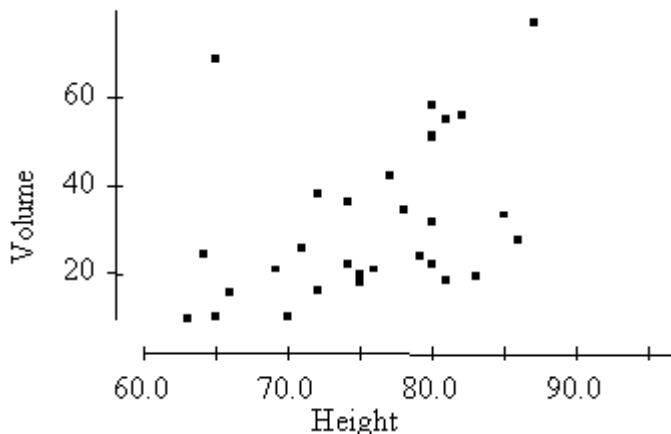


Use the following to answer questions 1 and 2:

A researcher measures the height (in feet) and volume of usable lumber (in cubic feet) of 32 cherry trees. The goal is to determine if the volume of a tree's usable lumber can be estimated from the height of the tree. The results are plotted below.



- In the study above, the response variable is
  - number of trees.
  - volume.
  - height or volume; it doesn't matter which is considered the response variable.
  - neither height nor volume; the measuring instrument used to measure height is the response variable.
  - height.
- The scatterplot above suggests that
  - there is a positive association between height and volume.
  - there is an outlier in the plot.
  - both A and B.
  - neither A nor B.
  - the relationship between height and volume is nonlinear.
- A company produces boxes of soap powder labeled "Giant Size 32 Ounces." The actual weight of soap powder in a box has a normal distribution with a mean of 33 ounces and a standard deviation of 0.7 ounces. What proportion of boxes are underweight (i.e., weigh less than 32 ounces)?
  - 0.0766.
  - 0.2420.
  - 0.4236.
  - 0.7580.
  - 0.9236.

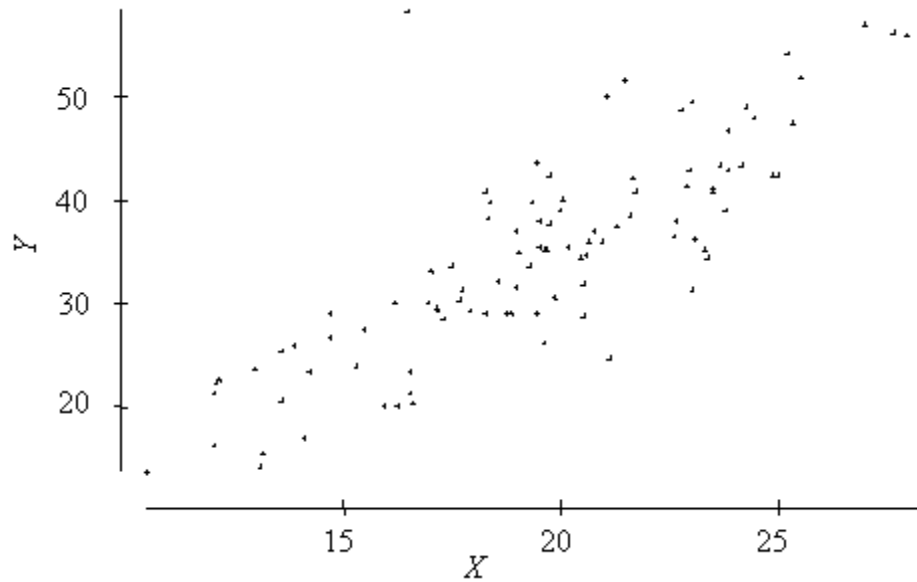
4. A school guidance counselor examines the number of extracurricular activities of students and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero." A correct interpretation of this statement would be that
- active students tend to be students with poor grades, and vice versa.
  - students with good grades tend to be students that are not involved in many activities, and vice versa.
  - students involved in many extracurricular activities are just as likely to get good grades as bad grades. The same is true for students involved in few extracurricular activities.
  - as a student becomes more involved in extracurricular activities, there will be a change in his/her grades.
  - involvement in many extracurricular activities and good grades go hand in hand.
5. A student wonders if people of similar heights tend to date each other. She measures herself, her dormitory roommate, and the women in the adjoining rooms; then she measures the next man each woman dates. Here are the data (heights in inches):

Women	66	64	66	65	70	65
Men	72	68	70	68	74	69

Which of the following statements is true about her sample?

