36-220 Lab #11
Multiple Regression

Please write your name below, tear off this front page and give it to a teaching assistant as you leave the lab. It will be a record of your participation in the lab. Please remember to include whether you are in Section A, B or C. Keep the rest of your lab write-up as a reference for doing homework and studying for exams.

Name:

Section:

- The symbol ♣ at the beginning of a question means that, after you answer that question, you should raise your hand and have either the TA or lab assistant review your answer. Once they have reviewed your work they will place a check in the appropriate space in the table below. The purpose of this check is to be sure you have answered the question correctly.
- **You should try to complete as much of the lab exercise as possible. We understand that students work at different paces and have tried to structure the exercise so that it can be completed in the allotted time. If you work systematically through the handout and still don't complete every question don't worry. The important thing is that you understand what you are doing. Nonetheless, you are encouraged to complete the lab on your own.**

Check-Problem ♣	Instructor's Initials
Question 6	
Question 11	
Question 13	

36-220 Lab #11
Multiple Regression

First let's download the dataset we will need for the lab. Download gas.MTW from the class website and save it to the desktop.
This data set contains information on passenger car gas mileage, based on measurements done by the US Environmental Protection Agency. The five variables are as follows.

- VOL: Cubic feet of cab space
- HP: Engine horsepower
- MPG: Average miles per gallon
- SP: Top speed (mph)
- WT: Vehicle weight (100 lb)

1 Exploratory Analysis

We begin by exploring the relationship between the *response* variable MPG and the four *explanatory* variables in the data set.

1. Select **File** > **Open Worksheet**. In the "Look in:" pull down menu, select "Desktop." Highlight "gas.MTW". Click **Open**.
2. Compute the correlation matrix using **Stat** → **Basic Stats** → **Correlation**.
3. Create scatterplots of MPG versus each of the explanatory variables (**Graph** → **Plot**).

   Question #1: Based on the scatterplots and the correlations, describe the relationships between the response variable and each of the explanatory variables.

Question #2: Do these plots give you information about the relation between the response variable and the *group* of explanatory variables? Why?

4. Close all the graphs (**Window → close all graphs**)

2 Simple Regression Analysis

1. Fit the least squares regression line to describe the relationship between MPG and each of the explanatory variables, *one at a time*. Choose **Stat → Regression → Regression** again, and enter the response and explanatory variable. *Click* the button for **graphs** and then check boxes next to "Normal plot for residuals" and "residuals vs. fits". *Click* **OK** twice.)
   Keep these residual plots, as you will need them in question 4.

   Question #3: Enter the requested results from the analyses in the summary table below. (Just use 2 decimal places.) To test whether each of the explanatory variable has a statistically significant linear relationship with the response variable, enter the T - statistic and P-value in the summary table.
   Circle those explanatory variables that have a significant linear relationship with the response variable.

   Summary Table: *Regression of* MPG *on*

Explanatory	$\hat{\beta}_1$	S.E.($\hat{\beta}_1$)	t-Statistic	p-value	R^2
VOL					
HP					
SP					
WT					

Question #4: For each of the simple regression models you fit in question 3, check if the model assumptions seem valid by examining the residual plots (plotting the residuals vs. fitted values,

and probability plot of the residuals). Describe the assumptions that seem to fail. What could possibly be done to remedy this?

2. Close all the graphs (**Window → close all graphs**)

3. Do the simple linear regression analyses (*including the residual analyses*) again. However, this time, use the log (base 10) of all the variables in your analyses. This requires that you create five new variables. For example, to create the log base 10 of "Vol", select **Calc → Calculator**. In the "Store result in variable:" field, enter your first available column (it should be C15). In the "Expression" field, enter "**logt**('Vol')". Click **OK**. Do the same for HP, MPG, SP, and WT, storing them in successive columns. Label these columns "log(Vol)", "log(HP)", "log(MPG)", "log(SP)", "log(WT)", respectively.

   Question #5: For an increase of 1 unit in $\log(\text{SP})$**, how much, and in which direction does the response,** $\log(\text{MPG})$**, change according to the model you created?**

♣ **Question #6: Briefly summarize in words your findings from the simple linear regression models you have fit.**

3 Multiple Regression Analysis

In this section we want to use a multiple linear regression model to explain as much of the variation in MPG as possible. The model that you will fit is:

$$\log \text{MPG} = \beta_0 + \beta_1 \log \text{Vol} + \beta_2 \log \text{HP} + \beta_3 \log \text{SP} + \beta_4 \log \text{WT} + \epsilon$$

This model considers the relationship between MPG and all four explanatory variables *simultaneously*.

Question #7: Fit this regression model and fill in the following summary table (choose Stat → Regression → Regression again, and enter the response and all four explanatory variables).

Parameter	Estimate	S.E.(Estimate)	t-Statistic	p-value
β_0				
β_1				
β_2				
β_3				
β_4				

Question #8: Write down the multiple linear regression equation with the estimated regression coefficients.

Question #9: Find the R^2 value and interpret it in the context of the problem.

Question #10: What is the null hypothesis being tested by the F-statistic (F-ratio) in the ANOVA table? What are the values of the F-ratio and its p-value? Test your hypothesis and interpret the results of the test.

♣ **Question #11: Explain what your value of $\hat{\beta}_3$ means. Compare this to your answer in Question 5. Explain how it is possible that $\log \mathrm{SP}$ had a strong explanatory value for $\log \mathrm{MPG}$ simple regression, whereas it is very insignificant in the multiple regression model.**

Question #12: Plot the residuals against the fitted values and against each of the four explanatory variables. Do the model's assumptions appear to be valid?

♣ **Question #13: Analyze what happens to the fit of the model if you remove $\log \mathrm{SP}$ from the model. Redo your analysis without this variable in the model.**