***Quiz 1.*** **All work must be the sole product of the student**. The quiz is due by 11:59 pm on Thursday, Feb. 20. The quiz should be uploaded to the CANVAS quiz 1 Assignment dropbox for this course.

1. A person’s muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 4 women from each 10-year age group, beginning with age 40 and ending with age 79, resulting in a total sample size of n=16. Some results follow, where X is the age, and Y is a measure of the muscle mass. Assume a simple linear regression model is appropriate.

For these data:

Mean( X) = 61.69 SD(X) = 14.67

Mean(Y) = 52.21 SD(Y) = 24.16

y-hat = 142.68 – 1.47 \*X

SE(Slope) = 0.200 Root MSE = 11.38

Each part is worth 2 points each.

a. Interpret the parameters of the regression model.

b. Find the expected muscle mass for a woman age 63.

c. Find the Pearson Correlation and interpret its value.

d. Test the hypothesis that muscle mass is linearly independent of age; report and interpret the p-value.

e. Find the 95% CI for the slope of the regression equation. Interpret.

f. Find the 95% CI for the expected muscle mass of a woman who is age 63. Interpret.

g. Suppose a woman who is age 60 had an observed muscle mass of 45. What is the residual value for this woman?

h. Summarize the findings of this study using the PRE and standardized slope as a measures of effect size.

i. Two women differ by 10 years in age. What is their expected difference in muscle mass?

j. Does the assumption of linearity seem reasonable to you? Why or why not?

2. The following SAS output contains an analysis in which 6-year graduation rates were collected over a 6-year period and analyzed for trends. The response variable Y is graduation rate (measured as a percentage). The explanatory variable X is year (beginning with year 1)

The REG Procedure

Model: MODEL1

Dependent Variable: rate

Number of Observations Read 6

Number of Observations Used 6

Analysis of Variance

Sum of Mean

Source DF Squares Square F Value Pr > F

Model 1 15.93657 15.93657 12.27 0.0248

Error 4 5.19676 1.29919

Corrected Total 5 21.13333

Root MSE 1.13982 R-Square 0.7541

Dependent Mean 64.86667 Adj R-Sq 0.6926

Coeff Var 1.75717

Parameter Estimates

Parameter Standard

Variable DF Estimate Error t Value Pr > |t|

Intercept 1 61.52667 1.06112 57.98 <.0001

year 1 0.95429 0.27247 3.50 0.0248

Using the Output above, please answer the following questions. Each part is worth 2 points.

a. Report and interpret the prediction equation for this analysis.

b. What is the predicted graduation rate in year 6?

c. Suppose year 5 had an actual graduation rate of 67.8. Find the residual for this year and interpret its meaning.

d. Find the PRE for this dataset and interpret.

e. Construct a 95% CI for the sample slope and interpret its meaning.

f. Find and report the test statistic and p-value for testing H0: β = 0; interpret this result.

g. What is value of the conditional SD; that is, what is SD(Y|x)?

h. How many years were in the study?

j. Find the number 15.93 in the output above. What statistical concept – *in model comparison terminology* - does this number refer to?

j. Suppose there was a drop in observed graduation rates from 67.8 in year 5 to 66.3 in year 6. Should we be worried about that size of a drop? Explain.

k. Someone has asked for the value of the standardized slope for this analysis. What is it?

3. A criminologist studying the relationship between population density and robbery rates in medium-sized US cities collected the following data for a random sample of 16 cities; X is the population density of the city (number of people per unit area) and Y is the robbery rate last year (number of robberies per 100,000 people). Assume that a simple linear regression model is appropriate.

I: 1 2 3 4 5 6 7 8

X: 59 49 75 54 78 56 60 82

Y; 209 180 195 192 215 197 208 189

I: 9 10 11 12 13 14 15 16

X: 69 83 88 94 47 65 89 70

Y: 213 201 214 212 205 186 200 204

a. What is the estimated regression equation for this data?

b. What is the Pearson correlation between X and Y?

c. What are the standardized estimates for the regression equation?

d. Interpret the slope parameter for the model.

e. What is the value of PRE for this model? Interpret the PRE in *model comparison language*.

f. What are the upper and lower 95% CLM values for the average city (city nearest the mean in density) in the sample?

g. What are the upper and lower 95% CLI values for the city with the highest density?

h. What are the upper and lower 95% CI values for the estimated slope?

i. Do you think the linearity assumption is valid for this dataset? Why or why not?

4. Parts of the output below have been tragically erased by an unscrupulous and unsavory character. Please fill in the missing material …. There are 14 fill-in-the-blanks here.

Grading : 1 point per correct answer

The REG Procedure

Model: MODEL1

Dependent Variable: Y

Number of Observations Read 50

Number of Observations Used 50

Analysis of Variance

Sum of Mean

Source DF Squares Square F Value Pr > F

Model \_\_ \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_ \_\_\_\_\_\_ 0.0001

Error \_\_ \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_

Corrected Total \_\_ 18.77

Root MSE 0.4305 R-Square \_\_\_\_\_\_

Dependent Mean 9.9268

Parameter Estimates

Parameter Standard Standardized

Variable DF Estimate Error t Value Pr > |t| Estimate

Intercept 1 9.9798 \_\_\_\_\_\_\_\_\_\_ 55.64 <.0001 \_\_\_\_\_\_

size 1 \_\_\_\_\_\_\_\_\_\_ 0.05684 7.30 \_\_\_\_\_ \_\_\_\_\_\_