

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

Video

Mine Cetinkaya-Rundel (internal)

cece Mercer

tierney bishop

Jacob Boissiere

Lili Ramirez

Amari Stokes

Sophie Cummins

Thomas Meister

Duke University

Bootstrapping

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.

90% 5% 5%

900 1000 1100 1200 1300 1400

Statistics 104 (Mine Cetinkaya-Rundel) U4 - L1: Bootstrap, paired, and two sample June 3, 2013 7 / 22

Mine
Çetinkaya-Rundel

Department of
Statistical
Science

background

data analysis
and statistical
inference

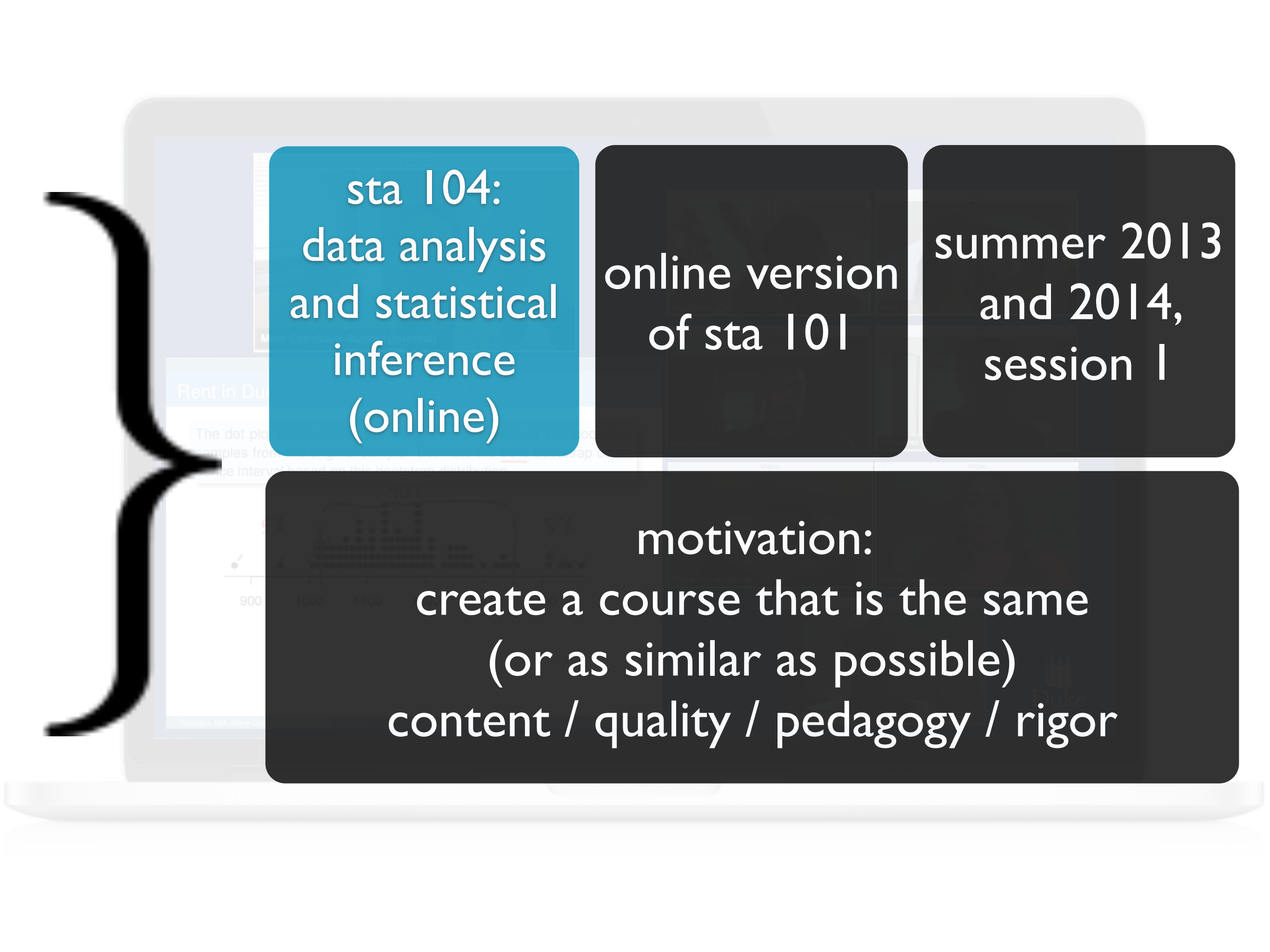
“lecture”
(2/week)
+
computing lab
(1/week)

