`

STAT 614 FINAL EXAM Fall 2020

1) Use and show R code to import the *VehicleData2* excel spread sheet. Assign your imported data the name vd2 3pts

2a) Use and show R code to create a correlation matrix for vd2. The response variable is **MilesperGallon.** 3pts

2b) Is the response variable highly correlated with all explanatory variables? If not, indicate

which explanatory variables are not highly correlated with the response variable. 3pts

2c) For which pairs of explanatory variables is multicollinearity a problem? 3pts

3a) Use and show R code to produce a multiple regression model for a response variable of

**MilesperGallon** and explanatory variables **CurbWeight** and **EngineSize**. 4pts

Write the model showing the intercept and the coefficients for both explanatory

variables. 2pts

3b) Now use your model to predict **Milespergallon** if **EngineSize** is 3.65 and **CurbWeight** is 4000. (Show your work) 3pts

4) Use and show R code to produce a summary table for the model that you have created. 3pts

5a) For this model, can you reject the null hypothesis that the population coefficient for **EngineSize** is 0? Justify your answer. 3pts

5b) According to the summary table, what is the proportion of variability in the response variable that is explained by the model using only explanatory variables that affect or impact on the response variable? 3pts

5c) What is typical distance between a sampled coefficient for **Curbweight** and the population coefficient for **Curbweight**? 3pts

5d) Find a 95% confidence interval for the **EngineSize** coefficient. (Show all steps and required R code) 5pts

6) Add another variable to your model ; **Cylinders**. Create a summary table for your model that now has three explanatory variables. Do you now have a model that does a better job predicting **MilesperGallon**? Justify your answer. 5pts

7) Now consider a regression model that only has one explanatory variable, **EngineSize**. Create a summary table for this model and compare the results to the previously developed model types with two and three explanatory variables. Pick the best model to use for predicting **MilesperGallon**. Give a detailed answer (6 or 7 sentences) using relevant information from each summary table. 7pts

8a) Use and show R code to import the excel spread sheet *VehicleData.* (This data table has a dummy or indicator variable). Assign the data table the name vd 3pts

8b) Use and show R code to develop a model produced to predict MilesperGallon using the variables CurbWeight, EngineSize and ForiegnDomestic. Indicate the intercept and the coefficients for all explanatory variables. 6pts

8c) For the categorical variable, let the level of Foreign be assigned 1 and the level of Domestic be assigned 0. Use your model to predict **Milespergallon** for a **CurbWeight** of 5000, an **EngineSize** of 3.7 and a vehicle that is Domestic. 3pts

9) Explain in detail the difference between One Way Anova and Two Way Anova (5 or 6 sentences). 4pts

10) A manufacturing researcher wants to determine if age or gender significantly affects the time required to learn an assembly line task. He randomly selected 24 adults aged 20 to 64 years old of whom 8 are 20 to 34 years old (4 males 4 females), 8 are 35 to 49 years old (4 males 4 females) and 8 are 50 to 64 years old (4 males 4 females). He then measured the time (in minutes required to complete a certain task. The data obtained is shown are organized as follows:

**Gender Age(years)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **20 - 34** | **35 - 49** | **50 - 64** |
| Male | 5.2 , 5.1 | 4.8 , 5.8 | 5.2 , 4.3 |
|  | 5.7 , 6.1 | 5.0, 4.8 | 5.5 , 4.7 |
| Female | 5.3 , 5.5 | 5.0 , 5.4 | 4.9 , 5.5 |
|  | 4.9 , 5.6 | 5.6 , 5.1 | 5.5 , 5.0 |

a) Prepare the data for R input as a data table. Give the input table the name xyz

The first four rows are given below: Complete the table !! 6pts

tribble(~Gender, ~Agerange, ~timeminutes,

"Male", "20to34", 5.2,

"Male", "20to34", 5.1,

"Male", "20to34", 5.7,

"Male", "20to34", 6.1,

. . .

. . .

b) Use and show R code to formally convert the character variables to factors (indicate the levels for each factor) 4pts

c) Use and show R code to produce a Two Way ANOVA Table. (be certain to introduce an interaction term) 4pts

d) Provide a four or five sentence explanation for interaction between two explanatory variables. In other words, if there is interaction between two explanatory variables, what does this mean. 4pts

e) Should the null hypothesis be rejected for the interaction term? Justify your answer. 3pts

f) Use and show R code to produce an interaction plot. Provide an interpretation of your interaction plot. 5pts

**Extra Credit**: 5pts

11) The logistic regression model below can be used to find the probability that a vehicle from the **mtcars** data table will have an automatic transmission (**am**):

Find the probability that a vehicle from the mtcars data table will have an automatic transmission if mpg = 18, hp = 115, cyl = 6, and vs = 1. (Show all of your work)

ln(P/1 – P) = 82.4961 + 3.3509**mpg** + 0.4187**hp** – 29.0369**cyl** – 78.2685**vs**