Homework Week #5

Problem 1

(a) Consider a sample of 20 people for a random variable X which is a feeling thermometer for President Biden. The variable can take on values 0-100, where 0 is "cold" feelings and 100 is "hot". In the particular sample that we received the mean of this variable is $\bar{x} = 53$ and the sample standard deviation $s_x = 7$.

Perform a hypothesis test for the null hypothesis that Americans on average are ambivalent about Biden, feeling neither warm nor cool on average (i.e. the population mean, μ, equals 50). In this test you should perform each of the 5 steps laid out in lecture, and make use of the qt()/pt() functions.

(b) Next, load in the 'ACSCountyData.Rdata' dataset. Run a regression where 'percent.college' is the independent variable and 'median.income' is the dependent variable.

(c) Looking at the output of the regression, first interpret what the estimated coefficeint for percent.college means.

(d) What null hypothesis is being tested automatically for percent.college? What is the alternative hypothesis? What is the result of that hypothesis test?

(e) Using the output of the model (i.e. the coefficient, the standard error, and the degrees of freedom) and the qt()/pt() functions, confirm the automatic hypothesis test is correct by completing the same 5-step hypothesis testing sequence you completed in (a).

Problem 2

(a)Create a new variable 'x' using the command rnorm(100, mean=0,sd=7). This creates a new sample that has an n of 100. Note that the expected value of the population we are drawing from is $\mu = 0$ and a population variance $\sigma^2 = 49$.

(b) Use the t.test function to perform a t.test on this new variable for the null hypothesis that $\mu = 0$. Report the p-value for this hypothesis test. (Note, the defualt for the t.test function is to test whether $\mu = 0$.)

(c) Next, run a loop that completes this same process 1000 times: sampling from rnorm(100, mean=0, sd=1), performing a t.test, and comptuing the p.value for the null hypothesis that $\mu = 0$. Each time through the loop save the p-value that was calcualted. (To do this, I first save the results of the t.test to an object, and then access the p value that's in that object: $t <- $ t.test(x); t$p.value).

(d) How many of the 1000 p-values you calculated were "statistically significant" (i.e. less than .05)? What does this tell you about α, the probability of getting a false positive?

(e) Complete the same process as above, but this time sample 1000 times from rnorm(100, mean=2, sd=7). That is, sample 100 observations 1000 times from the normal distribution with $\mu = 2$ and $\sigma^2 = 49$. You should again produce 1000 p-values which test the null hypothesis that $\mu = 0$.

(f) How many of the 1000 p-values you calculated were NOT statistically significant (i.e. are greater than .05)? What does this tell you about β, the probability of getting a false negative.

(g) Run the loop one more time, but modify the sampling command (rnorm(100,mean = 2,sd=7)) in one way that results in a lower β. Describe the change you made and why it leads to a lower false negative rate.