Classroom activities in applied regression and causal inference, accelerated one-semester course[1]

Andrew Gelman and Aki Vehtari

[1]For details, see *Active Statistics*, by Andrew Gelman and Aki Vehtari, Cambridge University Press (2023), http://www.stat.columbia.edu/∼gelman/active-statistics/.

Class 1a: Introduction to quantitative social science

Story

Wikipedia experiment

ABBA

A/B testing statistics

Label	Number of successes	Number of trials	
dsn_cnt	4861	954630	Remove
dsn_squareCorners	4695	1082180	Remove

Interval confidence level:

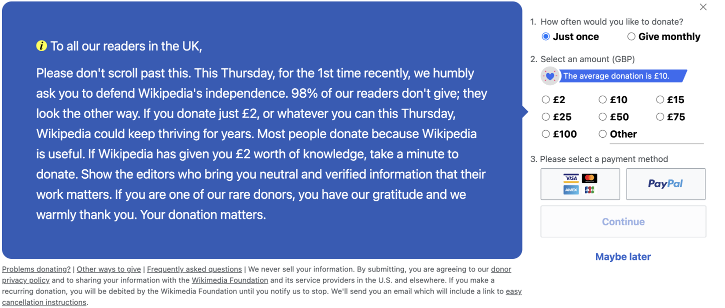
0.95

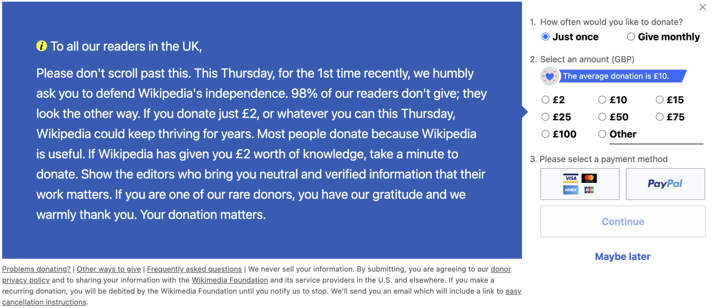
Use multiple testing correction: ☑

Compute Add another group

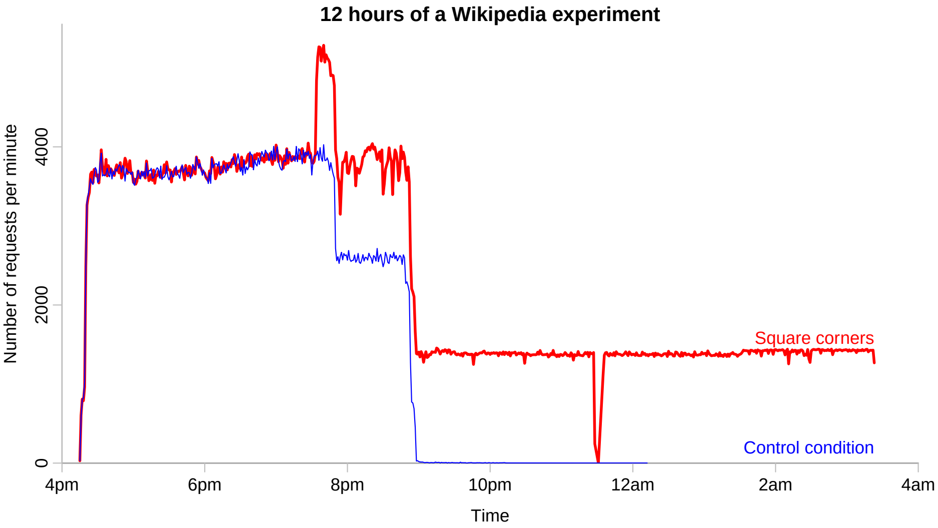
	Successes	Total	Success Rate	p-value	Improvement
dsn_cnt	4,861	954,630	0.5% – 0.52% (0.51%)	—	—
dsn_squareCorners	4,695	1,082,180	0.42% – 0.45% (0.43%)	< 0.0001	-19% – -11% (-15%)

Wikipedia experiment





Wikipedia experiment



Activity

Designing a social science study

1. Treatments
2. Population
3. Sample
4. Treatment assignment
5. Pre-test measurement
6. Outcome measurement

Introduction to the course

Topics

- Goals of the course
- Components of the course
- Structure of each class period
- Students' responsibilities
- Roles of mathematics, computing, and applications

Computer demonstration

Drill

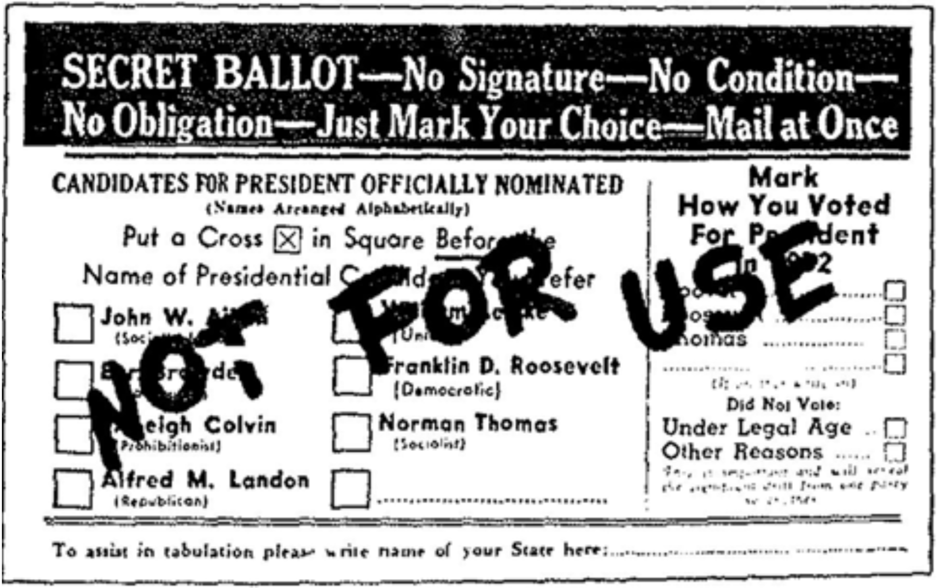
Generalizing

1. From sample to population
2. From treatment to control group
3. From measurement to underlying construct

Class 1b: Overview of applied regression

Story

Literary Digest poll



SECRET BALLOT—No Signature—No Condition—
No Obligation—Just Mark Your Choice—Mail at Once

CANDIDATES FOR PRESIDENT OFFICIALLY NOMINATED
(Names Arranged Alphabetically)

Put a Cross ☒ in Square Befor...
Name of Presidential C...

☐ John W. ...
(Soci...)

☐ ...

☐ ...eigh Colvin
(Prohibitionist)

☐ Alfred M. Landon
(Republican)

☐ ...
(Uni...)

☐ ...ranklin D. Roosevelt
(Democratic)

☐ Norman Thomas
(Socialist)

☐ ....

Mark
How You Voted
For P...dent
...

...omas ☐

Did Not Vote:
Under Legal Age ☐
Other Reasons ☐

To assist in tabulation please write name of your State here: ....

NOT FOR USE

Activity

Candy weighing

1. Pull 5 candies out of the bag
2. Weigh the candies
3. Write down the weight
4. Put the candies back in the bag!!
5. Pass the scale and bag to your neighbors
6. Silently multiply the weight of the 5 candies by 20

Discuss reading and homework

Computer demonstration

Drill

Describing a fitted regression in words

Summarize the slope for the underlined variable, first wrongly and then correctly.

Discussion problem

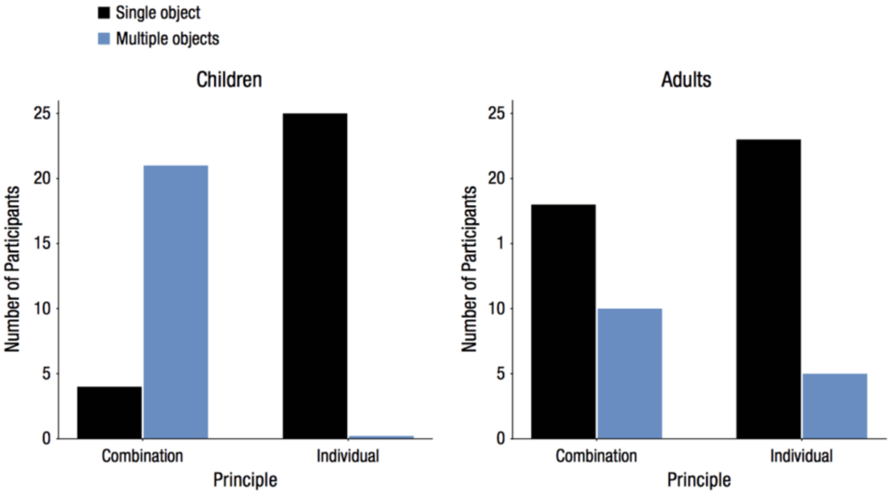
Height and earnings

$$\text{earnings} = -26000 + 600 * \text{height} + 10600 * \text{male} + \text{error}$$

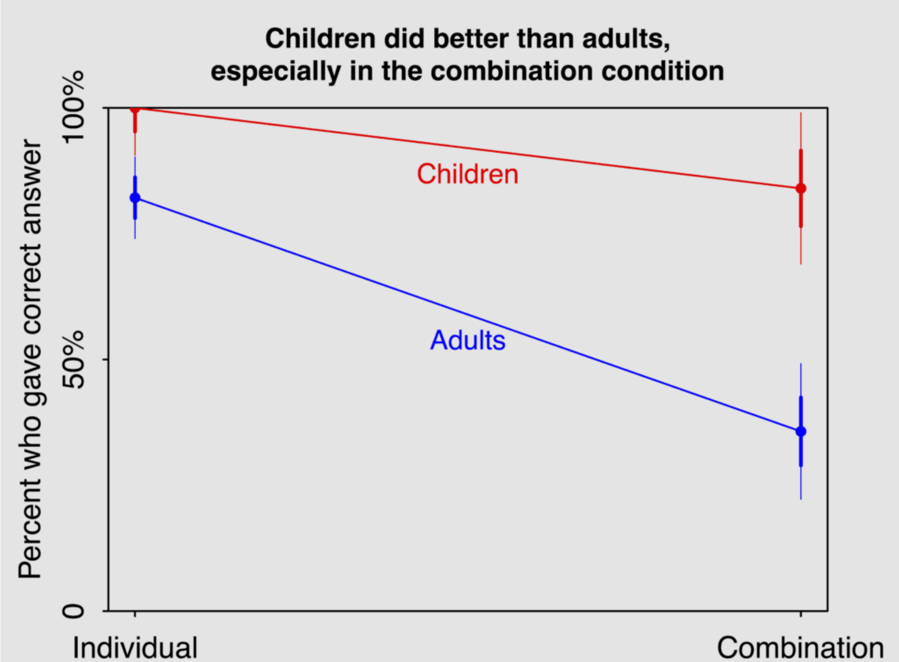
Class 2a: Data collection and visualization

Story

Using the "graphs as comparisons" idea to redraw a graph



Using the "graphs as comparisons" idea to redraw a graph



Activity

Measuring handedness

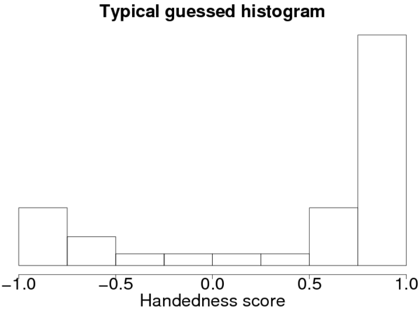
Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you use would never use the other hand for that activity. If in any case you are really indifferent, put + in both columns.
Some of the activities require both hands. In these cases the part of the task, or object, for which hand preference is wanted is indicated in parentheses.

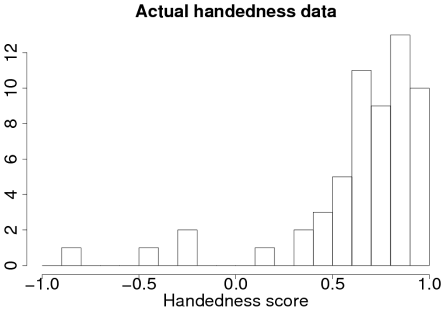
Task	Left	Right
Writing		
Drawing		
Throwing		
Scissors		
Toothbrush		
Spoon		
Total		

Right − Left: Right + Left: $\frac{\text{Right} - \text{Left}}{\text{Right} + \text{Left}}$:

Create a Left and a Right score by counting the total number of + signs in each column. Your handedness score is (Right − Left)/(Right + Left): thus, a pure right-hander will have a score of score $(12 - 0)/(12 + 0) = 1$, and a pure left-hander will score $(0 - 12)/(0 + 12) = -1$.

Measuring handedness





Measuring handedness

Please indicate which hand you use for each of the following activities by putting a + in the appropriate column, or ++ if you use would never use the other hand for that activity. If in any case you are really indifferent, put + in both columns.
Some of the activities require both hands. In these cases the part of the task, or object, for which hand preference is wanted is indicated in parentheses.

Task	Left	Right
Writing		
Drawing		
Throwing		
Scissors		
Toothbrush		
Knife (without fork)		
Spoon		
Broom (upper hand)		
Striking match (hand that holds the match)		
Opening box (hand that holds the lid)		
Total		

Right − Left: Right + Left: $\dfrac{\text{Right} - \text{Left}}{\text{Right} + \text{Left}}$:

Create a Left and a Right score by counting the total number of + signs in each column. Your handedness score is (Right − Left)/(Right + Left): thus, a pure right-hander will have a score of score $(20-0)/(20+0) = 1$, and a pure left-hander will score $(0-20)/(0+20) = -1$.

