Classroom activities in applied regression and causal inference, second semester[1]

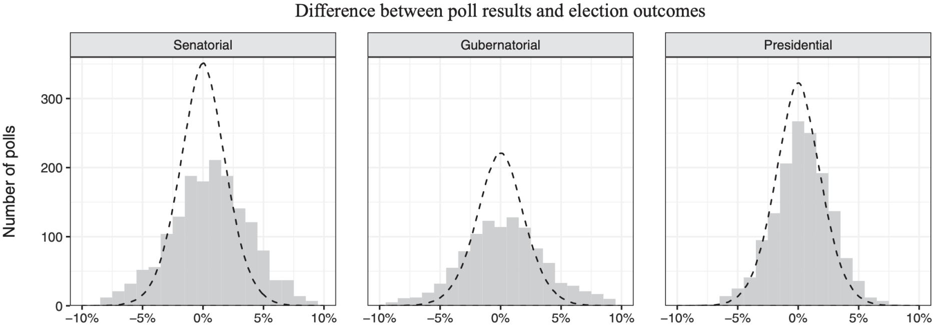
Andrew Gelman and Aki Vehtari

[1]For details, see chapter 4 of *Active Statistics*, by Andrew Gelman and Aki Vehtari, Cambridge University Press (2023), `http://www.stat.columbia.edu/~gelman/active-statistics/`.

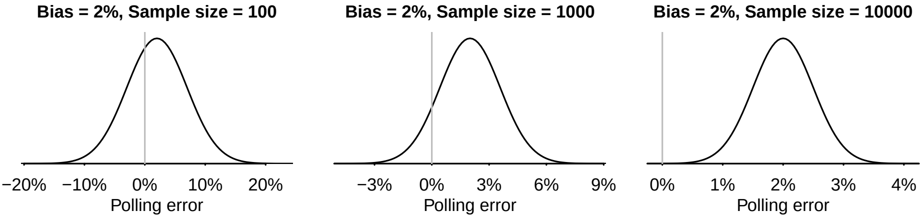
Class 1a

Story

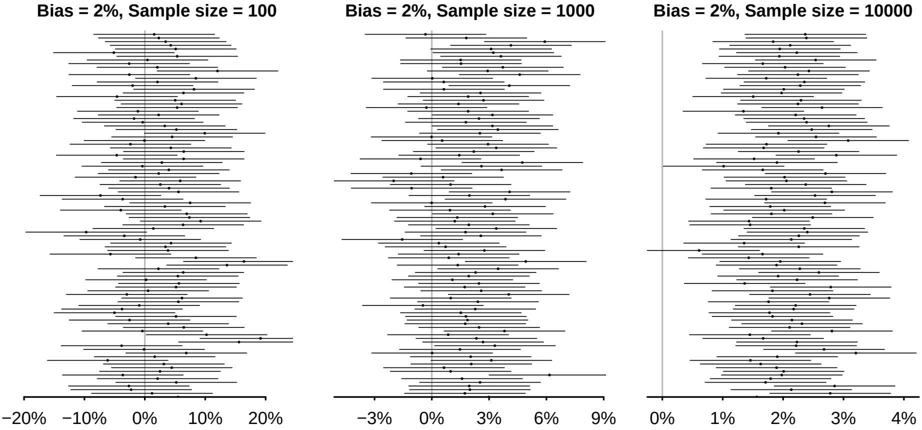
Biased samples and confidence interval coverage



Biased samples and confidence interval coverage



Biased samples and confidence interval coverage



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?

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

Regression coefficients as comparisons

For each example, express the underlined coefficient as a comparison.

Discussion problem

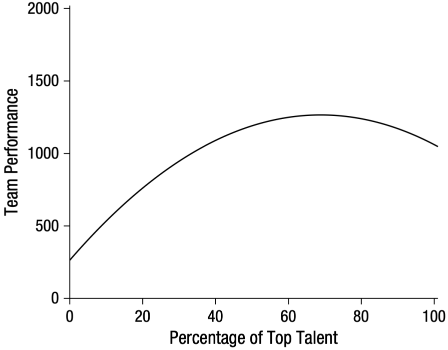
Sampling and adjustment

1. Consider a sampling problem of interest to you.
2. Consider ways in which the sample would not be representative of the population (undercoverage, unequal sampling probabilities, etc.).
3. What statistical adjustments could be done to account for these problems?

Class 1b

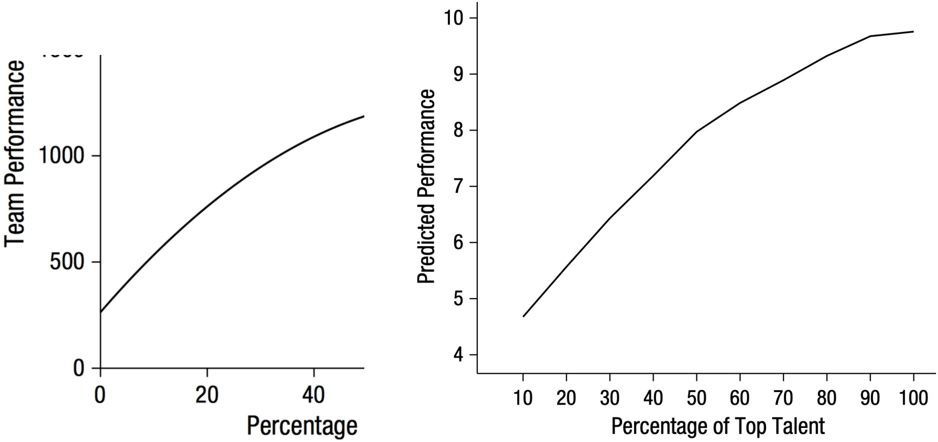
Story

The "problem of too much talent"

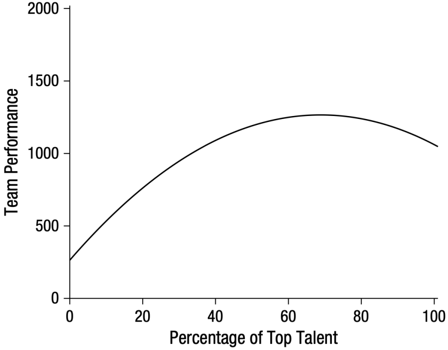


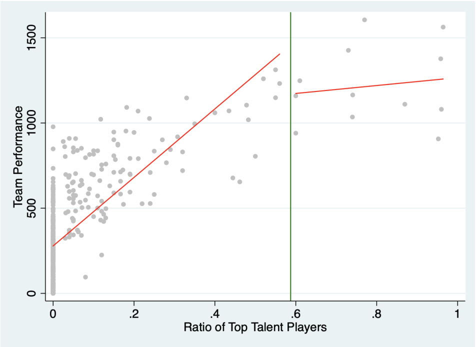
Measure	*M*	*SD*
1. Team performance (points)	393.30	320.12
2. Top-talent percentage	7%	16%
3. Roster size	18.53	6.79
4. Games played	8.90	4.65

The "problem of too much talent"

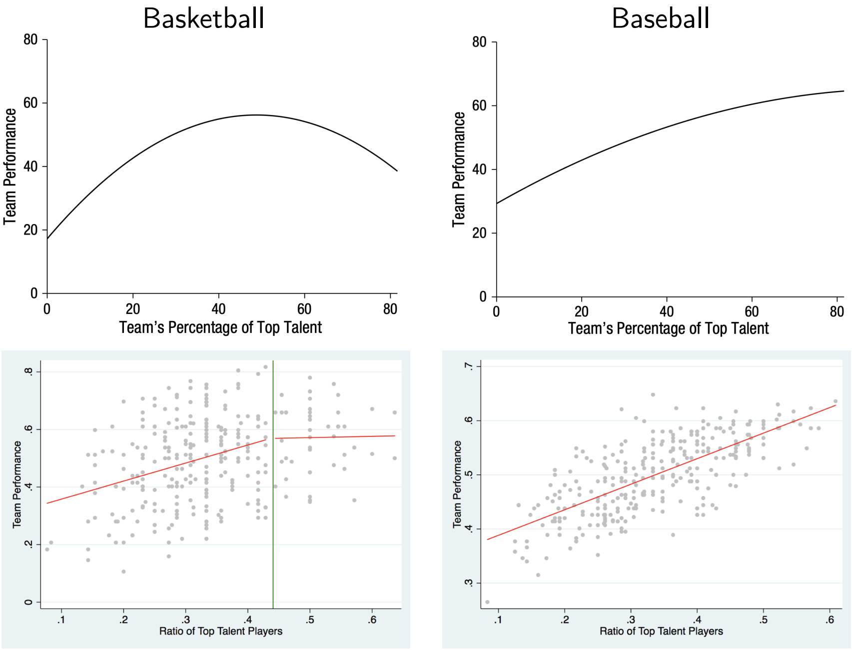


The "problem of too much talent"





The "problem of too much talent"



Activity

Designing a study to explore a nonlinear relation

- *Causal question:* If you add (or subtract) top-talent players from your team, what would you expect to happen to the team's performance?
- *Descriptive question:* How much better or worse are teams with more top talent, on average?

To answer the *descriptive question*, design a study:

- Measurement
- Data collection
- Analysis

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

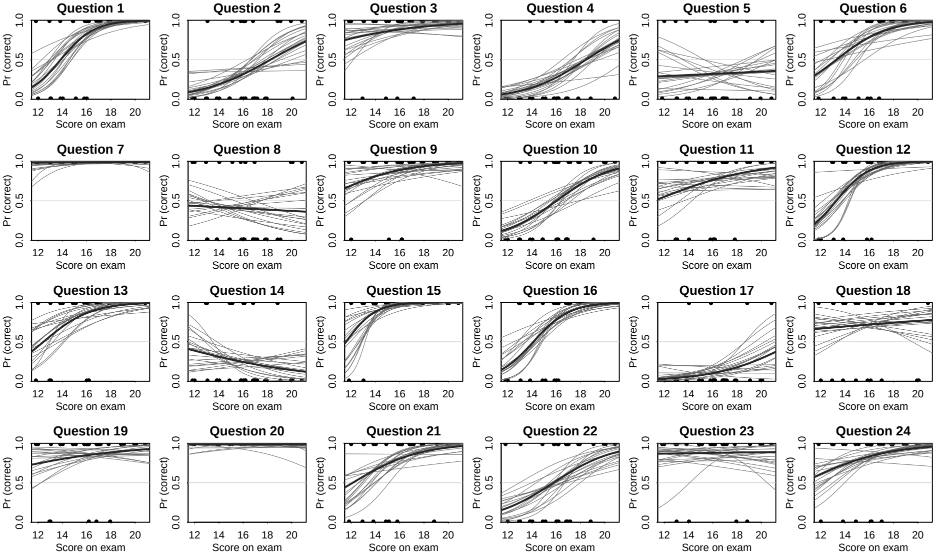
Causal inference and adjustment for context

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?

