AP Stat
CH 3 (3.2)
Practice

Name TEACHER

1. Data: GPA against IQ. IQ is the independent (explanatory) variable.

IQ Scores:

$\bar{x} = 108.9$ Sx= 13.17

GPA:

$\bar{y} = 7.447$ Sy = 2.10

r = 0.6337

$b = .6337\left(\frac{2.1}{13.17}\right)$

$b = .101$

$a = 7.447 - (.101)(108.9)$

$a = -3.5519$

$\hat{y} = -3.552 + .101x$

$GPA = -3.552 + .101(IQ)$

A) Find the equation of least squares regression for predicting GPA from IQ.(Write the equation with IQ and GPA as well as "x and y".) Which is the explanatory and which is the response variable.

B) What percent of observed variation in the student's GPAs can be explained by the linear relationship between GPA and IQ?

$r^2 = (.6337)^2 = .4015$ 40.15% of variation of GPA is accounted for by the linear relationship with IQ

C) One student has an IQ of 103 but has a very low GPA of 0.53. What is the predicted GPA of the student with this IQ? What is the residual for this student?

$\hat{y} = -3.552 + .101(103)$

$\hat{y} = 6.851$

residual $= 6.851 - .53 =$
$= 6.321$

2. Some people think that predicting the behavior for the stock market in January predicts the behavior for the rest of the year. Let the explanatory variable x be the percent change in the stock market index in January and the response variable y to be the percent change for the year. The correlation from data for the years 1960 to 1997 gives:

$\bar{x} = 1.75\%$ Sx= 5.36%

$\bar{y} = 9.07\%$ Sy = 15.35%

r = 0.596

$b = .596\left(\frac{15.35}{5.36}\right) = 1.707 = b$

$a = 9.07 - 1.707(1.75)$

$a = 6.08$

$\hat{y} = 6.08 + 1.707x$ OR Full yr change $= 6.083 + 1.707$(Jan change)

A. What percent of the observed variation (r-sq) in yearly change in the index is explained by a straight line relationship with the change during January.

$r^2 = .596^2 = .3552$ or 35.52% of variation in yearly changes is explained by the Linear relationship (given by the LSREG.)

B. What is the equation of the least squares regression line for predicting full year change from January change?

See previous page

C. The mean change in January is $\overline{x} = 1.75\%$. Use your regression line to predict the change in the index in the year in which the index rises 1.75% in January. Why could you not have used your regression line to get your answer without doing calculations.

$\hat{y} = 6.083 + 1.707x$

$\hat{y} = 6.083 + 1.707(1.75)$

$\hat{y} = 9.07 \Rightarrow$ This is $\overline{y}$ $\rightarrow$ $(\overline{x}, \overline{y})$ are on regression line

3. You are given 4 sets of data.

Set A

x	10	8	13	9	11	14	6	4	12	7	5	
y	8.04	6.95	7.58	8.81	8.33	9.96	7.24	4.26	10.82	4.82	5.68	

Set B

x	10	8	13	9	11	14	6	4	12	7	5	
y	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74	

Set C

x	10	8	13	9	11	14	6	4	12	7	5	
y	7.76	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73	

L4L6

Set D

x	8	8	8	8	8	8	8	8	8	8	19	
y	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.5	

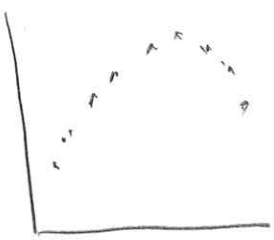
A. Without making scatter plots, find the correlation (r) and the least squares regression line for all four data sets. What do you notice? Use your regression line to predict y for x = 10 for all four data sets.

A	B	C	D
$r = .8168$	$r = .8162$	$r = .820$	$r = .817$
$\hat{y} = 3.003 + .4995x$	$\hat{y} = 3.000 + .5x$	$\hat{y} = 3.005 + .5x$	$\hat{y} = 3.00 + .499x$
$(10, 7.99)$	$(10, 8)$	$(10, 8.005)$	$(10, 7.99)$

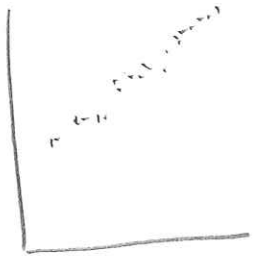
B. Now, make a scatter plot for each set of data and add your regression line to each plot.

