Week 5 Practice Exam (#2)

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Instructions: This is a “low stakes” (i.e., not graded) learning assessment of your comprehension of the first four weeks of this course*.* Compose brief answers to each of the following questions, typing your response in *italics* below each question.

1. Your boss at the social media marketing company asks you to conduct another A/B test on two different social media ad configurations. Each of the two ads is displayed on n=96 high traffic social media pages:   
     
   The A banner gets an average of 1373 clicks per hour.   
   The B banner gets an average of 1394 clicks per hour.   
     
   The 95% confidence interval is as follows:   
    -23 < (mean difference between A and B) < 17.   
     
   Answer the following questions about that confidence interval:   
   1. Does this particular confidence interval contain the population mean difference?
   2. Which banner ad do you prefer (A or B) and why?
   3. Your boss tells you to run the same experiment 99 more times, calculating a new confidence interval each time. Now you have a collection of 100 confidence intervals, each of which was constructed in the same way, but from new data samples: What can you say about this collection of confidence intervals?

* 1. Which command in R would you use to produce the confidence interval for each of the 100 that you constructed?

1. Some output appears from a t-test that compared annual U.K. driver fatalities for several years before and after a seat belt law was enacted. Interpret these results in a brief paragraph, making sure to explain as much of the statistical output as you can:  
     
   **Welch Two Sample t-test**

**data: FatalitiesPreLaw and FatalitiesPostLaw**

**t = 5.1253, df = 29.609, p-value = 1.693e-05**

**alternative hypothesis: true difference in means is not equal to 0**

**95 percent confidence interval:**

**15.39892 35.81899**

**sample estimates:**

**mean of x mean of y**

**125.8698 100.2609**

1. Explain the following diagram, which was created from these five lines of code:  
     
   **x <- seq(from=-3,to=3,by=.1)  
   plot(x, dt(x,df=30))  
   abline(v=-2.04)  
   abline(v=2.04)  
   abline(v=2.5,col="green")**  
     
   Hint: dt() is the probability density function of the t-distribution, so the total area under the curve equals 1. The -0.025 quantile for t, with 30 degrees of freedom, is -2.04. Make sure to explain what the green line might represent and the consequences of its position on the extreme right of the diagram.

*4.* Imagine that you just ran the following R code:

X1 <- c(32, 48, 23, 23, 23, 21, 28)

X2 <- c(51, 32, 33, 50, 26, 66, 27)

df <- data.frame(mpg=X1, wt=X2)

1. Fill in the data table below so that it resembles what you would see as a result of running the R-Studio command “View(df)”. Make sure to fill in the column labels!

|  |  |  |
| --- | --- | --- |
| *Observation* | *X1* | *X2* |
| *1* |  |  |
| *2* |  |  |
| *3* |  |  |
| *4* |  |  |
| *5* |  |  |
| *6* |  |  |
| *7* |  |  |

1. In the data set for the previous question, would it or would it not make sense to run a t-test comparing df$mpg and df$wt? Briefly explain why.