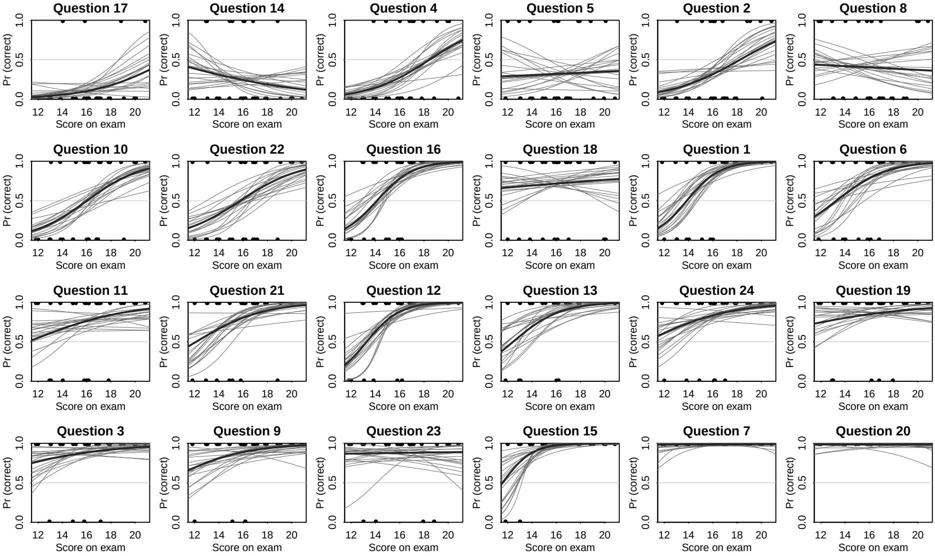
Class 2a

Story

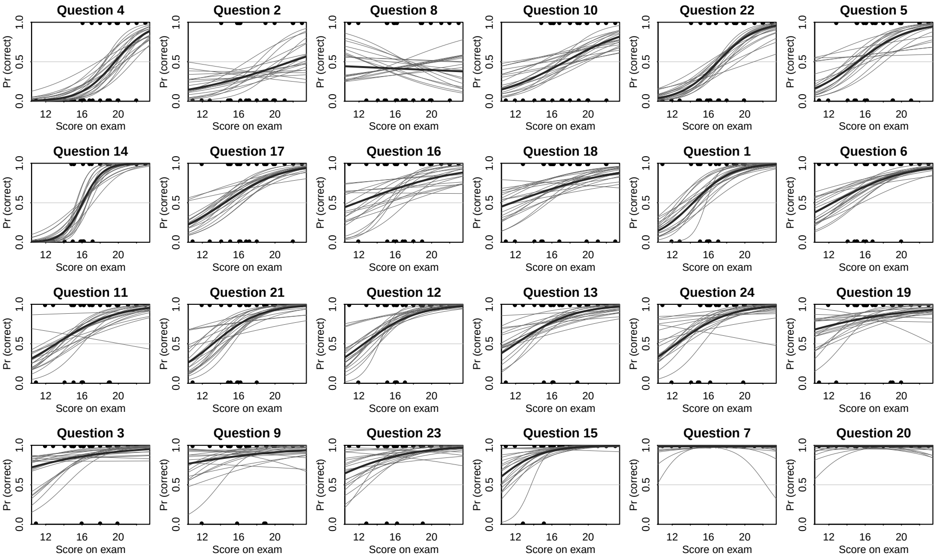
Item-response analysis of final exams



Item-response analysis of final exams



Item-response analysis of final exams



Activity

Two truths and a lie

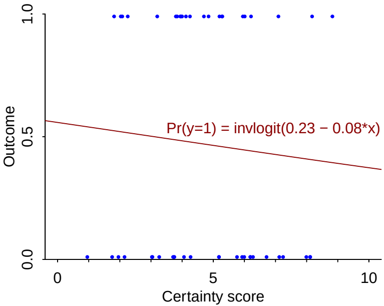
Within your group:

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

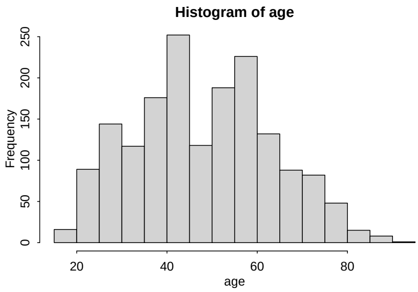
Real-world example of logistic regression

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

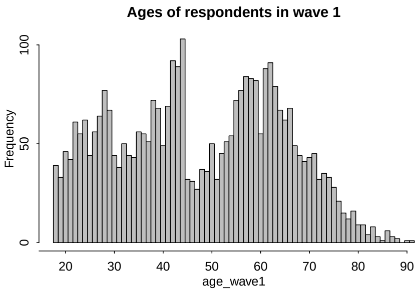
Class 2b

Story

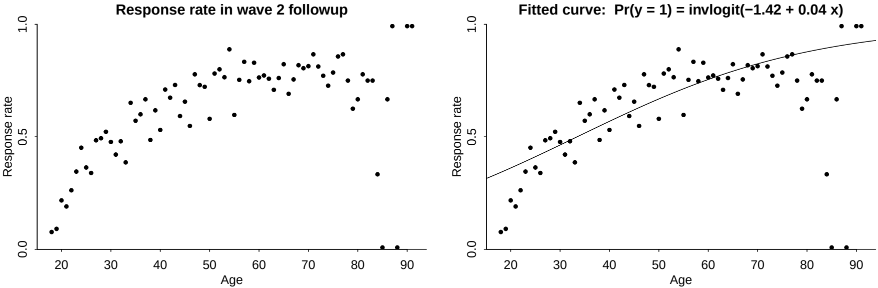
Survey nonresponse



Survey nonresponse



Survey nonresponse



Survey nonresponse

```
 family:       binomial [logit]
 formula:      respond ~ age + female + ethnicity
 observations: 3333
 predictors:   6
------
                  Median MAD_SD
(Intercept)       -1.32   0.14
age                0.04   0.00
female             0.28   0.08
ethnicityBlack    -0.55   0.12
ethnicityHispanic -0.56   0.13
ethnicityOther    -0.21   0.16
```

Activity

In-class survey

Come up with five questions to ask students in the class.

Discuss reading and homework

Computer demonstration

Drill

Interpreting logistic regression coefficients

For each of the following examples, interpret the constant and the slope estimate (where appropriate, you can use the divide-by-four rule). Plot the curves on an adequate range.

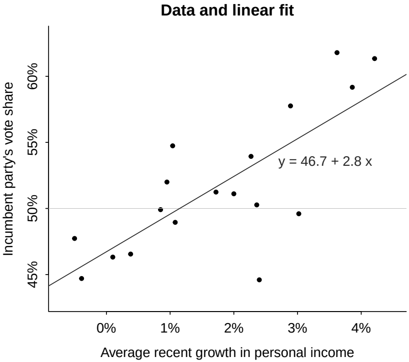
Discussion problem

What will happen if you fit a logistic regression to data that don't fit the model?

Class 3a

Story

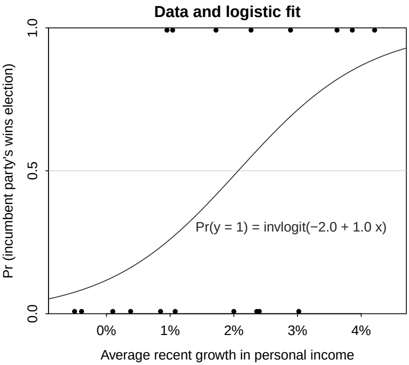
The "Keys to the White House"




```
 family:       gaussian [identity]
 formula:      vote ~ growth
 observations: 18
 predictors:   2
------
            Median MAD_SD
(Intercept) 46.7    1.4
growth       2.8    0.6

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

The “Keys to the White House”




```
 family:       binomial [logit]
 formula:      inc_win ~ growth
 observations: 18
 predictors:   2
------
            Median MAD_SD
(Intercept) -2.0    1.1
growth       1.0    0.5
```

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

Expected probabilities in logistic regression

	Median	MAD_SD
(Intercept)	-0.22	0.09
dist100	-0.90	0.11
arsenic	0.47	0.04
educ4	0.17	0.04

For each question, calculate the difference in expected probabilities of well-switching and interpret the result.

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 3b

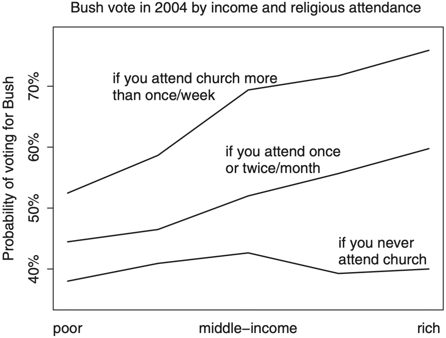
Story

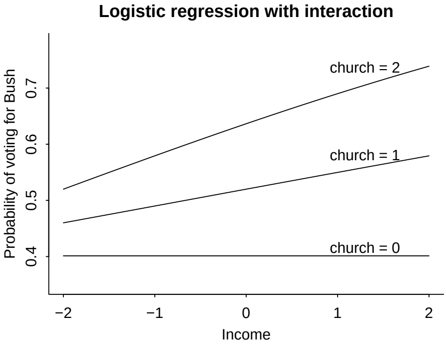
Opiate of the masses and post-materialism

Two models of voting on social issues:

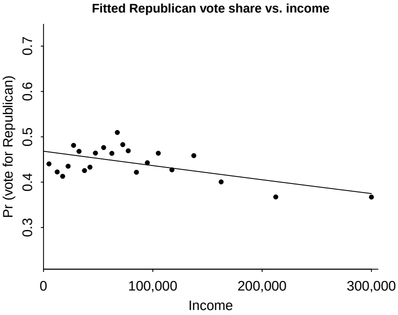
- *Opiate of the masses:* Rich people vote their interests; poor people vote "Gods, guns, and gays."
- *Postmaterialism:* Poor people vote based on economics; rich people have the luxury to vote on social issues.

Opiate of the masses and post-materialism





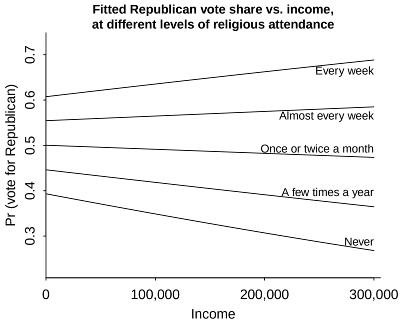
Opiate of the masses and post-materialism



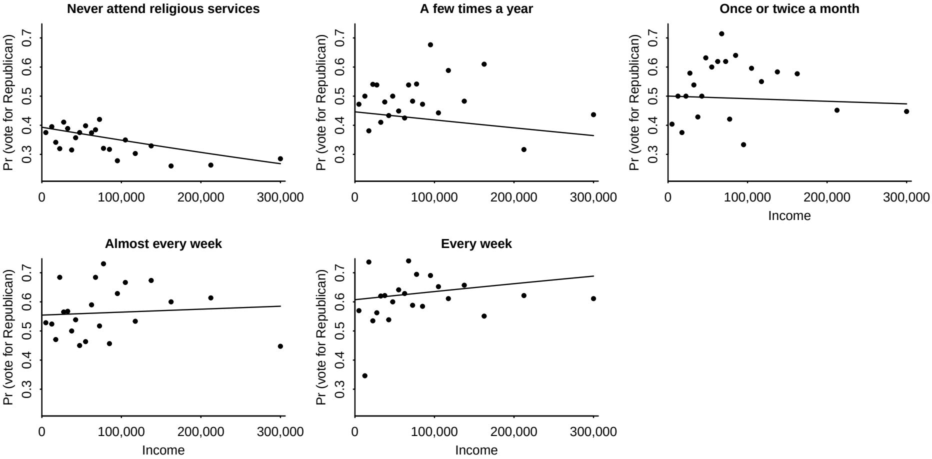

