

STA130 Winter 2020

(Materials used in this presentation are provided by the U of T Statistical Sciences Department.

This presentation was prepared by Vivian Ngo.)

[Github.com/vivianngo97/STA130-Winter-2020](https://github.com/vivianngo97/STA130-Winter-2020)

viv.ngo@mail.utoronto.ca

Agenda

- Reminders
- Vocabulary
- Group discussion
- Poster project discussion/prep time
- Individual* writing activity: poster project plan

Reminders

- No class or tutorial next week (reading week!)
- No OH, put Piazza will be monitored
- Midterm: Friday after reading week (Feb 28) during regular tutorial time in EX100
 - You MUST attend the correct section's term test.
 - **No calculators** are allowed.
 - See Quercus for other details and practice tests.
 - Take advantage of our TA office hours in HS381/390!

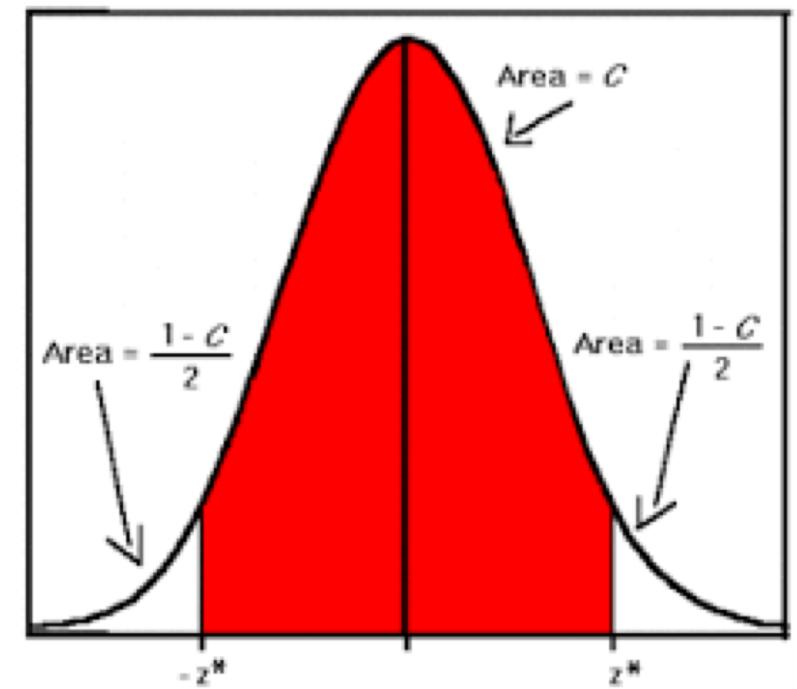


Vocabulary

- Parameter
- Statistic
- Population
- Sample
- Sampling distribution
- Random sampling
- Resampling
- Bootstrap
- Percentile (quantile)
- Confidence interval
- Confidence level
- Testing
- Estimation
- Representative

Vocabulary

- Always check that your CI range makes sense.
 - E.g. if reported the CI for a proportion, it needs to be bounded by zero and one. You can't have a probability less than zero or greater than 1.
- **Purpose of CIs:** to obtain an estimate the parameter that reflects sampling variability. E.g. estimate the proportion of people in Toronto who use the TTC, number of coffees people in this class drink each week, etc.
 - Wide CI: if you had taken a different sample from the population, you could arrive at a very different estimate.
 - Narrow CI: if you had taken a different sample from the population, you could expect to get a similar estimate.



Vocabulary

- **Purpose of bootstrap:** to estimate the sampling distribution of a statistics to; e.g. CI.
- The percentile bootstrap method (that we are using this week) works best for large samples and when the bootstrap distribution is approximately symmetric and continuous.
 - Therefore, your CIs should be roughly symmetric around the point estimate.
- You will see other versions of the bootstrap method in future statistics courses.