required for
(some) SS
majors

increasing
demand
↓
larger class sizes

on campus
summer
sta 101

study abroad
+ other
demands/
engagements



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

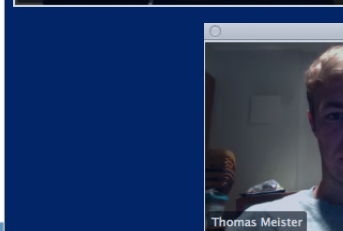
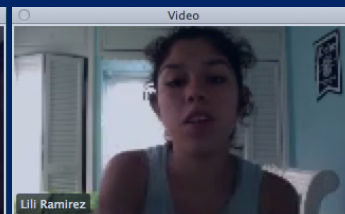
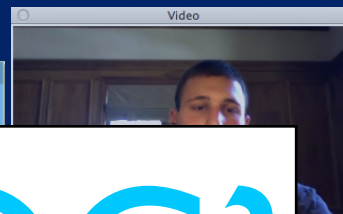
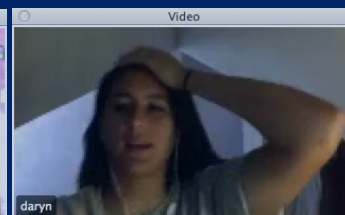
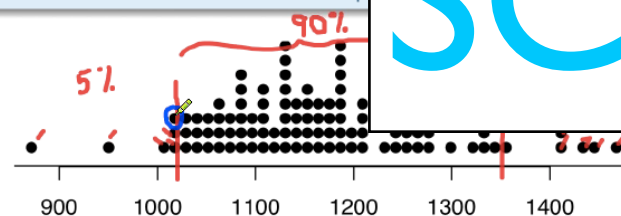
MOOC

SCOC?



Rent in Durham - bootstrap interval

The dot plot below shows the distribution of bootstrap samples from the original sample. Estimate a confidence interval based on this bootstrap distribution.



students

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
work load

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)

logistics

Network Recording Player - Sta 104 - S13 - June 24


Meeting number : 312 013 435 Date : Monday, June 24, 2013 Time : 1:00 PM, America/New_York (GMT -4:00)

Keynote File Edit Insert Slide Format Arrange View Play Window Share Help

Review

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.

Statistics 104 (Mine Cetinkaya-Rundel) Final Review June 24, 2013 3 / 24

Polling

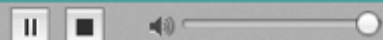
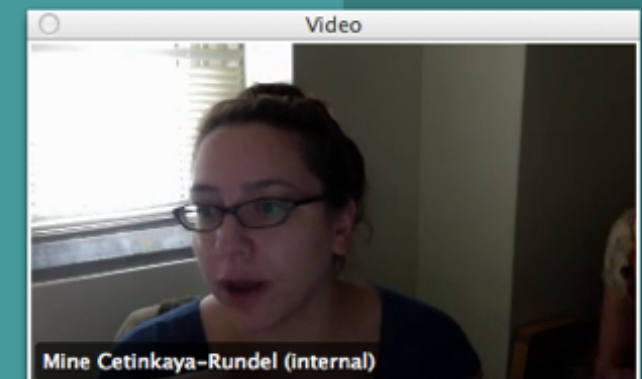
----- Start:01:06:34 PM-----

1. Your Question
(Apply:01:07:34 PM)

a.a	0/5 (0%)
b.b	3/5 (60%)
c.c	2/5 (40%)
d.d	0/5 (0%)
e.e	0/5 (0%)

Participants

Name	Joining Time / Leaving Time
	01:00 PM / 02:34 PM
re	01:00 PM / 02:34 PM
	01:00 PM / 02:34 PM
(l)	01:00 PM / 02:34 PM
g	01:00 PM / 02:34 PM
er	01:00 PM / 02:34 PM



00:08:30 / 01:33:37

logistics

Duke
UNIVERSITY

Department of
Statistical Science

home

syllabus

schedule

resources

support

links

faq

Sta 104 - Data Analysis and Statistical Inference - Online (Summer 2014) Dr. Çetinkaya-Rundel

Tentative schedule:

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

logistics

Introduction



Part 1: Designing studies

(1) Data basics

Unit 1, Part 1: (1) Data Basics

country	cr_req	cr_comply	ud_req	ud_comply	...	hemisphere	hdi
Argentina	21	100	134	32	...	southern	very high

pre-course
meeting

last week
of
classes

meet
&
greet

technical
details

Rent in Durham

The dot plot below shows 1000 samples from the original sample. The 95% bootstrap confidence interval based on this plot is



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

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

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 $\sqrt{\frac{p(1-p)}{n}}$.

course format

Single population proportion Hypothesis testing for a proportion

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 $\pm 3\%$. 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

$11\% \pm 3\% = (8\%, 14\%)$
 $H_0: p = 10\%$
null value is in CI, then fail to reject

Polling















----- Start:01:39:01 PM-----

2.Your Question
(Apply:01:40:45 PM)

a.Your Answer	2/7 (29%)
b.Your Answer	2/7 (29%)
c.Your Answer	2/7 (29%)
No Answer	1/7 (14%)

Statistics 104 (Mine Çetinkaya-Rundel) U5 - L1: Inf. for prop.s - theoretical June 10, 2013 14 / 30

course format

- Home 
- Announcements 
- Schedule 
- Piazza 
- Tests & Quizzes 
- Assignments **
- Resources 
- Syllabus 
- Gradebook 
- Statistics 
- Site Info 
- Polls 
- Email 
- Help 

STA.104.01.1Su14: Assignments

[Add](#) [Assignment List](#) [Grade Report](#) [Student View](#) [Reorder](#) [Permissions](#) [Options](#)

Application activity - 7.2 CI for Slope in MLR - Submissions

[Download All](#) | [Upload All](#) | [Release Grades](#)



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

Viewing 1 - 3 of 3 items

Please select default grade: [Apply](#)

[|<](#) [<](#) [Show 200 items...](#) [>](#) [>|](#)

▶ Select Group(s) and Allow Resubmission

<input type="checkbox"/>	 Group 	Submitted	Status	Grade	Release
<input type="checkbox"/>	Dumbo Octopus Kathleen Axelrod , Emma Bunting , Kara Fisher , Amari Stokes Grade	Jun 19, 2014 2:15 pm	Returned	Checked	✓
<input type="checkbox"/>	Mean Devils Hayley Bohart , Tyler Deane-Krantz , Leah Reisman , Shelby Wailes Grade	Jun 19, 2014 2:16 pm	Returned	Checked	✓
<input type="checkbox"/>	Stats Super Stars Tierney Bishop , Caleb Ellis , Kara Wilson Grade	Jun 19, 2014 2:15 pm	Returned	Checked	✓

▶ Assignment Details

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

$n = 30$	Jacob	Shelby	Daryn
\bar{x}	10.05	10.1	10.2
$p - value$			
$n = 5000$	Thomas	Cece	Lili
\bar{x}	10.05	10.1	10.2
$p - value$			

support

office hours
twice a week
on WebEx

summer 2014:
TA through
BASS Online
Apprentice
Program

impromptu
meetings
on Google
Hangouts

student
formed
study groups
on Google
Hangouts

student
feedback

13/16
pace is
about right

9/16
learned most
in sync.
sessions

“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
✓

scalable ?

exams on
Sakai ?

one-on-one
student
support ?

WebEx for teaching



video feed
from students

polling

recording

Training
Center:
breakout
sessions

WebEx for teaching



video feeds
limited to 6

annotation
tool buggy
[DocCam]

no auto
recording +
video
streaming
slow

proctoring
exams
[Lockdown
browser +
Video]

WebEx for teaching



stat.duke.edu/courses/Summer14/sta104.01-1/app_ex/app_5.1_Inference_single_prop.pdf

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 1,095 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 probable 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 (60% out of 1,200). Does this data provide convincing evidence that majority of Filipinos experience emotions daily?
 - (a) How do the conditions and the calculation of the standard error for a hypothesis test (Question 2) differ from the 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

Video

$\hat{p} = 0.36$

1. indep: random sample
1095 < 10% all Singapore

2. S/E: $1096 \times 0.36 > 10$ ✓
 $1096 \times 0.64 > 10$ ✓

$SE = \sqrt{0.36 \times 0.64}$

Mine Cetinkaya-Rundel

don't need to
be tech savvy,
but should be
flexible

remember
to record
[if not automatic]

especially if class
size is large, a
tech assistant
(at least)
first few days

on call
tech help
throughout
course

thank you!

questions?

Rent in Dur

The dot plot below shows the distribution of means of 100 bootstrap samples from the original sample. Estimate the 90% bootstrap confidence interval.

