Teaching Introductory Statistics: Ask Good Questions

Allan J. Rossman
Department 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"

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

Topics of questions in this presentation

- 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

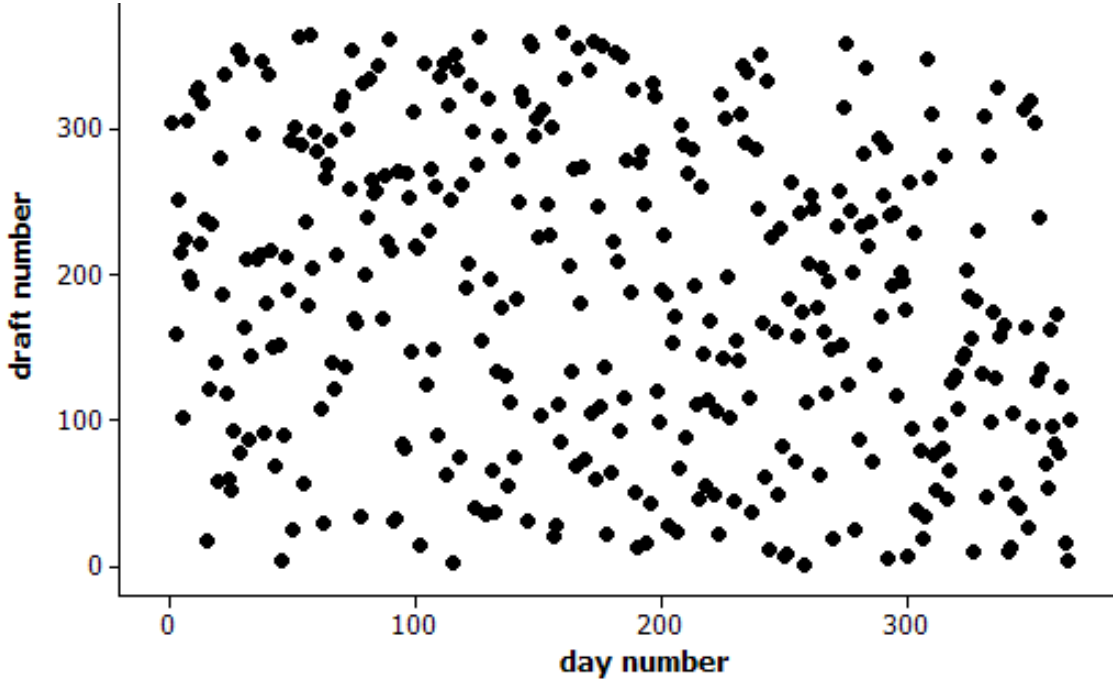
- Program A
  - Men: 511/825 ≈ .619
  - Women: 89/108 ≈ .824
- Program F:
  - Men: 22/373 ≈ .059
  - Women: 24/341 ≈ .070

