Central Limit Theorem Applications

YOUR NAME

21 October, 2020

Exercises

1. Suppose we roll a fair six-sided die and let X be the resulting number. The distribution of X is discrete uniform. (Each of the six discrete outcomes is equally likely.)

a. Suppose we roll the fair die 5 times and record the value of $\bar{X}$, the *mean* of the resulting rolls. Under the central limit theorem, what should be the distribution of $\bar{X}$?
b. Simulate this process in `R`. Plot the resulting empirical distribution of $\bar{X}$ and report the mean and standard deviation of $\bar{X}$. Was it what you expected?

(HINT: You can simulate a die roll using the `sample` function. Be careful and make sure you use it properly.)

c. Repeat parts a) and b) for $n = 20$ and $n = 50$. Describe what you notice. Make sure all three plots are plotted on the same x-axis scale. You can use facets if you combine your data into one `tibble`.

2. The nutrition label on a bag of potato chips says that a one ounce (28 gram) serving of potato chips has 130 calories and contains ten grams of fat, with three grams of saturated fat. A random sample of 35 bags yielded a sample mean of 134 calories with a standard deviation of 17 calories. Is there evidence that the nutrition label does not provide an accurate measure of calories in the bags of potato chips? The conditions necessary for applying the normal model have been checked and are satisfied.

The question has been framed in terms of two possibilities: the nutrition label accurately lists the correct average calories per bag of chips or it does not, which may be framed as a hypothesis test.

a. Write the null and alternative hypothesis.
b. What level of significance are you going to use?
c. What is the distribution of the test statistic $\frac{\bar{X}-\mu}{S/\sqrt{n}}$? Calculate the observed value.
d. Calculate a p-value.
e. Draw a conclusion.

3. Exploration of the chi-squared and t distributions.

a. In `R`, plot the pdf of a random variable with the chi-squared distribution with 1 degree of freedom. On the same plot, include the pdfs with degrees of freedom of 5, 10 and 50. Describe how the behavior of the pdf changes with increasing degrees of freedom.

b. Repeat part (a) with the t distribution. Add the pdf of a standard normal random variable as well. What do you notice?

4. In this lesson, we have used the expression *degrees of freedom* a lot. What does this expression mean? When we have sample of size n, why are there $n - 1$ degrees of freedom for the t distribution? Give a short concise answer (about one paragraph). You will likely have to do a little research on your own.

5. Deborah Toohey is running for Congress, and her campaign manager claims she has more than 50% support from the district's electorate. Ms. Toohey's opponent claimed that Ms. Toohey has **less** than 50%. Set up a hypothesis test to evaluate who is right.

a. Should we run a one-sided or two-sided hypothesis test?

b. Write the null and alternative hypothesis.

c. What level of significance are you going to use?

d. What are the assumptions of this test?

e. Calculate the test statistic.

f. Calculate a p-value.

g. Draw a conclusion.

Note: A newspaper collects a simple random sample of 500 likely voters in the district and estimates Toohey's support to be 52%.