

STA130 Fall 2019 – T0107

Week 7: Sampling, bootstrap, confidence intervals

(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-Fall-2019](https://github.com/vivianngo97/STA130-Fall-2019)

viv.ngo@mail.utoronto.ca

Agenda

- **Reminders**
- **Material & Vocabulary review**
- **Group discussion**
- **Poster Project Planning**
- **Individual* writing activity: poster project plan**
- **Last week's "ticket out of the door" submissions**

Reminders

- The term test is next Friday during their regularly scheduled tutorial time
 - You MUST attend the correct section's term test.
 - You MUST sit in the correct room.
 - No calculators are allowed.
 - See Quercus for other details and practice tests.

Material and Vocabulary Review

- Percentile (Quantile)
- **Parameter**
- Statistic
- Population
- Sample
- Sampling distribution
- Resampling
- **Bootstrap**
- **Confidence interval**
- Confidence level
- Testing
- Estimation
- Representative

Material and Vocabulary Review

- **Purpose of bootstrap:** to estimate the sampling distribution of a statistics
 - E.g. to get a confidence interval (CI)
- **Purpose of CIs:**
 - to obtain an estimate of the parameter that reflects sampling variability.
 - E.g. Wish to estimate the proportion of people living in Toronto who use the TTC, number of coffees people in this class drink each week, etc.
 - Always check that your CI range makes sense.
 - 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.
- 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.

Discussion in pairs

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.

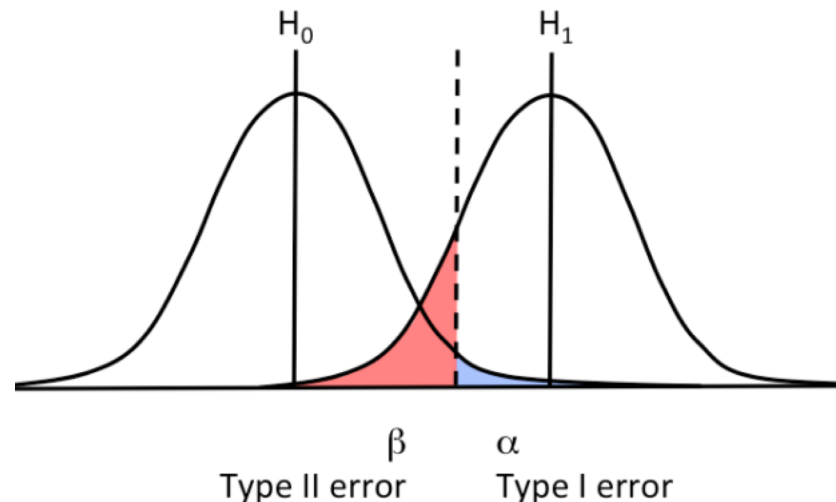
[5 mins]

Discussion in pairs

2a)

How do you expect the width (precision) of a 95% CI to compare to a 90% CI? Compared to a 99% CI? How do you think this relates to type I and/ or type II errors?

[5 mins]



Discussion in pairs

2b)

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?

[5 mins]

Discussion in pairs

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?

[5 mins]

Discussion in pairs

4)

Discuss question 1d from the practice problem set. Indicate whether or not each of the following statements is a correct interpretation of the confidence interval constructed in 1c) and justify your answers.

[5 mins]

- (i) We are 95% confident that between 56% and 73% of kissing couples in this sample tilt their head to the right when they kiss.
- (ii) We are 95% confident that between 56% and 73% of all kissing couples in the population tilt their head to the right when they kiss.
- (iii) There is a 95% chance that between 56% and 73% of all kissing couples in the population tilt their head to the right when they kiss.
- (iv) If we considered many random samples of 124 couples, and 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 right when they kiss.

Poster Project

- Time to start planning your poster project! **[30 min]**
- Form groups (3-4 people per group, no exceptions!)
- Suggestion:
 - Do NOT leave the poster project until the last minute as it requires a significant amount of work and you will need time to practice your presentations.
- You will learn more methods later in the term; e.g. classification trees for prediction.
 - You can incorporate these methods into your project (if appropriate to your question), but you are not required to use *all* of the methods. In fact, this is discouraged as not every method will help answer your research question.

Poster Project: Things to discuss

- The research question(s) you are interested in investigating. What is something that the Toronto Police might be interested in knowing?
- Any interesting visualizations. Remember, these should be interesting and useful for the audience. They should also be appropriately labelled and should stand alone (e.g. information title, axes, and a brief sentence or two explaining the main take away messages)
- Think about whether you will need to join datasets to answer your research question. You are able to use additional sources of data, if you like – and if relevant to your research question. This would be considered in the WOW factor category!)
- What will each group member be responsible for?
- How often will you meet?
- Research plan:
 - Research questions
 - The hypotheses
 - Aims/objectives
 - Research design: data, **methods***, visualizations, etc.
 - *** methods do not mean data manipulation, filtering, etc. E.g. calculating a confidence interval and using bootstrap method IS a method, but filtering and joining dataframes is not a method.

Writing Activity

- Individually write-up your research plan, based on what you discussed with your group. **[30 min]**
 - Consider formatting it to include the above sections (i.e., Research question; Hypothesis; Objective; Research design: data, methods, visualization, etc.)
- **** *The activity MUST be submitted by the end of tutorial – unless there is some extenuating circumstance.***

Last week's “ticket out of the door” submissions

- When do we reject p-value?
- Can we have review on vocabulary?
- How to translate findings on Rstudio to an oral presentation
- Where to find past tests?
- How to get a p-value from a test statistic?
- What is positive and negative? (H_a vs H_0)
- How to do well on the project?