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

- Example: 1970 Draft Lottery



- Any reason to doubt randomness?
- Calculate median draft number for each month

Statistical thinking

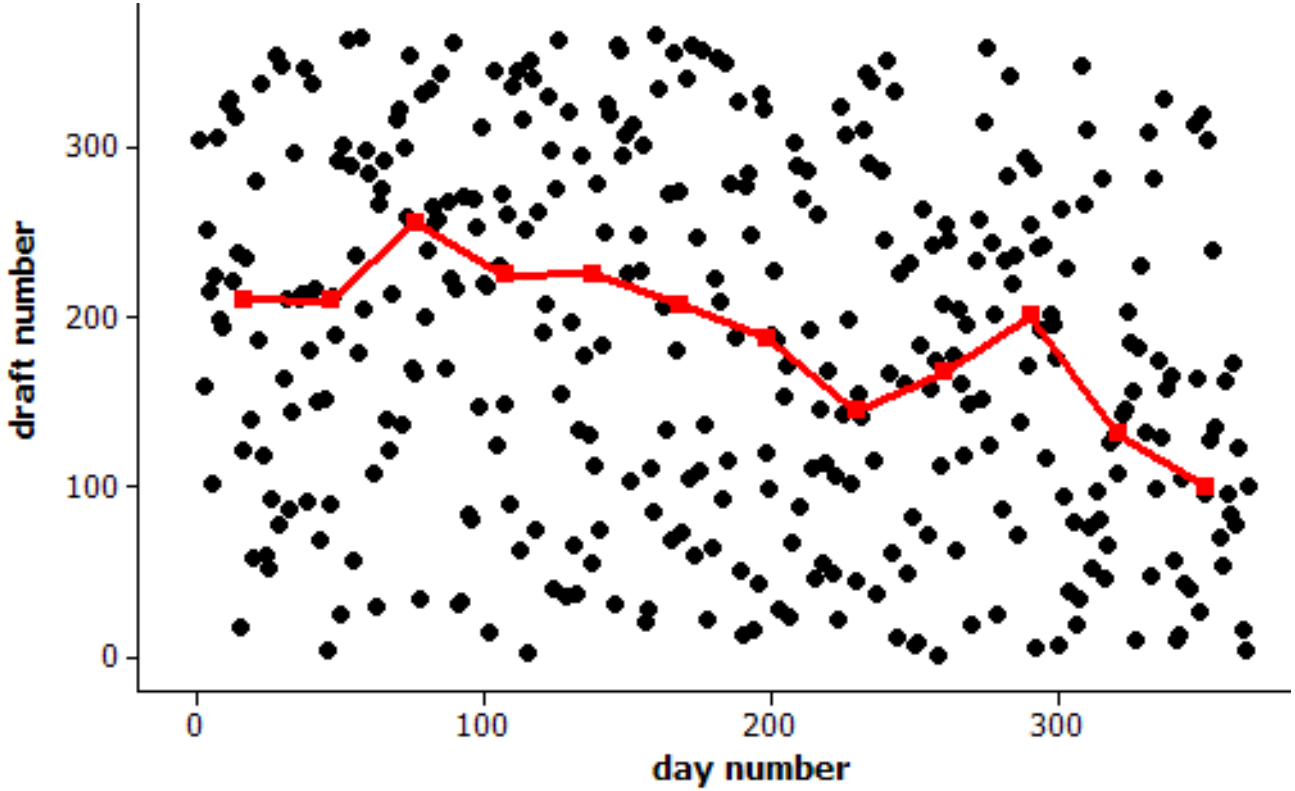
- Calculate median draft number for each month

Jan 211	Jul 188
Feb 210	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

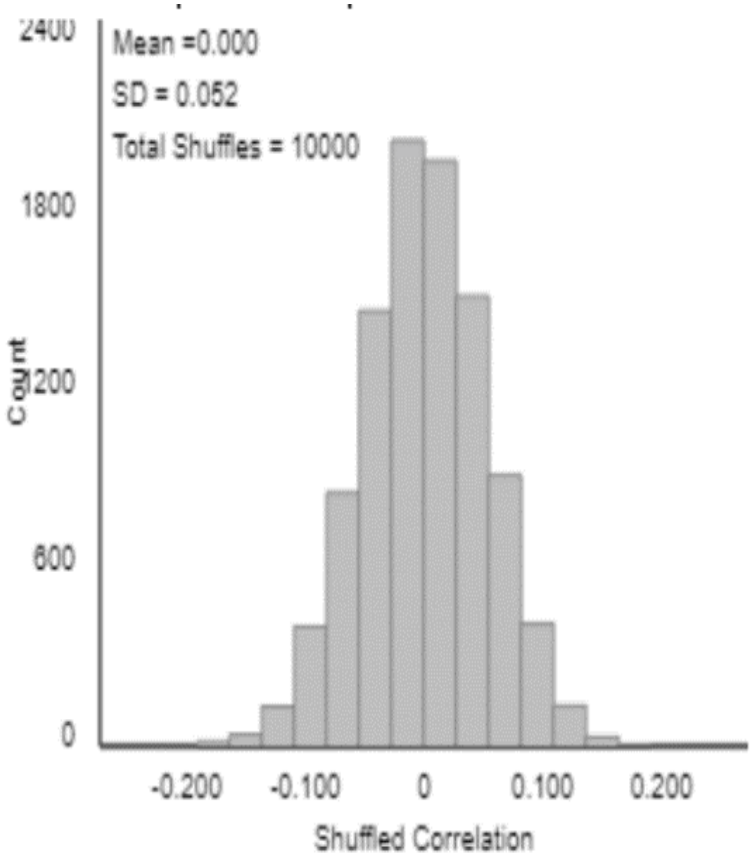
- How often would such an extreme outcome occur with a truly random lottery?

