Teaching Introductory Statistics: Ask Good Questions

Allan J. Rossman
Dept of Statistics, Cal Poly – San Luis Obispo
Email: arossman@calpoly.edu
Twitter: @allanjrossman
Blog: https://askgoodquestions.blog

Frank Sinatra

- What's the secret to being a successful singer?
  - "Sing Good Songs"

My similarly succinct suggestion

- What's the secret to being an effective teacher?
  - "Ask Good Questions"

I'm tempted to stop right now

- My take-home message
  - Ask Good Questions
- Quiz at end of presentation
  - Q: What was my point?
  - A: Ask Good Questions
- I have started writing a weekly blog
  - Name: Ask Good Questions
  - https://askgoodquestions.blog

What kinds/purposes of questions?

- Guide students to develop their understanding and skills
  - Formative assessment
  - Learning activities
- Assess what students have learned
  - Summative assessment
  - Quiz/exam questions

Examples of questions for teaching …

- Statistical thinking
- Descriptive statistics
- Sampling and experimentation
- Probability
- Simulation-based inference
- Confidence intervals
- Hypothesis tests

Statistical thinking

- Example: Sex discrimination?

	Men	Women
Accepted	533	113
Denied	665	336
Total	1198	449

- Men: 533/1198 ≈ .445 were accepted
- Women: 113/449 ≈ .252 were accepted
- Does this provide evidence of discrimination against women?

Statistical thinking

	Men		Women	
	Accepted	Denied	Accepted	Denied
Program A	511	314	89	19
Program F	22	351	24	317
Total	533	665	113	336

- Program A
  - Men: 511/825 ≈ .619
  - Women: 89/108 ≈ .824
- Program F:
  - Men: 22/373 ≈ .059
  - Women: 24/341 ≈ .070

Statistical thinking

- Describe and explain the oddity, using the data given
- Most men apply to the easy program to get in; most women apply to the hard program to get in
- Multivariable thinking
  - Simpson's paradox

Statistical thinking

Follow-up assessment:

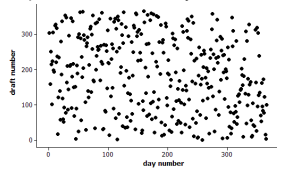
- Two softball players Amy (A) and Barb (B)
- Create an example in which
  - A has a higher proportion of hits than B in June
  - A has a higher proportion of hits than B in July
  - B has a higher proportion of hits than A when June and July are combined

Statistical thinking

- June: Amy gets 9 hits in 10 attempts (90%), Barb gets 80 hits in 100 attempts (80%)
- July: Amy gets 20 hits in 100 attempts (20%), Barb gets 1 hit in 10 attempts (10%)
- Combined: Amy gets 29 hits in 110 attempts (26.4%), Barb gets 81 hits in 110 attempts (73.6%)

Statistical thinking

- Example: 1970 Draft Lottery



- Any reason to doubt randomness?

Statistical thinking

- Calculate median draft number for each month

Statistical thinking

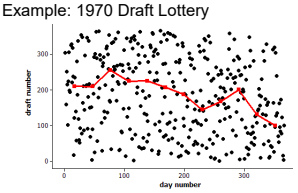
- Calculate median draft number for each month

Jan 211	Jul 188
Feb 201	Aug 145
Mar 256	Sep 168
Apr 225	Oct 201
May 226	Nov 131.5
Jun 207.5	Dec 100

- Do you see a pattern/trend?

Statistical thinking

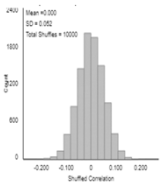
- Example: 1970 Draft Lottery



- Any reason to doubt randomness?

Statistical thinking

- How often would such an extreme outcome occur with a truly random lottery?
- What statistic might you use to determine "extreme"-ness?
  - Correlation ≈ -0.226



Undergraduate statistics enrollments

- 2015 CBMS Survey

TABLE S.2 Total enrollment (in 1000s), including distance-learning enrollment, by course level in undergraduate mathematics, statistics, and computer science courses taught in mathematics and statistics departments at four-year colleges and universities, and in mathematics programs at two-year colleges in fall 2000, 2005, 2010, and 2015.

	Mathematics Departments				Statistics Departments				Two-Year College Mathematics Programs			
Course level	2000	2005	2010	2015	2000	2005	2010	2015	2000	2005	2010	2015
Probability and Statistics courses												
Introductory level	136	148	231	253	54	54	81	94	74	117	137	280
Upper level	35	34	32	60	20	24	27	50	0	0	0	0

- Look at this growth! Especially at TYCs

Descriptive statistics

- I suspect that when I moved from PA to CA, the average IQ dropped in both states!
  - Is this possible?
  - What would have to be true (in principle) for this to happen?
- Yes, if (my IQ > average IQ in PA) and (my IQ < average IQ in CA)

Descriptive statistics

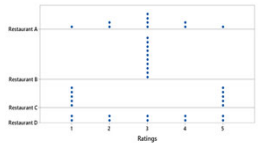
- Consider two new statistics:
  - Midrange = (maximum + minimum) / 2
  - Midhinge = (lower quartile + upper quartile) / 2
- Measure of center or variability?
- Resistant to outliers or not?