A

B



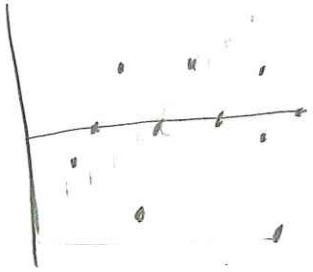
C



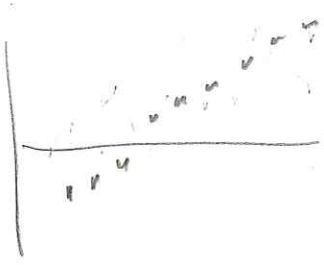
D

C. Now make a sketch of the residual plots for each data set.

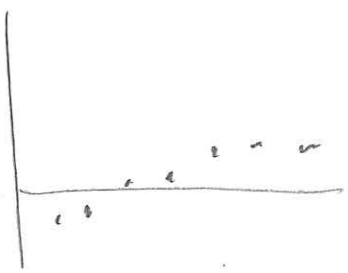
A



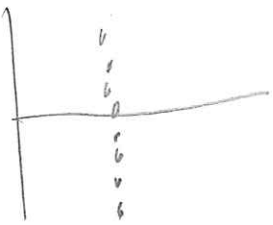
B



C



D



$\bar{x} = 0$

D. In which of the four cases would you be willing to use the regression line to describe the dependence of y on x. Explain.

Data set A

B- nonlinear

C- deviation pt (13, 12.74)

D- x=19 very influential

Chapter 3: Examining Relationships

1. A study is conducted to determine if one can predict the yield of a crop based on the amount of yearly rainfall. The response variable in this study is
   A) the yield of the crop.
   B) the amount of yearly rainfall.
   C) the experimenter.
   D) either bushels or inches of water.
   E) the month the crop is harvested.

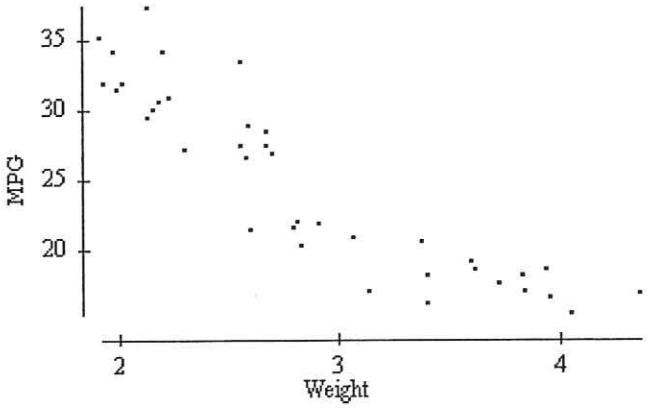
9. A school guidance counselor examines the number of extracurricular activities of students and their grade point average. The guidance counselor says, "The evidence indicates that the correlation between the number of extracurricular activities a student participates in and his or her grade point average is close to zero." A correct interpretation of this statement would be that
   A) active students tend to be students with poor grades, and vice versa.
   B) students with good grades tend to be students that are not involved in many activities, and vice versa.
   C) 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.
   D) as a student becomes more involved in extracurricular activities, there will be a change in his/her grades.
   E) involvement in many extracurricular activities and good grades go hand in hand.

11. Which of the following statements about the correlation coefficient is true?
   A) The correlation coefficient measures the proportion of variability between the two variables.
   B) The correlation coefficient will be equal to 1 only if all the data lie on a perfectly horizontal straight line.
   C) The correlation coefficient measures the fraction of outliers that appear in a scatterplot.
   D) The correlation coefficient has no unit of measurement and must always lie between –1 and 1, inclusive.
   E) The correlation coefficient equals the proportion of times two variables lie on a straight line.