Statistical thinking

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

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



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)

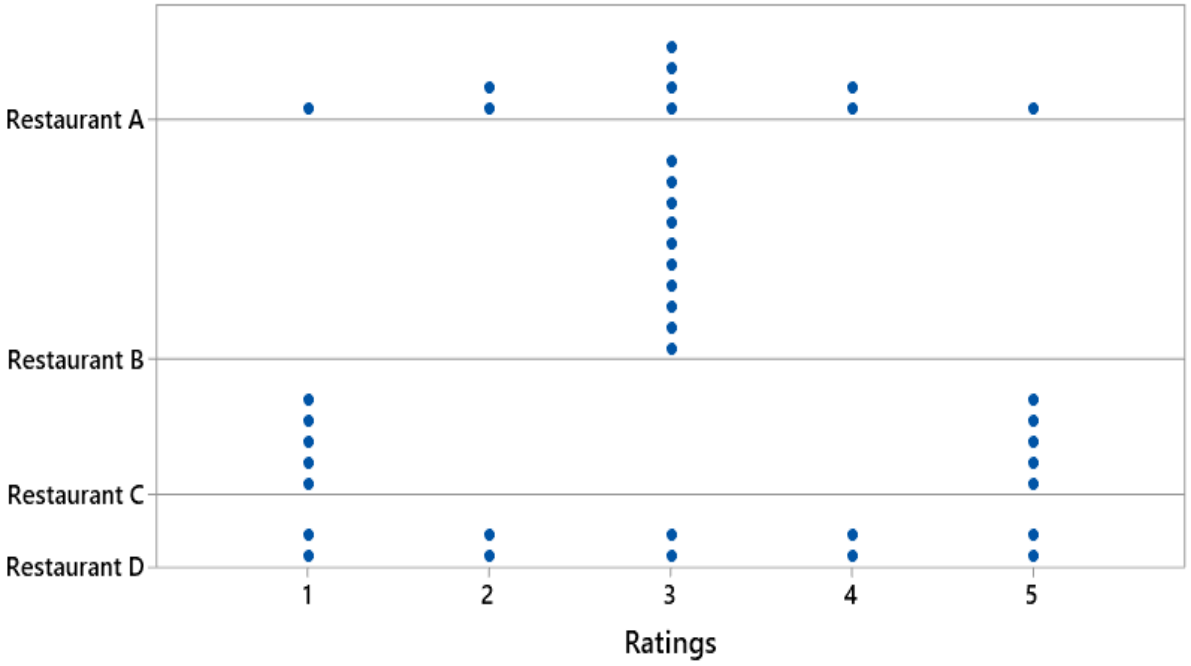
and (my IQ < average IQ in CA)

Descriptive statistics

- 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

- Arrange 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 that larger sample size produces smaller SD
- Do not realize that 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
- What is Bella's score on the exam?
  - Concept: Measuring distance as number of SDs away from *something*
  - Not necessarily away from *mean*

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

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.

