See you / see me:
An interactive real-time online course

Rent in Durham - bootstrap interval

The dot plot below shows the distribution of means of 100 bootstrap samples from the original sample. Estimate the 90% bootstrap confidence interval based on this bootstrap distribution.
Data analysis and statistical inference are required for (some) SS majors. "Lecture" (2/week) + computing lab (1/week) are needed. On campus, study abroad + other demands/engagements increase the demand, leading to larger class sizes.
sta 104: data analysis and statistical inference (online)

online version of sta 101

summer 2013 and 2014, session 1

motivation:
create a course that is the same (or as similar as possible) content / quality / pedagogy / rigor
MOOC

SCOC?
2013: 7 students

- First-year: 4
- Sophomores: 2
- Junior: 1
- Senior: 0

2014: 11 students

- First-year: 0
- Sophomores: 4
- Junior: 5
- Senior: 2

Most majoring in PubPol

Some dropped STA 101 during regular session due to workload

Some enjoy taking online courses

Others avoiding taking course during regular session or need to meet requirements before fall
logistics

virtual daily meetings on WebEx Training Center (recorded)

90 min / day
5 days / week
“lecture” + lab

breakout sessions

assignments and forums (Piazza) on Sakai

materials on public course website

videos hosted on YouTube (temporary solution)
Poll

Which of the following is true?

(a) If the sample size is large enough, conclusions can be generalized to the population.
(b) If subjects are randomly assigned to treatments, conclusions can be generalized to the population.
(c) Blocking in experiments serves a similar purpose as stratifying in observational studies.
(d) Representative samples allow us to make causal conclusions.
(e) Statistical inference requires normal distribution of the response variable.
**Sta 104 - Data Analysis and Statistical Inference - Online (Summer 2014)**
Dr. Çetinkaya-Rundel

**Tentative schedule:**

<table>
<thead>
<tr>
<th>Unit</th>
<th>Date</th>
<th>Topics</th>
<th>Slides / App. Ex. / Lab</th>
<th>Prep</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>Wed, 5/14</td>
<td>Introduction + Data collection</td>
<td>Lec 0 + Lec 1.1 + App 1.1</td>
<td>Unit 1 resources</td>
<td>Complete survey + pretest</td>
</tr>
<tr>
<td>Unit 1</td>
<td>Thur, 5/15</td>
<td>Observational studies &amp; experiments</td>
<td>Lec 1.2 + App 1.2</td>
<td></td>
<td>RA 1 in class (practice)</td>
</tr>
<tr>
<td>Unit 1</td>
<td>Fri, 5/16</td>
<td>Introduction to R</td>
<td>Lab 0 + Lab 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sat, 5/17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sun, 5/18</td>
<td></td>
<td></td>
<td></td>
<td>PS 1 due</td>
</tr>
<tr>
<td>Unit 1</td>
<td>Mon, 5/19</td>
<td>Exploratory data analysis</td>
<td>Lec 1.3 + App 1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>Tue, 5/20</td>
<td>Introduction to statistical inference</td>
<td>Lec 1.4 + App 1.4</td>
<td></td>
<td>PA 1 due (Practice)</td>
</tr>
</tbody>
</table>
Introduction

Unit 1: Introduction

populations and samples

Part 1: Designing studies

(1) Data basics
pre-course meeting

last week of classes

meet & greet

technical details
course format

flipped + team-based

7 learning units

pre-unit preparation: videos + learning objectives + textbook

unit readiness assessment: individual, then team

lecture, with polling questions

application activities in breakout sessions

performance assessment
Unit 5 - Inference for categorical variables

Suggested reading: OpenIntro Statistics, Chapter 6

Suggested exercises:
* Part 1 - Single proportion: 6.1, 6.3, 6.5, 6.9, 6.11, 6.15, 6.21
* Part 2 - Comparing two proportions: 6.23, 6.25, 6.27, 6.29, 6.31, 6.33, 6.35
* Part 3 - Inference for proportions via simulation: 6.47, 6.49, 6.51
* Part 4 - Comparing three or more proportions (Chi-square): 6.37, 6.39, 6.41, 6.43, 6.45

* Suggested Reading: Section 6.1 of OpenIntro Statistics

LO 1. Define population proportion $p$ (parameter) and sample proportion $\hat{p}$ (point estimate).

LO 2. Calculate the sampling variability of the proportion, the standard error, as

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

where $p$ is the population proportion.

- Note that when the population proportion $p$ is not known (almost always), this can be estimated using the sample proportion, $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

LO 3. Recognize that the Central Limit Theorem (CLT) is about the distribution of point estimates, and that given certain conditions, this distribution will be nearly normal.