```
 family:       binomial [logit]
 formula:      trump ~ income100
 observations: 6336
 predictors:   2
------
            Median MAD_SD
(Intercept) -0.13   0.04
income100   -0.13   0.03
```

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

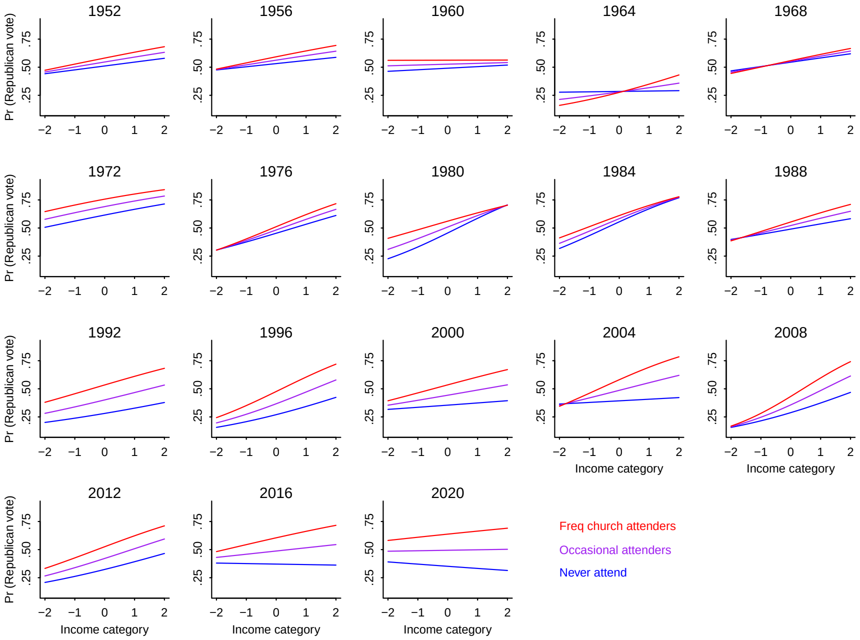


Opiate of the masses and post-materialism

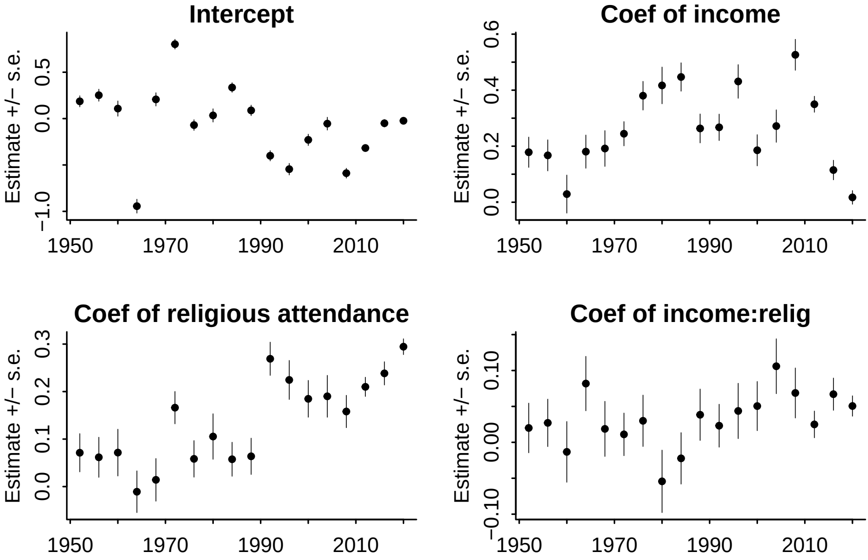
```
 family:       binomial [logit]
 formula:      trump ~ church +
   income100 + church:income100
 observations: 6313
 predictors:   4
------
                 Median MAD_SD
(Intercept)      -0.65   0.07
church            0.22   0.03
income100        -0.27   0.06
church:income100  0.08   0.02
```

```
 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



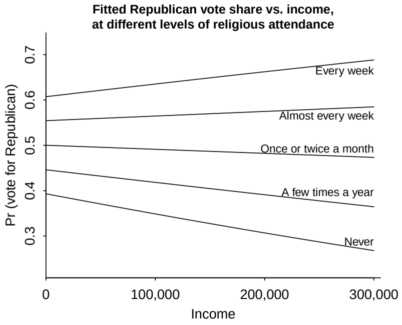
Opiate of the masses and post-materialism



Activity

Logistic regressions with interactions

```
 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
```



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 with pre-treatment predictor

Consider estimating the effect of a job-training program ($z = 1$ if a person gets the program and $z = 0$ otherwise) on a binary outcome y corresponding to whether he or she is employed within six months after the program, and a continuous pre-treatment predictor x representing employment history that is on a 1–5 scale. Suppose that, in the absence of the program, the probability of employment in six months ranges from 20% for people on the low end of the scale to 70% for people on the high end, and further suppose that the treatment increases these probabilities to 35% and 70%.
Express this hypothesized pattern as a logistic regression model with an interaction.
Next suppose this experiment is conducted with $n/2$ people getting the treatment and $n/2$ getting the control, with random assignment to a group of people whose values of x are uniformly distributed between 1 and 5. 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

Class 4a

Story

Understanding gun ownership using generalized linear models

From an article from 2021: "The gun ownership literature is vast, with dozens of studies seeking to explain who owns guns and why. We build on this literature in two key ways. First, we introduce a new variable into the fold: moral concern about harming others. We theorize that this concern actively inhibits gun ownership. Second, we direct theoretical and empirical attention to a predictor of gun ownership that has frequently been overlooked in the contemporary gun literature: childhood socialization . . ."

Understanding gun ownership using generalized linear models

"To measure Moral Harm, we used four items: 'If I saw a mother slapping her child, I would be outraged'; 'Compassion for those who are suffering is the most crucial virtue'; 'It can never be right to kill a human being'; 'The government must first and foremost protect all people from harm.' These items are drawn from previous measures ... We averaged the four items to create a Moral Harm index ..."

Understanding gun ownership using generalized linear models

"We measure Childhood Socialization using a five-item additive index that gauges both gun presence and gun socialization experiences. Specifically, the respondents were asked if when they were growing up, their family members or guardians did any of the following: 'Keep a firearm in the house'; 'Teach you how to shoot a firearm'; 'Teach you how to clean a firearm'; 'Take you hunting'; and 'Take you to a gun show.'"

Understanding gun ownership using generalized linear models

"Around 36% of the sample own a gun. Of these gun owners, the majority own multiple guns: around 70% own two or more guns and the average number of guns owned is greater than four."

Understanding gun ownership using generalized linear models

	Gun Ownership				Gun Quantity			
	Model 1		Model 2		Model 3		Model 4	
Variables	*b*	SE	*b*	SE	*b*	SE	*b*	SE
Moral Harm	-.543***	.104	-.400**	.146	-.565***	.090	-.325**	.103
Childhood Socialization	—	—	.666***	.057	—	—	.544***	.040
Fear of Crime	—	—	.003	.098	—	—	.035	.062
Victimization	—	—	.354	.223	—	—	.344*	.165
Dangerous World Beliefs	—	—	.193	.113	—	—	.129	.087
Confidence in the Police	—	—	-.149	.116	—	—	-.117	.082
Confidence in the Government	—	—	.112	.101	—	—	.157*	.068
Racial Resentment	—	—	-.416**	.124	—	—	-.371***	.090
Southerner	—	—	.760***	.182	—	—	.173	.140
Rightward Political Views	—	—	.486**	.181	—	—	.572***	.137
Born-Again Protestant	—	—	.125	.257	—	—	-.160	.161
Religiosity	—	—	.004	.132	—	—	-.007	.104
White	.254	.206	.208	.225	.346*	.161	.354*	.165
Male	.374*	.168	-.039	.196	.493**	.143	-.008	.142
Age	.004	.006	.002	.006	-.002	.005	.004	.004
Education	-.207***	.059	-.234**	.078	-.106	.055	-.134*	.062
Income	.047	.031	.075*	.037	.036	.029	.066*	.029
Married	.563**	.195	.456*	.203	.374*	.156	.201	.153
Child in Household	.057	.193	-.094	.227	.239	.184	.158	.183
N	1,065		1,051		1,057		1,044	

Activity

How similar are you to your friends?

- 10 questions (including some political attitudes and some personal views)
- Unique ID number for each pair of students
- Guess of your partner's attitudes

Discuss reading and homework

Computer demonstration

Drill

Interpret coefficients in a negative binomial or Poisson regression

1. For each example below, interpret the coefficients and standard errors of the fitted Poisson regression.
2. For each example it would be better to fit a negative binomial regression. If a negative binomial regression were fit instead of a Poisson regression, how would the estimated coefficients change and how would the standard errors change?

Discussion problem

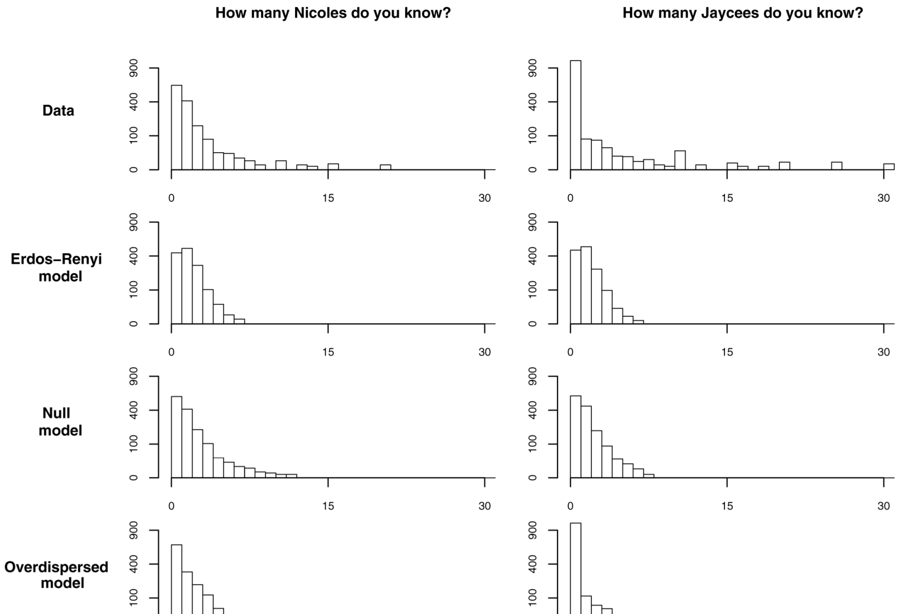
Identification in linear models

Discuss different ways that a linear model can be non-identified, and provide examples.

