

## Stat 104: Quantitative Methods for Economics

### Homework 6: Due Monday, October 23

**Homework policy for campus students:** Homework is due by 8:00 am on the due date. Homework is to be handed in via the course website in pdf format. You do not need to type the homework; there are many ways (scanner in the library or phone apps) to convert written homework into a pdf file. Ask the teaching staff if you need assistance.

Late homework will not be accepted. You are encouraged to discuss homework problems with other students (and with the instructor and TFs, of course), but you must write your final answer in your own words. Solutions prepared “in committee” or by copying someone else’s paper are not acceptable.

- Please submit your homework in pdf format; this can be done in Word, or OpenOffice or via cellphone apps that will scan and turn into pdf.
- Please make your homework solutions legible by **bolding** or using circles to identify your solution.
- Since we are not printing out anything, use lots of s p a c e for your solutions, and put each answer on a different page if it makes the solution easier to read.
- Please make sure your submitted solutions are in numerical order [problem 1, problem 2 and so on].
- Please keep your computer output to a minimum and focus on the required answer. The easiest way to put your computer output into your homework is to cut and paste it into a Word file and use the font “courier new”.
- Please keep in mind the course rules on Academic Honesty and Collaboration

- 1) An insurance company is interested in the average claim on its auto insurance policies. Using 36 randomly selected claims, it finds the mean claim to be \$1,270 with a standard deviation of \$421. Construct a 95 percent confidence interval for the mean claim on all policies.
- 2) A random sample of the luggage of 49 passengers of Jet Blue finds that the mean weight of the luggage is 47 pounds with a standard deviation of 8 pounds. Construct a 95 percent confidence interval for the mean weight of Jet Blue Airlines luggage.
- 3) A random sample of 250 credit card holders shows that the mean annual credit card debt for individual accounts is \$1592 with a standard deviation of \$997. Use this information to construct a 92% (yes that is not a typo) confidence interval for the mean annual credit card debt for the population of all accounts.
- 4) For this problem we are going to use class survey data from a previous offering of Stat 111. Enter the following commands into R:

```
mydata=read.csv("http://people.fas.harvard.edu/~mparzen/stat100/stat111_survey.csv")
weight=mydata$weight
female=mydata$female
sleep=mydata$sleep
haircut=mydata$haircut
texts=mydata$texts
```

Note that we can find confidence intervals in R using this data as follows. For number of texts someone sends a day:

```
t.test(texts) ## ci for everyone
t.test(texts[female==1]) ## ci for just females
t.test(texts[female==0]) ## ci for just males
```

- a) Find a 95% confidence interval for the sleep variable for men and women separately. Compare the results. Are you inside your respective interval?
- b) The variable haircut is what do you usually pay for a haircut. Find a 95% confidence interval for this variable for men and women separately. Do the intervals appear that different?
- c) Find a 95% confidence interval for the variable texts, the number of texts you send per day. Are you inside this interval? Do the separate intervals for men and women differ that much?

- 5) In a newspaper or magazine of your choice (can be online), find a statistical study that contains an estimated population proportion. Include the article with your homework (via a cut and paste or copy of the article). Calculate a confidence interval for the sample proportion given.
- 6) The paralyzed Veterans of America is a philanthropic organization that relies on contributions. They send free mailing labels and greeting cards to potential donors on their list and ask for voluntary contribution. To test a new campaign they recently sent letters to a random sample of 100,000 potential donors and received 4781 donations.
  - a) Give a 95% confidence interval for the true proportion of those from their entire mailing list who may donate.
  - b) A staff member thinks that the true rate is 5%. Given the confidence interval you found, do you find that percentage plausible?
- 7) A recent Gallup poll consisted of 1012 randomly selected adults who were asked whether “cloning of humans should or should not be allowed.” Results showed that 901 of those surveyed indicated that cloning should not be allowed. Construct a 95% confidence interval estimate of the proportion of adults believing that cloning of humans should not be allowed.
- 8) A national health organization warns that 30% of middle school students nationwide have been drunk. Concerned, a local health agency randomly and anonymously surveys 110 of the middle 1212 middle school students in its city. Only 21 of them report having been drunk.
  - a) What proportion of the sample reported having been drunk?
  - b) Does this mean that this city’s youth are not drinking as much as the national data would indicate?
  - c) Create a 95% confidence interval for the proportion of the city’s middle school students who have been drunk.
  - d) Is there any reason to believe that the national level of 30% is not true of the middle school students in the city?
  - e) To keep the margin of error at most 5%, how many middle school students do we need to survey? Assume we have no prior idea what the true proportion is.
- 9) A researcher wishes to be 95% confident that her estimate of the true proportion of individuals who travel overseas is within 3% of the true proportion.
  - a) Find the sample necessary if, in a prior study, a sample of 200 people showed that 40 traveled overseas last year.
  - b) If no estimate of the sample proportion is available, how large should the sample be?
- 10) Obesity is defined as a body mass index (BMI) of 30 kg/m<sup>2</sup> or more. A 95% confidence interval for the percentage of U.S. adults aged 20 years and over who were obese was found to be 22% to 24%. What was the sample size?