Discuss reading and homework

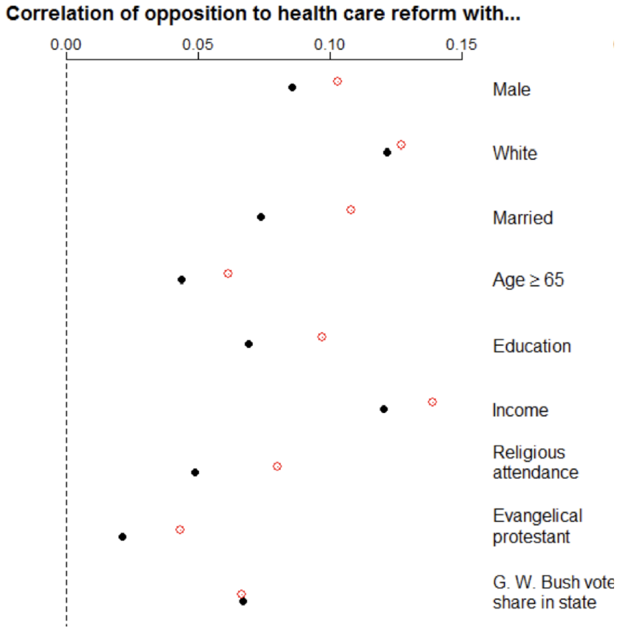
Computer demonstration

Drill

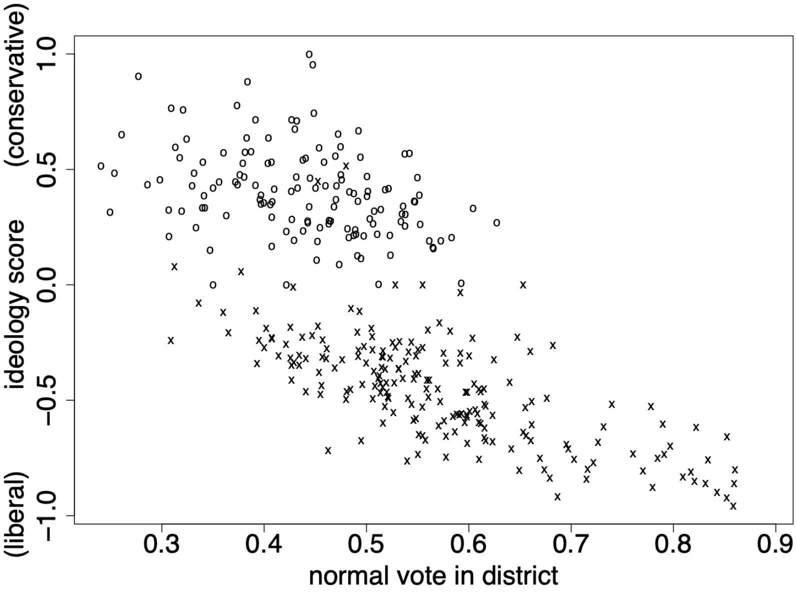
All graphs are comparisons

Identify the implicit or explicit comparison that is facilitated by each graph.

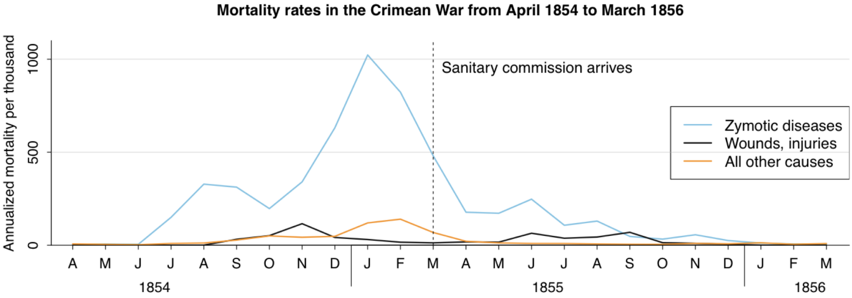
All graphs are comparisons



All graphs are comparisons



All graphs are comparisons

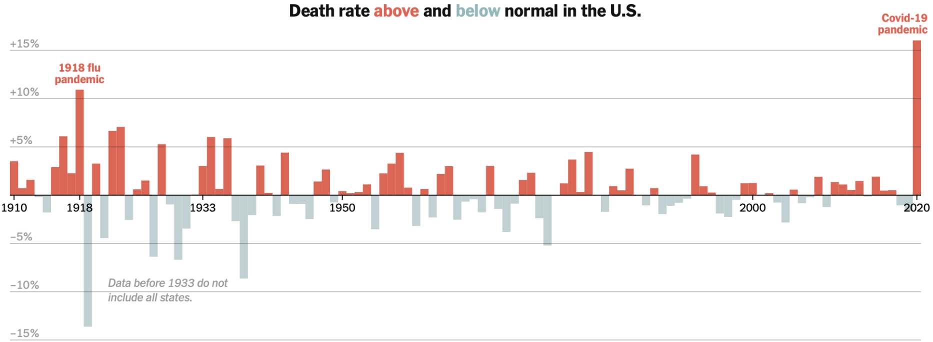


Discussion problem

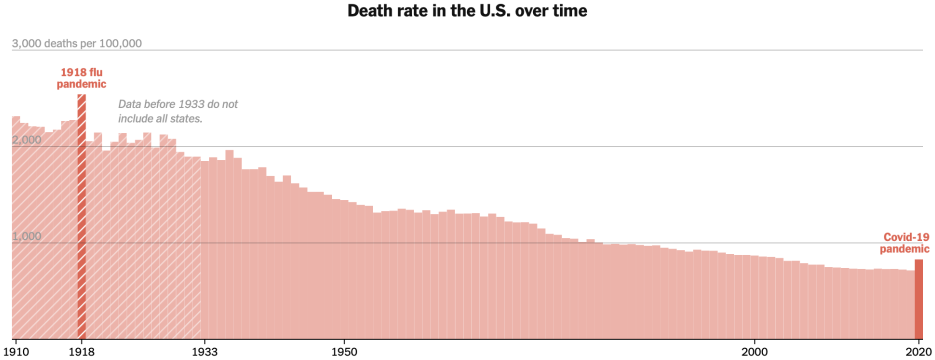
Class 2b: Basics of math and probability

Story

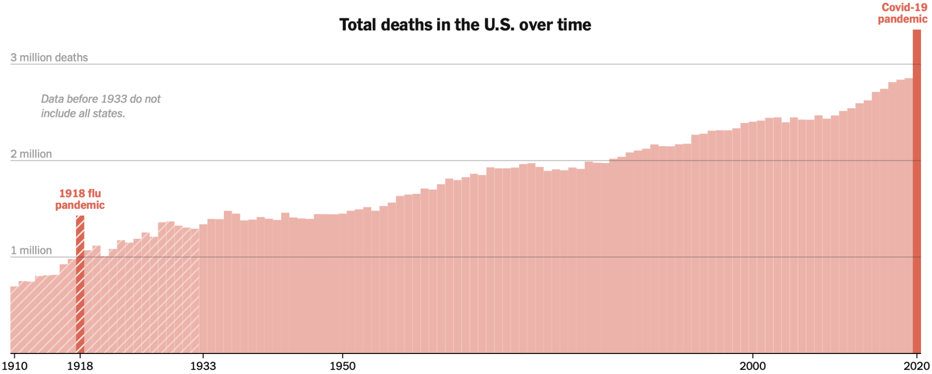
Death rate in the pandemic



Death rate in the pandemic



Death rate in the pandemic



Activity

Amoebas and population growth

- Exponential growth:

$$y = A\exp(bx)$$

$$\log y = a + bx$$

- Exponential decline:

$$y = A\exp(-bx)$$

$$\log y = a - bx$$

Discuss reading and homework

Computer demonstration

Drill

Straight lines

Give R code to graph these lines.

Discussion problem

Squares, cubes, and metabolic rates

- Power-law growth:

$$y = Ax^b$$

$$\log y = a + b \log x$$

- Power-law decline

$$y = Ax^{-b}$$

$$\log y = a - b \log x$$

Class 3a: Statistical inference

Story

They got the wrong standard error

	Response Yes	Margin of Error +/-
Was the SIU assigned to the case?	4.1%	0.5%
Were other anti-fraud professionals assigned or alerted?	2.0%	0.3%
PIP Claims		
Was there an indication in file of suspected fraud, particularly with regard to a staged accident or exaggerated medical care, medical bills, and loss or earnings?	45.7%	1.9%

Story

Claims of implausibly large effects

Labor Market Returns to Early Childhood Stimulation: a 20-year Followup to an Experimental Intervention in Jamaica

Paul Gertler, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, Sally Grantham-McGregor

We find large effects on the earnings of participants from a randomized intervention that gave psychosocial stimulation to stunted Jamaican toddlers living in poverty. The intervention consisted of one-hour weekly visits from community Jamaican health workers over a 2-year period that taught parenting skills and encouraged mothers to interact and play with their children in ways that would develop their children's cognitive and personality skills. We re-interviewed the study participants 20 years after the intervention. Stimulation increased the average earnings of participants by 42 percent. Treatment group earnings caught up to the earnings of a matched non-stunted comparison group. These findings show that psychosocial stimulation early in childhood in disadvantaged settings can have substantial effects on labor market outcomes and reduce later life inequality.

Psychol Sci. 2013 Sep 1;24(9):1837-41. doi: 10.1177/0956797613476045. Epub 2013 Jul 10.

Women are more likely to wear red or pink at peak fertility.

Beall AT, Tracy JL.

University of British Columbia.

Abstract

Although females of many species closely related to humans signal their fertile window in an observable manner, often involving red or pink coloration, no such display has been found for humans. Building on evidence that men are sexually attracted to women wearing or surrounded by red, we tested whether women show a behavioral tendency toward wearing reddish clothing when at peak fertility. Across two samples (N = 124), women at high conception risk were more than 3 times more likely to wear a red or pink shirt than were women at low conception risk, and 77% of women who wore red or pink were found to be at high, rather than low, risk. Conception risk had no effect on the prevalence of any other shirt color. Our results thus suggest that red and pink adornment in women is reliably associated with fertility and that female ovulation, long assumed to be hidden, is associated with a salient visual cue.

Psychological Science | aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE | Journal Indexing & M

Keep Your Fingers Crossed!: How Superstition Improves Performance

Lysann Damisch, Barbara Stoberock, Thomas Mussweiler

First Published May 28, 2010 | Research Article | Find in PubMed

https://doi.org/10.1177/0956797610372631

Article information | Altmetric 465

Abstract

Superstitions are typically seen as inconsequential creations of irrational minds. Nevertheless, many people rely on superstitious thoughts and practices in their daily routines in order to gain good luck. To date, little is known about the consequences and potential benefits of such superstitions. The present research closes this gap by demonstrating performance benefits of superstitions and identifying their underlying psychological mechanisms. Specifically, Experiments 1 through 4 show that activating good-luck-related superstitions via a common saying or action (e.g., "break a leg," keeping one's fingers crossed) or a lucky charm improves subsequent performance in golfing, motor dexterity,

The Fluctuating Female Vote: Politics, Religion, and the Ovulatory Cycle

Psychological Science
XX(X) 1–10
© The Author(s) 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797612466416
pss.sagepub.com
SAGE

Kristina M. Durante[1], Ashley Rae[1], and Vladas Griskevicius[2]

[1]College of Business, University of Texas, San Antonio, and [2]Carlson School of Management, University of Minnesota

Abstract

Each month, many women experience an ovulatory cycle that regulates fertility. Although research has found that this cycle influences women's mating preferences, we proposed that it might also change women's political and religious views. Building on theory suggesting that political and religious orientation are linked to reproductive goals, we tested how fertility influenced women's politics, religiosity, and voting in the 2012 U.S. presidential election. In two studies with large and diverse samples, ovulation had drastically different effects on single women and women in committed relationships. Ovulation led single women to become more liberal, less religious, and more likely to vote for Barack Obama. In contrast, ovulation led women in committed relationships to become more conservative, more religious, and more likely to vote for Mitt Romney. In addition, ovulation-induced changes in political orientation mediated women's voting behavior. Overall, the ovulatory cycle not only influences women's politics but also appears to do so differently for single women than for women in relationships.

Activity

Discuss effects in the context of a social science example

1. Consider a topic of interest
2. Consider an outcome measure and hypothesize a treatment effect
3. Construct a hypothetical experiment
4. Specify sample size
5. Hypothesize distribution of outcomes under control and treatment
6. Figure out estimate and standard error
7. Will the experiment give a reliable estimate?

Discuss reading and homework

Computer demonstration

Drill

Binomial distribution

A basketball player takes n shots. The shots are independent and she has a 30% chance of making each shot. Let y be the number of shots she makes. What are the mean and standard deviation of y? Sketch the distribution of y.

Drill

Sample size and standard errors

How large does n have to be so that your estimate has a standard error of . . . ?

Discussion problem

Approximate standard error for average "feeling thermometer" ratings

From American National Election Study: "I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person."

Class 3b: Simulation

Story

The proportion of identical twins in the population

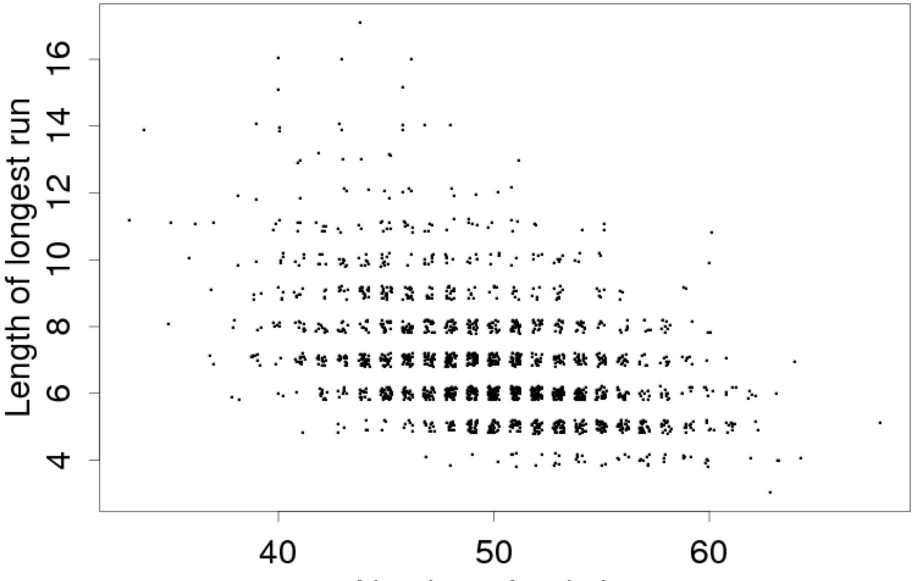
- Probability of fraternal twins: 1/125
- Probability of identical twins: 1/300
- How do we know this?

Activity

Real vs. fake coin flips

1. Instructor and two judges leave the room
2. Group A: Create a sequence of 100 real coin flips
3. Group B; Create a sequence of 100 real coin flips
4. Group C; Create a fake sequence of 0's and 1's that looks like 100 coin flips
5. Group D; Create a fake sequence of 0's and 1's that looks like 100 coin flips
6. Groups A and B write sequences on one board; groups C and D write on the other board
7. We return and figure out which is which!