12. 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
   A) women earn more than men on the average.
   B) women earn less than men on the average.
   C) an arithmetic mistake was made, since correlation must always be positive.
   D) this result is incorrect, because computing r makes no sense in this situation.
   E) on average, women earn 61% less than men.

B/c gender is not quantitative

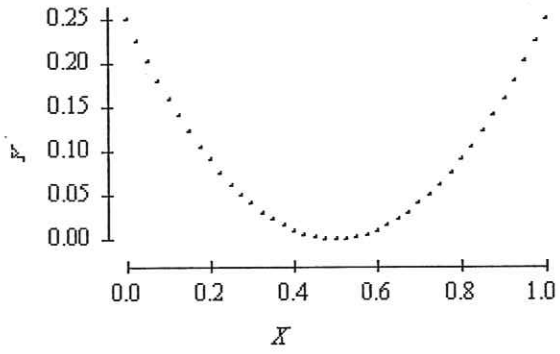
13. Consider the scatterplot below.



According to the scatterplot, which of the following is a plausible value for the correlation coefficient between weight and MPG?

A) −1.0. B) −0.9. C) −0.5. D) 0.2. E) 0.7.

15. Consider the scatterplot below.



We may conclude that

A) the correlation between X and Y must be close to 1 since there is a nearly perfect relationship between them.
B) the correlation between X and Y shows a quadratic relationship.
C) the correlation between X and Y is close to 0.
D) the correlation between X and Y could be any number between –1 and 1. Without knowing the actual values of X and Y we can say nothing more.
E) the correlation between X and Y must be close to –1 since there is a nearly perfect relation between them, but it is not a straight-line relation.

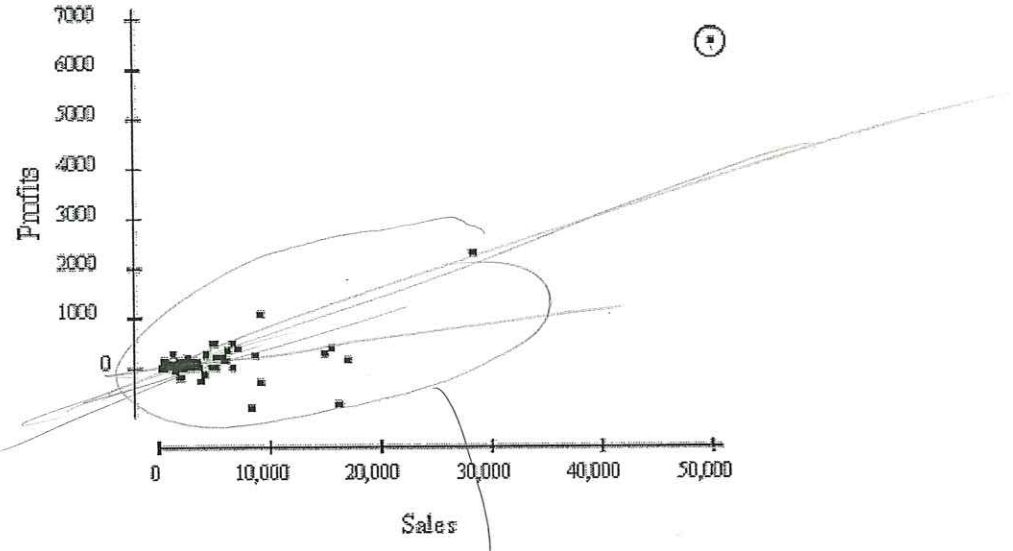
Use the following to answer questions 16 and 17:

I wish to determine the correlation between the height (in inches) and weight (in pounds) of 21-year-old males. To do this, I measure the height and weight of two 21-year-old men. The measured values are

