Linear Regression Word Problems

1. The relationship between hospital patient-to-nurse ratio and various characteristics of job satisfaction and patient care has been the focus of a number of research studies. Suppose x = patient-to-nurse ratio is the independent variable. For each of the following potential dependent variables, indicate whether you expect the slope of the least squares regression line to be positive or negative.
    a. y = a measure of a nurse's job satisfaction
    b. y = a measure of patient satisfaction with hospital care
    c. y = a measure of patient quality of care
        a) negative
        b) negative
        c) negative

2. Explain why it can be dangerous to use the least-squares regression line to obtain predictions for x values that are substantially larger or smaller than those contained in the sample.

    A least squares regression line might not be an accurate predictor for x values that fall outside those contained in the sample.

3. Explain why the slope of the least-squares regression line always has the same sign (positive or negative) as the correlation coefficient.

    A positive of negative correlation determines whether the dependent variable increase/ decreases in response to increasing the independent variable, which also happens in the interpretation of slope.

4. An article included findings regarding the use of SAT I scores, SAT II scores, and high school GPA to predict first year college GPA. The article states that "among these, SAT II scores are the best predictor, explaining 16 percent of the variance in first year college grades. GPA was second at 15.4 percent, and SAT I was last at 13.3 percent."
    a. If the data from this study were used to fit a least-squares regression line with y = first year college GPA and x = high school GPA, what would be the value of r^2?
    b. Do you think the predictions based on a least-squares line with y = first year college GPA and x = SAT II score would be very accurate? Why or why not?
        a) 0.154
        b) No, 0.16 is still a very low coefficient of determination, so not much of the variation is explained by the linear model.

5. A researcher took a sample of 25 electronics companies and found the following relationship between x and y where x is the amount of money (in millions of dollars) spent on advertising by a company in 1996 and y represents the total gross sales (in millions of dollars) of that company in 1996. $\hat{y} = 3.4 + 11.55x$ is the least-squares regression line.
    a. An electronics company spent $2 million on advertising in 1996. What are the expected gross sales for 1996?

b. Suppose four electronics companies spent $2 million each on advertising in 1996. Do you expect these four companies to have the same actual gross sales for 1996? Explain.
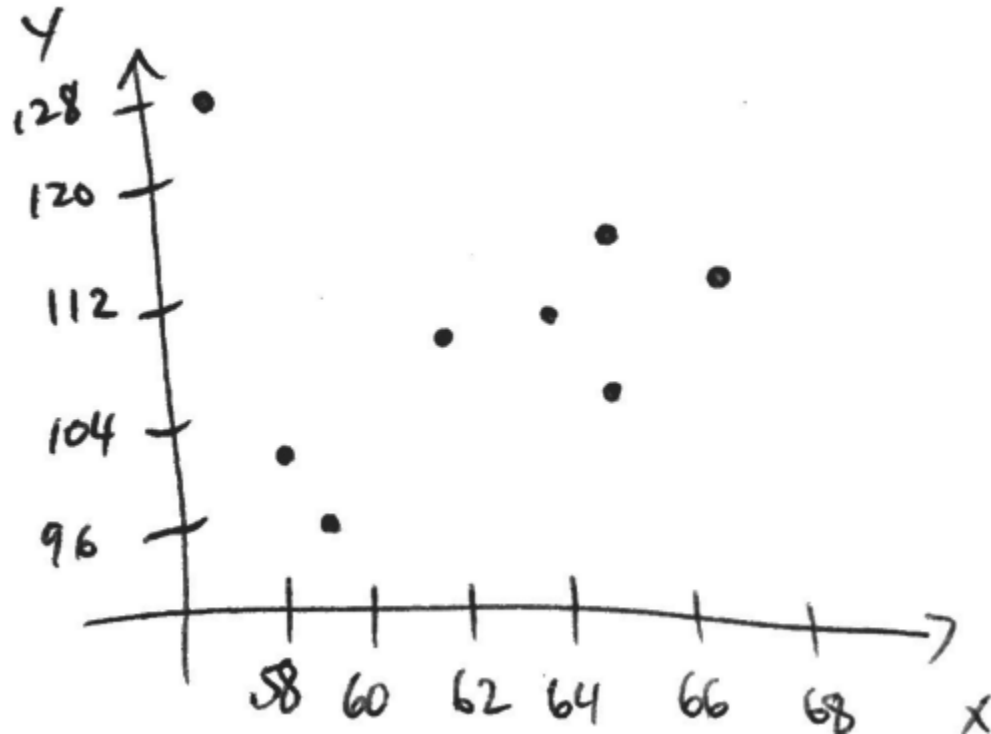 a) 26.5 million dollars
  b) No, each company might differ in gross sales due to randomness.

6. The height (in inches) of 8 high school girls and their scores on an IQ test is given below:

Height, x	62	58	65	67	59	64	65	57
IQ Score, y	109	102	107	114	96	110	116	128

  a. Display the data in a scatter plot.
  b. Describe the type of correlation, and interpret the correlation in context of the data.

a)

b) Positive correlation, as height increases, IQ score increases

7. Use the value of the correlation coefficient r to calculate the coefficient of determination r^2. What does this tell you about the explained variation of the data about the regression line?
  a. $r = 0.465$
  b. $r = -0.957$
    a) $r^2 = 0.216$
       21.6% of the variation is explained by the regression line

    b) $r^2 = 0.92$
       92% of the variation is explained by the regression line