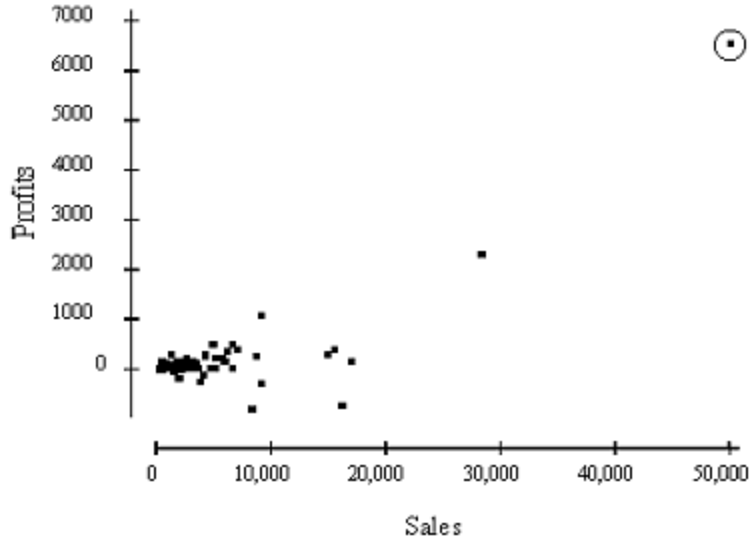
- The variables measured are all categorical.
  - There is a strong negative association between the heights of men and women, since the women are always smaller than the men they date.
  - Tall women tend to date short men.
  - Any height above 70 inches must be considered an outlier.
  - There is a positive association between the heights of men and women who date each other.
6. A study found a correlation of  $r = -0.61$  between the gender of a worker and his or her income. We may correctly conclude that
- women earn more than men on the average.
  - women earn less than men on the average.
  - an arithmetic mistake was made, since correlation must always be positive.
  - this result is incorrect, because computing  $r$  makes no sense in this situation.
  - on average, women earn 61% less than men.

7. Consider the scatterplot below.



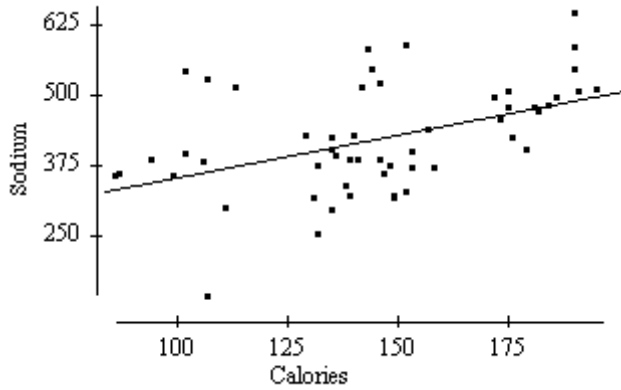
The correlation between  $X$  and  $Y$  is approximately  
 A) 0.999. B) 0.8. C) 0.5. D) 0. E)  $-0.7$ .

8. The profits (in multiples of \$100,000) versus the sales (in multiples of \$100,000) for a number of companies are plotted below. The correlation between profits and sales is 0.814. Suppose we removed the point that is circled from the data represented in the plot. The correlation between profits and sales would then be



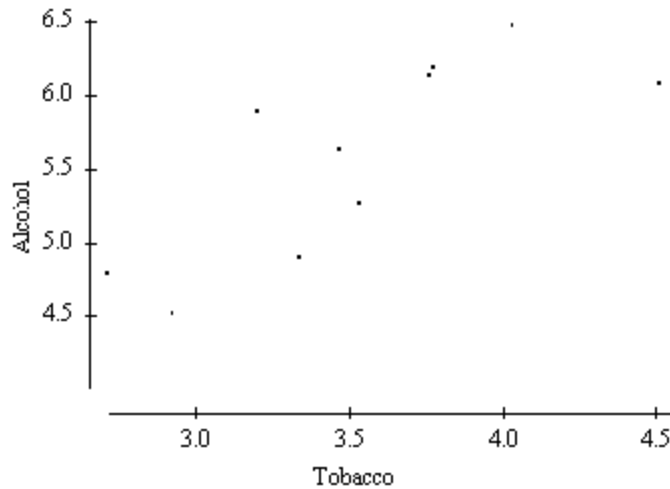
- A) 0.814.
- B) significantly larger than 0.814.
- C) significantly smaller than 0.814.
- D) slightly larger than 0.814.
- E) slightly smaller than 0.814.

9. Below is a scatterplot of the calories and sodium content (in milligrams) of several brands of meat hot dogs. The least-squares regression line has been drawn on the plot.



Based on the least-squares regression line in this scatterplot, one would predict that a hot dog containing 100 calories would have a sodium content (in milligrams) of about  
 A) 70.    B) 350.    C) 375.    D) 400.    E) 600.

10. The British government conducts regular surveys of household spending. The average weekly household spending on tobacco products and alcoholic beverages for each of 11 regions in Great Britain was recorded. A scatterplot of spending on alcohol versus spending on tobacco is given below.



Which of the following statements is true?

- A) The observation (4.5, 6.0) is an outlier.  
 B) There is clear evidence of a negative association between spending on alcohol and spending on tobacco.  
 C) The equation of the least-squares line for this plot would be approximately  $y = 10 - 2x$   
 D) The correlation coefficient for this data is 0.99.  
 E) The observation in the lower right corner of the plot is influential.