Real vs. fake coin flips



Discuss reading and homework

Computer demonstration

Drill

Propagation of uncertainty

A man applies for n jobs. For each job he has a p_1 chance of getting an interview. If he is interviewed, he has a p_2 chance of getting an offer. Write an R function to simulate this process and compute the number of offers he gets. The function should take n, p_1, and p_2 as inputs and return a single number.

Discussion problem

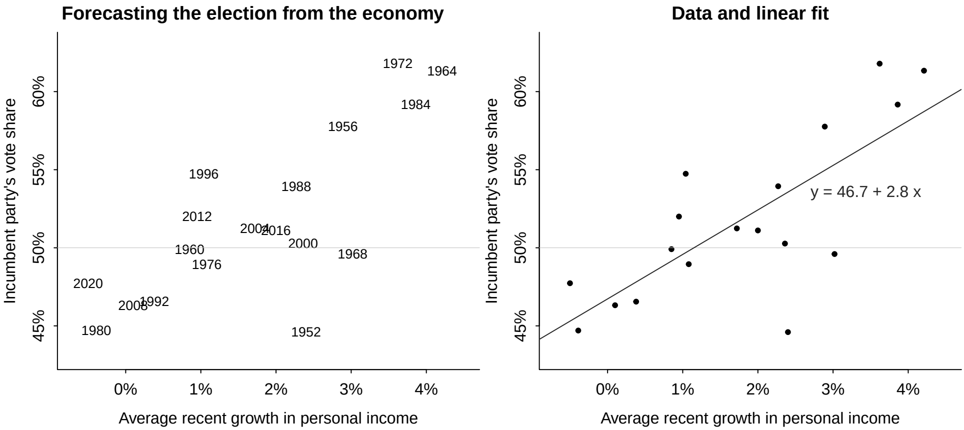
Simulate a mixed discrete/continuous distribution

Simulate the incomes of a hypothetical set of 100 people where there is a probability of zero income and a lognormal distribution otherwise.

Class 4a: Background on regression modeling

Story

Slope when predicting elections from the economy



Slope when predicting elections from the economy

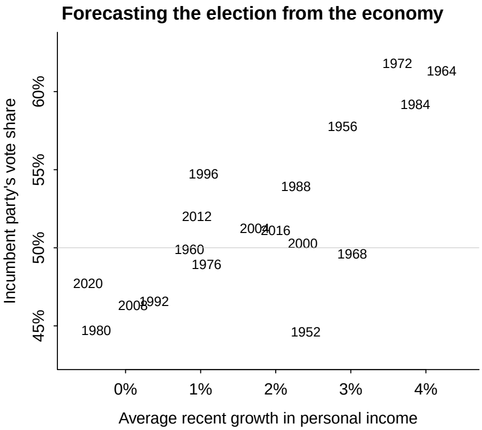
```
1948-2020:

            Median MAD_SD
(Intercept) 46.7    1.4
growth       2.8    0.6

Auxiliary parameter(s):
      Median MAD_SD
sigma 3.7    0.7
```

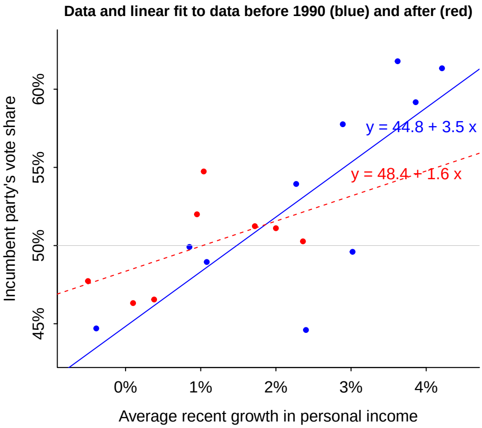
Slope when predicting elections from the economy

```
1948-1988:

             Median MAD_SD
(Intercept) 44.8    2.7
growth       3.5    1.0

Auxiliary parameter(s):
      Median MAD_SD
sigma 4.5    1.2
```

```
1992-2020:

             Median MAD_SD
(Intercept) 48.4    1.5
growth       1.6    1.1

Auxiliary parameter(s):
      Median MAD_SD
sigma 2.8    0.8
```

Slope when predicting elections from the economy





Activity

Before-after memory tests

brother	house
bat	theory
beginner	train
boy	prose
run	government
experience	art
end	song
wheel	nation
jeans	baseball
reward	flock

Before-after memory tests

Before-after memory tests

brother	house
bat	theory
beginner	train
boy	prose
run	government
experience	art
end	song
wheel	nation
jeans	baseball
reward	flock

Before-after memory tests

cloth	boundary
lizard	drain
hook	health
wheel	wax
school	car
fight	lace
string	class
wave	woman
garden	army
division	fold

Before-after memory tests

Before-after memory tests

cloth	boundary
lizard	drain
hook	health
wheel	wax
school	car
fight	lace
string	class
wave	woman
garden	army
division	fold

Discuss reading and homework

Computer demonstration

Drill

Regression to the mean

What is the student's expected score on the post-test?

Discussion problem

Understanding uniform partisan swing (considering regression to the mean)

National elections approximately follow uniform partisan swing at the national and local levels, typically with only small changes from year to year. But over a 20-year period there can be big changes. How can these patterns in the U.S. and elsewhere be understood? Is the concept of regression to the mean relevant here?

Class 4b: Linear regression with a single predictor

Story

$5^2 + 12^2 = 13^2$

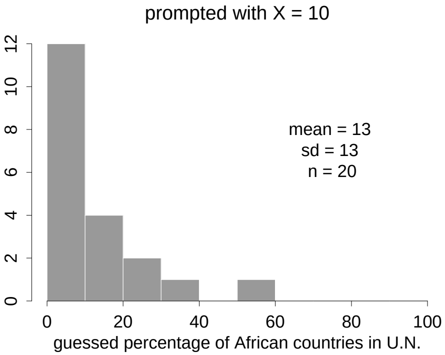
Two sources of uncertainty

Activity

African countries in the United Nations

Answer the two questions on the survey form, then enter your responses into the Google form.

African countries in the United Nations



Discuss reading and homework

Computer demonstration

Drill

Sketch a fitted regression model

Sketch the regression line and data that match both the fitted model and the residual standard deviation.

Discussion problem

How large was the sample size?

Regression predicting final from midterm exam score:

```
              Median  MAD_SD
(Intercept) 24.8    1.4
midterm     0.5     0.1

Auxiliary parameter(s):
        Median  MAD_SD
sigma   11.6    0.3
```

Approximately what was the sample size of this regression?

Class 5a: Fitting regression models

Story

Does having a girl make you more conservative or more liberal?

Study #1: "Using nationally-representative data from the [1994] General Social Survey, we find that female offspring induce more conservative political identification. We hypothesize that this results from the change in reproductive fitness strategy that daughters may evince."

Study #2: "We document evidence that having daughters leads people to be more sympathetic to left-wing parties. Giving birth to sons, by contrast, seems to make people more likely to vote for a right-wing party. Our data, which are primarily from Great Britain, are longitudinal. We also report corroborative results for a German panel."

Does having a girl make you more conservative or more liberal?

Headlines:

- "The Effect of Daughters on Partisanship and Social Attitudes Toward Women"
- "Does Having Daughters Make You More Republican?"
- "Parents With Daughters Are More Likely To Be Republicans, Says New Study"
- "Parents Of Daughters Lean Republican, Study Shows"
- "The Daughter Theory: Does Raising Girls Make Parents Conservative?"

What's missing there?

Activity

In pairs, simulate and recover regression lines

- Student #1:
  1. Create fake data from the model, $y = a + bx + \text{error}$
  2. Put `x` and `y` in a data frame called `data`
  3. Type `library("rstanarm")`
  4. Type ctrl-L to clear the R console
  5. Type `range(data$x)`
- Student #2:
  1. Take the computer
  2. Fit the regression of `y` on `x` using `stan_glm`
  3. Sketch (not on the computer) your guess of the scatterplot of x and y

Discuss reading and homework

Computer demonstration

Drill

Sample size and standard errors

```
stan_glm
 family:       gaussian [identity]
 formula:      earn ~ height
 observations: 1816
 predictors:   2
------
            Median MAD_SD
(Intercept) -85000   9000
height        1600    100

Auxiliary parameter(s):
      Median MAD_SD
sigma 22000    400
```

Averages and comparisons as regression models

For each statement, express it as a regression in R code and algebra, and give the estimated regression coefficients.

Discussion problem

Sample size and statistical significance

You run an experiment on 200 people and get an estimated treatment effect of 0.20 with standard error 0.15. So, not quite "statistically significant." What might you expect to see if you re-ran with 400 people? Would you expect statistical significance then?

Class 5b: Prediction and Bayesian inference

Story

Studying fairness of random exams

- Students randomly assigned to exams
- Average scores:
  - 65 for exam A
  - 71 for exam B
- Should we adjust the students' scores?

Activity

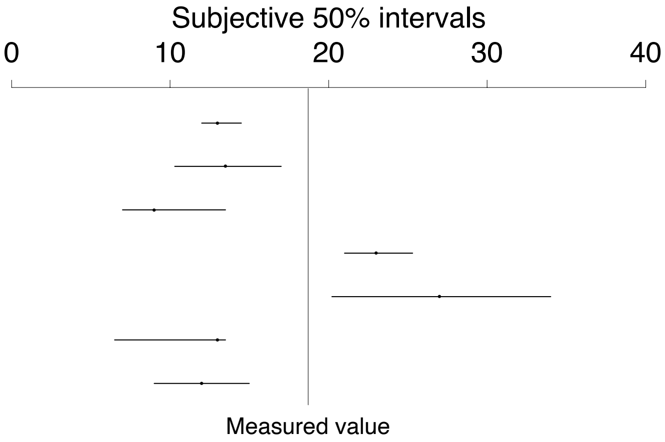
Coverage of prediction intervals

Uncertain quantity	25% lower bound	75% upper bound
% Black		
# eggs		
# airline deaths		
% girl births		
# babies born		
# abortions		
% degrees in CS		
# degrees		
# Super Bowl watchers		
$ median income		

Coverage of prediction intervals

Uncertain quantity	25% bound	75% bound	TRUTH!
% Black			12.4
# eggs			64.6 billion
# airline deaths			299
% girl births			48.8
# babies born			4.06 million
# abortions			857 000
% degrees in CS			4.4
# degrees			2.01 million
# Super Bowl watchers			101.3 million
$ median income			67 500

Coverage of prediction intervals



Discuss reading and homework

Computer demonstration

Drill

Prediction

```
stan_glm
 family:       gaussian [identity]
 formula:      earn ~ height
 observations: 1816
 predictors:   2
------
            Median MAD_SD
(Intercept) -85000   9000
height        1600    100

Auxiliary parameter(s):
      Median MAD_SD
sigma 22000    400
```

Approximately what is the predictive distribution from this regression of the earnings of a person who is . . . ?

Discussion problem

Interpreting statistically significant results given huge sample sizes

Suppose you run a regression with a huge sample size, for example a mega-poll with 100 000 respondents or an A/B test at a big company. With a large enough sample size, the standard error will be very small, and so even a very small effect can be statistically significant. How can you interpret such a result?

Class 6a: Linear regression with multiple predictors

Story

Incumbency advantage in congressional elections

Predicting Democratic vote share in U.S. House elections in 1988, given incumbency,

$$\begin{cases} +1 & \text{for districts where a Democrat was running for reelection} \\ 0 & \text{for open seats} \\ -1 & \text{for districts where a Republican was running for reelection} \end{cases}$$

```
            Median MAD_SD
(Intercept) 0.50   0.00
inc88       0.17   0.01

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.08   0.00
```

Incumbency advantage in congressional elections

Predicting Democratic vote share in U.S. House elections in 1988, given incumbency,

$$\begin{cases} +1 & \text{for districts where a Democrat was running for reelection} \\ 0 & \text{for open seats} \\ -1 & \text{for districts where a Republican was running for reelection} \end{cases}$$

and Democratic vote share in 1986:

```
            Median MAD_SD
(Intercept) 0.23   0.02
inc88       0.09   0.01
v86         0.53   0.04

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.07   0.00
```

Incumbency advantage in congressional elections

Compare the two models:

```
             Median MAD_SD
(Intercept) 0.50   0.00
inc88       0.17   0.01

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.08   0.00
```

```
             Median MAD_SD
(Intercept) 0.23   0.02
inc88       0.09   0.01
v86         0.53   0.04

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.07   0.00
```

Incumbency advantage in congressional elections

Predicting Democratic vote share in U.S. House elections in 2020, given incumbency,

$$\begin{cases} +1 & \text{for districts where a Democrat was running for reelection} \\ 0 & \text{for open seats} \\ -1 & \text{for districts where a Republican was running for reelection} \end{cases}$$

and Democratic vote share in 2018:

```
            Median MAD_SD
(Intercept) 0.03   0.01
inc2020     0.02   0.00
v2018       0.89   0.02

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.03   0.00
```

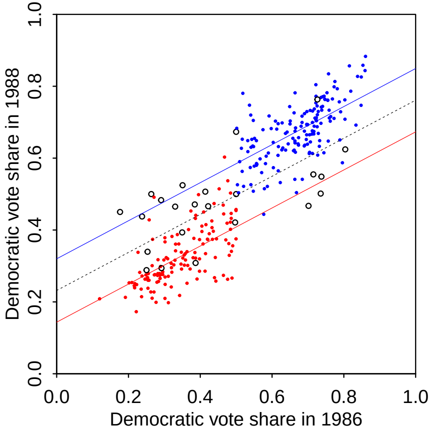
Incumbency advantage in congressional elections

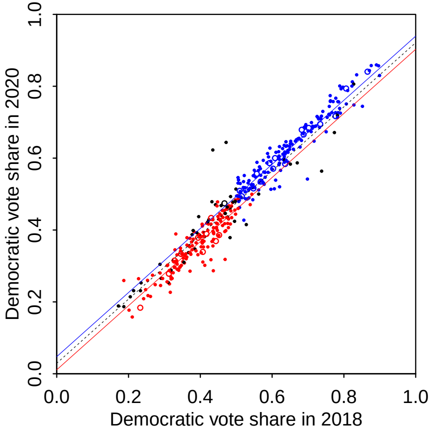
Compare the two time periods:

```
         1986-1988

             Median MAD_SD
(Intercept) 0.23   0.02
inc88       0.09   0.01
v86         0.53   0.04

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.07   0.00
```

```
         2018-2020

             Median MAD_SD
(Intercept) 0.03   0.01
inc2020     0.02   0.00
v2018       0.89   0.02

Auxiliary parameter(s):
      Median MAD_SD
sigma 0.03   0.00
```

Incumbency advantage in congressional elections





Activity

Memory quiz with pre-test, treatment, and outcome

friend	verse
cloth	curtain
metal	attempt
comparison	size
balloon	match
tiger	form
cable	maid
dress	expansion
worm	goose
mother	liquid

Memory quiz with pre-test, treatment, and outcome

For second memory quiz:

- If the last digit of your Social Security number is *odd*, you'll get *30 seconds*
- If the last digit of your Social Security number is *even*, you'll get *60 seconds*

Memory quiz with pre-test, treatment, and outcome

friend	verse
cloth	curtain
metal	attempt
comparison	size
balloon	match
tiger	form
cable	maid
dress	expansion
worm	goose
mother	liquid

Memory quiz with pre-test, treatment, and outcome

power	snail
screw	cake
curve	unit
writing	driving
sister	hair
baby	scarecrow
cry	discussion
collar	channel
trousers	sheep
brick	ocean

Memory quiz with pre-test, treatment, and outcome

- x: Score on first memory quiz
- z: Treatment ($z = 1$ if you got 60 seconds, $z = 0$ if you got 30 seconds)
- y: Score on second memory quiz

Consider two regression models predicting y from x and z:

- `y ~ x + z`
- `y ~ x + z + x*z`

Memory quiz with pre-test, treatment, and outcome

power	snail
screw	cake
curve	unit
writing	driving
sister	hair
baby	scarecrow
cry	discussion
collar	channel
trousers	sheep
brick	ocean

Discuss reading and homework

Computer demonstration

Drill

Interpret interaction coefficients

For each model, describe each coefficient in words.

