This quiz will make use of an R dataset called ‘mpg’.

You will first need to load in the ggplot2 library with the following code:

library(ggplot2)

mpg<-as.data.frame(mpg)

Answer 5 of the following 7 questions, each question is worth two points

1. Calculate the mean, variance, standard deviation, and standard error of the mean of city millage of all cars (‘cty’ column). *Answer in R script*
2. What does α refer to in statistics? What trade-offs do we make when we change the value of α? Be sure to use the terms ‘Type I’ and ‘Type II’ error in your answer. *Answer in .doc*
3. Use a Chi-Square test to answer the following question: are the proportions of the different classes of cars (2seater, compact, midsize, minivan, pickup, subcompact, suv) the same in year 1999 and 2008? *Answer in R script*

When I ran this chi-square I got a warning that said “…Chi-square approximation may be incorrect” This is because some of the numbers are small. Don’t worry if you get that warning, it doesn’t mean you’ve done something wrong.

1. Use a t-test to test for differences between city millage and highway millage. Make a boxplot to visualize this data. What is the null hypothesis of this test? Should you accept or reject the null hypothesis? Since the t-test is relatively robust to the failed assumption of normality, you may ignore that assumption in this case if you need to. *Answer in R script*
2. Use an ANOVA to test for differences between city millage of cars with different numbers of cylinders. Make a boxplot to visualize this data. What is the null hypothesis of this test? Should you accept or reject the null hypothesis? *Answer in R script*
3. Make a bar chart of the mean city millage (cty column) of cars split by class (2seater, compact, midsize, minivan, pickup, subcompact, suv). Make sure the bars are labeled. *Answer in R script*
4. Design a simple experiment to test effect of temperature on plants. What are the independent, dependent, and control variables? Are your independent and dependent variables categorical, numeric, or ranked? Briefly explain. What are your treatments? What statistical test will you use to analyze your data? *Answer in .doc*

Please e-mail me your completed document (s).