**Statistical Inference for Means and Proportions of a Single Population**

The purpose of this activity is to practice producing and interpreting confidence intervals and hypothesis testing (test statistics, p-values) for both means and proportions, both by hand and in R.

**Inference for Proportions**

**Example #1**:

**a.** Earlier in the course, we used the formula to determine the margin of error for a 95% confidence interval. That formula provides what is called the “conservative” estimate of the margin of error.

Suppose that we have a survey based on a randomly selected sample of 400 people. Find the value of the conservative estimate of the margin of error. Don’t convert the answer to a percent.

**b.** The technically precise determination of the margin of error for a 95% confidence interval is

Margin of error = 2 standard errors =. Suppose that we have a sample of *n* = 400 and the sample proportion with a trait of interest is = 0.5. Calculate.

**c.** Now suppose that we have n = 400 and = 0.8. Calculate 

**d.** Now suppose that we have n = 400 and= 0.35. Calculate 

**e.** Compare the answers to parts b, c, and d to the answer for part a.

(i) When does the margin of error given by the conservative formula (part a) match the more technically precise calculation?

(ii) What does the comparison of parts b, c and d to part a show about why is called a conservative margin of error.

**Example #2**:

A randomly selected sample of 900 U.S. adults is surveyed to determine the proportion that favors stricter penalties for drunk driving.

**a.** Describe the parameter of interest in this situation.

**b.** In the sample of 900 people, 657 individuals are in favor of stricter penalties for drunk driving.

Determine the sample proportion that favors stricter penalties.

**c.** Calculate the standard error of the sample proportion. The formula is .

**d.** Calculate a 95% confidence interval that estimates the population proportion in favor of stricter penalties for drunk driving. [Sample proportion ± 2 standard errors]

**e.** Write a sentence that interprets the 95% confidence interval calculated in the previous part.

**f.** On the basis of the 95% confidence interval computed in this activity, explain whether it is reasonable to conclude that more than 60% of American adults favor stronger penalties for drunk driving.

**g.** Now produce all of the formal hypothesis testing for whether the proportion of American adults who favor stronger penalties for drunk driving exceeds 60%.

H0:

HA:

z\* =

p-value (use pnorm(z\*,lower.tail=FALSE)) =

conclusion:

**h.** Do you get the same p-value from R? Do this test via prop.test(x=657,n=900,p=.6,alternative=’greater’).

**Example #3**:

In the RStudio website folder for today, you’ll find a dataset called CIs\_ex3.R, which you can load using

load(file.choose())

and then navigating to the .R datafile on the server. The data now exists in R, and you can see it in the Environment tab of the upper-right hand corner pane in RStudio.

The data are from a survey (54 variables) of 1485 students from PSU in 2007. Let’s explore a quantitative variable: HrsStudy, which of course asks students how many hours they study per week.

**a.** We can get a confidence interval that estimates the unknown value of a population mean. We’ll get the interval and interpret it. Use

t.test(CIs\_ex3$HrsStudy)

The output will give a 95% confidence interval. Report the interval.

**b.** On the basis of the confidence interval calculated in part a, can we say that, on average, Penn State students study less than 15 hours per week? Explain.

**c.** Now produce all of the formal hypothesis testing for whether or not, on average, Penn State students study less than 15 hours per week.

H0:

HA:

z\* =

p-value (use pnorm(z\*,lower.tail=FALSE)) =

conclusion:

**d.** Have R confirm your calculations, via t.test(CIs\_ex3$HrsStudy,mu=15,alternative=’less’).

**c**. Use R to estimate the mean amount that Penn State students spent on textbooks in the Spring 2007 semester. The relevant variable in the dataset is **TextSpd**.

What is the sample mean? What is the correct statistical notation for this sample mean?

Write the 95% confidence interval given on the output.

Write a sentence that interprets the 95% confidence interval in this case. The sentence structure is “We are 95% confident that the mean {describe the characteristic} for {describe the population} is between \_\_\_ and \_\_\_.”