Discussion problem

What is gained by including a pre-test?

Consider a randomized experiment:

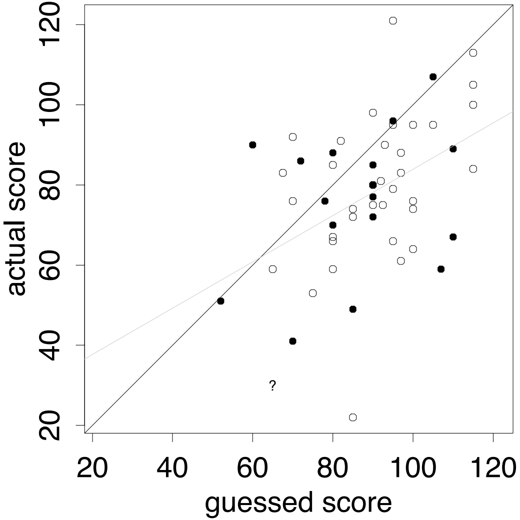
- Regression of post-test on treatment: `y ~ z`
- Regression of post-test on treatment and pre-test: `y ~ z + x`

What is gained by adjusting for pre-test?

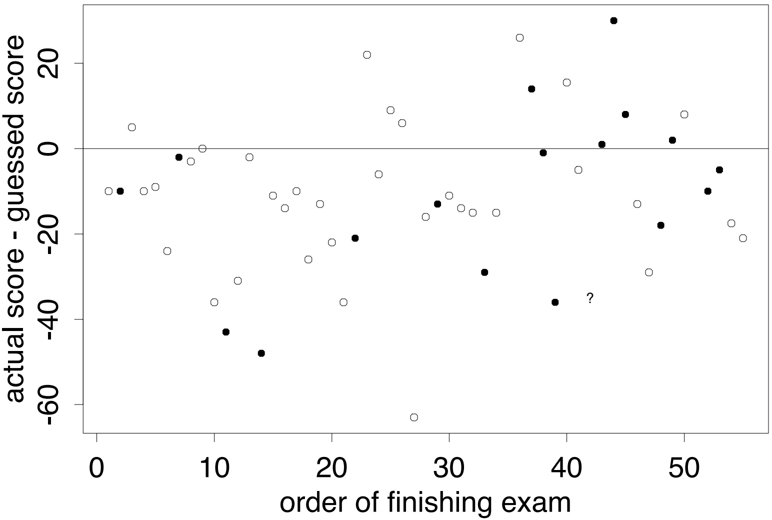
Class 6b: Assumptions, diagnostics, and evaluation

Story

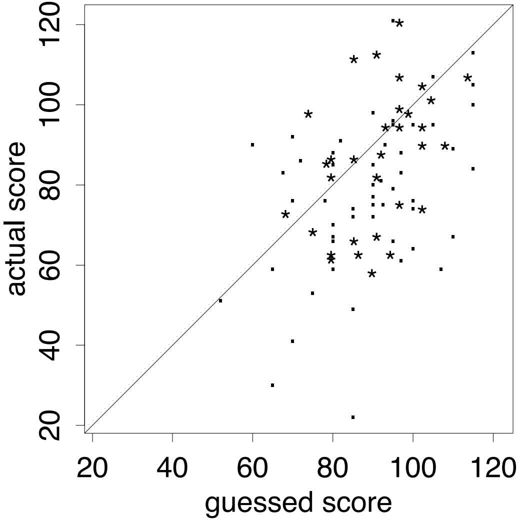
Actual vs. guessed exam scores



Actual vs. guessed exam scores



Actual vs. guessed exam scores



Activity

Assumptions of regression

1. Validity
2. Representativeness
3. Additivity
4. Linearity
5. Independence of errors
6. Equal variance of errors
7. Normality of errors

Discuss reading and homework

Computer demonstration

Drill

Assumptions of regression and and how they can fail

Consider a regression fit to a set of different countries, predicting the rate of some illegal behavior (for example, tax evasion or speeding) given country-level predictors (per-capita income, average education level, etc.). For each assumption, give an example of how it can fail.

Discussion problem

Consider the implications of regression assumptions for a real-world study

1. Validity
2. Representativeness
3. Additivity
4. Linearity
5. Independence of errors
6. Equal variance of errors
7. Normality of errors

Class 7a: Transformations and regression

Story

Price elasticity of demand

$$\log(\text{demand}) = a + b * \log(\text{price})$$

Activity

Combining predictors to create a total score

1. Construct several questions to measure an underlying construct of interest
2. Use these to create a combined score

Discuss reading and homework

Computer demonstration

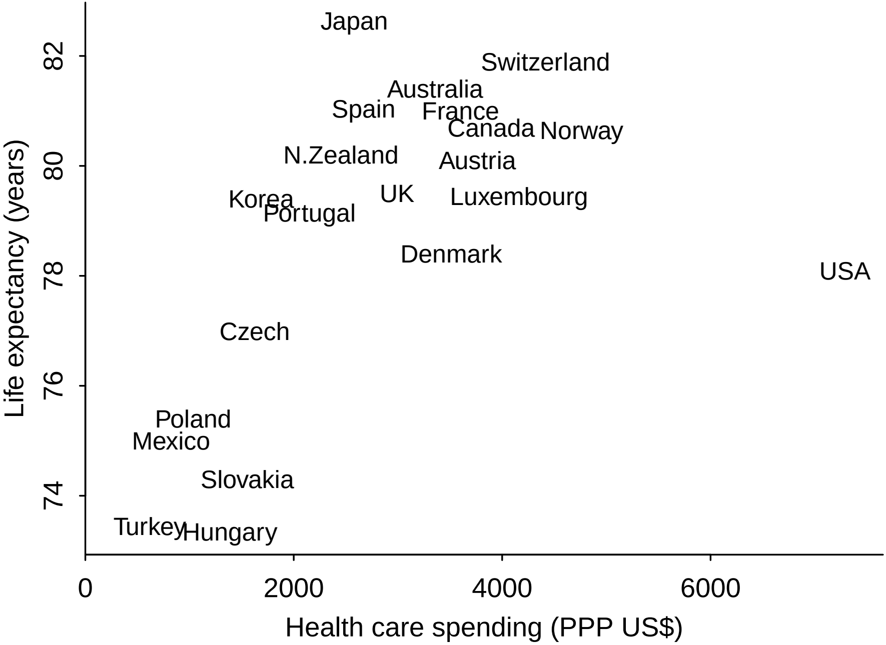
Drill

Examples of exponential growth and decline

Examples of exponential growth and decline

Discussion problem

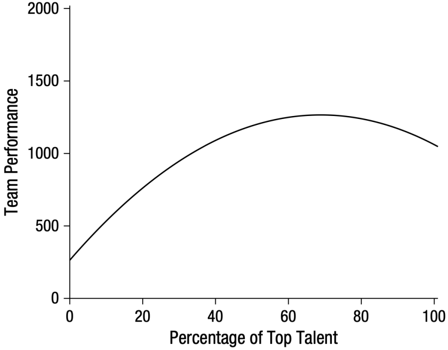
General rules of when to use the log scale



Class 7b: Review of statistics and linear regression

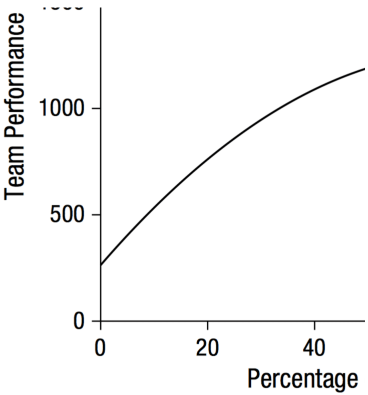
Story

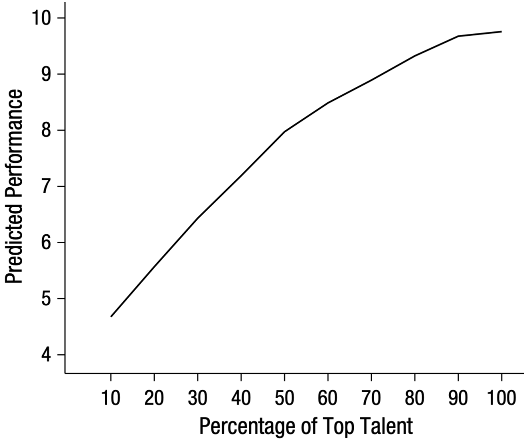
The "problem of too much talent"



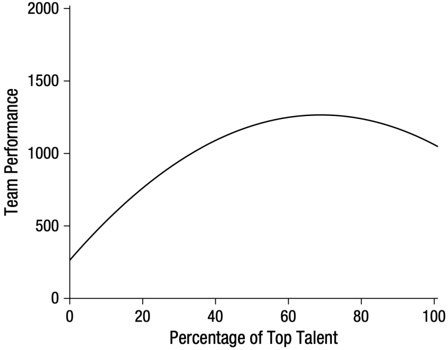
Measure	M	SD
1. Team performance (points)	393.30	320.12
2. Top-talent percentage	7%	16%
3. Roster size	18.53	6.79
4. Games played	8.90	4.65

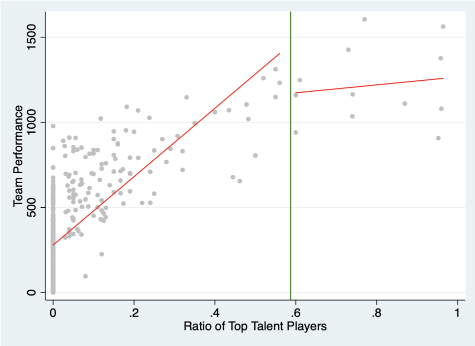
The "problem of too much talent"



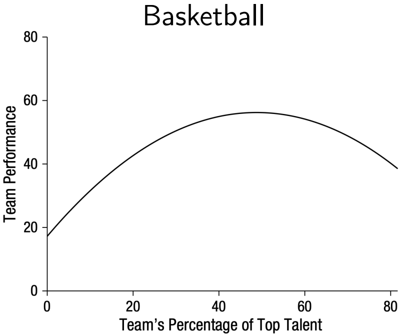


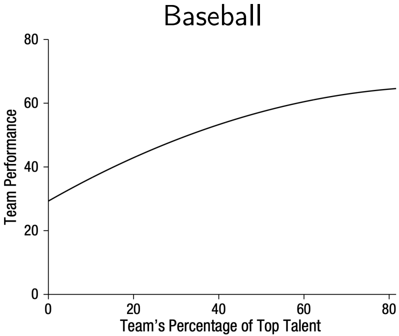
The "problem of too much talent"

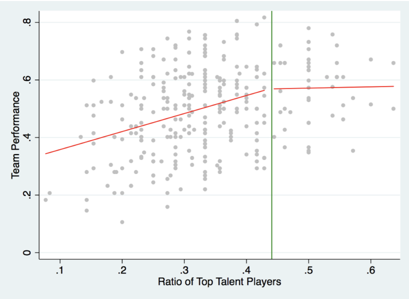


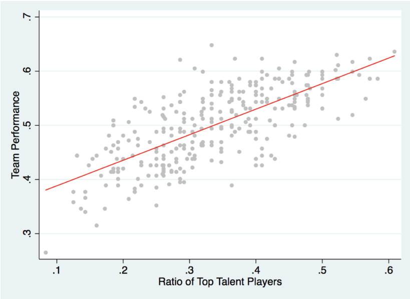


The "problem of too much talent"









Activity

Experiment with self-selected treatment assignment

Age, sex, interest in sports, theater, cooking, politics

Choice of vignette:

Sports vignette:

William is on the varsity soccer team. The night before an important final exam, the coach calls up to remind him of an upcoming practice. The next day, William goes to the practice and misses the exam without notifying the instructor. With a zero on the final exam, William would fail the class. The instructor allows him to take a makeup exam but will only give him partial credit. If you were the instructor, how much credit (between 0 and 100%) would you give for the makeup exam?

Theater vignette:

William is in the university theater program. The night before an important final exam, the director calls up to remind him of an upcoming rehearsal. The next day, William goes to the rehearsal and misses the exam without notifying the instructor. With a zero on the final exam, William would fail the class. The instructor allows him to take a makeup exam but will only give him partial credit. If you were the instructor, how much credit (between 0 and 100%) would you give for the makeup exam?

Discuss reading and homework

Computer demonstration

Drill

Log transformations

For each of example, express the underlined coefficient as a comparison, first on the transformed scale, then on the untransformed scale.