Group discussion

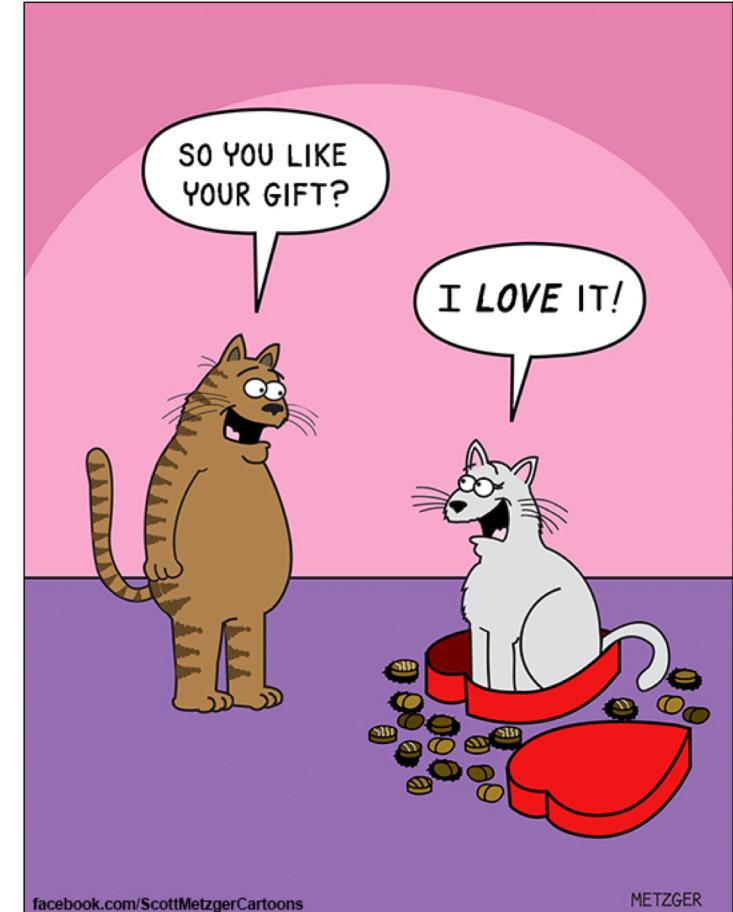
- 1. Are the use of p-values and confidence intervals mutually exclusive? What do the two have in common? How do they differ? Think about under which circumstances you may want to use each of these.
- 2. If we want to be more confident in correctly capturing the correct proportion of our outcome, such as the percent of couples who tilt their heads right when kissing, should we use a larger or smaller confidence interval? How do you think this relates to type I and/ or type II errors?

Group discussion

- 3. If you and your partner both applied the same bootstrap sampling method to the same data, do you expect that you both arrive at the same estimate and CI? What are some factors that you would need to consider (and hold constant) to ensure that you both arrived at the same answer?