- In the case of the proportion the CLT tells us that if (1) the observations in the sample are independent, (2) the sample size is sufficiently large (checked using the success/failure condition: $np \geq 10$ and $n(1-p) \geq 10$), then the distribution of the sample proportion will be nearly normal, centered at the true population proportion and with a standard error of $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. 
Poll

11% of 1,001 Americans responding to a 2006 Gallup survey stated that they have objections to celebrating Halloween on religious grounds. At 95% confidence level, the margin of error for this survey is ±3%. A news piece on this study’s findings states: “More than 10% of all Americans have objections on religious grounds to celebrating Halloween.” At 95% confidence level, is this news piece’s statement justified?

(a) Yes
(b) No
(c) Cannot tell
STA.104.01.1Su14: Assignments

Application activity - 7.2 CI for Slope in MLR 🚀 Submissions

Found 3 participant(s). Assign this grade to participants without a grade:

- Please select default grade: 

Select Group(s) and Allow Resubmission

<table>
<thead>
<tr>
<th>Group</th>
<th>Submitted</th>
<th>Status</th>
<th>Grade</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dumbo Octopus</td>
<td>Jun 19, 2014 2:15 pm</td>
<td>Returned</td>
<td>Checked</td>
<td>✔️</td>
</tr>
<tr>
<td>Kathleen Axerod, Emma Bunting, Kara Fisher, Amari Stokes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Devils</td>
<td>Jun 19, 2014 2:16 pm</td>
<td>Returned</td>
<td>Checked</td>
<td>✔️</td>
</tr>
<tr>
<td>Hayley Bohart, Tyler Deane-Krantz, Leah Reisman, Shelby Wales</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stats Super Stars</td>
<td>Jun 19, 2014 2:15 pm</td>
<td>Returned</td>
<td>Checked</td>
<td>✔️</td>
</tr>
<tr>
<td>Terriey Bishop, Caleb Ellis, Kara Wilson</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assignment Details
Test the hypothesis $H_0 : \mu = 10$ vs. $H_A : \mu > 10$ for the following 8 samples. Assume $\sigma = 2$.

<table>
<thead>
<tr>
<th></th>
<th>Jacob</th>
<th>Shelby</th>
<th>Daryn</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 30$</td>
<td>$\bar{x}$</td>
<td>10.05</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>$p$-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>$\bar{x}$</td>
<td>10.05</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>$p$-value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
office hours twice a week on WebEx

impromptu meetings on Google Hangouts

summer 2014: TA through BASS Online Apprentice Program

student formed study groups on Google Hangouts


“I like the convenience of the online class but also the web chat structure makes it feel as if you are actually in a classroom. So it is the best of both worlds.”

“I really enjoy the videos! They are a very helpful learning tool. It is also nice to have them to go back to at any time to clear up a concept.”

“I like that the class is discussion-based and interactive. I enjoy working with my classmates on application exercises and we are able to explain concepts to each other in terms that we understand. I also like the various polls that we do during lectures because they keep you engaged and you can learn a lot from hearing other students explain the reasoning behind their answers.”
- Great experience ✔
- Synchronous sessions worthwhile ✔
- Assessment submission and grading on Sakai ✔
- Teams on Sakai for application activity submission & reveal ✔
- Performance assessments ✔
WebEx for teaching

- Video feed from students
- Recording
- Training Center: breakout sessions
- Polling
WeEx for teaching

- Video feeds limited to 6
- Annotation tool buggy
- No auto recording + video streaming slow
- Proctoring exams [Lockdown browser + Video]
Dr. Çetinkaya-Rundel

Data Analysis and Statistical Inference

Application exercise: 5.1

Inference for a single proportion

Emotionless and emotional

1. According to a 2012 Gallup poll “Singaporeans are the least likely in the world to report experiencing emotions of any kind on a daily basis.” Only 36% out of the 1,695 Singaporeans polled report feeling either positive or negative emotions, lowest in the world.

(a) You are asked to write a newspaper article about this finding, and provide a plausible range of values (a 95% confidence interval) for the true proportion of Singaporeans who experience emotions on a daily basis. What is the parameter of interest and what is the sample statistic?

(b) Are conditions for inference satisfied?

(c) Calculate the 95% confidence interval, and interpret it in context.

(d) If we wanted to decrease the margin of error by 20% (i.e. make it 80% of what you’ve calculated above), how many Singaporeans would we need to sample?

2. The same Gallup poll also found that the Philippines had the highest percentage of people reporting experiencing emotions of any kind on a daily basis (86% out of 1,200). Does this data provide convincing evidence that majority of Filipinos experience emotions daily?

How do the conditions and the calculation of the standard error for a hypothesis test (Question 2) differ from those for a confidence interval (Question 1)? Explain why there is a discrepancy in how we check the conditions and how we calculate the standard error for the two methods when
[Technical]
don’t need to be tech savvy, but should be flexible
especially if class size is large, a tech assistant (at least) first few days

remember to record [if not automatic]
on call tech help throughout course
thank you!

questions?