Discussion problem

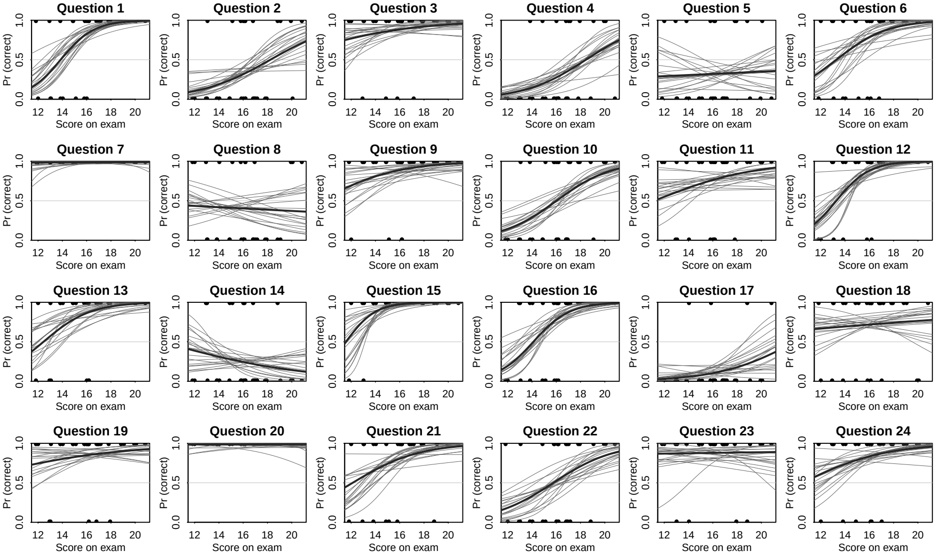
Causal inference and adjustment for context

1. Consider a causal question of interest to you.
2. Consider how the effect of interest could vary across the population, over time, or based on context.
3. What statistical adjustments could be done when generalizing from an experiment to more general settings?

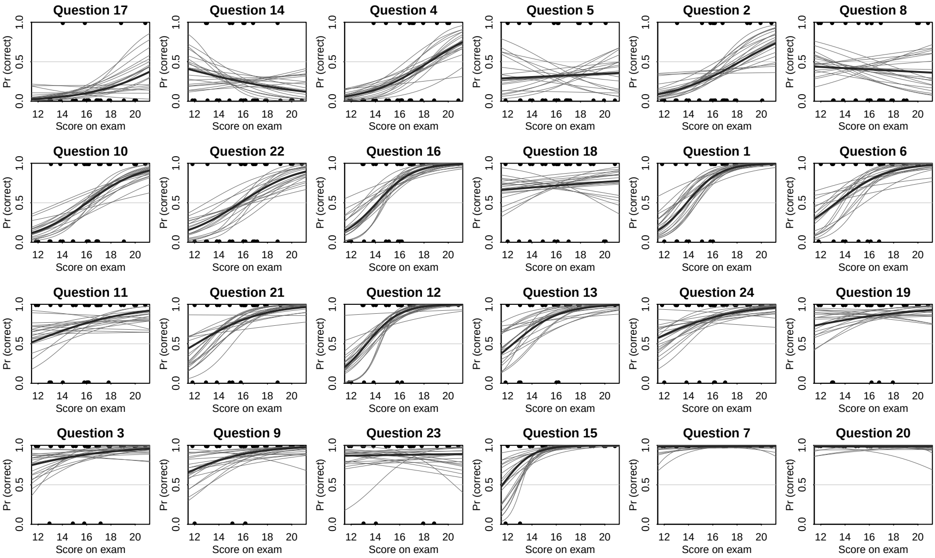
Class 8a: Logistic regression

Story

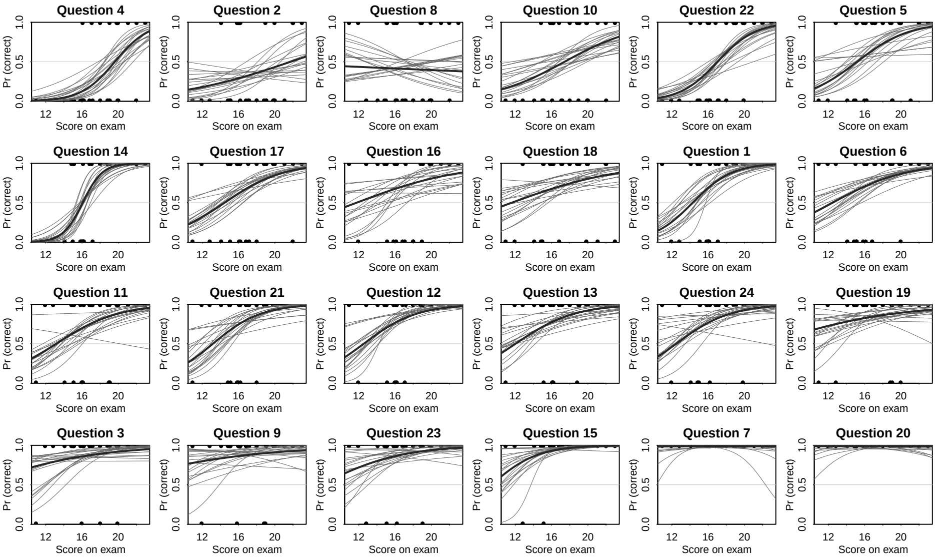
Item-response analysis of final exams



Item-response analysis of final exams



Item-response analysis of final exams



Activity

Two truths and a lie

Within your group:

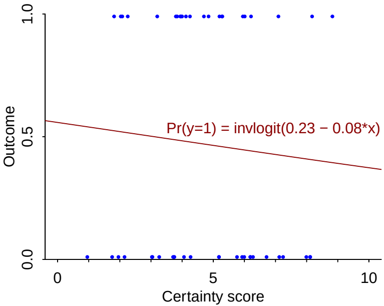
1. One person tells three personal statements, one of which is a lie.
2. Others discuss and guess which statement is the lie, and they give an estimate of their certainty in the guess (on a 0–10 scale).
3. The storyteller reveals which was the lie.
4. Enter the certainty estimate and the outcome (success or failure) and submit in the Google form.

Rotate through everyone in your group so that each person plays the storyteller role once.

Example data:

Certainty	Outcome
8	Success
4	Success
7	Failure
5	Success

Two truths and a lie



Coefficient	Estimate (s.e.)
Intercept	0.23 (0.77)
Slope	−0.08 (0.15)

Discuss reading and homework

Computer demonstration

Drill

Divide-by-4 rule

For each of the following models, calculate the halfway point (where the predicted probability is $\frac{1}{2}$) as well as the curve's steepest slope, using the divide-by-4 rule.

Discussion problem

Real-world example of logistic regression

Give an example of logistic regression, other than the examples we've considered in class or the readings.

Class 8b: Working with logistic regression

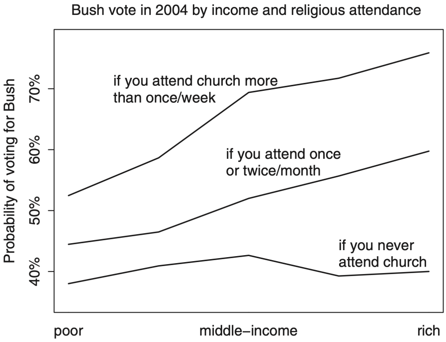
Story

Opiate of the masses and post-materialism

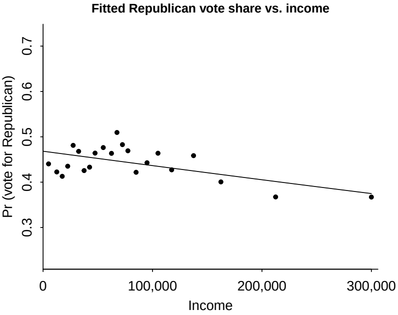
Two models of voting on social issues:

- *Opiate of the masses:* Rich people vote their interests; poor people vote "Gods, guns, and gays."
- *Postmaterialism:* Poor people vote based on economics; rich people have the luxury to vote on social issues.

Opiate of the masses and post-materialism



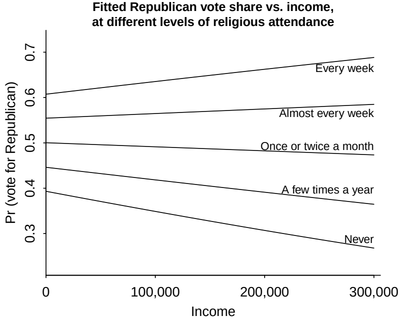
Opiate of the masses and post-materialism



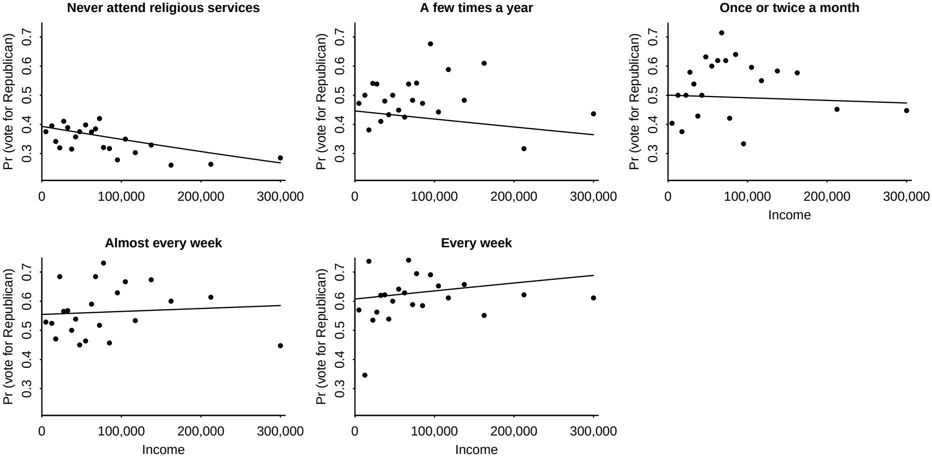

```
 family:       binomial [logit]
 formula:      trump ~ income100
 observations: 6336
 predictors:   2
------
            Median MAD_SD
(Intercept) -0.13   0.04
income100   -0.13   0.03
```

Opiate of the masses and post-materialism

```
family:        binomial [logit]
formula:       trump ~ church +
  income100 + church:income100
observations: 6313
predictors:   4
------
                 Median MAD_SD
(Intercept)      -0.65   0.07
church            0.22   0.03
income100        -0.27   0.06
church:income100  0.08   0.02
```



Opiate of the masses and post-materialism

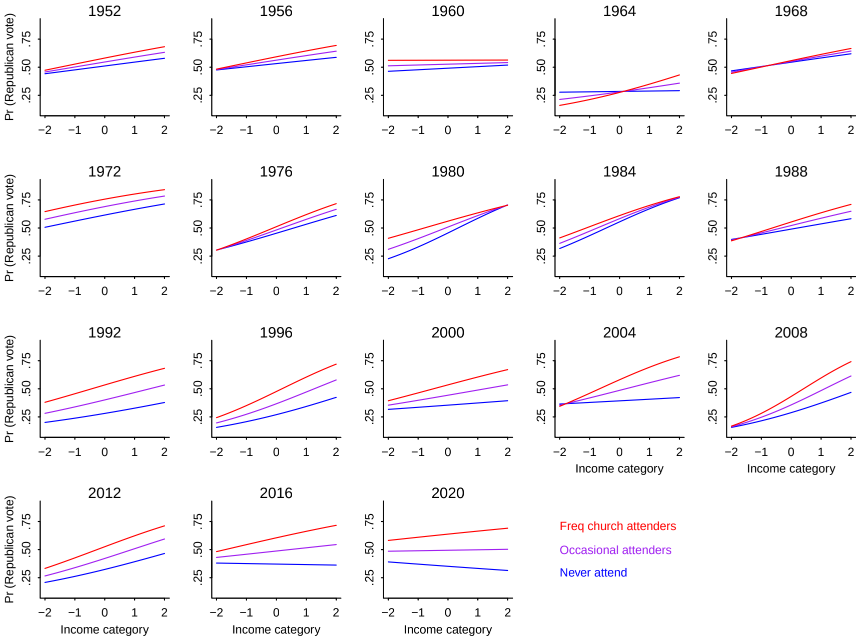


Opiate of the masses and post-materialism

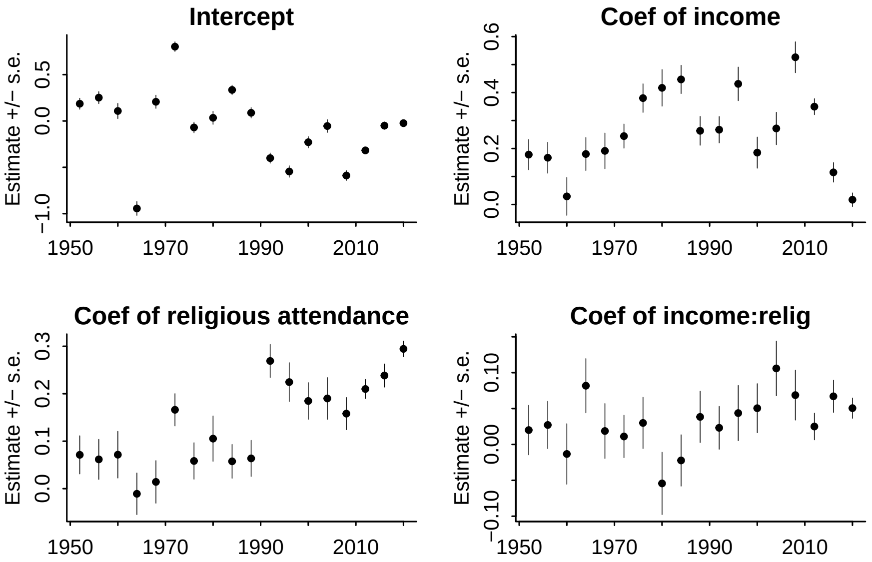
```
 family:        binomial [logit]
 formula:       trump ~ church +
   income100 + church:income100
 observations: 6313
 predictors:   4
------
                 Median MAD_SD
(Intercept)      -0.65   0.07
church            0.22   0.03
income100        -0.27   0.06
church:income100  0.08   0.02
```

```
 family:        binomial [logit]
 formula:       trump ~ c_church +
 c_income100 + c_church:c_income100
 observations: 6313
 predictors:   4
------
                     Median MAD_SD
(Intercept)          -0.25   0.03
c_church              0.28   0.02
c_income100          -0.10   0.04
c_church:c_income100  0.08   0.02
```

Opiate of the masses and post-materialism



Opiate of the masses and post-materialism



Activity

Job training programs and average predictive comparisons

- x: Pre-treatment employment history (0–10 scale)
- z: Treatment indicator
- y: Outcome: 1 if employed at the end of the study or 0 otherwise

Logistic regression, $\Pr(y = 1) = \text{logit}^{-1}(a + bx + \theta z)$

Discuss reading and homework

Computer demonstration

Drill

Understanding the logistic function

For each example, fill in the blanks in the formula,
$\Pr(y = 1) = \text{logit}^{-1}(\_ + \_x)$.

