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?

*Maybe , We are 95 percent confident that the population mean is contained -3*

* 1. Which banner ad do you prefer (A or B) and why?

*We cant really tell which one to prefer because the interval crosses 0. Because the differences do not appear to be statistically significant*

* 1. 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?

*If we created 100 confidence intervals, we know that roughly 5 of them would not contain the true population mean click difference between A and B and we also know that roughly 95 of them would contain the true population mean click difference between A and B.*

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

*t.test*

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** *p values is less than .05 so it is statistically significant*

T stastitic is 5 can tell it is sifniticant

Df 29 combined number of observations

*We can say with 95 percent confidence the population mean will fall between 15.39 and 35.81   
  
mean of fatalitiesprelaw – 125.87 is mean of fatalities pre law*

*FatalititiesPostLaw – 100.26 is mean*

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.

*X Sequence that goes from -3 to 3 by .1*

*Plot x on the x axis, take values of x and plug it into the t distribution*

*Abline plots the confidence intervals*

*Degrees of freedom raises the graph and makes it thinner*

*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* | *mpg* | *wt* |
| *1* | *32* | *51* |
| *2* | *48* | *32* |
| *3* | *23* | *33* |
| *4* | *23* | *50* |
| *5* | *23* | *26* |
| *6* | *21* | *66* |
| *7* | *28* | *27* |

1. Next, review the R code in each of these boxes and write in the box what you would see if you ran that code at the console after creating df with the code above. There is no need for a calculator for any of these items. There are no “trick” questions. All commands below run without error.

|  |  |  |  |
| --- | --- | --- | --- |
| length(X1)  7 | length(df$mpg)  7 | median(X1)  23 | max(X2)  66 |
| length(df$mpg)== length(df$wt)  *True* | min(df$wt)  26 | X2[1]  51 | df$wt[1]  51 |

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.