For the data analyzed in CO2_2014Vehicles.sas. Here are the solutions to the lab questions.

1. Use SAS to run the multiple linear regression model with Engine, Cylinder and City MPG. Use the output to answer the following questions.
   a. Give the equation for predicting the cityCO2 values from the three explanatory variables.

   $$\hat{Y}_i = 635.55 + 10.80 * \text{Engine} + 16.94 * \text{Cylinder} - 14.66 * \text{CityMPG}$$

   b. Conduct an F-test for the overall model in helping to explain the cityCO2 values. Report the null and alternative hypotheses, test statistic, p-value, decision and conclusion.

   Here is the ANOVA Table:

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	2564779	854926	965.91	<.0001
Error	196	173480	885.09952		
Total	199	2738259			

   The null and alternative hypotheses are:

   $$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$
   $$H_a: at\ least\ one\ \beta_j \neq 0$$

   F-test statistic: $F = 965.91$ with p-value < 0.0001. Since the p-value is so small, we will reject the null hypothesis and conclude at least one of the explanatory variables is significant in explaining the response variable City CO2.

   c. Give the value of R^2 for this model and interpret its value.

   The value of $R^2 = 93.66\%$. This means that 93.66% of the variation in City CO2 can be explained by the multiple linear regression model with Engine, Cylinder and City MPG.

   d. Conduct a t-test for the significance of Engine in the model that includes Cylinder and City MPG. Report the null and alternative hypotheses, test statistic, p-value, decision and conclusion.

   The table of parameter estimates is:

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	635.54671	22.58844	28.14	<.0001
engine	1	10.80080	4.46873	2.42	0.0166
cylinder	1	16.94226	3.25099	5.21	<.0001
cityMPG	1	-14.66201	0.67837	-21.61	<.0001

The null and alternative hypotheses are:

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

The t-test statistic is $t = 2.42$ with p-value = 0.0166. Since the p-value is small, we will reject the null hypothesis and conclude the variable Engine is statistically significant in the model that also includes Cylinder and City MPG.

2. Use SAS to run the multiple linear regression model with Engine, Cylinder, CityMPG and Gears as explanatory variables. Use the output to answer the following questions.
   a. How much does adding Gears to the multiple linear regression model with Engine, Cylinder and City MPG reduce the SS for Error?

      The difference in the Sums of Squares for Error is $173480 - 164229 = 9251$.

   b. How much does adding Gears to the multiple linear regression model with Engine, Cylinder and City MPG increase the value of of R^2?

      The difference in the R^2 values is: 94.00% – 93.66% = 0.34%.

   c. Conduct a F-test for the effect of adding Gears to the multiple linear regression model with Engine, Cylinder and City MPG. Report the null and alternative hypotheses, test statistic, p-value, decision and conclusion.

      The null and alternative hypotheses for the F-test for Gears are:

$$H_0: \beta_4 = 0$$
$$H_a: \beta_4 \neq 0$$

      The F-test statistic can be found as:

$$F = \frac{9251}{842.20144} = 10.98$$

      The p-value for this test will be the same as the p-value for the t-test. This is 0.0011. Since the p-value is so small, we will reject the null hypothesis and conclude that Gears is statistically significant in the model that includes Engine, Cylinder and City MPG.

3. Use SAS to run the multiple linear regression model with Engine, Cylinder, City MPG, Gears and Intake as explanatory variables. Use the output to answer the following questions.
   a. Give the equation for predicting the cityCO2 values from the four explanatory variables for vehicles with two intake valves per cylinder and for vehicles that do

not have two intake valves per cylinder. What is the difference in these two equations?

The two equations are:

$$\hat{Y}_i = \begin{cases} 682.35 + 1.38 * Engine + 22.32 * Cylinder - 15.33 * City\ MPG - 5.48 * Gears - if\ intake = 1 \\ 704.97 + 1.38 * Engine + 22.32 * Cylinder - 15.33 * City\ MPG - 5.48 * Gears - if\ intake = 0 \end{cases}$$

The difference in the two equations is the intercept. They are different by the amount of the coefficient for Intake in the model.

b. Conduct a t-test for the significance of Intake in the model that includes Engine, Cylinder, City MPG and Gears. Report the null and alternative hypotheses, test statistic, p-value, decision and conclusion.

Here is the parameter estimates table:

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	704.96723	28.13299	25.06	<.0001
engine	1	1.37730	5.27482	0.26	0.7943
cylinder	1	22.31665	3.48894	6.40	<.0001
cityMPG	1	-15.32579	0.67577	-22.68	<.0001
gears	1	-5.48026	1.62899	-3.36	0.0009
intake	1	-22.60939	10.18088	-2.22	0.0275

The null and alternative hypotheses for the F-test for Gears are:

$$H_0: \beta_5 = 0$$
$$H_a: \beta_5 \neq 0$$

The t-test statistic is -2.22 with a p-value of 0.0275. Since the p-value is small, we will reject the null hypothesis and conclude that Intake is statistically significant in the model that also includes Engine, Cylinder, City MPG, and Gears.