Descriptive statistics

- Suppose that Abby records the ages of customers at The Avenue (on-campus snack bar) from 11am-2pm today, while Mary records ages of customers at McDonald's (near freeway).
- Who will have the larger standard deviation of customer ages: Abby or Mary? Explain.

Descriptive statistics

- Put the four restaurants in order from smallest SD of ratings to largest SD of ratings



Descriptive statistics

- Example (adapted from Jay Lehmann):

a) Which would be larger – the mean weight of 10 randomly selected people or the mean weight of 1000 randomly selected cats? Explain briefly.

b) Which would be larger – the standard deviation of the weights of 1000 randomly selected people or the standard deviation of the weights of 10 randomly selected cats? Explain briefly.

Descriptive statistics

- Many of my students think I said that larger sample size means smaller SD
- I really said that the SD **of the sample mean** (or sample proportion) gets smaller as sample size increases
- Advice: When you talk about SD, always emphasize SD **of what**

Descriptive statistics

- Exam scores have mean 70, SD 8
- Arturo's score is 75
- Bella's score is 1.5 SDs above Arturo's score.
- What is Bella's score on the exam? Show your work.

Descriptive statistics

- Suppose that every student in our class scored 5 points lower on the second exam than on the first exam.
- What would be the value of the correlation coefficient between (exam1 score) and (exam2 score)?
  - Options: -5, -1, -0.5, 0, 0.5, 1, 5

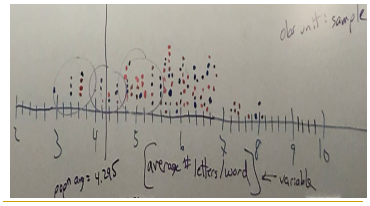
Sampling and experimentation

- Example: Gettysburg Address
- Select a sample of 10 words from the population of 268 words in the Gettysburg Address. (Just circle 10 words.)
- Record the length (# of letters) of each word.
- Calculate the average length for your sample.
- Produce graph of sample averages.

Sampling and experimentation

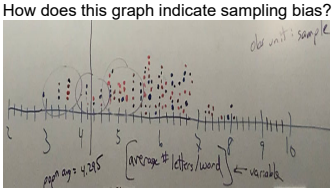
Is this a reasonable sampling method for estimating the average word length in the population?

Sampling and experimentation



Sampling and experimentation

- How does this graph indicate sampling bias?



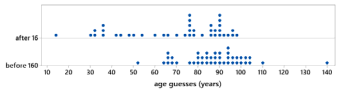
- Would closing eyes and pointing be unbiased?

Sampling and experimentation

- Example: Mandela's age
  - Group 1: Nelson Mandela was the first president of South Africa following apartheid.
    - Do you think he was older or younger than 16 years old when he died?
    - Make a guess for how old he was when he died.
  - Group 2: Nelson Mandela was the first president of South Africa following apartheid.
    - Do you think he was older or younger than 160 years old when he died?
    - Make a guess for how old he was when he died.

Sampling and experimentation

- Example: Mandela's age



- t = -4.17; p-value ≈ .0000
- Very strong evidence of anchoring effect

Sampling and experimentation

- Random sampling and random assignment are different tools with different purposes
- Random sampling
  - Goal: Select representative sample
  - Benefit: Generalize results
- Random assignment
  - Goal: Produce similar groups
  - Benefit: Draw cause-effect conclusions

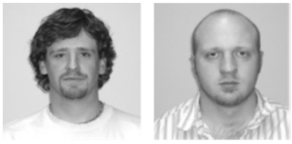
Probability

- 2018 General Social Survey
  - 47% had a pet dog
  - 25% had a pet cat
- Does it follow that 72% (which is 47% + 25%) had a pet dog or a pet cat?
  - What would be required for this to happen?

Simulation-based inference

- Example: Facial prototyping

Do people tend to associate names with faces?



Who is on the left: Bob or Tim?

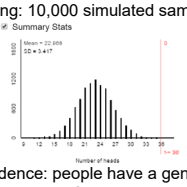
Simulation-based inference

Example: Facial prototyping (cont)

- 36 of 46 students put Tim on the left
  - What are two possible explanations for our observed sample result?
  - Which explanation can we investigate/model? How?
  - How often would such an extreme sample result occur by chance alone (if there were no facial prototyping)?
  - Have students flip coins to investigate

Simulation-based inference

- Facial prototyping: 10,000 simulated samples of 46 students



- Very strong evidence: people have a genuine tendency to put Tim on left

Confidence intervals

- Suppose that an alien lands on earth and sets out to estimate the proportion of human beings who are female
- The alien took a good statistics course on its home planet and knows to take a sample, produce a confidence interval
- Sample: the 2019 U.S. Senate, which has 25 women (the most ever!) and 75 men

Confidence intervals

- Calculate the alien's 95% CI
  - (.165 → .335)
- Interpret the CI for the alien
  - The alien is 95% confident that between 16.5% and 33.5% of all humans are female.
- Is this interval consistent with your experience as a long-time resident of this planet?
  - Duh!

