

STA130 T0209

Week 6: 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.)



Agenda

- Reminders
- Material & Vocabulary review
- Discussion in pairs
- Oral presentations

Reminders

- No class or tutorial next week (Reading week)
 - No OH, but Piazza will be regularly checked
 - New College Stats Aid Centre will be open during reading week
- The midterm is the following week during your usual tutorial time (March 1st)
 - You MUST attend the correct section's midterm.
 - AM section: EX200
 - PM Section:
 - MS 3154: Last names from A – Lo
 - WB 116: Last names from Lu – Z
 - Includes all material up to & including Feb 25th (mostly a review class)
 - Format: Multiple choice, fill in the blanks, written answers (make sure to write complete sentences)
 - Example midterms have been posted to Quercus

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.
- 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.

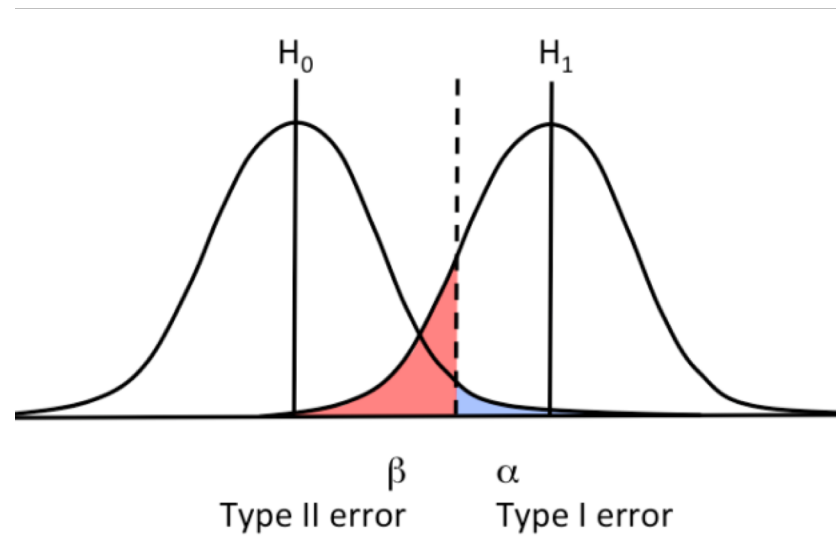
[5 mins]

Discussion in pairs

2)

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

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]

Group presentation

- Last time:

Keep it up!	Let's work on this!
Confidence	Fidgeting
Preparedness	Reading off of large devices
Volume	Eye contact
Answering questions	

Group presentation

- Prepare a 5 minute oral presentation based on the following topics regarding Question 3
- Pick the topic you would like (next slides)
 - The bullet points are an outline of topics that you can use to prepare for your presentation
- When not presenting:
 - One person from each group evaluate other students, upload rubric to Quercus
 - Write down any questions you have
 - Presentation rubric on Quercus and Github

Group presentation: Topic 1

- Describe the content of the graph (mention the variable of interest)
 - What does the x-, y-axis represent?
 - Describe the distribution (range, center, symmetry, skewness, number of points).
- How was one dot calculated:
 - how a single bootstrap sample is produced (e.g. size, with/without replacement)?
 - what statistic did you calculate from this particular bootstrap sample?
- How can the generated distribution of mean age be used for inference?
 - Since the observed data were generated ... (this was a key word from last week), we can ... (a key word from today's vocab list) from the observed data by sampling with replacement.
 - In other words, if the data resemble the ... (a key word from today's vocab list), the bootstrap samples will also resemble the (the same key word from today's list).
 - Using the statistic calculated from each bootstrap sample, we can obtain a distribution of sample statistic and it gives an estimate of the ... (a key phrase from today's vocab list) of the statistic.