	Male #1	Male #2
Height	70	75
Weight	160	200

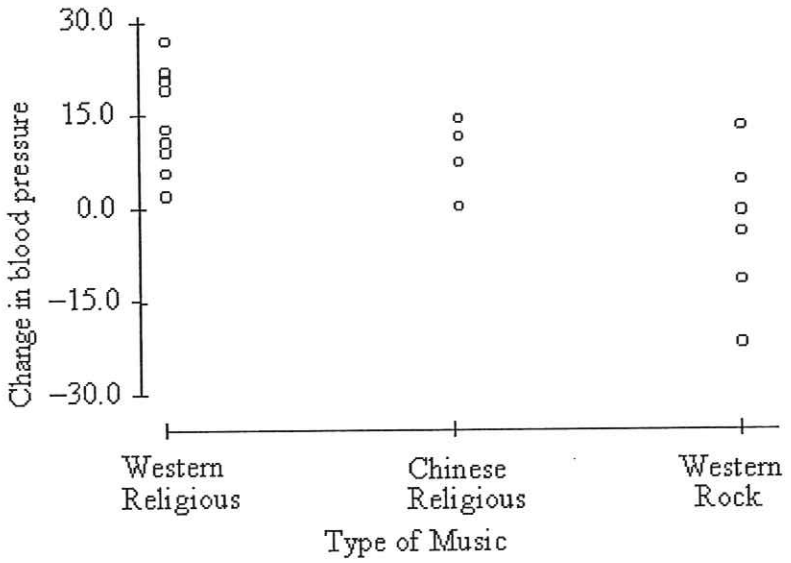
16. Referring to the information above, the correlation r computed from the measurements on these males is (calc)
    A) equal to 1.
    B) positive and between 0.25 and 0.75.
    C) near 0, but could be either positive or negative.
    D) exactly 0.
    E) Meaningless, since the slope is greater than 1.

21. 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. more clustered & random
    D) slightly larger than 0.814.
    E) slightly smaller than 0.814.

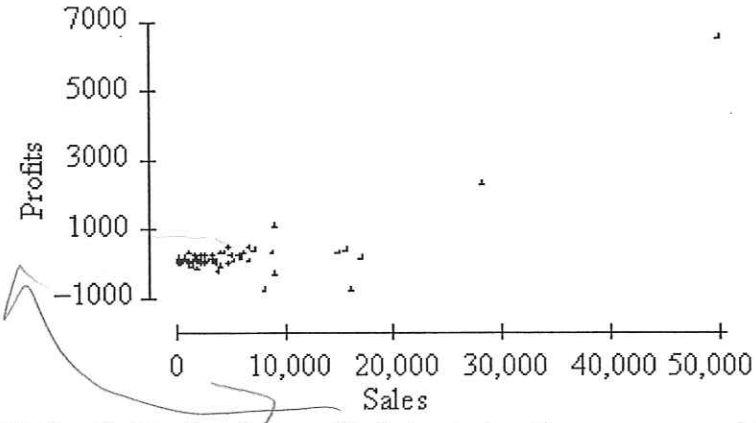
22. Volunteers for a research study were divided into three groups. Group 1 listened to Western religious music, group 2 listened to Western rock music, and group 3 listened to Chinese religious music. The blood pressure of each volunteer was measured before and after listening to the music, and the change in blood pressure (blood pressure before listening minus blood pressure after listening) was recorded. A scatterplot of change in blood pressure versus type of music listened to is given below.



The correlation between change in blood pressure and type of music listened to is
A) negative. B) positive. C) first negative, then positive. D) nearly 0.
E) none of the above.

music not quantitative

23. The profits (in multiples of $100,000) versus the sales (in multiples of $100,000) for a number of companies are plotted below.



Sales↑ Profits↑

profits↑ Sales↑

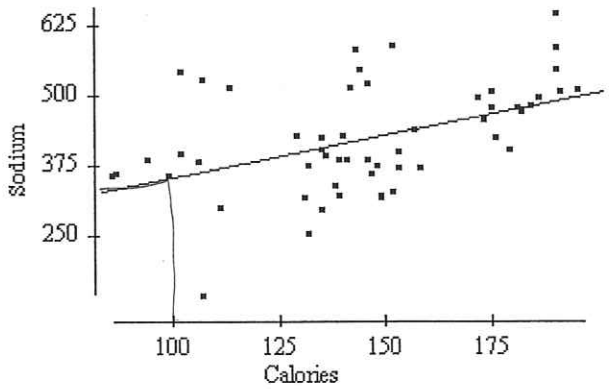
Notice that in the plot, profits is treated as the response variable and sales as the explanatory variable. The correlation between profits and sales is 0.814. Suppose we had taken sales to be the response variable and profits to be the explanatory variable. In this case, the correlation between sales and profits would be

Still pos. correlation

A) 0.814.
B) −0.814.
C) 0.
D) any number between −0.814 and 0.814, but we can't state the exact value.
E) 1, since the direction of the data doesn't change.

