

QUIZZ Worksheets

Linear Regression and Correlation 24

Total questions: 21

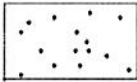
Worksheet time: 40mins

Instructor name: Linda Jackson

Name

Class

Date



1.

Estimate the correlation coefficient.

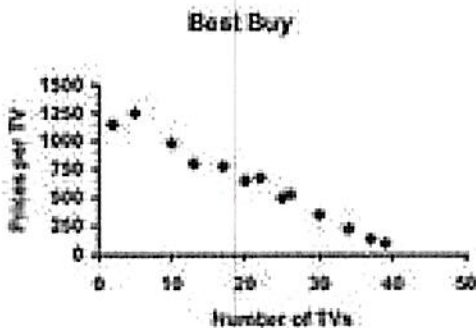
a) $r=0$

c) $r=2$

b) $r=-1$

d) $r=1$

See correlation handout



2.

Describe the correlation in the graph shown.

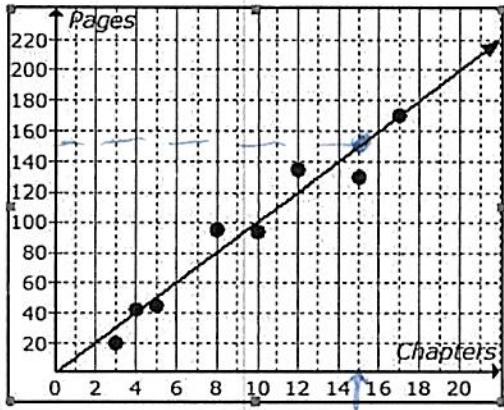
a) Strong Negative

c) Weak Positive

b) Strong Positive

d) Weak Negative

3.



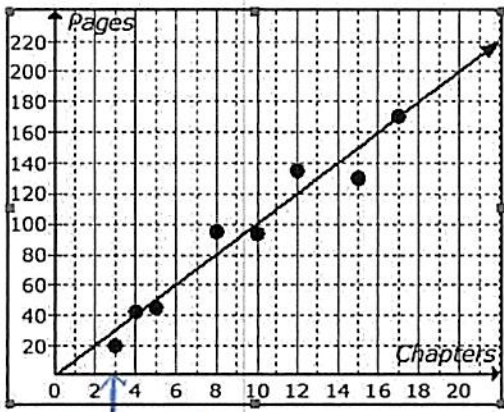
Prediction is on the line.

The scatter plot shows the relationship between the number of chapters and the total number of pages for several books. Use the least-squares regression line to predict how many pages would be in a book with 15 chapters.

- a) 180 pages
- b) 100 pages
- c) 150 pages
- d) 130 pages

← This is the observed value, not the predicted value.

4.



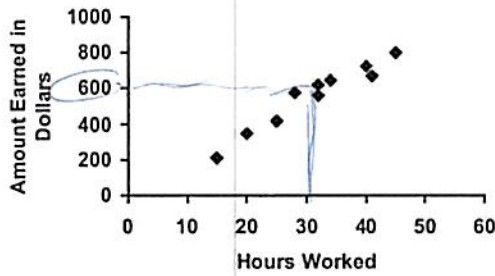
$$\begin{aligned} \text{Residual} &= \text{actual} - \text{predicted} \\ &= 20 - 30 \\ &= -10 \end{aligned}$$

The scatter plot shows the relationship between the number of chapters and the total number of pages for several books. The data set includes a 3 chapter book with 20 pages. Calculate and interpret the residual for this book.

- a) residual = 10; the book has 10 more pages than predicted by the lsrl.
- b) residual = -10; the book has 10 fewer pages than predicted by the lsrl.
- c) residual = 30; the book has 30 pages.

Observed value is below the lsrl ⇒ residual is negative

ABC Company

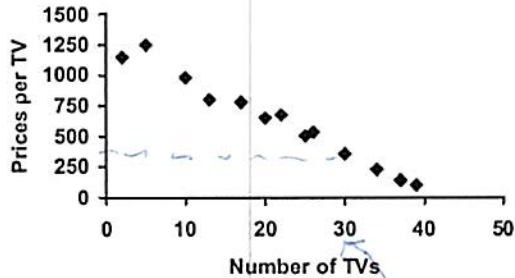


5.

A regression equation, $y = -75 + 22.5x$ estimates the dollars earned (y) for working x hours. If Joe earned \$600, how many hours would you estimate that he worked?

- a) 40 hours
- b) 20 hours
- c) 50 hours
- d) 30 hours

Best Buy



6.

If Gold's Gym is looking to buy 30 TVs, how much is each one expected to cost?

- a) 1250
- b) 750
- c) 400
- d) 0

Between 250 + 500

7. What type of correlation do you think there is between: Outside Temperature & Heating Bill

- a) NONE
- b) Negative
- c) Not enough information
- d) Positive

As temperatures go down (colder), heating bill increases on average

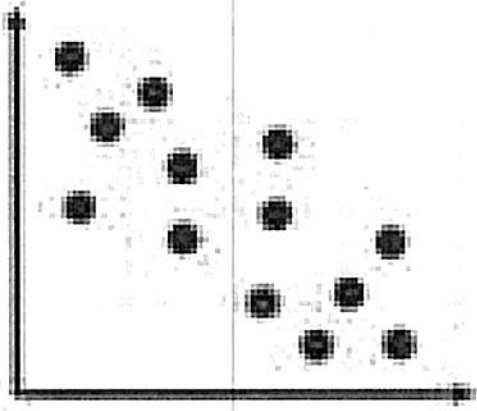


8. What type of correlation do you think there is between: The amount of time spent driving and the number of miles driven

- a) positive
- b) negative
- c) none
- d) not enough info



9.