Class 4b

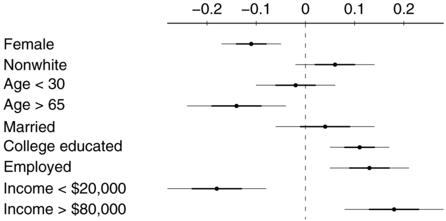
Story

Using overdispersion in count data to estimate social structure in networks

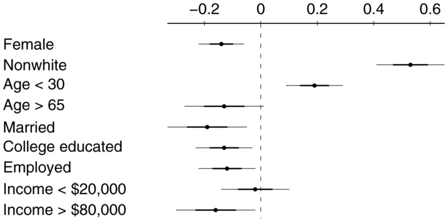


Using overdispersion in count data to estimate social structure in networks

Regression of estimated log gregariousness parameters



Regression of residuals from the "How many men do you know in prison?'q́uestion



Activity

Real-life overdispersion

Alternative model choices:

- Linear regression
- Thresholding and binary logistic regression
- Poisson regression
- Overdispersed Poisson regression

Discuss reading and homework

Computer demonstration

Drill

Interpret the parameters in ordered logistic regression

For each example, interpret the estimated coefficients and cutpoints of the fitted ordered logistic regression.

Discussion problem

Functional forms for nonlinear models

What is the problem with predicting y from x using the model, $y = a + b * \exp(-cx + d) + \text{error}$? This is an entry point to a discussion of different possible nonlinear functional forms.

Class 5a

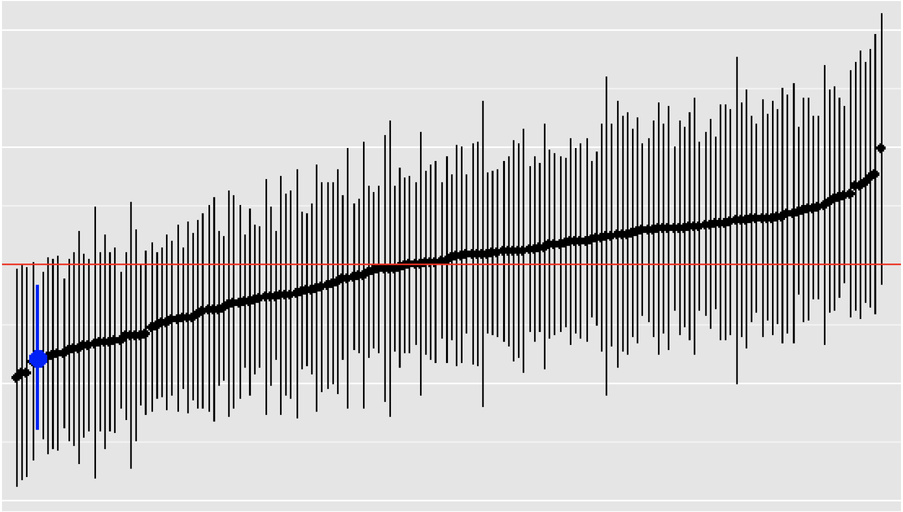
Story

The multiverse, the statistical significance filter, and the feedback loop

1. Exclusion criteria based on cycle length (3 options)
2. Exclusion criteria based on "How sure are you?" response (2)
3. Cycle day assessment (3)
4. Fertility assessment (4)
5. Relationship status assessment (3)

168 possibilities (after excluding some contradictory combinations)

The multiverse, the statistical significance filter, and the feedback loop



Activity

Design analysis for an experiment

1. Set a distribution for the pre-test measurement x
2. Assume 50/50 randomized assignment for the treatment, z
3. Specify values for the coefficients in the model
   $\mathrm{E}(y) = \exp(\beta_0 + \beta_1 x + \beta_2 z)$
4. Specify a reciprocal overdispersion parameter ϕ for the negative binomial regression
5. Choose a provisional sample size n
6. Simulate x, z, and y (in that order)
7. Plot the data and see if they make sense
8. Fit the negative binomial regression and look at the parameter estimates and uncertainties
9. Look at the standard error for the treatment effect and alter the sample size accordingly
10. Loop it: perform 100 simulations, save the results, and look at the distribution of estimates and uncertainties

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 5b

Story

Lucky golf balls and implausible effect sizes

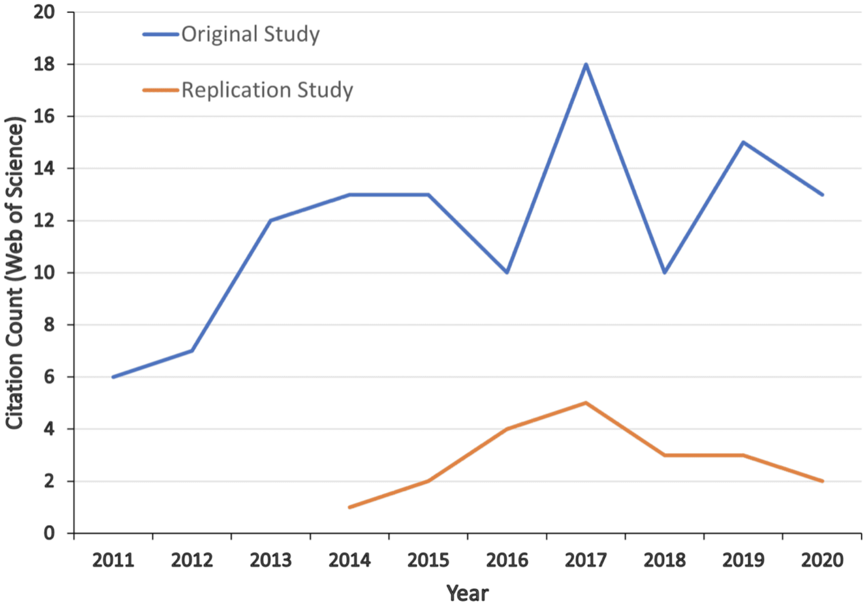
From a published paper: "We recruited 28 university students (12 males, 16 females) as participants and randomly assigned them to a superstition-activated or a control condition. Participants were asked to engage in a 10-trial putting task. A pretest revealed that more than 80% of our participant population 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 calculation for averages

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

Discussion problem

Implications and potential escapes from the trap of designing future studies based on past "statistically significant" estimates

Consider a world in which researchers report "statistically significant" estimates and then use these to design future studies. Work out the implications of this decision. Is there a way for the self-correcting nature of science to escape from this trap?

Class 6a

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

Generalizing from the class to a larger population

Discuss reading and homework

Computer demonstration

Drill

Poststratification and weighted averages

Calculate population-level averages from the information about group-level averages as well as the groups' sizes.

Discussion problem

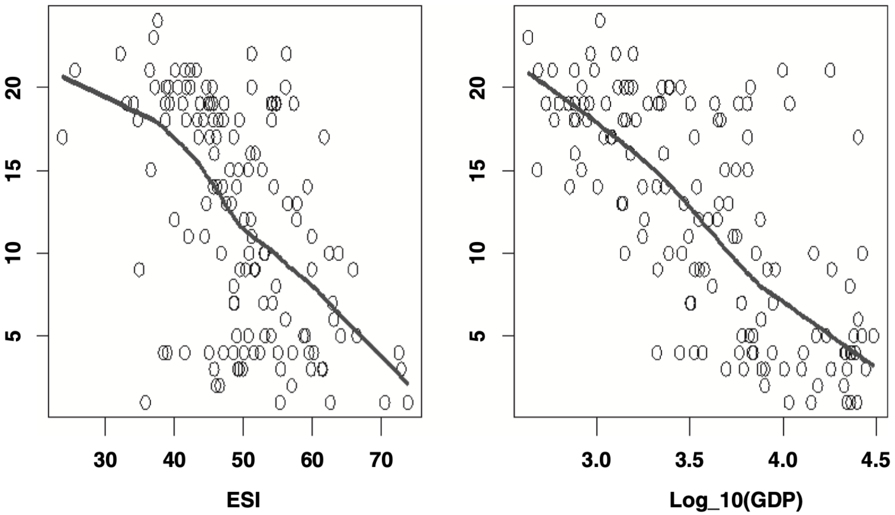
Network sampling

- 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.

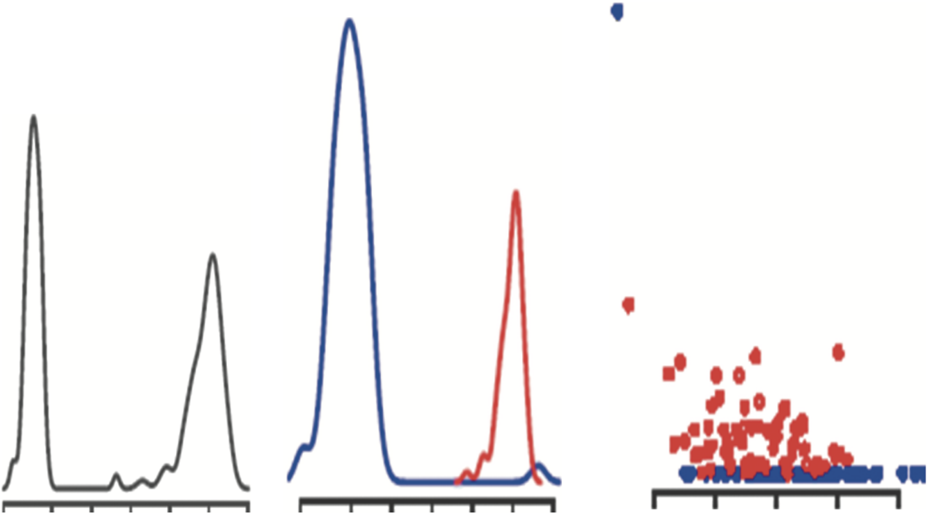
Class 6b

Story

Challenges with imputations of the Environmental Sustainability Index



Challenges with imputations of the Environmental Sustainability Index



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

Missing data imputation problems

In groups of 4:

1. Come up with a problem you care about involving missing data.
2. Missingness can come from nonresponse, censoring, incomplete measurement, ...

Class 7a

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 7b

Story

Effects of ballot order on voting

From BBC News in 2017: "One of the world's leading political scientists believes Donald Trump most likely won the US presidential election for a very simple reason . . . his name came first on the ballot in some critical swing states. . . .
'There is a human tendency to lean towards the first name listed on the ballot,' says Krosnick, a politics professor at Stanford University. 'And that has caused increases on average of about three percentage points for candidates, across lots of races and states and years.' . . . When an election is very close the effect can be decisive, Krosnick says—and in some US states, such as Pennsylvania, Michigan and Wisconsin, the 2016 election was very close."