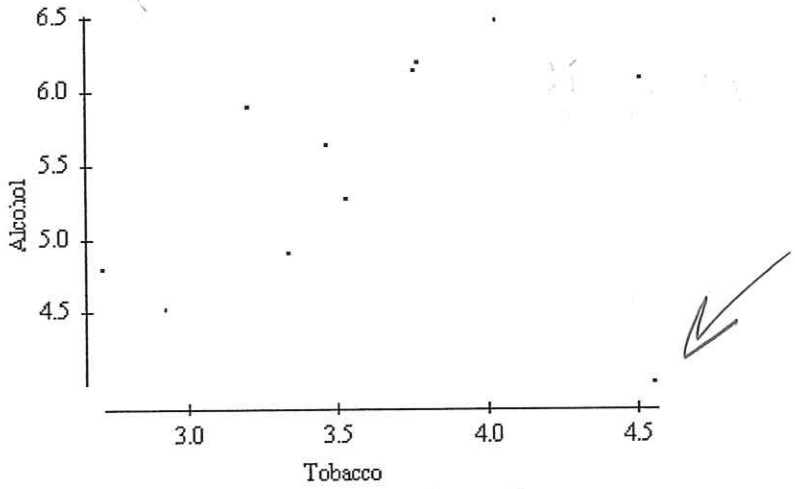
no neg #'s

24. 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.

25. 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.

26. The fraction of the variation in the values of y that is explained by the least-squares regression of y on x is

A) the correlation coefficient.
B) the slope of the least-squares regression line.
C) the square of the correlation coefficient.
D) the intercept of the least-squares regression line.
E) the residual.

27. In a statistics course, a linear regression equation was computed to predict a student's final exam score from his/her score on the first test. The equation of the least-squares regression line was

$$\hat{y} = 10 + 0.9x$$

where y represents the final exam score and x is the score on the first exam. Suppose Joe scores a 90 on the first exam. What would be the predicted value of his score on the final exam?

$\hat{y} = 10 + .9(90)$

A) 91.
B) 90.
C) 89.
D) 81.
E) It cannot be determined from the information given. We also need to know the correlation coefficient.

28. John's parents recorded his height at various ages up to 66 months. Below is a record of the results.

H depends on A

Age (months)	36	48	54	60	66
Height (inches)	35	38	41	43	45

Which of the following is the equation of the least-squares regression line of John's height on age? (NOTE: You do not need to directly calculate the least-squares regression line to answer this question.)

A) $\widehat{\text{Height}} = 12 \times (\text{Age})$.
B) $\widehat{\text{Height}} = 0.34 + 22.3 \times (\text{Age})$.
C) $\widehat{\text{Height}} = \text{Age}/12$.
D) $\widehat{\text{Height}} = 60 - 0.22 \times (\text{Age})$.
E) $\widehat{\text{Height}} = 22.3 + 0.34 \times (\text{Age})$.

put in L1, L2 + calc #8

33. 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 $Y = 1.2$

H = y
A = x

$b = .9\left(\frac{1.2}{3.6}\right)$

$b = .3$

The slope of the least-squares line is
A) 0.30. B) 0.88. C) 1.01. D) 2.7. E) 3.0.

34. 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.
B) the intercept will also be –0.44.
C) the intercept will be 1.0.
D) the correlation will be 1.0.
E) the correlation will also be –0.44.

Standardizing won't change it.

35. 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

r^2

A) is 0.80.
B) is 0.64.
C) is 0.41.
D) is –0.80.
E) cannot be determined from the information given.