Discussion problem

Experimental design for logistic regression

Suppose a certain disease has a 20% mortality rate, and a new drug is hypothesized to reduce the mortality rate to 10%. Frame this as a logistic regression, and suppose a randomized experiment is performed with $n/2$ people getting the treatment and $n/2$ getting the control. How large must n need to be so that the uncertainty in the estimated treatment effect is low enough that we can be nearly certain of correctly identifying its beneficial effect?

Class 9a: Design and sample size decisions

Story

Lucky golf balls and implausible effect sizes

From a published paper: "We recruited 28 university students (12 males, 16 females) as participants and randomly assigned them to a superstition-activated or a control condition. Participants were asked to engage in a 10-trial putting task. A pretest revealed that more than 80% of our participant population believed in good luck, so to activate the superstition, we linked the concept of good luck to the ball participants used during the task. Specifically, while handing the ball over to the participants, the experimenter said, 'Here is your ball. So far it has turned out to be a lucky ball' (superstition-activated condition) or 'This is the ball everyone has used so far' (control condition). Finally, participants performed the required 10 putts from a distance of 100 cm.

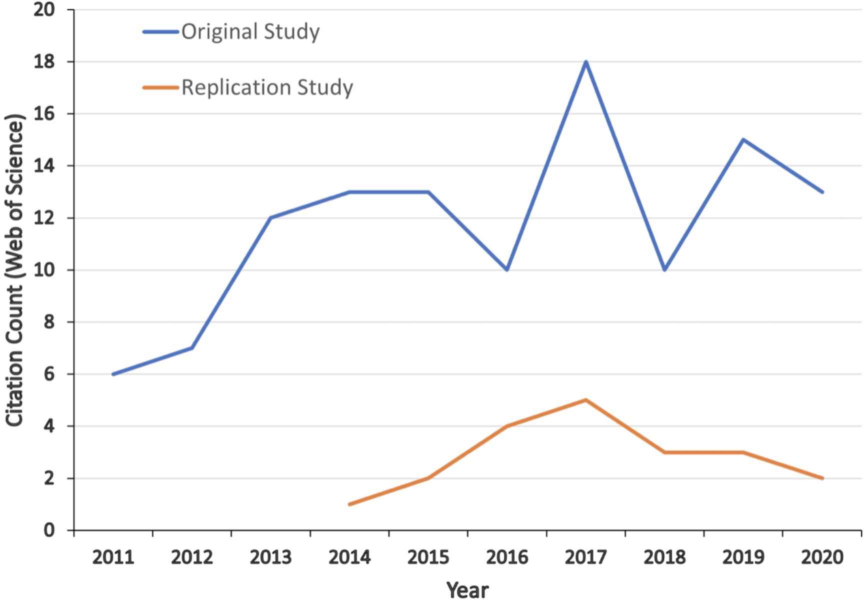
We used the number of hits as our central dependent measure, with 'hits' defined as successful putts (when the ball actually ended up where it was supposed to be). As predicted, participants performed better when playing with an ostensibly lucky ball ($M = 6.42$, $SD = 1.88$) rather than a neutral ball ($M = 4.75$, $SD = 2.15$)."

Lucky golf balls and implausible effect sizes

Original: "Participants were asked to engage in a 10-trial putting task . . . from a distance of 100 cm. We used the *number of hits* as our central dependent measure. As predicted, *participants* performed better when playing with an ostensibly lucky ball rather than a neutral ball."

Alternative: "All participants were asked to engage in a *15-trial* putting task from a distance of *150 cm*. As our central dependent measure, we used *total success*, defined as more than half the attempted putts ending up where they were supposed to be. Consistent with our hypothesis, participants *who believed in luck* performed better when playing with an ostensibly lucky ball, while there was no difference *among those expressed no belief in luck*."

Lucky golf balls and implausible effect sizes



Activity

Sample size calculation for a hypothetical study of left-handedness

Course of study	Sample size	Percentage left-handed
Behavioral sciences	90	8.89%
Humanities	51	9.80%
Sciences	92	4.35%
Other arts and sciences	156	7.05%
Business	241	9.54%
Music	47	14.89%
Design and art	147	12.24%
Engineering	75	10.67%
Nursing	71	4.23%
Other	75	9.38%

Sample size calculation for a hypothetical study of left-handedness

Course of study	Sample size	Percentage left-handed	± standard error
Behavioral sciences	90	9%	±3%
Humanities	51	10%	±4%
Sciences	92	4%	±2%
Other arts and sciences	156	7%	±2%
Business	241	10%	±2%
Music	47	15%	±5%
Design and art	147	12%	±3%
Engineering	75	11%	±2%
Nursing	71	4%	±4%
Other	75	9%	±3%

Discuss reading and homework

Computer demonstration

Drill

Sample size calculations: proportions

Calculate the sample size needed to achieve the stated statistical goal.

Discussion problem

Choosing a survey design to estimate question-wording effects

Compare two options:

- *Within-subject design*: Put the two different wordings on the same survey form (randomizing the order of the two questions) and compare responses to the two wordings.
- *Between-subject design*: Randomly give one wording to half the respondents and the other wording to the other half.

Class 9b: Poststratification and missing-data imputation

Story

Using MRP to estimate public opinion by state

THE TIMES
FRIDAY JANUARY 7 2022

Log in | Subscribe | Search

VIDEO | ELECTION 2019

MRP election poll: Boris Johnson heads for big majority

Corbyn to reshape election strategy as survey predicts Tories will win 359 seats, Labour 211

The MRP poll predicts a big win for the Tories – is it right?

Major YouGov poll points to comfortable 68-seat majority for Boris Johnson on 12 December

Using MRP to estimate public opinion by state

Consider 4 MRP models:

1. No state-level predictors
2. Republican vote share as a state-level predictor
3. Percent rural as a state-level predictor.
4. Include Republican vote share and percent rural as two state-level predictors

Activity

Experimental design and effect sizes

1. z <- sample(rep(c(0, 10), c(n/2, n/2)), n)
2. z <- sample(rep(c(0, 50), c(n/2, n/2)), n)
3. z <- sample(rep(c(0, 200), c(n/2, n/2)), n)
4. z <- sample(rep(c(0, 5, 10), c(n/3, n/3, n/3)), n)
5. z <- sample(rep(c(0, 25, 50), c(n/3, n/3, n/3)), n)
6. z <- sample(rep(c(0, 100, 200), c(n/3, n/3, n/3)), n)
7. z <- sample(runif(n, 0, 10))
8. z <- sample(runif(n, 0, 50))
9. z <- sample(runif(n, 0, 200))

Discuss reading and homework

Computer demonstration

Drill

Methods for imputation

Give an advantage and a disadvantage of each of the following approaches for imputing missing responses for a question in a survey, and come up with a scenario in which you would be comfortable using it.

Discussion problem

Network sampling

- A researcher at Columbia University's School of Social Work wanted to estimate the prevalence of drug abuse problems among American Indians (Native Americans) living in the New York City area.
- She did not have a list of the population, so instead she planned to obtain a sample using network sampling.
- Challenges of sampling and adjustment to the population.

Class 10a: Causal inference and randomized experiments

Story

Treatment effect depends on the population: coronavirus example

Hypothetical scenario of 1000 people:

- 300 would live either way
- 450 would die either way
- 250 would be saved by the treatment

Average treatment effect: 25 percentage points

Activity

Potential outcomes and treatment assignments for basketball training

Potential outcomes, y:

- y^1: number of free throws you make out of 50 tries, if $z = 1$
- y^0: number of free throws you make out of 50 tries, if $z = 0$

Treatment, z:

- $z = 1$: practice for 15 minutes each day for a month
- $z = 0$: no practice

Pre-treatment predictors, x:

- Age
- Self-assessed athleticism (on 1–10 scale)

Discuss reading and homework

Computer demonstration

Drill

Average treatment effects

```
 formula:      post_test ~ z + pre_test + z:pre_test
 observations: 100
 predictors:   3
 ------
             Median MAD_SD
(Intercept) 23.6   10.9
z           10.4    4.0
pre_test     0.7    0.2
z:pre_test  -0.4    0.3

Auxiliary parameter(s):
      Median MAD_SD
sigma 20.1    1.4
```

You also have a data frame, `pop`, representing the population.

Give R code to estimate different average treatment effects:

Discussion problem

Randomization and ethics

Ethical challenges in a medical experiment:

- The experimental treatment could be risky.
- If the new treatment is believed to be better, it could seem unfair to give someone the control.
- What standard of evidence should be required for the treatment to be deemed effective enough to be approved for public use?
- How to balance risks and benefits?
- What if a new treatment is very slightly more effective but much more expensive?

Class 10b: Causal inference using regression on treatment

Story

Pest control experiment: estimating a multiplicative treatment effect

Potential outcomes, y:

- y^1: number of roaches in your apartment, if $z = 1$
- y^0: number of roaches in your apartment, if $z = 0$

Treatment, z:

- $z = 1$: cleaning/poison/sealing and pest control advice
- $z = 0$: pest control advice

Pre-treatment predictor, x:

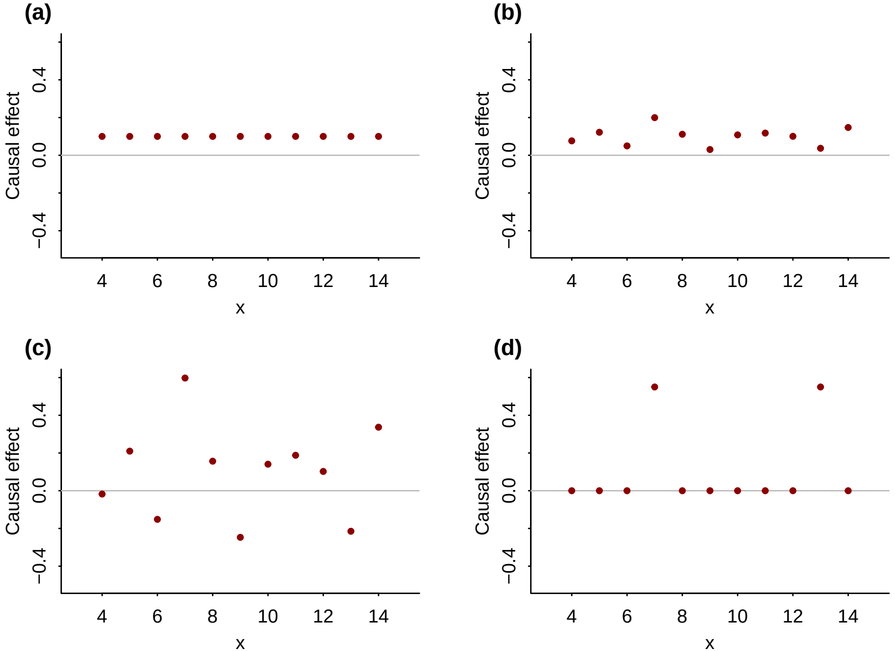
- Number of roaches measured before treatment

Activity

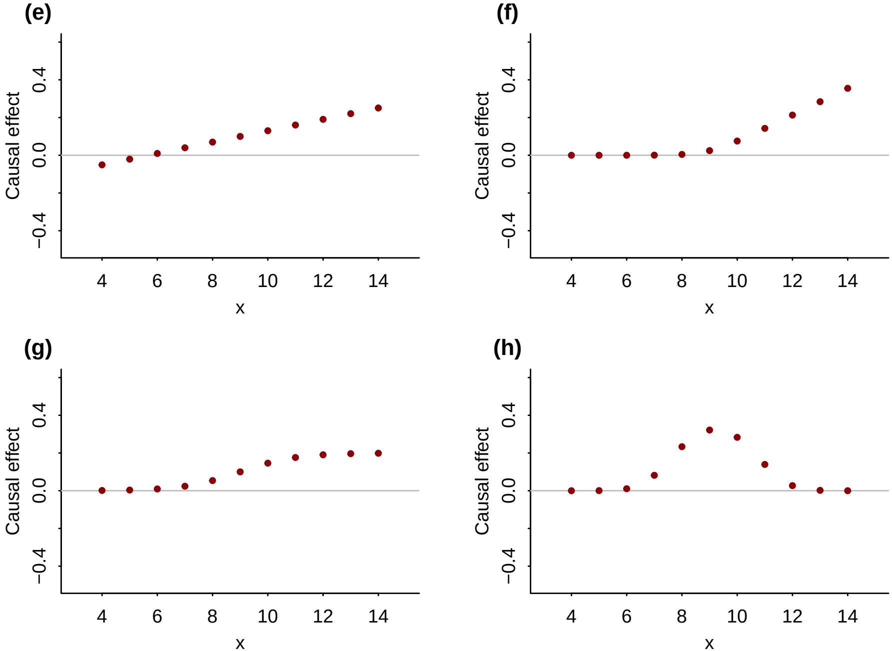
Understand the "average treatment effect"

- Reported effect size: 0.1 points of grade point average
- Equivalent to a 1 point effect on 10% of people
- In pairs, come up with examples of individual and average effects

Understand the "average treatment effect"



Understand the "average treatment effect"



Drill

Average treatment effect and poststratification

For each model,

- Give R code to compute the sample average treatment effect, ignoring any uncertainty in the coefficient estimates
- z is a binary variable and you want to compare $z = 0$ to $z = 1$
- Assumee that the data for the regression are in a data frame, `expt`

Discussion problem

Holding all else equal?

A published claim: "Education is an important determinant of income—one of the most important—but it is less important than most people think. If everyone had the same education, the inequality of income would be reduced by less than 10%. When you focus on education you neglect the myriad other factors that determine income. The differences of income among people who have the same education are huge."

What's wrong with the above statement?

Class 11a: Causal inference

Story

The Freshman Fallacy and interactions

Me: "[There is a problem with] representativeness. What color clothing you wear has a lot to do with where you live and who you hang out with. Participants in an Internet survey and University of British Columbia students aren't particularly representative of much more than ... participants in an Internet survey and University of British Columbia students."

An angry psychology professor: "Complaining that subjects in an experiment were not randomly sampled is what freshmen do before they take their first psychology class. I really *hope* you [know] why that is an absurd criticism—especially of authors who never claimed that their study generalized to all humans. (And please spare me 'but they said men and didn't say THESE men' because you said there were problems in social psychology and didn't mention that you had failed to randomly sample the field. Everyone who understands English understands their claims are about their data and that your claims are about the parts of psychology you happen to know about)."