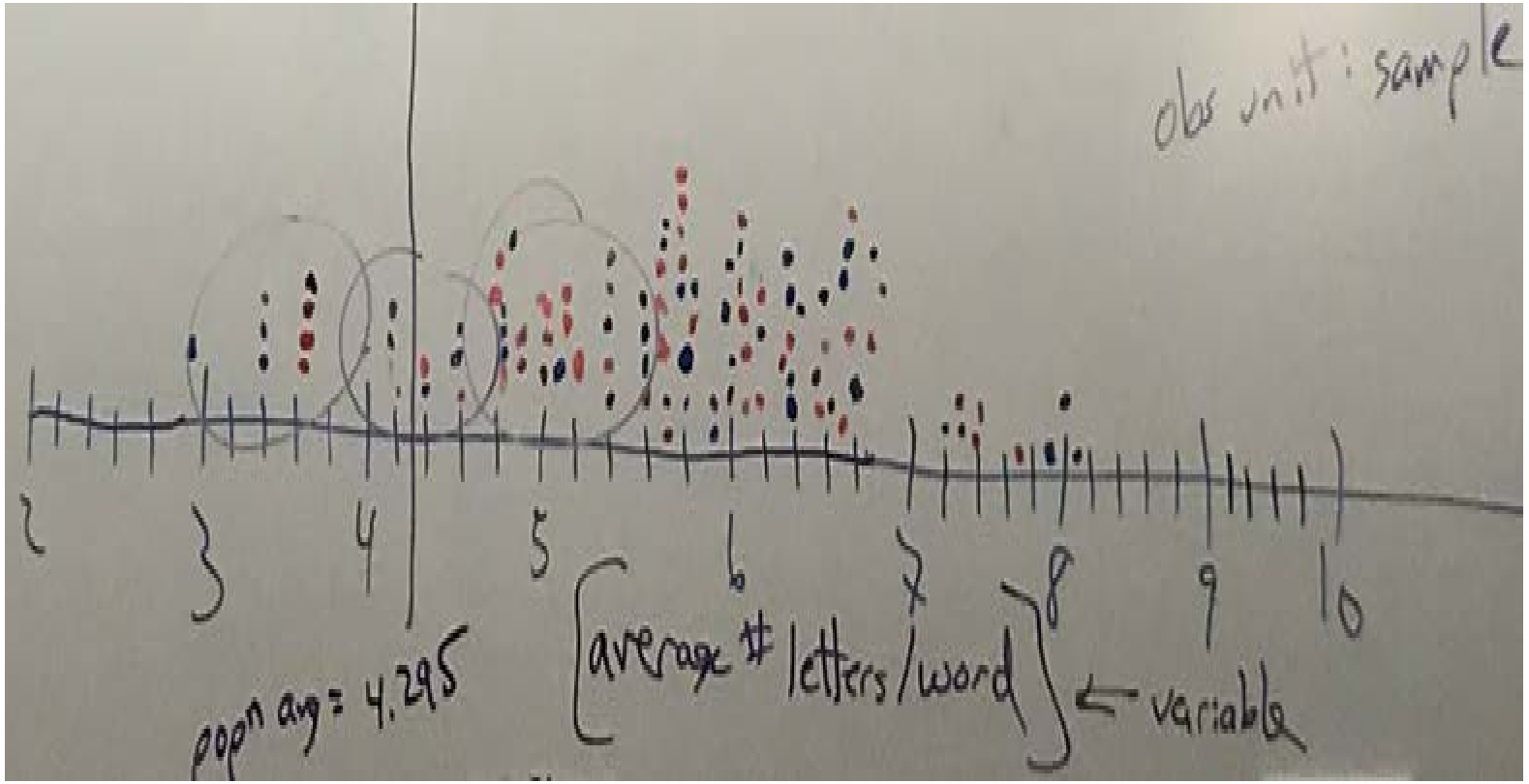
	Mathematics Departments				Statistics Departments				Two-Year College Mathematics Programs			
Course level	2000	2005	2010	2015	2000	2005	2010	2015	2000	2005	2010	2015
Probability and Statistics courses												
Introductory level	136	148	231	253	54	54	81	94	74	117	137	280
Upper level	35	34	32	60	20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

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 average length for your sample.
  - Produce graph of sample averages.
- Is this a reasonable sampling method for estimating average word length in population?

Sampling and experimentation

- How does this graph indicate sampling bias?



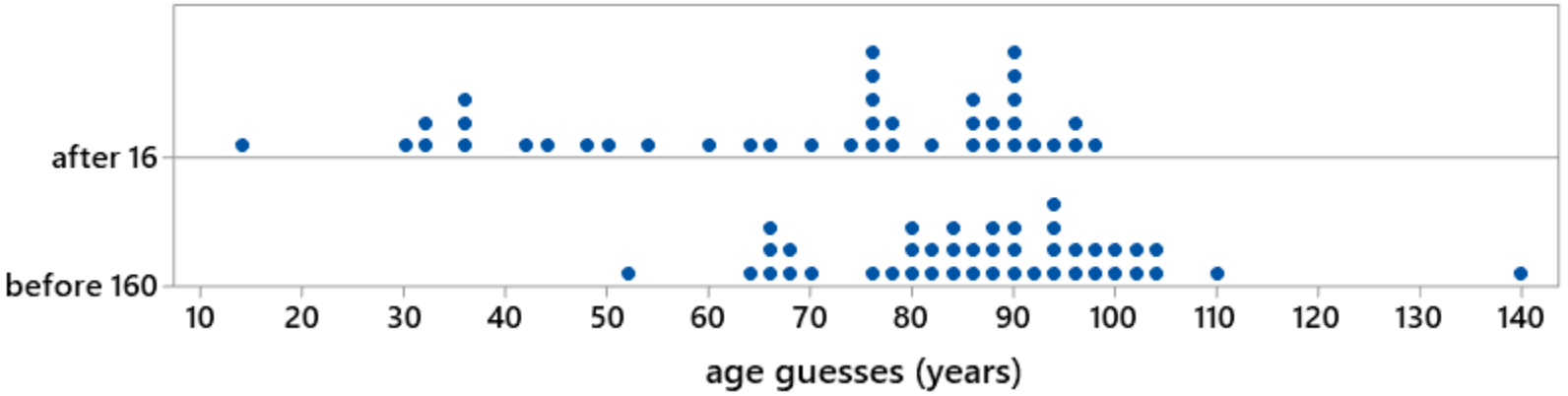
- Would closing eyes and pointing be unbiased?