$\sqrt{.64} = r = .8$

Yates TPS 3e Chapter 03

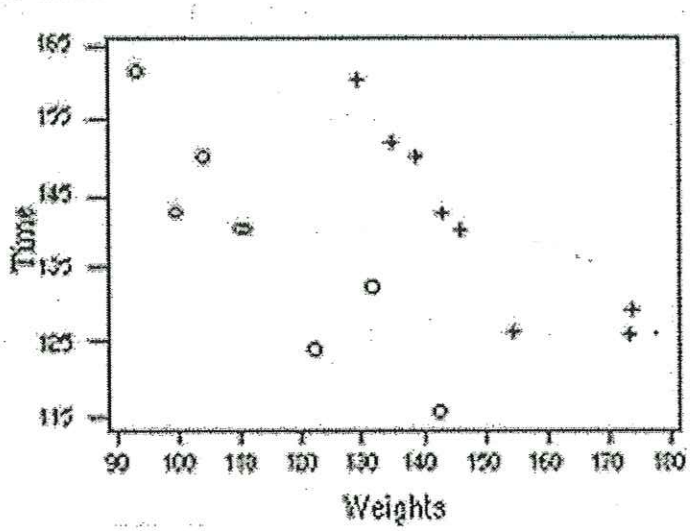
1. You wish to conduct a study to examine how the religious affiliation of a person may influence his or her opinion about the permissibility of partial-birth abortions. Which of the following statements about the variable "religious affiliation" is correct?

A. It is the explanatory variable and is quantitative.

B. It is the response variable and is categorical.

C. It is the explanatory variable and is categorical.

2. The points in the scatterplot represent paired observations (x, y) where x is an individual's weight and y is the time (in seconds) it takes for walking on a treadmill to raise the individual's pulse rate to 140 beats per minute. The o's correspond to females and the +'s to males.



From the scatterplot, we can make which conclusion?

A. There is a positive correlation r between gender and weight, since men tend to weigh more than women.

B. There is a negative correlation r between weight and time for both males and females.

C. In general, males tend to take less time to have their pulse rate raised to 140 bpm while walking on the treadmill.

3. Which of the following statements *does not* contain a blunder? categorical

A. There is a correlation of $r = 0.54$ between the position a football player plays and his or her weight.

B. The correlation between amount of fertilizer and yield of tomatoes was found to be $r = 0.33$.

C. The correlation between the gas mileage of a car and its weight is $r = -0.71$ gallon-pounds.

Label?

---

4. There is a strongly linear association between the weight of a football player and the time in seconds it takes for that player to run a 100-yard dash. Knowing this, a reasonable value for the correlation r between weight and 100-yard dash time would be

A. $r = 0.8$.

B. $r = 0$.

C. $r = -0.8$.

pos. corr.
more weight more time

---

5. Foresters use linear regression to predict the volume of timber in a tree using easily measured quantities such as diameter. Let y be the volume of timber in cubic feet produced by a tree and let x be the tree's diameter in feet (measured at a height of 3 feet above the ground). One set of paired data gives the prediction equation

✓ depends on draw

$\hat{y} = -30 + 60x$

The predicted volume of timber for a tree of diameter 18 inches is

A. 1050 cubic feet.

B. 90 cubic feet.

C. 60 cubic feet.

$\hat{y} = -30 + 60(18)$

---

6. Foresters use linear regression to predict the volume of timber in a tree using easily measured quantities such as diameter. Let y be the volume of timber in cubic feet produced by a tree and let x be the tree's diameter in feet (measured at a height of 3 feet above the ground). One set of paired data gives the prediction equation

$\hat{y} = -30 + 60x$

$R = \hat{y}_1 - y_1$

The residual of a tree of diameter 2 feet that yields 120 cubic feet of timber is

A. 30 cubic feet.

B. -30 cubic feet.

C. 90 cubic feet.

$y - \hat{y} = 120 =$

$-30 + 60(2)$

90

$90 - 120$

-30

---

7. Which of the following statements about the slope of the least-squares regression line is true?

A. It has the same sign as the correlation coefficient r.

B. The square of the slope equals the proportion of the variation in the response variable that is explained by the explanatory variable.