The Freshman Fallacy and interactions

1. Essentially no effect, with patterns in data coming from noise or measurement artifacts
2. Large and variable effects that depend strongly on the person and context
3. Large and consistent effects

Activity

Components of an observational study

1. Population
2. Sample
3. Pre-treatment measurement, x
4. Treatment or exposure, z
5. Treatment assignment rule
6. Outcome, y

Discuss reading and homework

Computer demonstration

Drill

Adjusting for post-treatment variables

For each hypothetical analysis, explain the problem with adjusting for post-treatment variables, and how this could be fixed.

Discussion problem

Individual and average effects

A published claim: "By some estimates, one or two patients died per week in a certain smallish town because of the lack of information flow between the hospital's emergency room and the nearby mental health clinic. In other words, if the records had been easier to match, they'd have been able to save more lives. On the other hand, if it had been easy to match records, other breaches of confidence might also have occurred. Of course it's hard to know exactly how many lives are at stake, but it's nontrivial."

How plausible is the above claim?

Class 11b: Observational studies with measured confounders

Story

Retrospective controlled evaluation of a policy experiment

Our summary: "The MVP had favourable impacts on outcomes in all MDG [Millennium Development Goal] areas, consistent with an integrated rural development approach. The greatest effects were in agriculture and health, suggesting support for the project's emphasis on agriculture and health systems strengthening."

A different group: "Our study finds that the impact of MVP on the MDGs was limited, and that core welfare indicators such as monetary poverty, child mortality and under-nutrition were not affected. . . . despite some positive impacts, we found mostly null results, suggesting that the intervention was ineffective."

Activity

Imbalance and lack of overlap

Discuss an example of real-world imbalance and lack of overlap.

Discuss reading and homework

Computer demonstration

Drill

Ignorability of treatment assignment

In the following cases of observational studies, discuss possible problems with the assumption of ignorability of the treatment assignment, conditional on the pre-treatment predictors and with respect to the potential outcomes.

Discussion problem

Variation in social science patterns

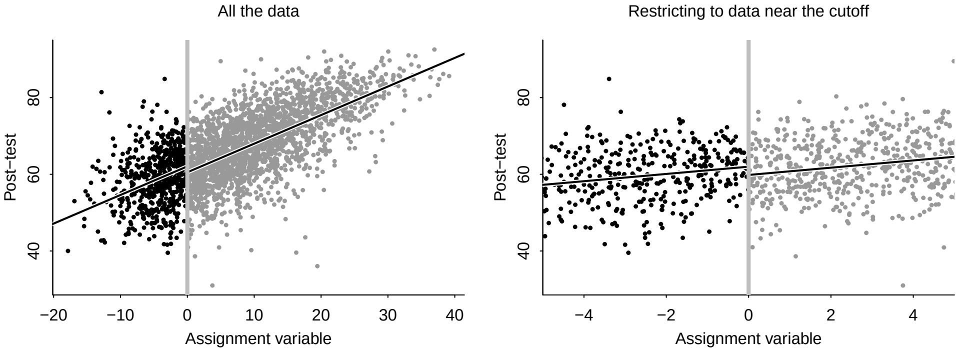
A published claim: "Sports participation causes women to be less religious, more likely to have children, and, if they do have children, more likely to be single mothers."

From authors' discussion: "It is true that many successful women with professional careers, such as Sheryl Sandberg and Brandi Chastain, are married. This fact, however, is not necessarily opposed to our hypothesis. Women who participate in sports may 'reject marriage' by getting divorces when they find themselves in unhappy marriages. Indeed, Sheryl Sandberg married and divorced before marrying her current husband."

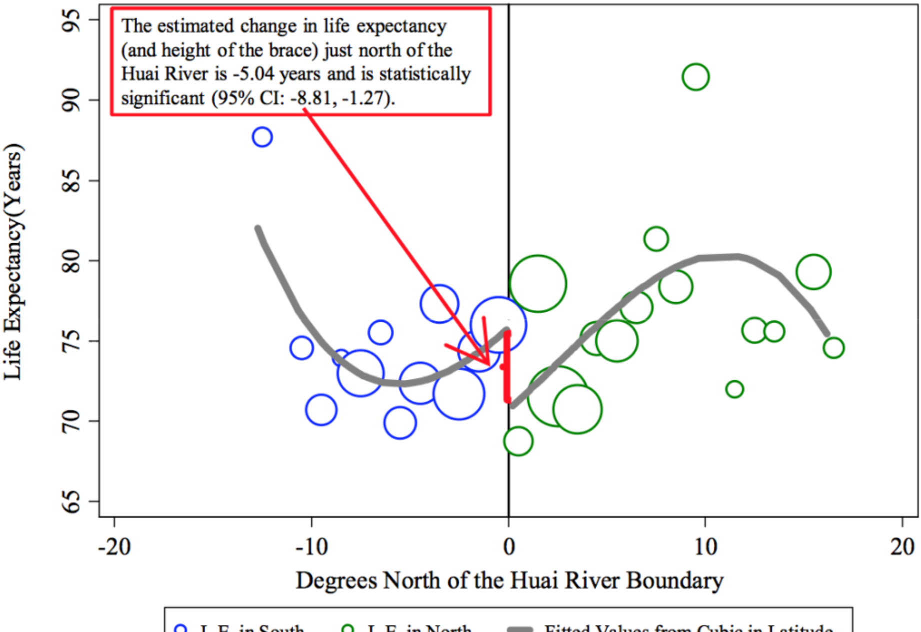
Class 12a: Additional topics in causal inference

Story

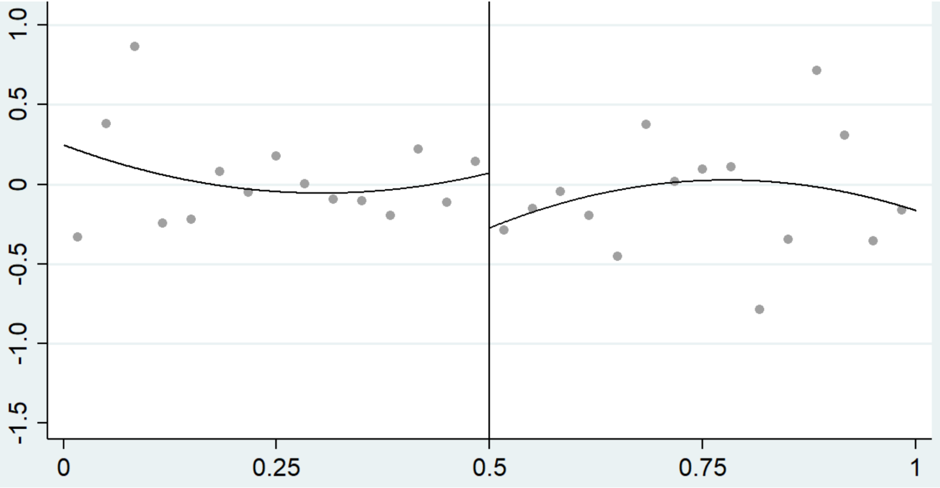
Regression discontinuity mishaps



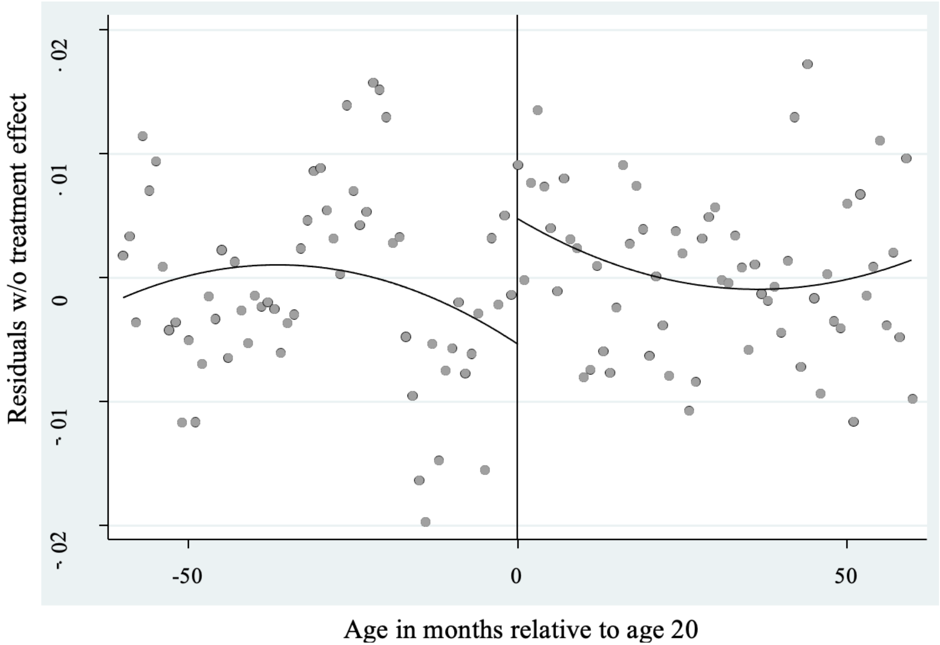
Regression discontinuity mishaps



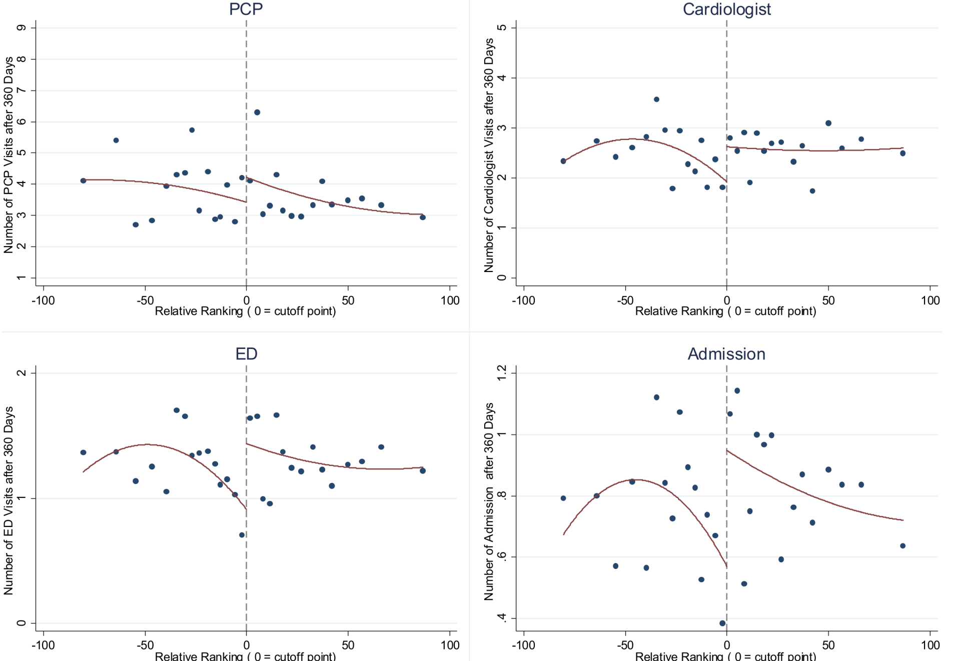
Regression discontinuity mishaps



Regression discontinuity mishaps



Regression discontinuity mishaps



Activity

Gather, plot, and discuss two measurements of the same underlying quantity from students

1. Come up with two ways of measuring a single characteristic using continuous or approximately continuous scales.
2. Everyone enters their data on to the Google form.
3. In pairs, discuss what the scatterplot might look like. Sketch guesses.
4. Compare to your actual data.

Discuss reading and homework

Computer demonstration

Drill

Assumptions for instrumental variables estimation

For each example, evaluate these assumptions for instrumental variables estimation:

- Ignorability of the instrument
- Monotonicity and nonzero association between instrument and treatment variable
- Exclusion restriction.

Discussion problem

Estimating the effects of masks and social distancing

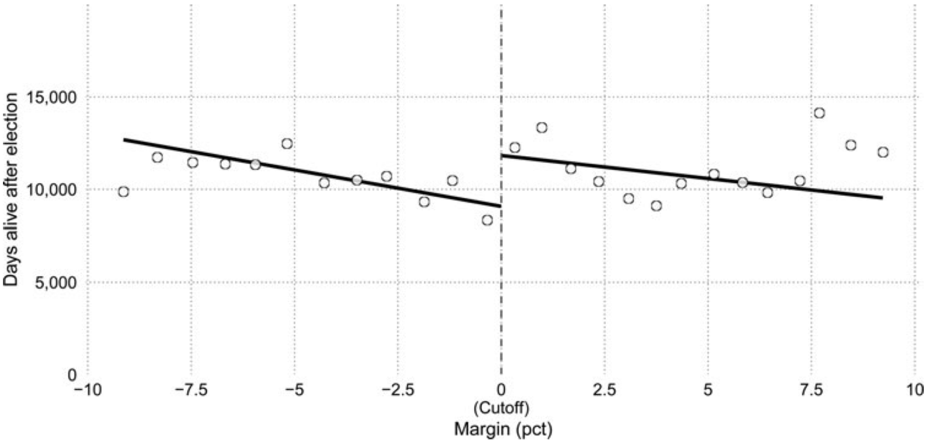
During the covid epidemic, people were not assigned at random to wear masks or to practice social distancing, but they were indirectly affected by national, state, and local policies mandating these actions.

1. Discuss how you might use instrumental variables to estimate these effects from available data on state-level policies, compliance, and outcomes.
2. Consider potential objections to such analyses.
3. Discuss possible data that could be gathered to better estimate the effects of interest.

