**BIOS 591P, Spring 2023: HW 6**

**Read Chapter 14 in the textbook and answer the questions below, submitting the answers on Canvas by 3 PM on Thursday, 4/6/23. You may write or type your answers into this document.**

Q1. Using a single data set containing data on dependent variable *Y* and independent variables *X*1, *X*2 and *X*3, with no missing values for any of the variables, three regressions were run. The regression models, and their associated ANOVA tables are shown below:

Model 1: *Y* = **0 + **2*X*2 + *E*

Source DF SS MS *F*

----------- ---- ------------- ------------- ------

Regression

Error 100

---------------------------------------------

Total 51 125

Model 2: *Y* = **0 + **1*X*1 + **2*X*2 + *E*

Source DF SS MS *F*

----------- ---- ------------- ------------- ------

Regression 75

Error

---------------------------------------------

Total

Model 3: *Y* = **0 + **1*X*1 + **2*X*2 + **3*X*3 + *E*

Source DF SS MS *F*

----------- ---- ------------- ------------- ------

Regression

Error 25

---------------------------------------------

Total

1. Fill in the missing ANOVA table values. Final answers must be accurate to 2 decimal places.

For parts b and c, carefully show all of the following details of the tests in the spaces provided on the

next three pages

* Hypotheses
* Test statistic computation (show your work!)
* Distribution of the test statistic under H0
* P-value determination (show your work!)
* Decision (reject H0 or fail to reject H0) (use α=0.05)
* Conclusion: elaborate –what does the decision mean?

b. Perform the overall F-test for model 3

c. Three partial F-tests would be performed for model 3. Perform the tests, if possible; if not possible,

state why.

**Q2b Overall F-test for Model 3:**

**H0:**

**HA:**

**α = 0.05**

**Test Statistic: F =**

**P-value = Pr(F > value calculated above) From Kleinbaum’s F-tables (using 3 and 40 df to be conservative, since 3 and 48 are not available in our tables), p-value < 0.001**

**Decision:**

**Conclusion:**

**Q2c. Partial Tests**

**There are three partial tests that one would want to perform for model 3. Show the details below, where possible.**

**If not possible, explain why.**

**First Partial F test:**

**H0:**

**HA:**

**F =**

**p-value =**

**Decision:**

**Conclusion:**

**Second Partial F test:**

**H0:**

**HA:**

**F =**

**p-value =**

**Decision:**

**Conclusion:**

**Third Partial F test:**

**H0:**

**HA:**

**F =**

**p-value =**

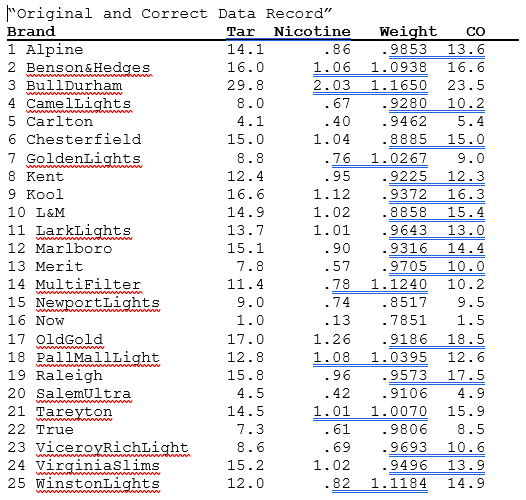
**Decision:**

**Conclusion:**

**For Q2 below, write or type your answers into this document**

**Upload the completed document into the ‘HW 6’ assignment item on Canvas**

# Q2. The Federal Trade Commission annually rates varieties of domestic cigarettes according to their tar, nicotine, and carbon monoxide content. The US Surgeon General considers each of these substances hazardous to a smoker's health. Past studies have shown that increases in the tar and nicotine content of a cigarette are accompanied by an increase in the carbon monoxide emitted from the cigarette smoke. The data set for this question (text file ‘cigarette.dat’ –available on Canvas) contains measurements of weight and tar, nicotine, and carbon monoxide (CO) content for 25 brands of cigarettes. The data were entered into the text file from the original source data table, which is shown below (and which you will need for parts of this exercise).



1. A multiple linear regression of carbon monoxide regressed on tar content, nicotine content and weight was performed. The separately provided output includes key results for this regression. (Note: for your own benefit, you should try to reproduce this output using SAS and R).

Use the provided output to identify outliers, and to list the outliers in the table below. Also, complete the “Criterion” column by writing in the cutoff value for each statistic:

Outliers (write the BRAND value for each outlying observation and

Statistic Criterion indicate whether any action is warranted)

======= =========== =================================================

Leverage hi ≥

Jacknife

Resids. |r-i | ≥

Cooks

Distance Di ≥

1. Using the provided output, check for multicollinearity problems and gross violations of the linearity, homoscedasticity and normality assumptions for the regression described in part a, then complete the following table:

Assumption/ Gross Violation or Other Plots and/or statistics Used To

Problem Problem? (Circle Answer) Assess Assumption

============ =================== ===========================

Multicollinearity Yes No

Linearity Yes No

Homoscedasticity Yes No

Normality Yes No

1. For any problems or **gross** violations that you detected in part (b), **briefly** name a potential remedy (other than ‘consult a statistician’)