Group presentation: Topic 2

- Rationale of using the bootstrap sampling distributions:
 - Does distribution of bootstrap sample statistic tend to capture the population value (mean/median)?
 - Where does the population value tend to be in the range of the bootstrap distribution?
- Construct the confidence interval
 - What is the range of values (in terms of percentile) taken to construct the 90% CI?
 - Describe how you did the above in R (are there ties? Was it easy or difficult to do this in R and why?)
- State the interval with reference to the data and variable (E.g. “A 90% CI for the mean (median) of the mother’s age is”).
- Interpret the interval you produced.
- (Bonus) Could you check if the interval produce is indeed the 90% CI? Why or why not?
 - Do you know the population parameter value (mean)?
 - Do you need to do more calculations?
 - How many times do you check if the population mean is captured by the CI?

Group presentation: Topic 3

- Describe how to produce the plot
 - How a single bootstrap sample is produced (e.g. number of data points used, with/without replacement)?
 - What statistic did you calculate from each bootstrap sample?
- Describe the content of the graph (mention the variable of interest)
 - What does the x-, y-axis represent?
 - Describe the shape of the distribution (range, center, symmetry, skewness, number of points).
- Rationale of using the bootstrap sampling distributions:
 - Does distribution of bootstrap sample statistic tend to capture the population value (median)?
 - Where does the population value tend to be in the range of the bootstrap distribution?
- Construct the confidence interval
 - What is the range of values (in terms of percentile) taken to construct the 99% CI?
 - Describe how you did the above in R (are there ties? Was it easy or difficult to do this in R and why?)
- State the interval with reference to the data and variable (E.g. “A 99% CI for the median of the mother’s age is”).
- Interpret the interval you produced.

	4 (Excellent)	3 (Good)	2 (Adequate)	1 (Poor)
Context	The context and connection to the problem are clear.	Some context was provided and all variables/concepts were mentioned. Some aspects were not clear.	Very little context was provided and only some variables/ concepts were mentioned.	No context and mentioning of any variables/ concepts covering in this week's materials.
Structure	Well organized, follows a logical structure.	The organization follows some logical structure.	Some structure but difficult to follow.	There is no structure, very difficult to follow.
Conclusion	There is a clear central idea and the conclusion is correct.	A central idea or conclusion is present. The conclusion might be incorrect.	The central idea or conclusion is weak and not supported.	The central idea or conclusion is missing. Incorrect conclusion.
Transitions	The progression is logical. Effective use of transitions.	The progression is controlled. The use of transitions is mostly meaningful.	Minor disruptions in flow and weak transitions.	Weak progression and lack of transitions.
Vocabulary	Good use of statistical terms and appropriate choice of words.	Use of statistical terms and phrases mostly correct, demonstrates understanding of concepts.	Some use of statistical terms/ phrases and some understanding of concepts demonstrated.	Inaccurate or incorrect use of statistical terms or phrases and a lack of understanding statistical concepts.
Presentation Skills	<p>Regular eye contact with all parts of the audience.</p> <p>The audience was engaged.</p> <p>The presenter held the audience's attention.</p> <p>Appropriate speaking volume & body language.</p> <p>Good pace.</p>	<p>Somewhat regular eye contact or eye contact with some of the audience</p> <p>The audience was mostly engaged.</p> <p>The presenter mostly spoke at a suitable volume.</p> <p>Spoke too quietly at times.</p> <p>Some fidgeting.</p> <p>Going too fast/slow.</p>	<p>Focused on only one or two members of the audience.</p> <p>Sporadic eye contact.</p> <p>The audience was not engaged.</p> <p>Speaker could be heard by only some of the audience.</p> <p>Body language was distracting.</p>	<p>Minimal (or no) eye contact.</p> <p>The audience was never engaged.</p> <p>The presenter did not speak clearly.</p> <p>Presenter was very difficult to hear.</p>
Preparedness/ Participation	<p>Extremely prepared and rehearsed.</p> <p>The presenter was confident.</p>	<p>Mostly prepared but some dependence on or reading off of notes.</p> <p>The presenter seemed fairly confident.</p>	<p>The presenter was not well prepared.</p> <p>The presenter did not seem confident.</p>	<p>Evident lack of preparation/rehearsal.</p> <p>Complete dependence on notes.</p>