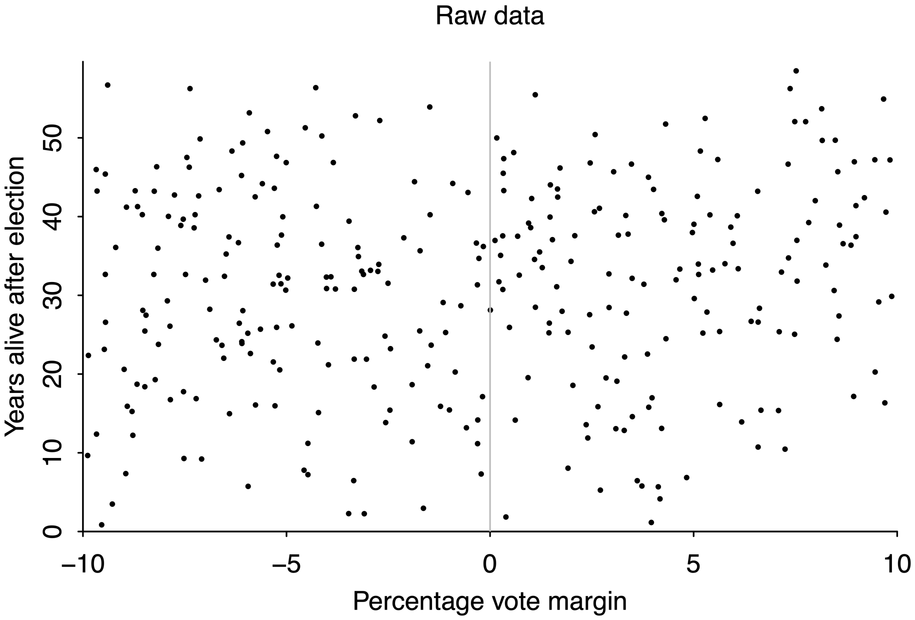
Class 12b: Advanced regression and multilevel models

Story

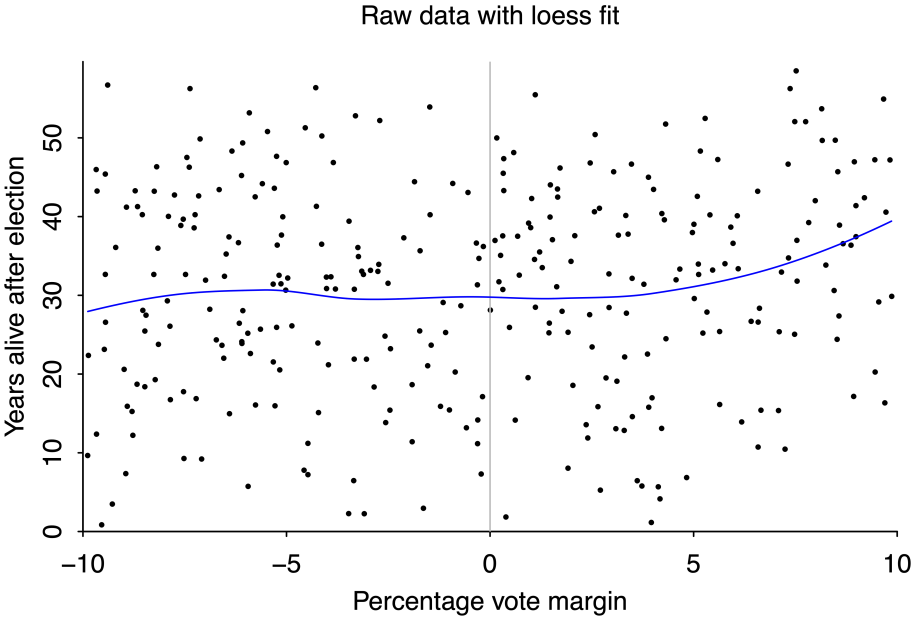
Nonlinear modeling for data exploration



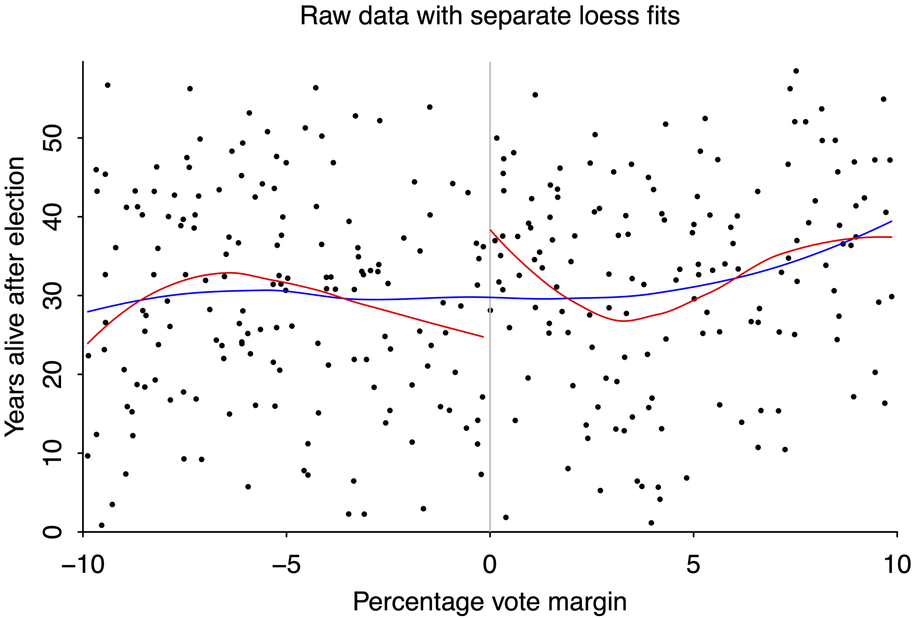
Nonlinear modeling for data exploration



Nonlinear modeling for data exploration



Nonlinear modeling for data exploration



Activity

Nonlinear treatment effect

1. Come up with a story with a pre-test measurement x, a treatment effect that is a non-monotonic effect of x, and potential outcomes y_0, y_1, under the control and treatment.
2. In pairs, sketch the expected value of y_0 and y_1 given x.
3. Come up with mathematical formulas to approximate these curves.
4. Simulate x, then z, then y.
5. Estimate the average treatment effect using a linear model fit to different subsets of the data.

Discuss reading and homework

Computer demonstration

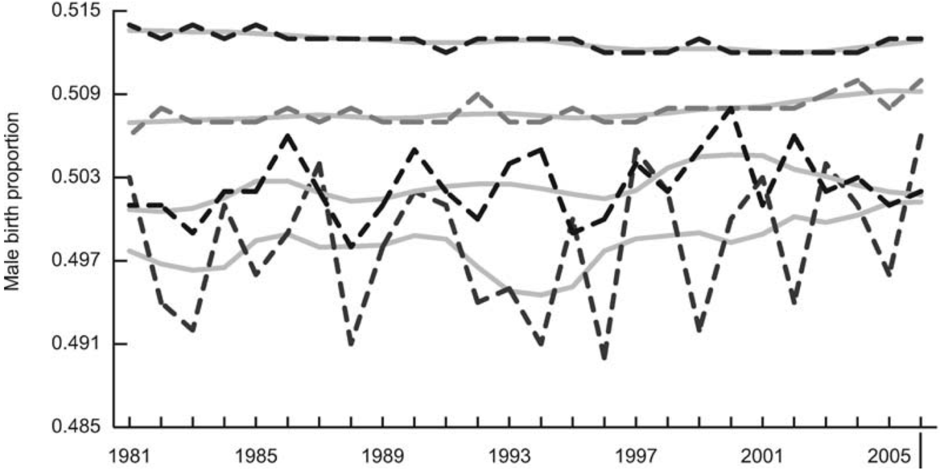
Drill

Nonlinear models

Each of these functions is defined for positive values of x. Sketch each function using pen on paper, labeling the axes appropriately.

Discussion problem

Noisy time series



Class 13a: Review of the course

Story

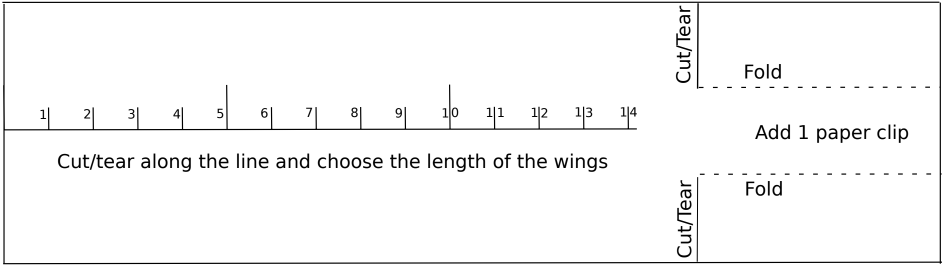
The rise and fall and rise of randomized controlled trials in international development

From an article from 2019: "What could explain the rise of RCTs [randomized controlled trials] in international development? . . . we are witnessing now a second wave of RCTs in international development, while a first wave of experiments in family planning, public health, and education in developing countries began in the 1960s and ended by the early 1980s. . . .

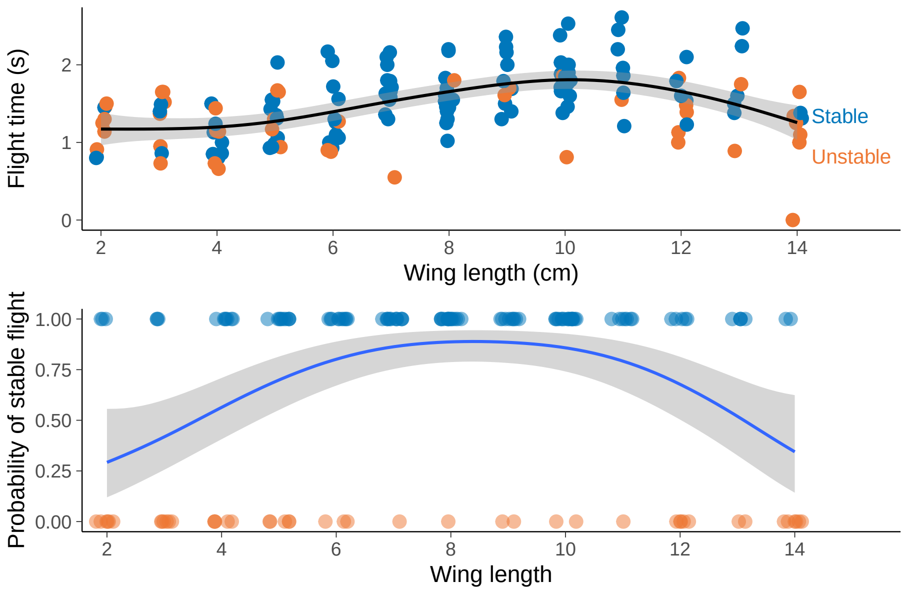
Instead of asking, 'why are RCTs increasing now?' we ask, 'why didn't RCTs spread to the same extent in the 1970s, and why were they discontinued?' In other words, how we explain the success of the second wave must be consistent with how we explain the failure of the first."

Activity

Design a paper helicopter



Design a paper helicopter



Discuss reading and homework

Computer demonstration

Drill

Consider the following model of price x and sales y: when the price is \$20, sales are 2000 units, and for every 1% increase in price, sales decrease by 0.8%. Write this model as a formula.

You are planning to conduct a random sample survey of n people in a country in which 80% of the population are native born and 20% are immigrants. As part of the analysis you plan to compare these two groups in their percentage who support more restrictive immigration laws. Suppose you want to estimate this difference to within a standard error of 5 percentage points. How large does n need to be?

Basic statistics and linear regression

Write R code to fit a linear regression with predictors x_1, x_2, x_3, and all their two-way interactions.

Basic statistics and linear regression

List at least four of the assumptions of linear regression, in decreasing order of importance.

Discussion problem

Designing an experiment using simulation

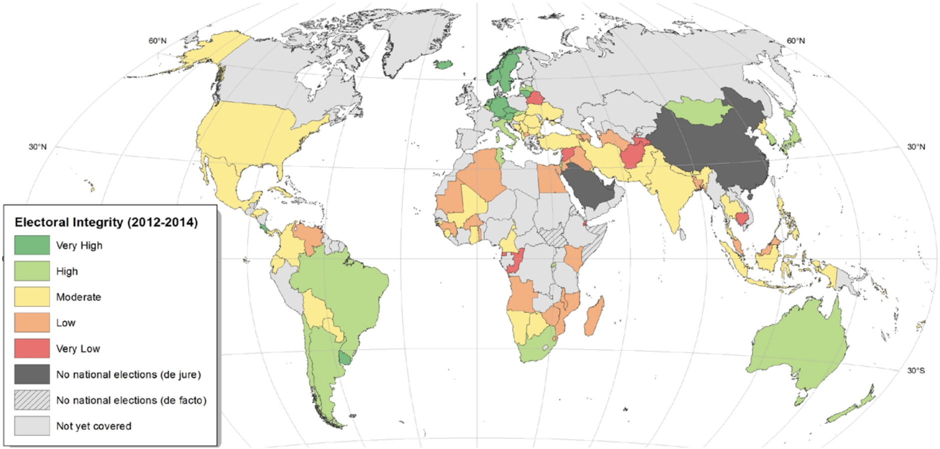
Suppose you want to design an experiment to estimate the effects of canvassing on voter turnout in an upcoming election. You have records on a large number of registered voters, with data on their past voter turnout, and your plan is to randomly select n people from this database and randomly chose $n/2$ to be contacted and encouraged to vote. You will then follow up after the election and see who actually voted.

1. Start with a guess of n.
2. Run a simulation, making assumptions as needed.
3. Use the simulations to compute the standard deviation of your estimate.
4. Scale up or down to pick n for your study.

Class 13b: Review of the course

Story

The Harvard study claiming North Carolina is less democratic than North Korea



The Harvard study claiming North Carolina is less democratic than North Korea

Electoral laws	53
Electoral procedures	73
District boundaries	73
Voter registration	83
Party and candidate registration	54
Media coverage	78
Campaign finance	84
Voting process	53
Vote count	74
Results	80
Electoral authorities	60

The Harvard study claiming North Carolina is less democratic than North Korea

From the project director: "The map identifies North Korea and Cuba as having moderate quality elections. The full report online gives details on how to interpret this. It does not mean that these countries are electoral or liberal democracies. The indicators measure expert perceptions of the quality of an election based on multiple criteria derived from international standards."

Activity

Semester review

1. Choose a method you have learned during the semester
2. Review the method
3. Discuss where the method works and where it fails
4. Discuss relevance to your applied interests
5. Points of confusion and open questions

Discuss reading and homework

Computer demonstration

Drill

From logistic regression through causal inference

Here is the result from a fitted logistic regression:

```
 family:       binomial [logit]
 formula:      y ~ x
 observations: 100
 predictors:   2
------
            Median MAD_SD
(Intercept)  1.0    0.5
x           -0.3    0.1
```

Suppose you define $z = 20 + 10 * x$. What would be the estimated coefficients of the logistic regression of y on z?

You are planning to conduct a randomized experiment with 100 people in the treatment group and 100 controls. The outcome is test scores, in a population where scores have a mean of 60 and standard deviation 15. You have a pre-test measurement, and you expect that the model fit to estimate the treatment effect will have an R^2 of 50%. Approximately what will be the standard error of the estimated treatment effect?

From logistic regression through causal inference

In an experiment you have outcome y, treatment indicator z, and a pre-test variables x in a data frame called `sample`. You also have x for a population of interest in a data frame called `pop`. Give R code to estimate the average causal effect in the population, allowing for the treatment effect to vary with x.

Discussion problem

Creating a better electoral integrity index

1. How would you create a better "electoral integrity index"?
2. How could you put North Carolina and North Korea on the same scale?
3. Consider issues of definition, measurement, and validation.