Confidence intervals

- Is the problem that 5% of all 95% CIs fail to capture actual value of population parameter?
  - No!
- Then what went wrong???
  - Biased sampling method!
- If the alien was only interested in estimating the proportion of 2019 U.S. Senators who are female, would this 95% CI make sense?
  - No, exactly 25% of current senators are female!

Intervals and tests

- Survey of 47,000 U.S. households in 2006 found that 32.4% had a pet cat
- Does this provide ***very strong*** evidence that the population proportion with a cat is different from one-third?
- Does this provide strong evidence that the population proportion with a cat is ***very different*** from one-third?

Intervals and tests

- Does this provide ***very strong*** evidence that the population proportion with a cat is different from one-third?
  - Yes! Test stat $z \approx -4.29$, p-value ≈ .00002
- Does this provide strong evidence that the population proportion with a cat is ***very different*** from one-third?
  - No! 99.9% CI: (.317 → .331)

Intervals and tests

- Hypothesis test and confidence interval give *consistent* results
  - Value one-third is rejected, does not appear in CI
- Statistical *significance* is different from practical *importance*
  - Especially relevant with large sample sizes

My cats

Repeated questions

- Example: Anchoring (Mandela's age)

a) What are the observational units in this study?
b) What are the variables in this study? Which type is each variable? Which variable plays which role?
c) Did this study make use of random sampling, random assignment, both, or neither?
d) Is this an observational study or an experiment?

Repeated questions

- Example: Anchoring (Mandela's age)

e) Summarize your conclusion from the (approximate) p-value.
f) Estimate magnitude of effect with confidence interval.
g) Is it reasonable to draw a cause-and-effect conclusion? Explain why or why not.
h) Is it reasonable to generalize the results to all people? Explain why or why not.

A question for you

- In the last 20 years, the proportion of the world population living in extreme poverty has …?
  - Almost doubled
  - Remained more or less the same
  - Almost halved

Human progress

"Over the past twenty years, the proportion of the global population living in extreme poverty has halved. This is absolutely revolutionary. I consider it to be the most important change that has happened in the world in my lifetime." – Hans Rosling

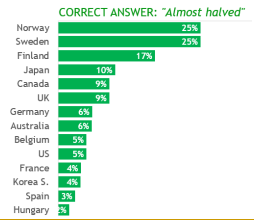
Human progress

"Here is a shocker: *The world has made spectacular progress in every single measure of human well-being.*

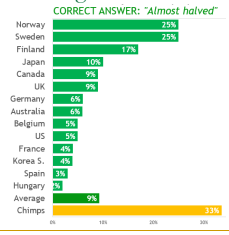
Here is a second shocker: *Almost no one knows about it.*"

– Steven Pinker

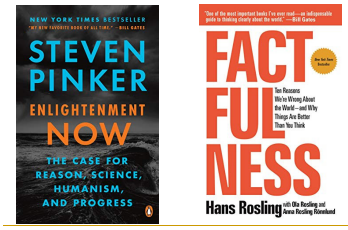
Ignorance about human progress



Worse than ignorance!



Human progress



Human progress (more on inference)

- U.S. survey: 5% of 1005 answered correctly
- Do the sample data provide strong evidence that less than one-third of all Americans would answer correctly?
- Test statistic: $z = \frac{.05-.3333}{\sqrt{\frac{.3333 \times .6667}{1005}}} \approx -19.1$ (!!!)
- Overwhelming evidence that Americans do worse than random guessing on this question

Human progress (more on inference)

- U.S. survey: 59% of 1005 answered "doubled"
- Estimate the proportion of all U.S. adults who would answer "doubled" with 95% confidence
  .59 ± .030: (.560, .620)
- Interpret this interval.
  - We are 95% confident that between 56% and 62% of all U.S. adults would give the most wrong answer to this question.

Some final questions for you

- Have you ever attended a presentation titled
  - Ask Bad Questions?
- Have you ever attended a presentation titled
  - Don't Ask Questions?
- Of course not! So, …
  - Why am I wasting your time with such obvious advice as "Ask Good Questions"??

Yeah, why am I wasting your time?

- I respectfully suggest that the next time you (or I) feel compelled to invest our valuable time, thought, energy, and creativity on …
  - Preparing crystal-clear lectures
  - Writing lucid paragraphs of exposition
  - Crafting beautiful presentations

What should we do?

What should we do?

- Resist this temptation!

So, what do I suggest instead?

- Instead we should invest these precious commodities (time, thought, energy, creativity) on
  - Developing engaging classroom activities
  - Preparing thought-provoking assignments
  - Crafting effective assessment items
- In other words, we should be sure to …

What's the take-home message?

Ask Good Questions!

Thanks very much!

- Please follow my weekly blog with many more examples and advice:

https://askgoodquestions.blog

- Please contact me with questions and comments:

arossman@calpoly.edu