C. It is unitless.

---

8. For 10 pairs of data (x, y), we obtain the following summary statistics:

The 10 x-values have sample mean 0.30 and sample standard deviation 0.02.
The 10 y-values have sample mean 0.28 and sample standard deviation 0.04.
The correlation coefficient $r = 0.896$.

$b = .896\left(\frac{.04}{.02}\right) = 1.792$

The equation of the least-squares regression line of y on x is

A. $\hat{y} = 0.1456 + 0.448x$.

B. $\hat{y} = -0.2576 + 1.792x$.

C. $\hat{y} = -0.20176 + 1.792x$.

$a = .28 - 1.792(.3)$
$= -.2576$

$\hat{y} = -.2576 + 1.792x$

---

9. A study showed that students who spend more time studying for statistics tests tend to achieve better scores on their tests. In fact, the number of hours studied turned out to explain 81% of the observed variation in test scores among the students who participated in the study. What is the value of the correlation between number of hours studied and test score?

$\sqrt{.81} = .9$

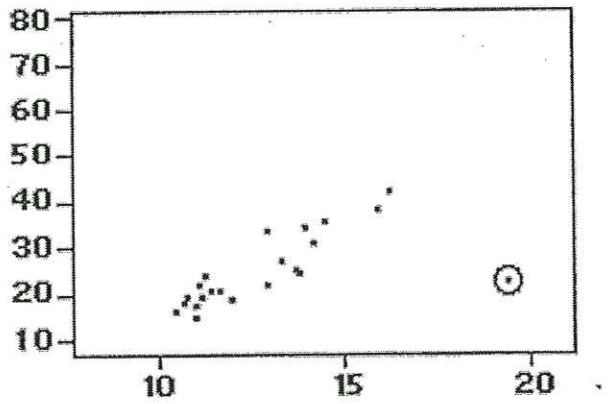
A. $r = 0.81$

B. $r = 0.656$

C. $r = 0.9$.

---

10. Suppose that the circled point were removed from the scatterplot. Which of the following would happen as a result?



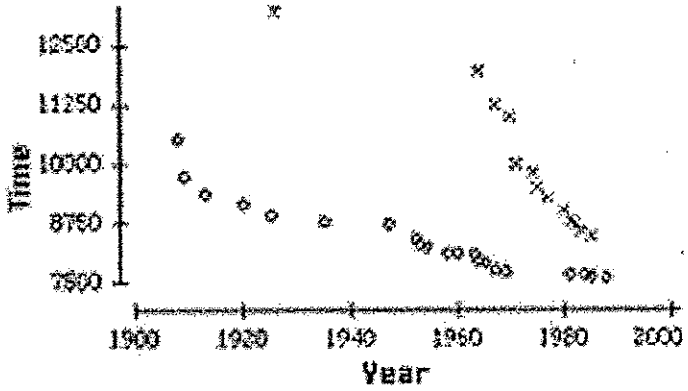
A. The slope of the regression line would increase (in the positive direction).

B. The correlation coefficient would decrease.

C. The fit of the regression line to the data would be worse.

---

11. In the scatterplot, the world record time (in minutes) in the marathon is plotted against the year in which the record was set. The plotting symbol o is used for men and x for women. The data include only records set between 1908 and 1988.



gotten faster

Based on the plot, which statement would be a valid conclusion?

A. We can expect the world record time for women to be lower than that for men sometime before the year 2010.

B. The world record times for women show a greater rate of improvement (a more rapid decrease) than the world record times for men.

C. By the year 2010, the world record time for men will reach a plateau beyond which no improvement will be possible.

---

12. A residual plot displays a "reverse fan" arrangement, with the spread of points about the line (residual = 0) gradually decreasing from left to right (that is, as x increases). Which statement would be a correct interpretation of this plot?

A. The original data display a nonlinear relationship (curved pattern of association).

B. Predictions using the regression line will be more reliable for small x than for large x.

C. Predictions using the regression line will be more reliable for large x than for small x.