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 for 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

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?

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?
- No!
  - Would require that *no* households have both

Simulation-based inference

- Example: Facial prototyping

Do people tend to associate names with faces?

Who is on the left: Bob or Tim?

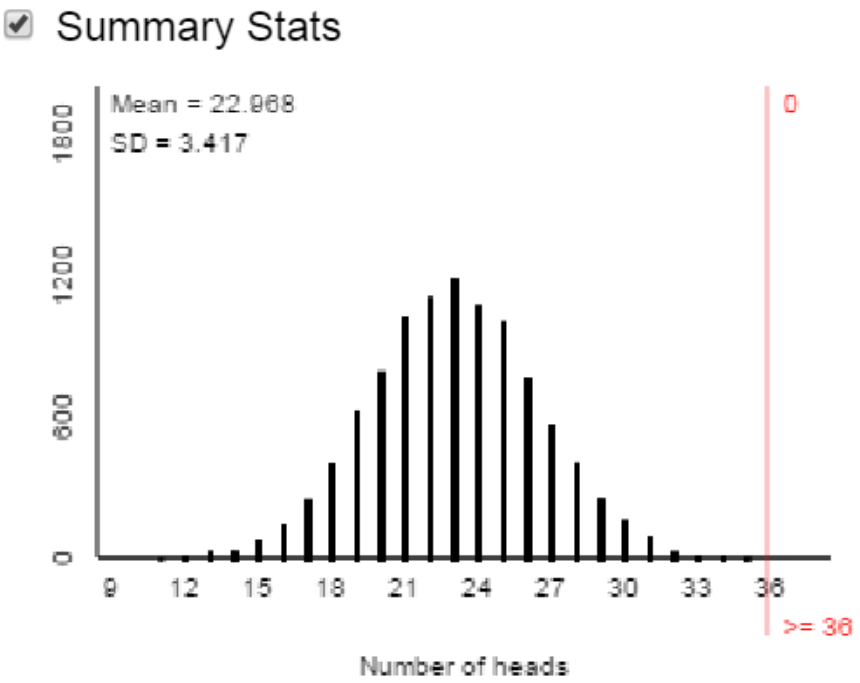
Simulation-based inference

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



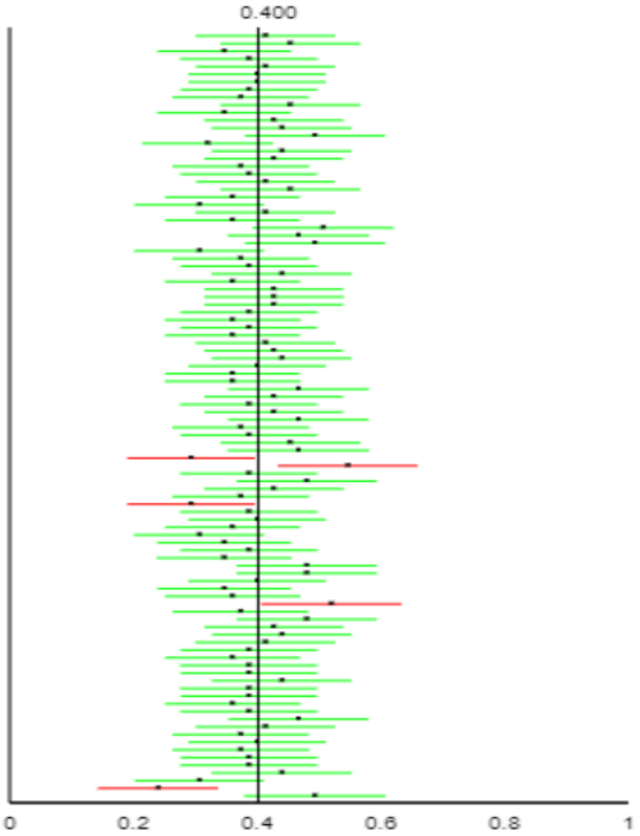
- Very strong evidence: people have a genuine tendency to put Tim on left

Confidence intervals

- What does 95% confidence mean?
- Simulation:

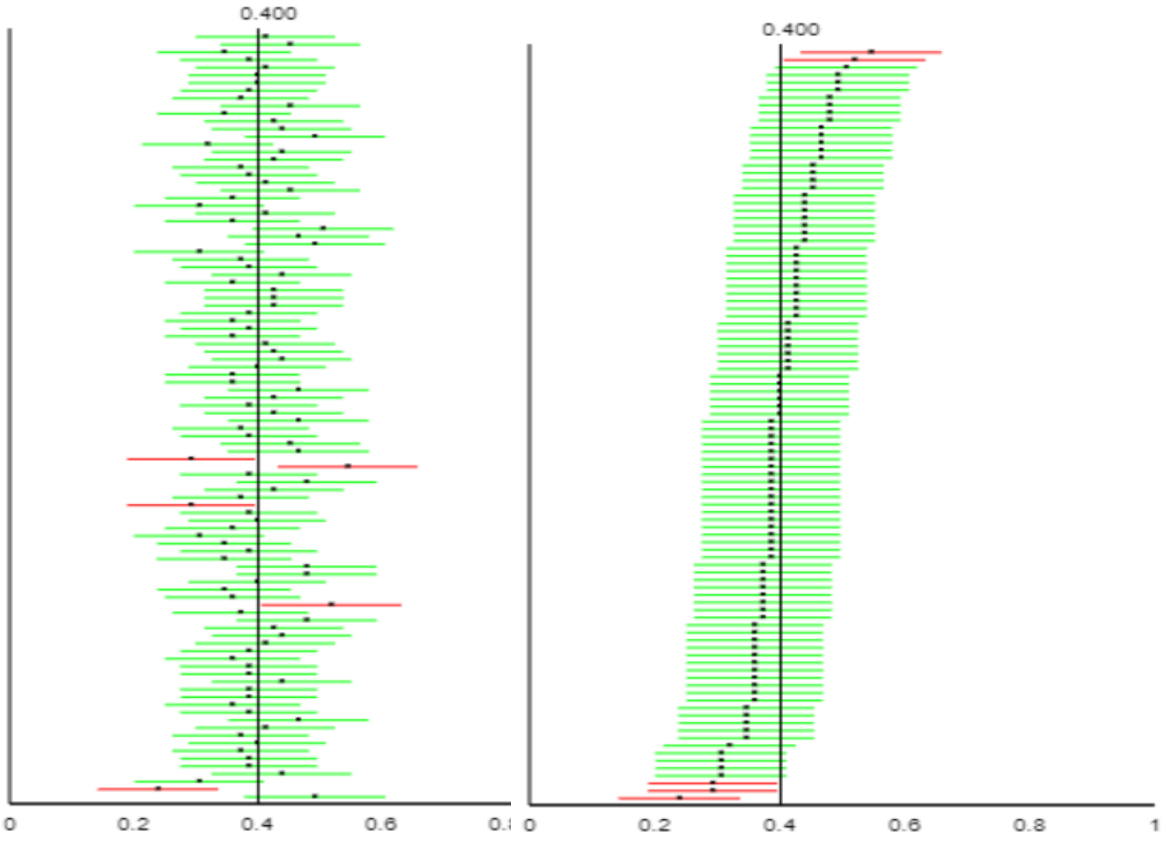
Confidence intervals

- What does 95% confidence mean?
- Simulation:



Confidence intervals

- What does 95% confidence mean?
- Simulation:



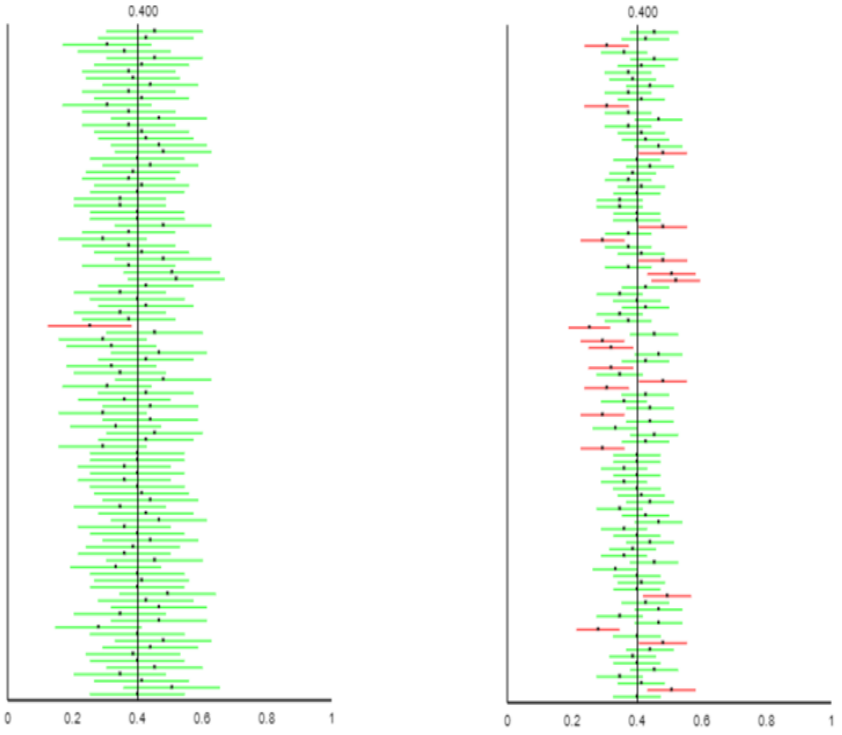
Confidence intervals

- What would change (two things) if the confidence level was increased or decreased?

Confidence intervals

- What would change (two things) if the confidence level was increased or decreased?

- 99% (left)
- 80% (right)



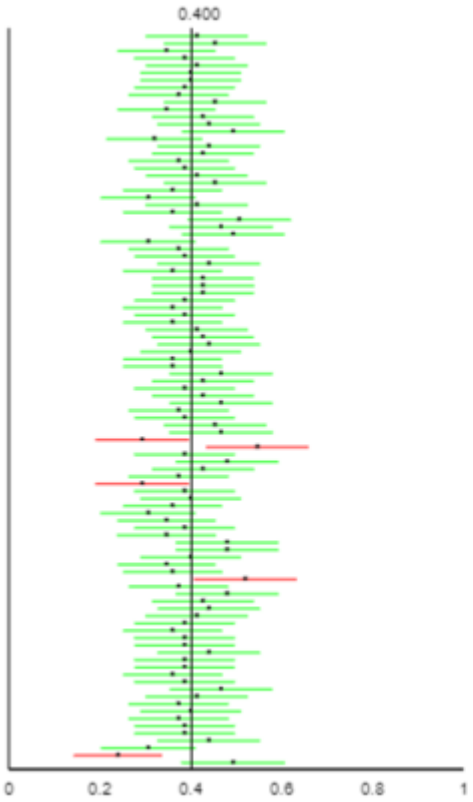
Confidence intervals

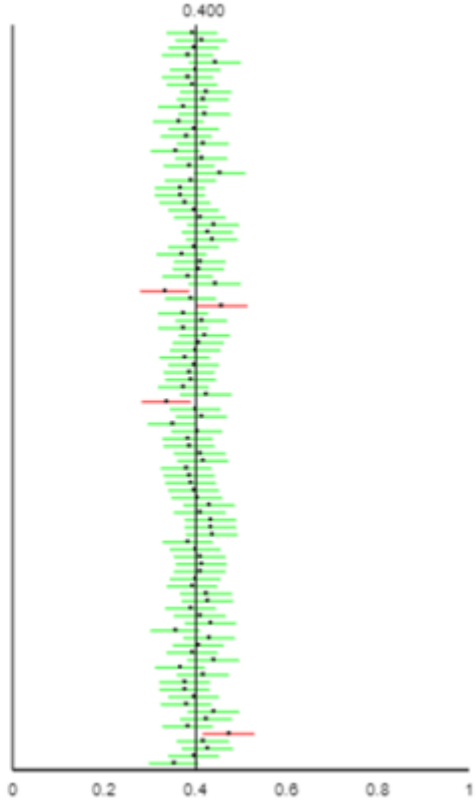
- What would change if the sample size was increased?

Confidence intervals

- What would change if the sample size was increased?

- Smaller (left)
- Larger (right)





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 2020 U.S. Senate, which has 26 women (the most ever!) and 74 men

Confidence intervals

- Calculate the alien's 95% CI
  - $.26 \pm 1.96\sqrt{\frac{.26\times.74}{100}}$, which is $.26 \pm .086$
- Interpret the CI for the alien
  - The alien is 95% confident that between 17.4% and 34.6% 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 2020 U.S. Senators who are female, would this 95% CI make sense?
  - No, exactly 26% of senators in 2020 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

"Ask Good Questions" blog

- Ideas, examples, activities, assessments and advice for teaching introductory statistics
- Two goals
  - Practical, for direct use with students
  - Thought-provoking, for discussion with peers
- One post/essay per week
  - Published on Mondays at 11am ET, 8am PT
  - 2000-3000 words per essay
  - 28 so far, 24 (at least) more to come

This week's post: A pervasive pet peeve

- 2017 Youth Risky Behavior Surveillance Survey

Seat belt use when riding with someone else driving	Arizona	California
Rarely or never	173	103
Sometimes, most of the time, or always	1966	1675
Total	2139	1778
Proportion who responded "rarely or never"	0.081	0.058

- Were Arizona youths 2.3% more likely to respond "rarely or never" than California youths?
- No! The Arizona percentage was 2.3 *percentage points* higher than California's.