Activity

Potential outcomes and randomization

Potential outcomes, y:

- y^1: Republican vote share in congressional election, if $z = 1$
- y^0: Republican vote share in congressional election, if $z = 0$

Treatment, z:

- $z = 1$: Democrat is listed first on the ballot
- $z = 0$: Republican is listed first on the ballot

Pre-treatment predictor, x:

- Republican vote share in most recent contested congressional election

Discuss reading and homework

Computer demonstration

Drill

Average treatment effects with uncertainty

```
fit <- stan_glm(post_test ~ z + pre_test + z:pre_test, data=
```

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

Give R code to compute the estimate and standard error of different average treatment effects:

Discussion problem

Assumptions in randomized experiments

Take a particular example of a randomized experiment in an area of interest to you—this could be a real experiment that has been done, or a hypothetical experiment on some treatment or exposure—and consider each of the following issues:

- Ignorability
- Efficiency
- Stable unit treatment value assumption
- External validity
- Internal validity
- Missing data
- Noncompliance

Class 8a

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

Regression to adjust for pre-treatment characteristics

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 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

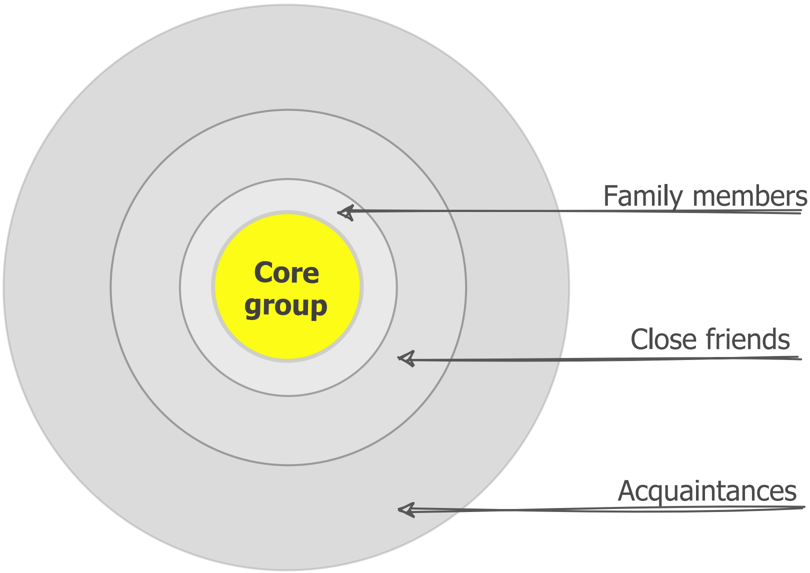
Causal inference using logistic regression for a binary outcome

We have focused on using linear models to estimate causal effects, starting with the additive model, $y = \beta_0 + \beta_1 x + \beta_2 z + \text{error}$, and continuing to the interaction model, $y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + \text{error}$, with pre-treatment predictor x, treatment indicator z, and outcome measurement y. But what if your outcome is binary? Then it would seem natural to model y given x and z using logistic regression. How would it work to estimate the treatment effect in this way? What is the estimated causal effect?

Class 8b

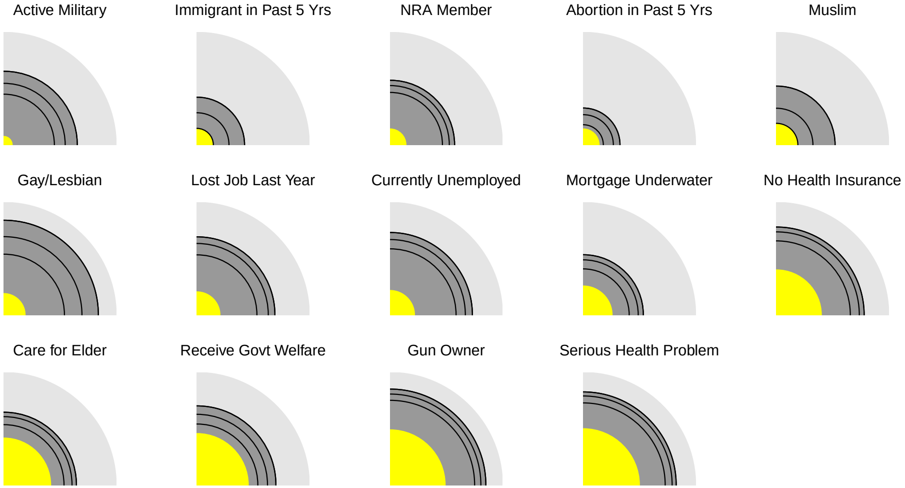
Story

Social penumbras and causal effects

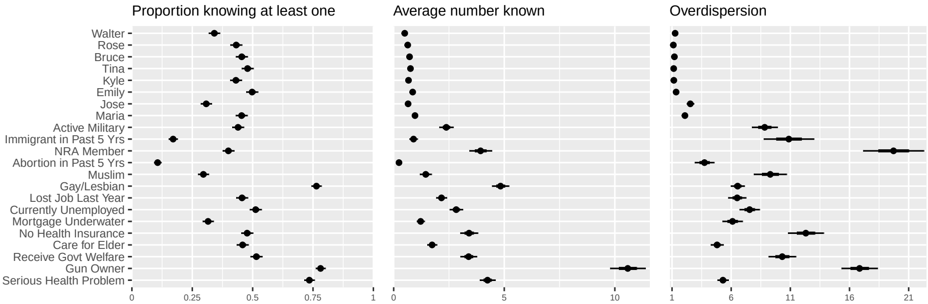


Social penumbras and causal effects

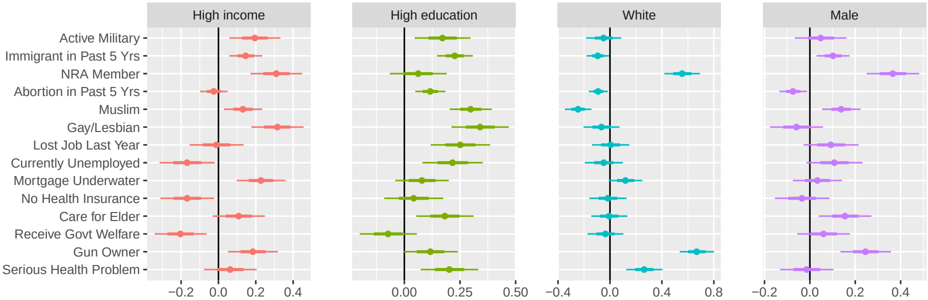
Group size (area of yellow circle) and penumbras (areas of black circles indicate close family, close family + close friends, and close family + close friends + others known) compared to the population of American adults (light gray circle)



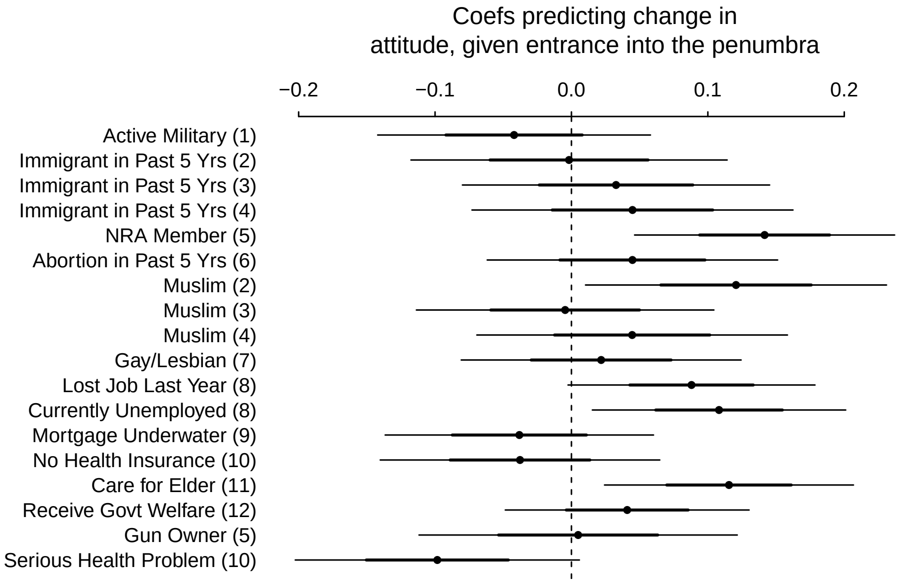
Social penumbras and causal effects



Social penumbras and causal effects



Social penumbras and causal effects

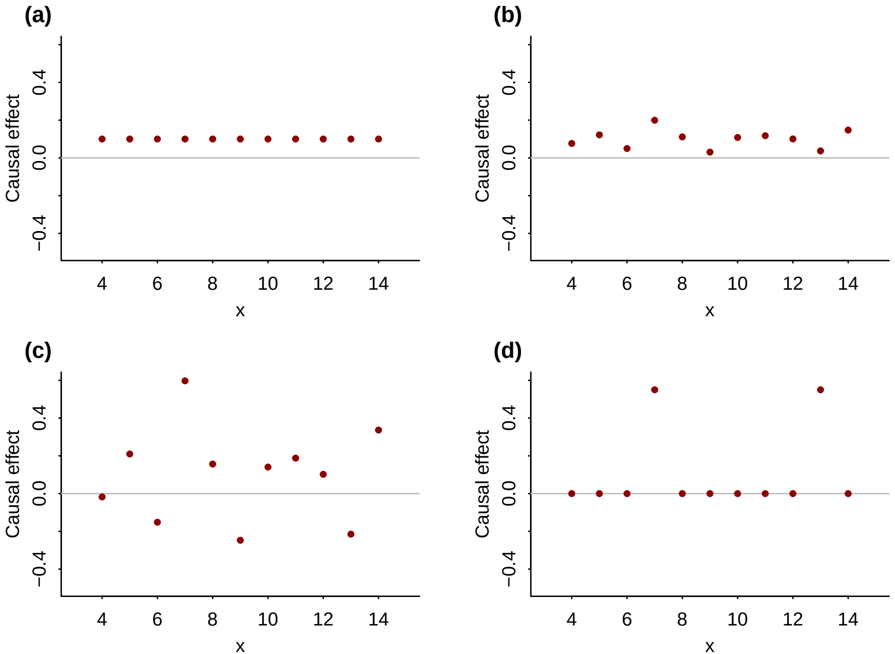


Activity

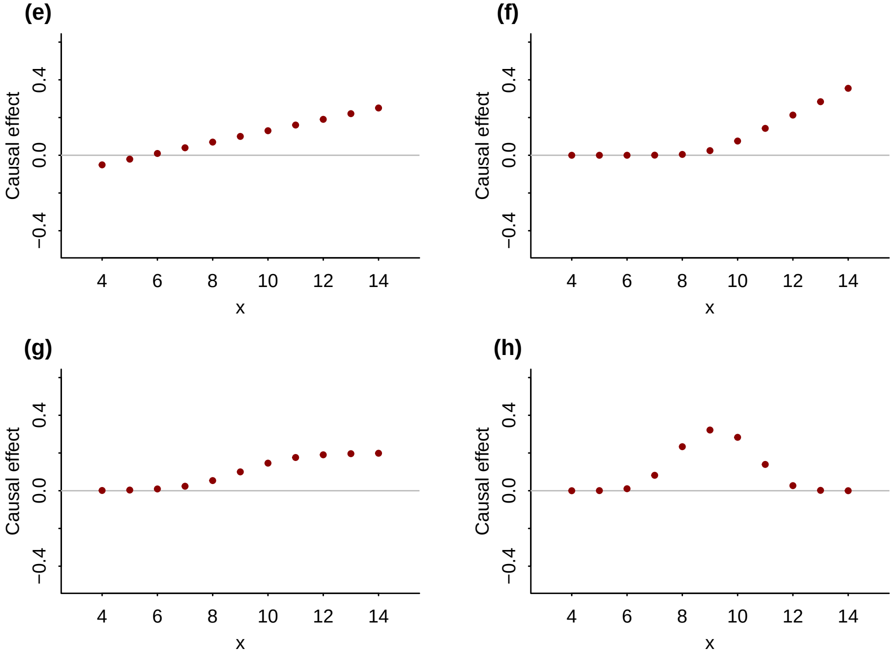
Understand the "average treatment effect"

- 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"



Discuss reading and homework

Computer demonstration

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
- is a binary variable and you want to compare $= 0$ to $= 1$
- 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 that statement?