11. Which of the following is **false** about the least-squares regression line?
- A) The slope is the change in the response variable that would be predicted by a unit change in the explanatory variable.
  - B) It always passes through the point  $(\bar{X}, \bar{Y})$ , the means of the explanatory and response variables, respectively.
  - C) It will only pass through all the data points if  $r = \pm 1$ .
  - D) No more than 50% of the residuals will be positive values.
  - E) All of the above.
12. A researcher wishes to study how the average weight  $Y$  (in kilograms) of children changes during the first year of life. He plots these averages versus the children's age  $X$  (in months) and decides to fit a least-squares regression line to the data with  $X$  as the explanatory variable and  $Y$  as the response variable. He computes the following quantities.
- $r =$  correlation between  $X$  and  $Y = 0.9$
  - $\bar{X} =$  mean of the values of  $X = 6.5$
  - $\bar{Y} =$  mean of the values of  $Y = 6.6$
  - $s_X =$  standard deviation of the values of  $X = 3.6$
  - $s_Y =$  standard deviation of the values of  $Y = 1.2$
- The slope of the least-squares line is
- A) 0.30.    B) 0.88.    C) 1.01.    D) 2.7.    E) 3.0.
13. Recall that when we standardize the values of a variable, the distribution of standardized values has mean 0 and standard deviation 1. Suppose we measure two variables  $X$  and  $Y$  on each of several subjects. We standardize both variables and then compute the least-squares regression line of  $Y$  on  $X$  for these standardized values. Suppose the slope of this least-squares regression line is  $-0.44$ . We may conclude that
- A) the correlation will be  $1/-0.44$ .                      D) the correlation will be 1.0.
  - B) the intercept will also be  $-0.44$ .                      E) the correlation will also be  $-0.44$ .
  - C) the intercept will be 1.0.
14. In a study of 1991 model cars, a researcher found that the fraction of the variation in the price of cars that was explained by the least-squares regression on horsepower was about 0.64. For the cars in this study, the correlation between the price of the car and its horsepower was found to be positive. The actual value of the correlation
- A) is 0.80.
  - B) is 0.64.
  - C) is 0.41.
  - D) is  $-0.80$ .
  - E) cannot be determined from the information given.

15. Which of the following statements concerning residuals is true?
- A) The sum of the residuals is always 0.
  - B) A plot of the residuals is useful for assessing the fit of the least-squares regression line.
  - C) The value of a residual is the observed value of the response minus the value of the response that one would predict from the least-squares regression line.
  - D) If the data are linear, then the plot of the residuals should have no discernible pattern.
  - E) All of the above.

16. A normal density curve has which of the following properties?
- A) It is symmetric
  - B) The median is equal to the mean
  - C) The spread of the curve is proportional to the standard deviation
  - D) It has a peak centered above its mean
  - E) All of the above.

17. Leonardo da Vinci, the renowned painter, speculated that an ideal human would have an armspan (distance from outstretched fingertip of left hand to outstretched fingertip of right hand) that was equal to his height. The following computer regression printout shows the results of a least-squares regression on height and armspan, in inches, for a sample of 18 high school students.

Dependent variable is: **Height**  
 No Selector  
 R squared = 87.1%      R squared (adjusted) = 86.3%  
 s = 1.613 with 18 - 2 = 16 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	280.631	1	280.631	108
Residual	41.6185	16	2.50116	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	11.5474	5.6	2.06	0.0558
Armspan	0.840424	0.00091	10.4	≤ 0.0001

- a. Write the equation of the least squares regression line.
- b. Interpret the slope in context to the problem
- c. Calculate the residual for the student with height 70.5 and an armspan of 68.
- d. Calculate the correlation coefficient. What does it tell you about the scatterplot?

18. Because elderly people may have difficulty standing to have their heights measured, a study looked at predicting overall height from height to the knee. Here are data (in centimeters) for five elderly men:

---

Knee height $x$	57.7	47.4	43.5	44.8	55.2
Height $y$	192.1	153.3	146.4	162.7	169.1

---

- Construct a scatterplot on your calculator. Describe the form, direction, and strength of the relationship that you see.
- Use your calculator to determine the least-squares regression line. Write the equation below. Be sure to define any variables you use.
- Interpret the slope and  $y$ -intercept of the regression line in the context of the problem.
- Interpret the value of  $r^2$  in the context of the problem.
- We define the standard deviation of the residuals,  $s$ , by  $s = \sqrt{\frac{\sum \text{residuals}^2}{n-2}}$ . Calculate  $s$  and interpret its value in the context of the problem.
- Create a residual plot (on your calculator). What does the residual plot tell you about the LSR line?
- Should you use your regression line from Part b. above to predict the height of an elderly man whose knee height is 70 centimeters? If so, do it. If not, explain why not.