Statistics   Name _____
Review LSR

1. Does fidgeting keep you slim. Some people don't gain weight when they overeat. Perhaps fidgeting and other "non-exercise activity" (NEA) help explain why – the body might spontaneously increase non-exercise activity when fed more. Researchers deliberately overfed 16 healthy young adults for 8 weeks. They measured fat gain (in kilograms) and increase in energy use (in calories) from activity other than deliberate exercise. The non-exercise activity is your explanatory variable.

NEA increase	-94	-57	-29	135	143	151	245	355	392	473	486	535	571	580	620	690
Fat Gain (kg)	4.2	3.0	3.7	2.7	3.2	3.6	2.4	1.3	3.8	1.7	1.6	2.2	1.0	0.4	2.3	1.1

      a. Make a scatterplot. State the direction, form and strength of the data.
      b. Find the LSR. State what the slope means in context to the data
      c. List r. Interpret the meaning of the correlation in context to the data.
      d. Calculate the residual for x = 143. Discuss what the value of the residual tells you about the point in relationship to the LSR line.
      e. Predict the amount of weight gain if your NEA is 800 calories.
      f. How reliable is your prediction? Is it appropriate to extrapolate here? Explain
      g. Sketch the residual plot for this data. What does the plot tell you about the data?
      h. Are there any outliers? If so, are they influential and how do you know?

2. The following data is a sample of 21 University of Miami faculty and their years since receiving their PhD (years) and the number of publications they have (pubs).

Years	3	6	3	8	9	6	16	10	2	5	5	8	25	6	6	2	1	4	5	12	11
Pubs	7	3	4	17	11	6	24	29	9	18	19	19	2	11	8	3	4	15	9	30	31

      a. Make a scatterplot. State the direction, form and strength of the data.
      b. Find the LSR. State what the slope means in context to the data
      c. List r. Interpret the meaning of the correlation in context to the data.
      d. Are there any outliers? If so, are they influential and how do you know? If they are influential, remove them and redo a – c.
      e. Calculate the residual for (6, 11). (Use the new LSR, if you recalculated it in d). Discuss what the value of the residual tells you about the point in relationship to the LSR line.
      f. Predict the number of publications a faculty member would have who completed his/her PhD 15 years ago. How confident are you with this prediction?
      g. Sketch the residual plot for this data. What does the plot tell you about the data?

3. The number of heart disease deaths per 100,000 people in the U.S. for the years 1950 – 1980 as reported by the National Center for Health Statistics has a regression line with equation: (Years since 1950 is the explanatory and heart disease death rate is the response). The scatterplot showed a clear linear pattern.

$$\hat{y} = -3.627x + 7386.87 \qquad r = -.981$$

a. Identify the statement below that refers to the slope of the line ____

b. Identify the statement below that refers to the y-intercept ____

c. Identify the statement that refers to r.

I. Heart disease will be cured in the year 2036.
II. The baseline heart disease death rate per 100,000 people is 7386.87
III. The regression line explains 98.1% of the variation in heart disease death rates over
    the years.
IV. The number of heart disease deaths per 100,000 people has been dropping about 3.627 deaths per year.
V. There is a strong, negative, linear association between the years since 1950 and the number of heart disease deaths per 100,000 people in the U.S.

d. Based on the regression line, what is the predicted death rate for the year 1973?

e. Would it be wise to predict the death rate for the year 2012 from the LSR line?
   Support your answer with a reason.

4. The original data was a list of the value of export (in billions of dollars) and the value of imports (in billions of dollars) for several different years based on data from the U.S. Department of Commerce.
   a. Use the information below, which is based upon the original data, to write the least squares regression line.

   $\bar{x} = 121.67, \quad \bar{y} = 167.5, \quad s_x = 107.165, \quad s_y = 165.39, \quad r = .967$

   b. Predict imports when exports are 150 (in billions of dollars)

5. The linear scatterplot of P, what your cell phone plan's monthly charge is versus m, the number of minutes you used has the linear regression line $\hat{P} = 15 + 0.1m$

Explain, in context, what the slope and y-intercept mean.