Class 9a

Story

Apparent null effects in a study of heart stents

"From a news article: Heart Stents Fail to Ease Chest Pain . . . When the researchers tested the patients six weeks later, both groups said they had less chest pain, and they did better than before on treadmill tests. But there was no real difference between the patients, the researchers found. Those who got the sham procedure did just as well as those who got stents. . . . 'It was impressive how negative it was,' Dr. Redberg said of the new study . . ."

Apparent null effects in a study of heart stents

Measurement	Treatment			Control			Comparison
	Pre	Post	Gain	Pre	Post	Gain	
	$\bar{y}$ (sd)	$\bar{y}$ (sd)	diff (ci)	$\bar{y}$ (sd)	$\bar{y}$ (sd)	diff (ci)	est (ci)
Exercise time (seconds)	528.0 (178.7)	556.3 (178.7)	28.4 (11.6, 45.1)	490.0 (195.0)	501.8 (190.9)	11.8 (−7.8, 31.3)	16.6 (−8.9, 42.0)
Peak oxygen uptake (mL/min)	1715.0 (638.1)	1713.0 (583.7)	−2.0 (−54.1, 50.1)	1707.4 (567.0)	1718.3 (550.4)	10.9 (−47.2, 69.0)	−12.9 (−90.2, 64.3)
SAQ-physical limitation	71.3 (22.5)	78.6 (24.0)	7.4 (3.5, 11.3)	69.1 (24.7)	74.1 (24.7)	5.0 (0.5, 9.5)	2.4 (−3.5, 8.3)
SAQ-angina frequency	79.0 (25.5)	93.0 (26.8)	14.0 (9.0, 18.9)	75.0 (31.4)	84.6 (27.7)	9.6 (3.6, 15.5)	4.4 (−3.3.12.0)
SAQ-angina stability	64.7 (25.5)	60.5 (23.7)	−4.2 (−10.7, 2.4)	68.5 (24.3)	63.5 (25.6)	−5.1 (−11.7, 1.6)	0.9 (−8.4, 10.2)
EQ-5D-5L QOL	0.80 (0.21)	0.83 (0.21)	0.03 (0.00, 0.06)	0.79 (0.22)	0.82 (0.20)	0.03 (0.00, 0.07)	0.00 (−0.04, 0.04)
Peak stress wall motion index score	1.11 (0.18)	1.03 (0.06)	−0.08 (−0.11, −0.04)	1.11 (0.18)	1.13 (0.19)	0.02 (0.03, 0.06)	−0.09 (−0.15, −0.04)
Duke treadmill score	4.24 (4.82)	5.46 (4.79)	1.22 (0.37, 2.07)	4.18 (4.65)	4.28 (4.98)	0.10 (−0.99, 1.19)	1.12 (−0.232.47)

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

Experimental design

For each hypothetical experiment, state whether it is a randomized block experiment, a matched pair experiment, a cluster-randomized experiment, or a simple randomized experiment.

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 this claim?

Class 9b

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

Role play: study makers vs. study breakers

In groups of 4:

- 2 "study makers": come up with an idea for an observational study, including plans for gathering data and estimating the treatment effect of interest.
- 2 "study breakers," raise concerns about the proposed idea.

Have a discussion, not a debate!

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

Meta-analysis of nudge experiments

A published claim: "Over the past decade, choice architecture interventions or so-called nudges have received widespread attention from both researchers and policy makers. ... Drawing on more than 200 studies ... we present a comprehensive analysis of the effectiveness of choice architecture interventions .... Our results show that choice architecture interventions overall promote behavior change with a small to medium effect size of Cohen's d = 0.45 (95% CI [0.39, 0.52]).... Food choices are particularly responsive to choice architecture interventions, with effect sizes up to 2.5 times larger than those in other behavioral domains.... Our analysis further reveals a moderate publication bias toward positive results in the literature...."

Class 10a

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

Causal challenges in estimating the effects of campaign contributions on politicians' behaviors

Class 10b

Story

U.S. Postal Service consulting: What are inferences used for?

Goals:

1. Estimate the volume of all sorts of mail (first-class mail of different weights, second-class mail, packages, etc.) handled by the Postal Service, along with the costs of processing these.
2. Assess the sources of uncertainty in these estimates.

Activity

Observational study on students: Crime victimization and policy views

- Pre-treatment predictors x: Background variables
- Treatment (or exposure) z: Victim of a crime
- Outcome y: Tough-on-crime attitudes

Discuss reading and homework

Computer demonstration

Drill

Imbalance and lack of complete overlap

In the following cases of observational studies, discuss possible concerns about imbalance or lack of complete overlap in the following examples.

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 11a

Story

Deterrent effect of the death penalty

From an article from 2006: "No clear correlation between homicides and executions emerges from this long time series. In the first decade of the twentieth century, execution and homicide rates seemed roughly uncorrelated, followed by a decade of divergence as executions fell sharply and homicides trended up. Then for the next forty years, execution and homicide rates again tended to move together—first rising together during the 1920s and 1930s, and then falling together in the 1940s and 1950s. As the death penalty fell into disuse in the 1960s, the homicide rate rose sharply. The death penalty moratorium that began . . . in 1972 and ended . . . in 1976 appears to have been a period in which the homicide rate rose. The homicide rate then remained high and variable through the 1980s while the rate of executions rose. Finally, homicides dropped dramatically during the 1990s."

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:

1. Ignorability of the instrument
2. Monotonicity and nonzero association between instrument and treatment variable
3. Exclusion restriction.

Discussion problem

Estimating the effects of masks and social distancing

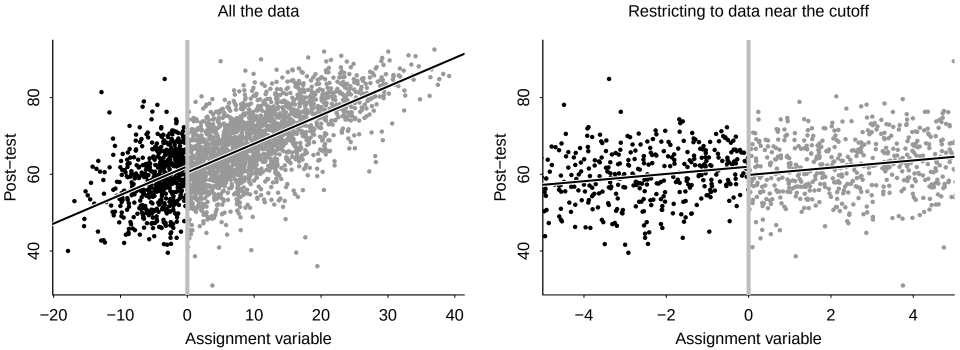
During the covid epidemic, people were not assigned at random to wear masks or to practice social distancing, but they were indirectly affected by national, state, and local policies mandating these actions.

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.