- 11) When 14 different second-year medical students at Bellevue Hospital measured the blood pressure of the same person, they obtained the results listed below.

138 130 135 140 120 125 120 130 130 144 143 140 130 150

You can read this data into R by entering the command:

```
mydata=c(138, 130, 135, 140, 120, 125, 120, 130, 130, 144, 143, 140, 130, 150)
```

- (a) Using R, find the 95% confidence interval for the mean blood pressure (use the `t.test` command).
  - (b) By hand, and using the  $t$  distribution, find the 95% confidence interval for the mean score. You can use the summary statistics from R. In R, the command `qt(.975, df)` will calculate the appropriate  $t$  cut-off value, where  $df=n-1$ .
  - (c) By hand, and using the normal distribution, find the 95% confidence interval for the mean score (i.e. use “1.96”). You can use the summary statistics from R.
  - (d) Discuss the difference between (a) and (b) and (c).
- 12) Answer true or false to the following statement and give a reason for your answer: If a 95% confidence interval for a population mean,  $\mu$ , is from 33.8 to 39.0, the mean of the population must lie somewhere between 33.8 and 39.0.
- 13) If you obtained one thousand 95% confidence intervals for a population mean,  $\mu$ , roughly how many of the intervals would actually contain  $\mu$ ?
- 14) Suppose we know that “a 95% confidence interval for the mean age of all U.S. millionaires is from 54.3 years to 62.8 years.” Decide which of the following sentences provide a correct interpretation of the statement in quotes. Justify your answer.
- a) Ninety-five percent of all U.S. millionaires are between the ages of 54.3 years and 62.8 years.
  - b) There is a 95% chance that the mean age of all U.S. millionaires is between 54.3 years and 62.8 years.
  - c) We can be 95% confident that the mean age of all U.S. millionaires is between 54.3 years and 62.8 years.
  - d) The probability is 0.95 that the mean age of all U.S. millionaires is between 54.3 years and 62.8 years.
- 15) Why is a sample proportion generally used to estimate a population proportion instead of obtaining the population proportion directly?

- 16) A worker at a car manufacturer invented a new device that he believes will increase gas mileage. The current car averages 28 miles per hour. The CEO decides to put the new device on 100 of its vehicles and measure the average from that sample. If the average gas mileage from the 100 cars is significantly greater than the current average of 28, the CEO will buy 100,000 devices for its new line of cars.
- a) Is this a one or two tailed test? Explain.
  - b) Write the null and alternative hypothesis.
  - c) In this context, what would happen if the CEO made a Type I error?
  - d) In this context, what would happen if the CEO made a Type II error?

- 17) The real estate industry claims that it is the best and most effective system to market residential real estate. A survey of randomly selected home sellers in Illinois found that a 95% confidence interval for the proportion of homes that are sold by a real estate agent is 69% to 81%. Interpret the interval in this context.
- a) In 95% of the years, between 69% and 81% of homes in Illinois are sold by a real estate agent.
  - b) 95% of all random samples of home sellers in Illinois will show that between 69% and 81% of homes are sold by a real estate agent.
  - c) If you sell a home in Illinois, you have a  $75\% \pm 6\%$  chance of using a real estate agent.
  - d) We are 95% confident that between 69% and 81% of homes in this survey are sold by a real estate agent.
  - e) We are 95% confident, based on this sample, that between 69% and 81% of all homes in Illinois are sold by a real estate agent.
- 18) Each of the following paragraphs calls for a statistical test about a population mean  $\mu$ . State the null hypothesis  $H_0$  and the alternative hypothesis  $H_a$  in each case.
- (a) The diameter of a spindle in a small motor is supposed to be 5 mm. If the spindle is either too small or too large, the motor will not work properly. The manufacturer measures the diameter in a sample of motors to determine whether the mean diameter has moved away from the target.
  - (b) Census Bureau data show that the mean household income in the area served by a shopping mall is \$42,500 per year. A market research firm questions shoppers at the mall. The researchers suspect the mean household income of mall shoppers is higher than that of the general population.
  - (c) A study in 2002 established the mean commuting distance for workers in a certain city to be 15 miles. Because of the westward spread of the city, it is hypothesized that the current mean commuting distance exceeds 15 miles. A traffic engineer wishes to test the hypothesis that the mean commuting distance for workers in this city is greater than 15 miles.
- 19) The fundraising officer for a charity organization claims the average donation from contributors to the charity is \$250.00. To test the claim, a random sample of 100 donations is obtained. The sample yielded a sample mean of \$234.85 and sample standard deviation of \$95.23. State and run the appropriate hypothesis test using the confidence interval approach. Clearly state your conclusion.
- 20) You want to test whether your candidate's approval rating has changed from the previous dismal 40% after a major policy announcement. You run a survey and 170 out of a random sample of 500 voters approve of your candidate. ( $\hat{p} = 34\%$ ). Construct a hypothesis test using a two sided confidence interval to test if the approval rating is now different from 40%. Clearly state your conclusion