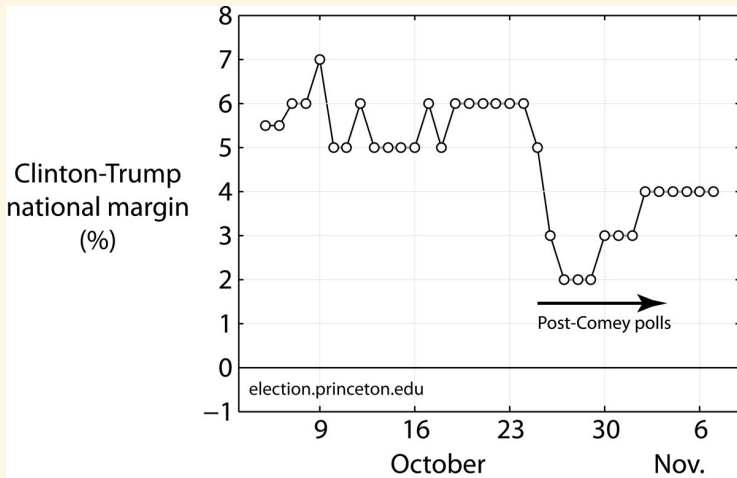
Moving Targets

Figure: Is our data about what we want it to be about? Source: Princeton Election Consortium http://election.princeton.edu

Statistics: An Alternative to "Alternative Facts"

"You're saying it's a falsehood. ... Sean Spicer, our press secretary, gave alternative facts to that." – Kellyanne Conway, White House counselor, on *Meet the Press*, 1/22/2017

"You know, the very powerful and the very stupid have one thing in common," the Doctor said. "They don't alter their views to fit the facts. They alter the facts to fit their views." – Doctor Who, *The Face of Evil, Part 4, aired 1977*

"In God we trust, all others must bring data." – attributed to statistician W. Edwards Deming, 1900-1993.

Statistics: A Highly Employable Skill

"[A] new analysis of help-wanted postings for entry-level jobs suggests that [liberal arts] graduates can improve their job prospects markedly by acquiring a small level of proficiency in one of eight specific skill sets, such as social media or data analysis. For example, the analysis found an additional 137,000 entry-level jobs for liberal-arts graduates who had data-analysis or management skills.

It also found that such data-analysis jobs paid an average of $12,700 above the average salary for jobs traditionally open to liberal-arts graduates without such skills." – Chronicle of Higher Ed., 6/9/16

Defensive Statistical Literacy

> "There are three kinds of lies: lies, damned lies, and statistics." – Unknown (questionably attributed by Mark Twain to Benjamin D'Israeli)

- An overarching course goal: become a literate *consumer* of statistics

Course Themes

- Elements of Data, Study Design and Confounding ($\sim$2 weeks)
- Description and Visualization of Data ($\sim$3 weeks)
- Generalizing from the Data to the Phenomenon ($\sim$4 weeks)
- Specific Analysis Techniques ($\sim$4 weeks)
- Application and Communication (throughout)

Meeting Schedule

- MWF classes in person starting Monday
- Thursdays, lab day on Zoom

Key Locations on the Web

- Course Website: `colindawson.net/stat113`
- Slack Workspace: `stat113s2022.slack.com` (need to join via the invitation link on the course website)
- RStudio Server (will be introduced in lab): `rstudio.oberlin.edu` (fill out Pre-Course survey at the website by Wednesday so I can set up an account for you)

Textbook

- A free and open source book: *Introduction to Modern Statistics* by Mine Cetinkaya-Rundel and Johanna Hardin
- PDF, HTML (free) and paperback ($20) versions available from `https://openintro.org/book/ims/`
- Some homework problems will be drawn from the book; others posted as pdfs

Graded Components

- Homework (due Monday nights 11:59 PM, starting 2/28)
- Quizzes (due Friday afternoons by 5 PM)
- Two Takehome Exams (due Fridays 04/01 and 05/06)
- Small Group Project (draft due 5/13, final version 5/31 during finals)
- Optional takehome final (also due 5/31)

"Mastery"-Based Grading System

- Non-traditional grading system, designed to emphasize
  - Progress, rather than immediate results
  - Understanding over memorization
  - Embracing mistakes as an opportunity to improve
- Tracks competence, not points
- High standards, but ample grade replacement opportunities (try, fail, and try again!)

To-Dos (See Webpage for Details)

- Join Slack Workspace
- Fill Out Background survey
- Read Chs. 1 and 2.1
- Homework 1 due Monday 02/28