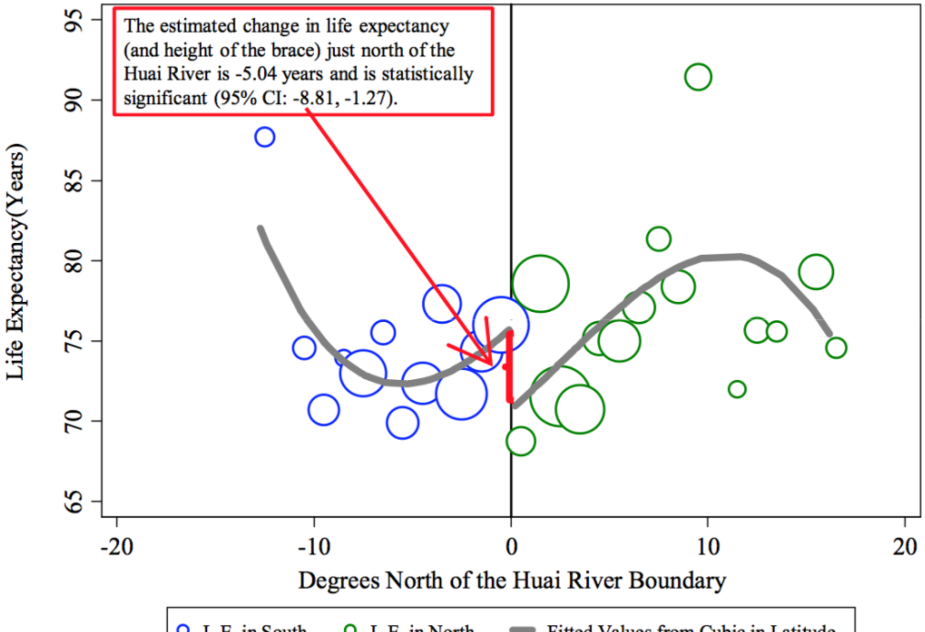
Class 11b

Story

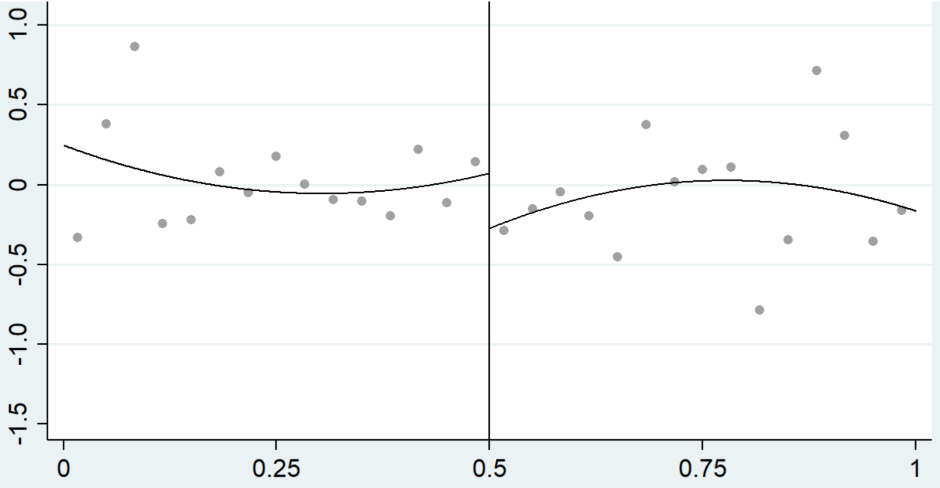
Regression discontinuity mishaps



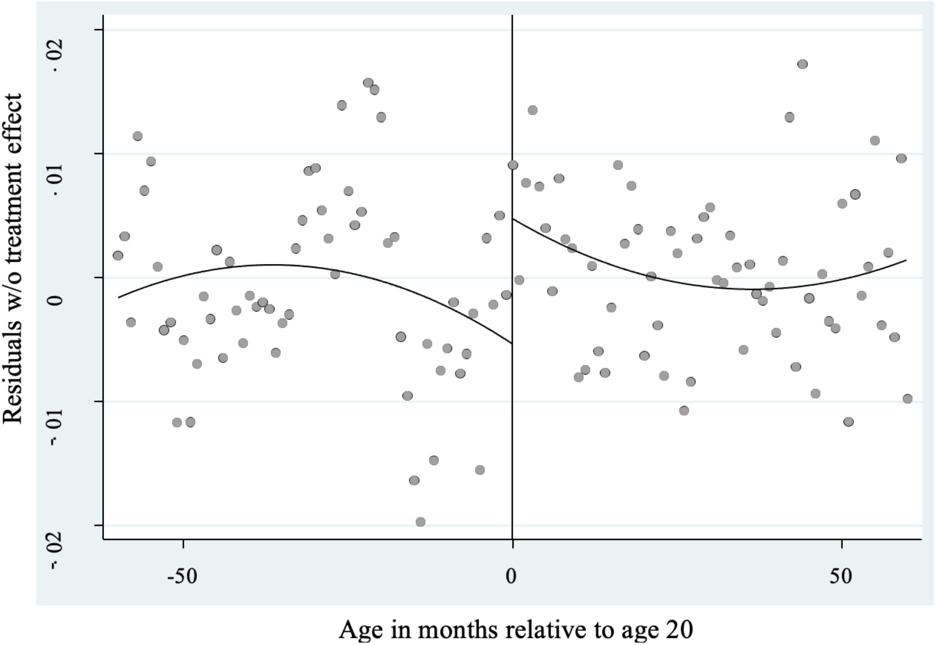
Regression discontinuity mishaps



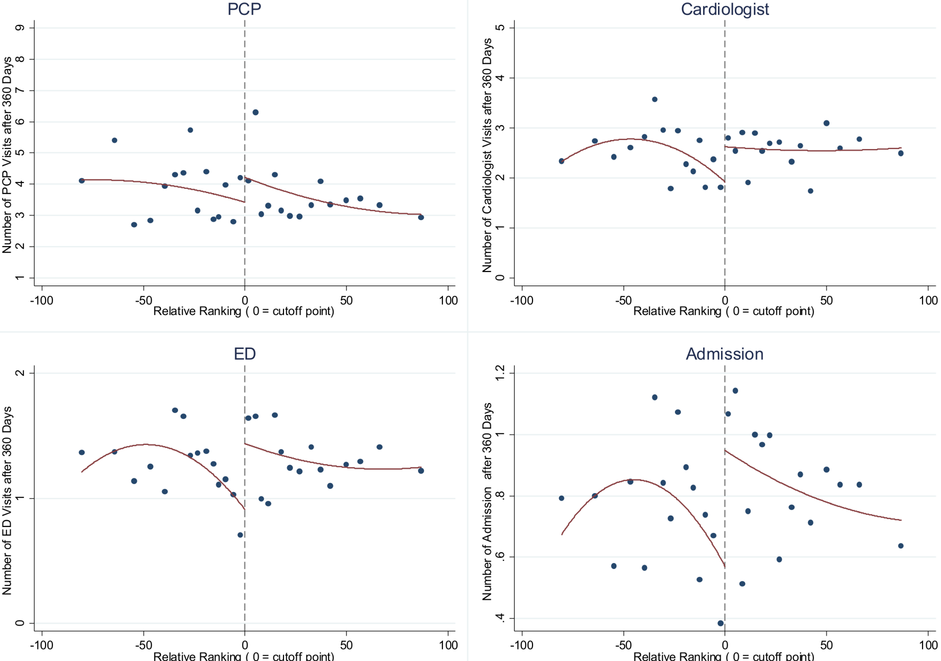
Regression discontinuity mishaps



Regression discontinuity mishaps

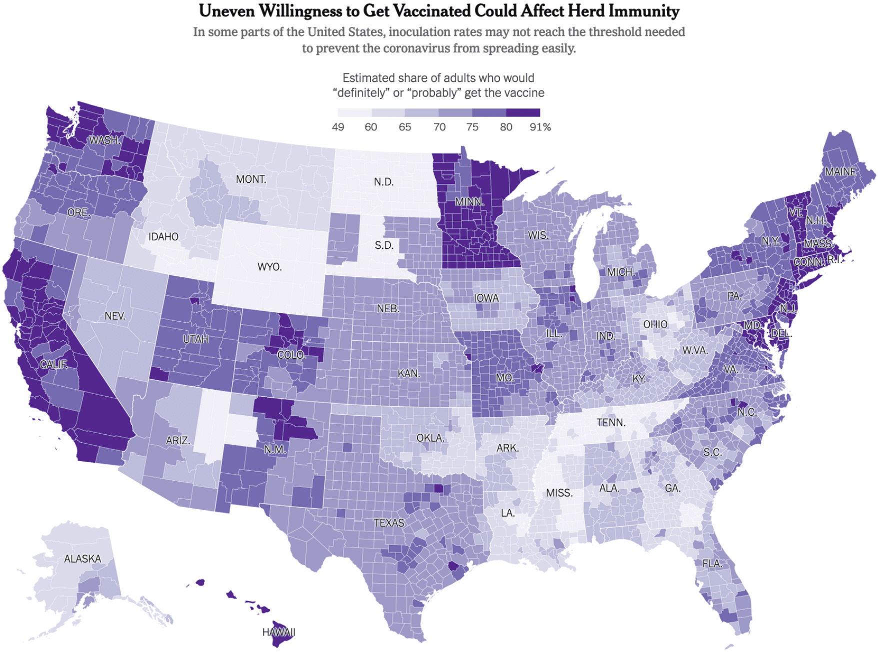


Regression discontinuity mishaps



Activity

"Why" questions and causal inference



Discuss reading and homework

Computer demonstration

Drill

Regression discontinuity

For each of these examples of a discontinuity design, which would plausibly allow you to estimate a causal effect? What variables would you want to adjust for?

Discussion problem

Coaching for college admissions tests

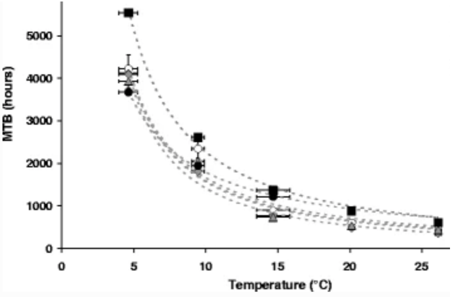
Kaplan Works	According to a recent study, Kaplan's SAT students have an average score improvement of 120 points, from actual PSAT to actual SAT. 28 percent of students surveyed improved by at least 170 points! Come to Kaplan and get a higher score!**

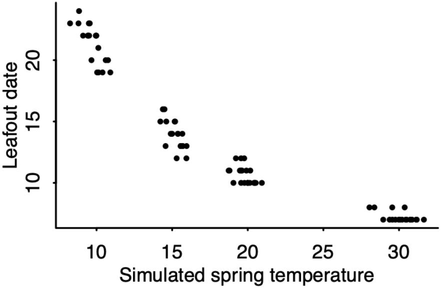
Higher Scores Guaranteed Our average score improvement for the SAT is **140 points**. If you complete our course, take the SAT, and do not improve your total score by at least 100 points over your prior PSAT or SAT, we'll work with you again for FREE.