This week's post: A pervasive pet peeve

- 2017 Youth Risky Behavior Surveillance Survey

Seat belt use when riding with someone else driving	Arizona	California
Rarely or never	173	103
Sometimes, most of the time, or always	1966	1675
Total	2139	1778
Proportion who responded "rarely or never"	0.081	0.058

- Percentage difference:

  $(.081-.058)/.058 \times 100\% \approx 39.6\%$

- Arizona youths were 39.6% more likely than California youths to respond "rarely or never"

This week's post: A pervasive pet peeve

- 2017 Youth Risky Behavior Surveillance Survey

Seat belt use when riding with someone else driving	Arizona	California
Rarely or never	173	103
Sometimes, most of the time, or always	1966	1675
Total	2139	1778
Proportion who responded "rarely or never"	0.081	0.058

- Relative risk: .081/.058 ≈ 1.396
- Arizona youths were 1.396 times more likely than California youths to respond "rarely or never"
- % difference = (relative risk – 1) × 100%

This week's post: A pervasive pet peeve

- Potential title: *A persnickety post that preaches about a pervasive, persistent, and pernicious pet peeve concerning percentages* (60% P-words)
- Actual title: *A pervasive pet peeve* (75%)

- So, is this a 15% increase in % of P-words?
  - No, that's the pet peeve!

"Ask Good Questions" blog

- Sign up to receive weekly email messages containing post
- Peruse annotated list of posts
- Post #1 answers ten questions

https://askgoodquestions.blog/posts

"Ask Good Questions" blog

#28 A pervasive pet peeve (Jan 13)

- Percentage difference, relative risk

#27 Simulation-based inference part 2 (Jan 6)

- Comparing two groups with categorical response

#26 Group quizzes, part 2 (Dec 30)

#25 Group quizzes, part 1 (Dec 23)

#24 Random rendezvous, part 2 (Dec 16)

- Normal distributions, combining random variables

#23 Random rendezvous, part 1 (Dec 9)

- Probability, simulation, geometry

#22 Four more exam questions (Dec 2)

- Free response questions

#21 Twenty final exam questions (Nov 25)

- Multiple choice conceptual questions

#20 Lincoln and Mandela, part 2 (Nov 18)

- Random assignment vs. random sampling

#19 Lincoln and Mandela, part 1 (Nov 11)

- Sampling bias, random sampling

#18 What do you expect? (Nov 4)

- Expected value, long-run average

#17 Random babies (Oct 28)

- Simulation, probability, long-run

#16 Questions about cats (Oct 21)

- Statistical significance vs. practical importance

"Ask Good Questions" blog

#15 How confident are you? part 2 (Oct 14)

- Confidence interval vs. prediction interval

#14 How confident are you? part 1 (Oct 7)

- Confidence level, effect of sample size

#13 A question of trust (Sept 30)

- Simulation-based inference, z-test, binomial p-value

#12 Simulation-based inference, part 1 (Sept 23)

- Strength of evidence, p-value

#11 Repeat after me (Sept 16)

- Observational units, variables

#10 My favorite theorem (Sept 9)

- Reverse conditional proportions, Bayes Thm

#9 Statistics of illumination, part 3 (Sept 2)

- Statistical thinking: signal amidst noise

#8 End of alphabet (Aug 26)

- z-scores, measuring distance by number of SDs

#7 Two dreaded words, part 2 (Aug 19)

- Standard deviation (real data)

#6 Two dreaded words, part 1 (Aug 12)

- Standard deviation (conceptual questions)

#5 A below-average joke (Aug 5)

- Properties of averages

#4 Statistics of illumination, part 2 (July 29)

- Statistical thinking in terms of distributions

#3 Statistics of illumination, part 1 (July 22)

- Statistical thinking, confounding variable, Simpson's paradox

#2 My favorite question (July 15)

- Measuring skewness, simulation-based inference

- Thanks very much!
- Take-home message: Ask Good Questions!
- Questions, comments: arossman@calpoly.edu