Name: *Caleb McWhorter — Solutions* **MAT 222**
Quiz 7 **Spring 2017**

Problem 1: Mark the following statements as T for True or F for False. Explain why the statement is True or False in the space provided.

(i) ___F___: A Simple Linear Regression is modeled simply as $\beta_0 + \beta_1 x$.

No, this is just for the mean. We have $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

(ii) ___T___: The population models for a Simple Linear Regression are β_0, β_1, σ.

Yes. The model is $\beta_0 + \beta_1 x_i + \epsilon_i$. The β's and σ (to create the ϵ_i's) are all that are needed.

(iii) ___T___: Decreasing the confidence level or a decrease in the MSE *both* result in a decrease in the width of a confidence interval.

Decreasing the MSE results in a decrease in confidence interval width (less 'error' means you can narrow your prediction) and decreasing the confidence level lowers the t^*-value which decreases the width of the confidence interval.

(iv) ___F___: At a significance level of $\alpha = 0.05$ while testing the hypothesis that $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$, a p–value of 0.001 is found. This means there is a strong linear relationship between the explanatory and response variables.

This just shows that there is some relationship between the two, but does not inherently mean that the relationship is linear.

(v) ___T___: Error in the measured x's for the model can have a great impact on the validity of a Simple Linear Regression.

Error in the measurements for the x's throw off every other portion of the model.

Problem 2: A company is trying to decide if there is a relationship between the number of advertisements they run throughout the country and the number of new customers they saw that year. They create a linear model to examine the relationship. Below is the data outputted by a computer program which created the model.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	25370455	25370455	92.99	0.000
Advertisement Number	1	25370455	25370455	92.99	0.000
Error	8	2182670			
Total	9	27553125			

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
522.335	92.08	91.09	85.37

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1400	357	3.92	0.004	
Advertisement Number	11.09	1.15	9.64	0.000	1.00

(i) Which is the explanatory variable and which is the response variable?

Advertisement number is the explanatory variable (x) and number of new customers is the response variable (y).

(ii) How many different data values did the company use to create this model?

We have degrees of freedom of the total $= n - 1 = 9$ *so that* $n = 10$.

(iii) What is the regression equation?

$$\textit{No. New Customers} = 11.09\,\textit{Advertisement Number} + 1400$$

(iv) Use the model to predict the number of customers if 443 advertisements were run that year or to predict the number of advertisements run that year if there was an increase of 5,227 customers that year. [Predict whichever you decided was the response variable in (i).]

$$11.09(443) + 1400 = 6312.87$$

(v) Using a significance level of $\alpha = 0.01$ to test the hypothesis $H_0 : \beta_1 = 0$ against $\beta_1 \neq 0$, what is the p–value and the conclusion? Can we say there is a linear relationship between the variables?

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{cases}$$

We have $b_1 = 11.09$*,* $SE_{b_1} = 1.15$*, test statistic* $t = 9.64$*, and* p*-value* 0.000*. Because* $p < \alpha$*, we reject the null hypothesis so that* $\beta_1 \neq 0$*, which implies that there is some relationship between advertisement number and number of new customers. This does not directly imply that the relationship is linear; however, the high* R^2 *value suggests that this is the case.*