Class 12a

Story

Nonlinear patterns in leafout dates





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

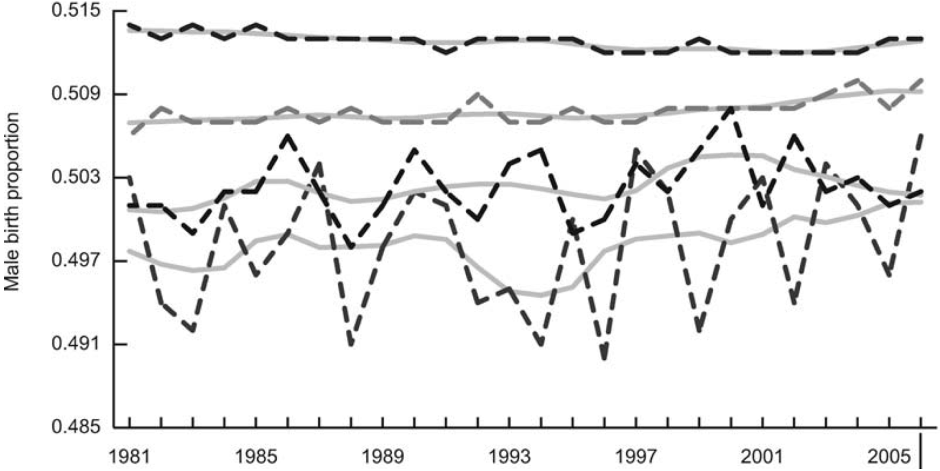
Drill

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

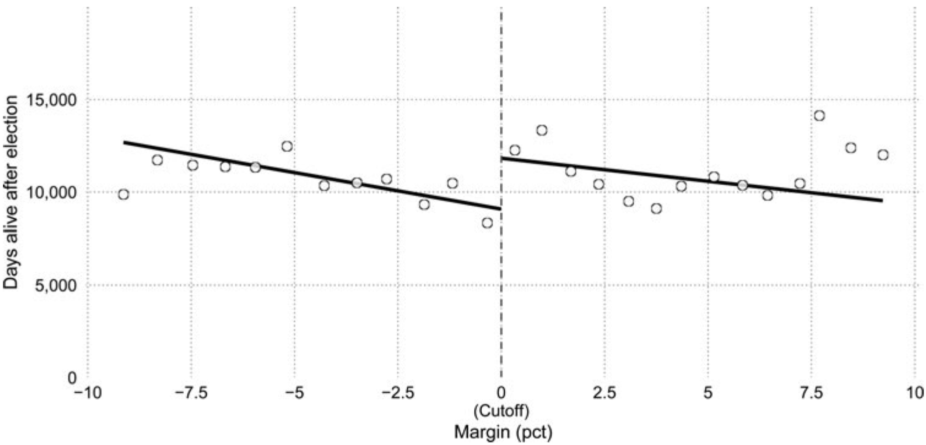
Noisy time series



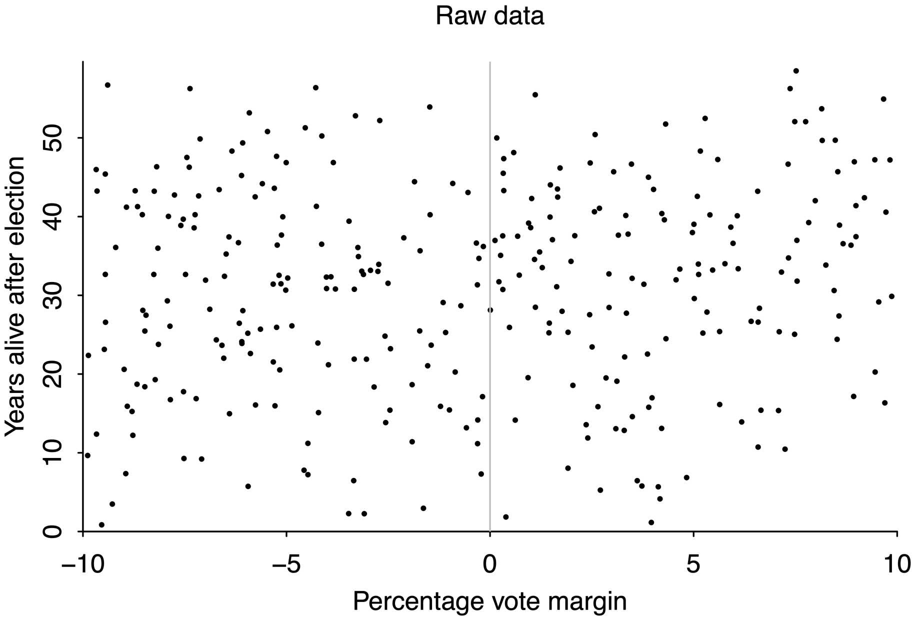
Class 12b

Story

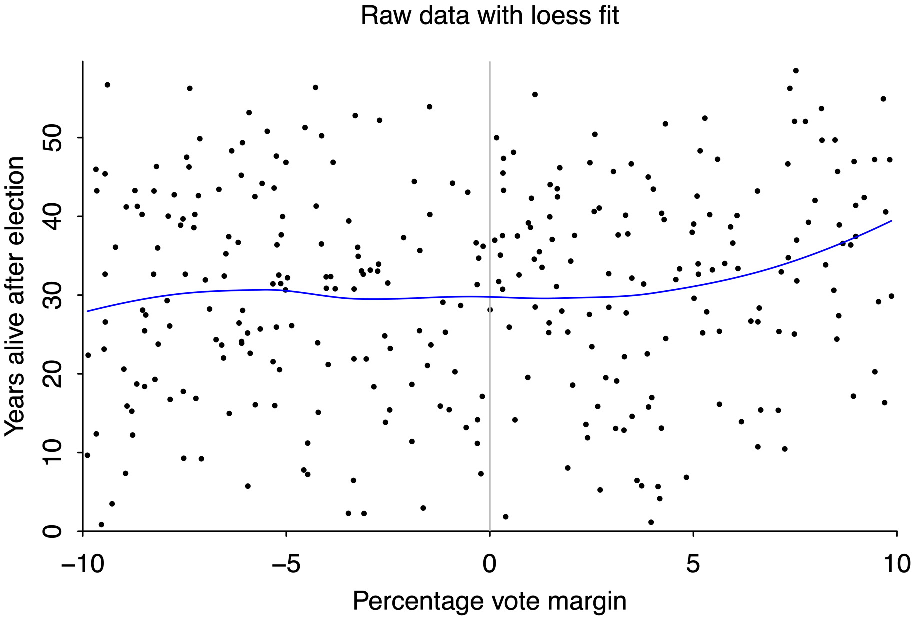
Nonlinear modeling for data exploration



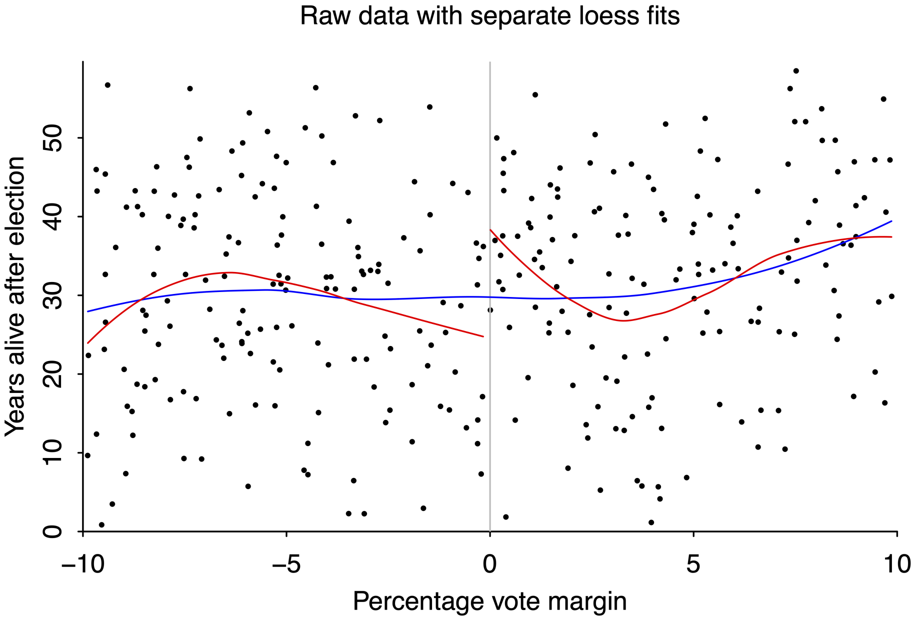
Nonlinear modeling for data exploration



Nonlinear modeling for data exploration

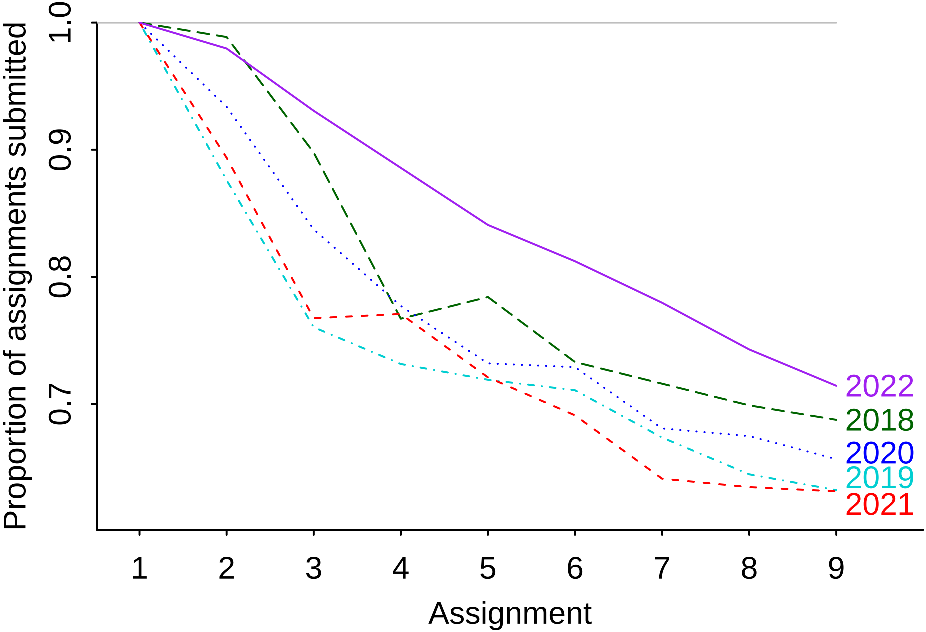


Nonlinear modeling for data exploration



Activity

When do students quit a class?



Discuss reading and homework

Computer demonstration

Drill

Error terms for nonlinear models

Using pen on paper, sketch scatterplots of each of the following models, plotting roughly 100 points with x uniformly distributed between 0 and 20.

Discussion problem

Regression with 21 data points and 16 predictors

- How to think about a regression with 21 data points and 16 predictors?
- Seems hopeless, but ... it can't be worse than having no data at all!

Class 13a

Story

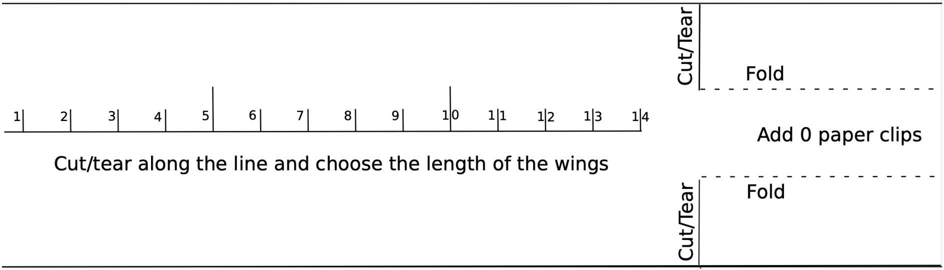
The rise and fall and rise of randomized controlled trials in international development

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. . . .

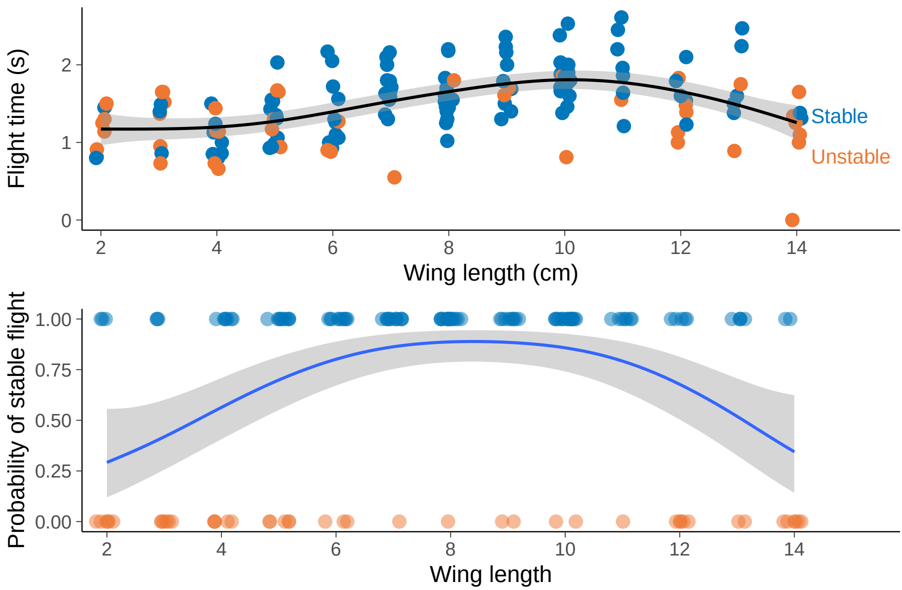
Instead of asking, 'why are RCTs increasing now?' we ask, 'why didn't RCTs spread to the same extent in the 1970s, and why were they discontinued?' In other words, how we explain the success of the second wave must be consistent with how we explain the failure of the first."

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.

Basic statistics and linear regression

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 an R function to compute the average and standard deviation of 1000 random draws from a Poisson distribution with parameter θ.

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

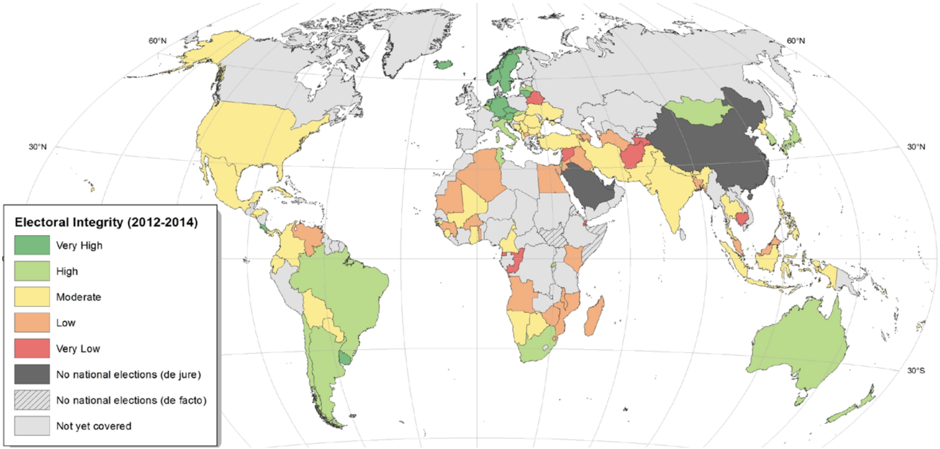
Suppose you want to design an experiment to estimate the effects of canvassing on voter turnout in an upcoming election. You have records on a large number of registered voters, with data on their past voter turnout, and your plan is to randomly select n people from this database and randomly chose $n/2$ 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

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?

From logistic regression through causal inference

Give R code for fitting an ordered logistic regression predicting an outcome y that can take on the values 1,2,3,4,5 from predictors x_1 and x_2.

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.