Practice Problems Section 4G

1. How can tell which variable should be the explanatory variable and which variable should be the response variable?

2. How can we use the correlation coefficient (r) to determine if there is strong positive correlation? How can we use the correlation coefficient (r) to determine if there is strong negative correlation? How can we use the correlation coefficient (r) to determine if there is no correlation?

3. What is the definition of the coefficient of determination (r^2)?

4. What are the two definitions for the standard deviation of the residual errors (s_e)?

5. What is the definition of the slope of the regression line?

6. What is the definition of the y-intercept of the regression line?

7. What is extrapolation? Why should we avoid extrapolation?

(#8-16) Directions: Go to www.matt-teachout.org, click on "Statistics" and then "Data Sets". Open the indicated data. Copy and paste the two indicated columns of quantitative data next to each other on a new Excel spreadsheet. Then copy the two columns together. Now go to www.lock5stat.com and click on StatKey. Under the "Descriptive Statistics and Graphs" menu click on "Two Quantitative Variables". Click on "Edit Data" and paste the two columns together into StatKey. Then answer indicated questions.

8. Open the cigarette data. Let the explanatory variable (X) represent the amount of nicotine (milligrams) and the response variable (Y) represent the amount of tar (milligrams).

   a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
   b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
   c) Find the slope of the regression line. Write a sentence to explain the slope.
   d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
   e) The standard deviation of the residual errors was 1.3 mg. Explain the two meanings of this statistic.
   f) Use the regression line formula to predict the amount of tar if a cigarette contains 1.2 mg of nicotine. How much error could there be in this prediction.

9. Open the cigarette data. Let the explanatory variable (X) represent the amount of nicotine (mg) and the response variable (Y) represent the amount of carbon monoxide in parts per million (PPM).

   a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
   b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
   c) Find the slope of the regression line. Write a sentence to explain the slope.
   d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
   e) The standard deviation of the residual errors was 2.3 PPM. Explain the two meanings of this statistic.
   f) Use the regression line formula to predict the amount of carbon monoxide if a cigarette contains 1.2 mg of nicotine. How much error could there be in this prediction.

10. Open the health data. Let the explanatory variable (X) represent the systolic blood pressure (mm of Hg) and the response variable (Y) represent the diastolic blood pressure (mm of Hg). Use the combined columns with 80 randomly selected adults. Do not separate by gender.

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.

c) Find the slope of the regression line. Write a sentence to explain the slope.
d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 7.4579 mm of Hg. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the diastolic blood pressure of a person who has a systolic blood pressure of 130. How much error might there be in that prediction?

11. Open the health data. Let the explanatory variable (X) represent the waist size in centimeters and the response variable (Y) represent the weight in pounds. Use the combined columns with 80 randomly selected adults. Do not separate by gender.

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
c) Find the slope of the regression line. Write a sentence to explain the slope.
d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 14.6809 pounds. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the weight of a person who has a waist size of 100 cm. How much error might there be in that prediction?

12. Open the health data. Let the explanatory variable (X) represent the age in years and the response variable (Y) represent the cholesterol in milligrams per deciliter (mg/dL). Use the combined columns with 80 randomly selected adults. Do not separate by gender.

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
c) Find the slope of the regression line. Write a sentence to explain the slope.
d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 255.3625 mg/dL. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the cholesterol of a person who is 40 years old. How much error might there be in that prediction?

13. Open the bear data. Let the explanatory variable represent the age of the bear in months and the response variable represent the length of the bear in inches.

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
c) Find the slope of the regression line. Write a sentence to explain the slope.

d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 7.51 inches. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the length of a bear that is 24 months old. How much error might there be in that prediction?

14. Open the bear data. Let the explanatory variable represent the neck circumference of the bear and the response variable represent the weight of the bear in pounds.

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
c) Find the slope of the regression line. Write a sentence to explain the slope.
d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 43.9 pounds. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the weight of a bear that has a neck circumference of 24 inches. How much error might there be in that prediction?

15. Open the car data. Let the explanatory variable (X) represent the weight of the car in tons and the response variable (Y) represent the gas mileage in miles per gallon.

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
c) Find the slope of the regression line. Write a sentence to explain the slope.
d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 2.8516 mpg. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the mpg for a car that weighs 3 tons. How much error might there be in that prediction?

16. Open the car data. Displacement is the amount of liquid in cubic centimeters forced out by the piston. Let the explanatory variable (X) represent the horsepower of the car and the response variable (Y) represent the displacement of the car (cc).

a) Look at the scatterplot and the correlation coefficient (r). Describe the strength and direction of the linear relationship.
b) Square the correlation coefficient in StatKey to calculate r^2. This is also called the coefficient of determination. Write a sentence to explain r^2.
c) Find the slope of the regression line. Write a sentence to explain the slope.
d) Find the y-intercept. Write a sentence to explain the y-intercept. Does the y-intercept make sense in the context of this data?
e) The standard deviation of the residual errors was 44.138 cubic centimeters. Explain the two meanings of this statistic.
f) Use the regression line formula to predict the number of cc's of displacement for a car with 120 horsepower. How much error might there be in that prediction?
