Problem Set #8

Econ 103

Midterm Announcements

- ☐ Midterm will only cover material **since the first midterm** (Chapter 4-7 slides)
- ☐ The only R code you are responsible for writing is the commands related to distributions (rnorm, dnorm, qnorm, pnorm, rchisq, etc.). Make sure you know what arguments each of those takes and how to decide when to use "r", "d", "q", or "p".
- ☐ You are not responsible for any R material outside of what has been covered in the lecture slides. For example, you will not be tested on RMarkdown. I want you to focus on studying the slides and getting comfortable with the material.

Lecture Progress

We made it to slide 70 of the Chapter 7 slides.

Homework Checklist

- ☐ **Book Problems (Chapter 8):** 1, 3, 5, 7, 9
- ☐ **Additional Problems:** See below
- ☐ **Practice Midterms:** Do them!
- ☐ **Ask questions on Piazza**
- ☐ **Review slides**
- ☐ **R Tutorial:** Use RMarkdown to write a description of your data (will be a section of your R Project) as well as load your data and summarize some statistics of interest. (Not due until the week after the midterm)

Additional Problems

1. For this question assume that we have a random sample from a normal distribution with unknown mean but *known* variance.
   (a) Suppose that we have 36 observations, the sample mean is 5, and the population variance is 9. Construct a 95% confidence interval for the population mean.
   (b) Repeat the preceding with a population variance of 25 rather than 9.
   (c) Repeat the preceding with a sample size of 25 rather than 36.
   (d) Repeat the preceding but construct a 50% rather than 95% confidence interval.
   (e) Repeat the preceding but construct a 99% rather than a 50% confidence interval.

2. In this question you will carry out a simulation exercise similar to the one I used to make the plot of twenty confidence intervals from lecture 16.
   (a) Write a function called `my.CI` that calculates a confidence interval for the mean of a normal population when the population standard deviation is known. It should take three arguments: `data` is a vector containing the observed data from which we will calculate the sample mean, `pop.sd` is the population standard deviation, and `alpha` controls the confidence level (e.g. `alpha = 0.1` for a 90% confidence interval). Your function should return a vector whose first element is the lower confidence limit and whose second element is the upper confidence limit. Test out your function on a simple example to make sure it's working properly.
   (b) Write a function called `CI.sim` that takes a single argument `sample.size`. Your function should carry out the following steps. First generate `sample.size` draws from a standard normal distribution. Second, pass your sample of standard normals to `my.CI` with `alpha` set to 0.05 and `pop.sd` set to 1. Third, return the resulting confidence interval. Test your function on a sample of size 10. (What we're doing here is constructing a 95% confidence interval for the mean of a normal population using simulated data. The population mean is in fact zero, but we want to see how our confidence interval procedure works. To do this we "pretend" that we don't know the population mean and only know the population variance. Think about this carefully and make sure you understand the intuition.)
   (c) Use `replicate` to construct 10000 confidence intervals based on simulated data using the function `CI.sim` with `sample.size` equal to 10. (Note that `replicate` will, in this case, return a matrix with *2* rows and *10000* columns. Each column corresponds to one of the simulated confidence intervals. The first row contains the lower confidence limit while the second row contains the upper confidence limit.) Calculate the proportion of the resulting confidence intervals contain the true population mean. Did you get the answer you were expecting?

(d) Repeat the preceding but rather than using `CI.sim` write a new function called `CI.sim2`. This new function should be identical to `CI.sim` except that, when calling `my.CI`, it sets `pop.sd = 1/2` rather than 1. How do your results change? Try to provide some intuition for any differences you find.

3. Oranges sold at Iovine Brothers Produce in Reading Terminal Market have weights that follow a normal distribution with a mean of 12 ounces and standard deviation of 2 ounces.

   (a) If we choose an orange at random, what is the probability that it will weigh less than 10 ounces?

   (b) If we choose 25 oranges at random, what is the probability that they will have a total weight of less than 250 ounces?

4. All other things equal, how would the following change the width of a confidence interval for the mean of a normal population? The population standard deviation is unknown. Explain.

   (a) The sample mean is smaller.

   (b) The population mean is smaller.

   (c) The sample standard deviation is smaller.

   (d) The sample size is smaller.

5. Do you agree or disagree with the following statement: "the household unemployment survey is hardly flawless; its 60,000 families constitute less than 0.1% of the workforce." Explain your answer.

6. Suppose you want to construct a 99% confidence interval for the average height of US males above the age of 20. Based on past studies you think the standard deviation of heights for this population is around 6 inches. How large a sample should you gather to ensure that your confidence interval has a width no greater than 1 inch?

7. A well-known weekly news magazine once wrote that the width of a confidence interval is inversely related to sample size: for example, if a sample size of 500 gives a confidence interval of plus or minus 5, then a sample of 2500 would give a confidence interval of plus or minus 1. Explain the error in this argument.