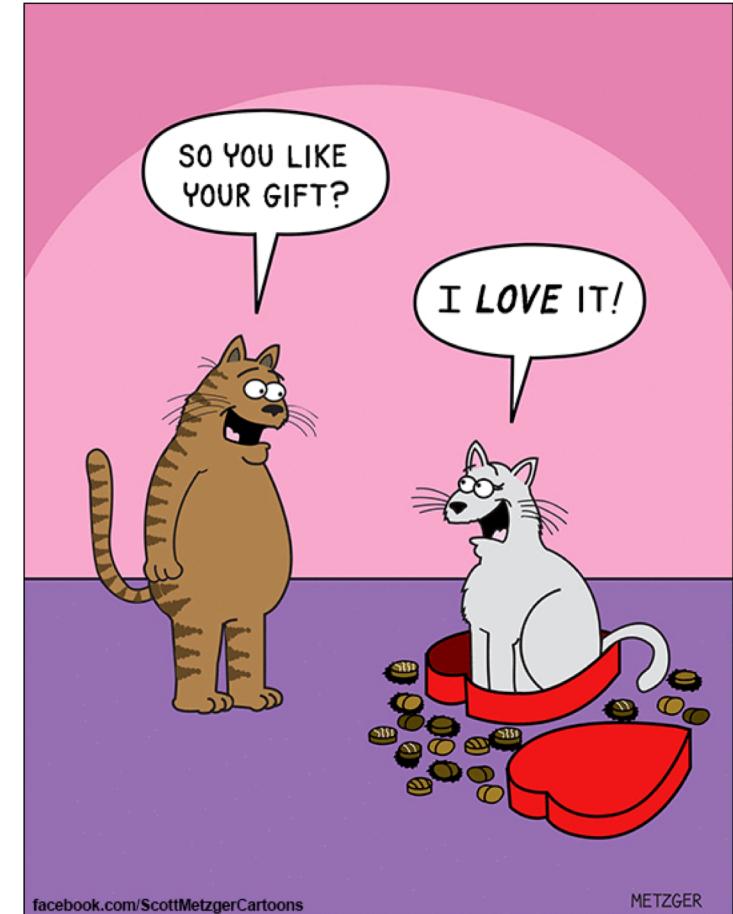
Group discussion

- 4. It's Valentine's Day! You are interested in whether there is a difference in the proportion of couples who tilt their heads to the right or left when they kiss! You survey several students on campus and find that 35.5% tilt their heads to the left when kissing; 95% CI: (27%, 44%). Which of the below (Question 1d) is the correct interpretation of the 95% CI?
 - (i) We are 95% confident that between 27% and 44% of kissing couples in this sample tilt their head to the left when they kiss
 - (ii) There is a 95% chance that between 27% and 44% of all kissing couples in the population tilt their head to the left when they kiss.
 - (iii) We are 95% confident that between 27% and 44% of all kissing couples in the population tilt their head to the left when they kiss.
 - (iv) If we considered many random samples of 124 couples, and we calculated 95% confidence intervals for each sample, 95% of these confidence intervals will include the true proportion of kissing couples in the population who tilt their heads to the left when kissing.



Group discussion

- 4. It's Valentine's Day! You are interested in whether there is a difference in the proportion of couples who tilt their heads to the right or left when they kiss! You survey several students on campus and find that 35.5% tilt their heads to the left when kissing; 95% CI: (27%, 44%). Which of the below (Question 1d) is the correct interpretation of the 95% CI?
 - (i) We are 95% confident that between 27% and 44% of kissing couples in this sample tilt their head to the left when they kiss
 - (ii) There is a 95% chance that between 27% and 44% of all kissing couples in the population tilt their head to the left when they kiss.
 - (iii) **We are 95% confident that between 27% and 44% of all kissing couples in the population tilt their head to the left when they kiss.**
 - (iv) **If we considered many random samples of 124 couples, and we calculated 95% confidence intervals for each sample, 95% of these confidence intervals will include the true proportion of kissing couples in the population who tilt their heads to the left when kissing.**



Poster project details

- **DATA:** Toronto Police Service (TPS) data to investigate break & enter robberies (B&Es) over the past few years.
- **TASK:** provide meaningful research results to TPS which could inform improvements to their ability to prevent/combat B&Es.
- **More info:** on Quercus
- Somebody from TPS will be coming to class next week to talk more about this project.
- **You will receive the data AFTER the midterm**
- The project will officially launch (on Quercus) on March 2nd.

Poster project

- Form project groups (3-4 people per group) / Think about who you want to work with
- Think about interesting research questions over reading week
- You will learn more **methods** after reading week that you can use in your project
- *Methods are not exploratory analyses **
- No need to use *every single method*. Pick the appropriate ones

Poster project: Key things to discuss

- What are you interested in researching?
- What might the police be interested in?
- What visualizations (graphs, plots, etc) will be useful?
- Will you join datasets?
- What about external datasets from the internet? (WOW factor!)
- What will each group member do?
- How often to meet?



Writing activity

- **Individually** write-up your research plan based on what you discussed with your group.
- **Suggested format:**
 - Aims/objectives
 - 2-3 research question(s)
 - The hypotheses
 - Research design: data/ variables, methods*, visualizations, etc.
- (This is not a final research plan)