Estimate the value of r :

a) 0.03

c) -0.99

← almost no correlation

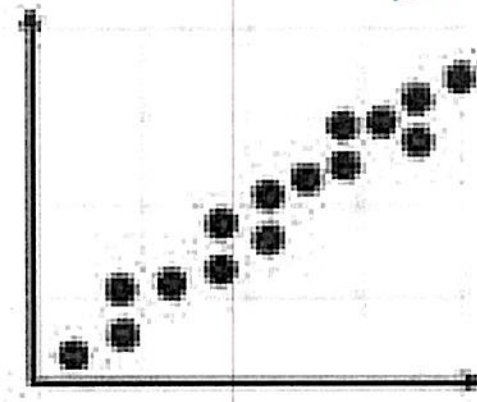
← Almost perfect line

b) -0.72

d) 0.79

← positive

10.

Estimate the value of r :

a) 0.35

c) -0.7

Weak negative

b) 0.89

11. If there is NOT a linear association between 2 quantitative variables, the correlation coefficient will be close to

a) 0

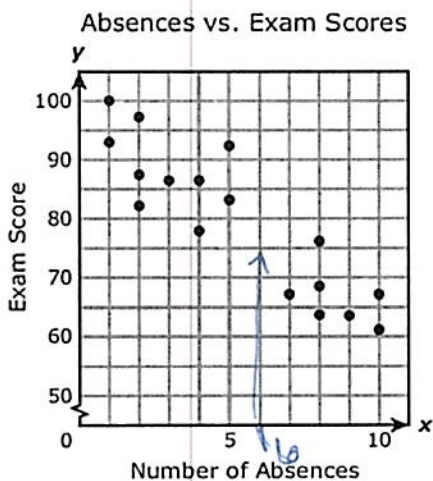
c) -1

b) 1

12. Which of the following statements shows a causal relationship and not just a correlated one? (Think cause and effect)

- a) As a child's weight increases so does her vocabulary. *This is because the child is getting older.*
- b) An individual's decision to work in a public school and his diagnosis of skin cancer.
- c) The number of minutes spent exercising and the amount of calories burned.
- d) A decrease in temperature and the increase in attendance at an roller skating rink.

Exercise causes calories to be burned



13.

Mrs. Collins made a scatterplot to show the relationship between the number of absences and a student's final exam score. Based on this scatterplot, a least-squares regression line would predict that a student with 6 absences would earn what score on the final exam?

- a) 65
- b) 76
- c) 92
- d) 70

14. A regression equation, $y = -75 + 22.5x$ estimates the dollars earned (y) for working x hours. Write a sentence interpreting the slope.

For each additional hour worked, one expects to be paid an additional \$22.50

15. A regression equation, $y = -75 + 22.5x$ estimates the dollars earned (y) for working x hours. Write a sentence interpreting the y -intercept. Is it meaningful?

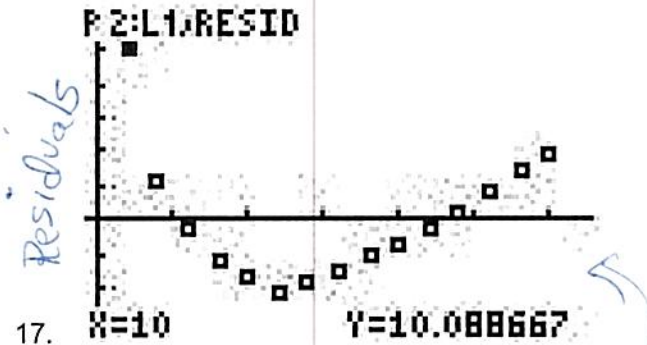
If someone works 0 hours, they will be paid -75 dollars (If someone doesn't work any hours, they will be in debt \$75). Not meaningful!

16. $\widehat{\text{final exam score}} = 10 + .9(\text{first test score})$

A linear regression equation was computed to predict the final-exam score from the score on the first test (see image). Carla scored 95 on the first test. What is the predicted value of her score on the final exam?

$$\hat{y} = 10 + .9(95)$$

- a) 95.5 b) 90
c) 95 d) 85.5



17. A study of the fuel economy for various automobiles plotted the fuel consumption (in liters of gasoline used per 100 kilometers traveled) vs. speed (in kilometers per hour). A least-squares line was fitted to the data. What does the residual plot tell you about the linear model? Choose the BEST answer.

- a) The residual plot does not confirm the linearity of the data. b) The evidence is inconclusive
c) The residual plot confirms the linearity of the fuel economy data. d) The residual plot clearly contradicts the linearity of the data.

See regression notes

18. You have data for many families on the parents' income and the years of education their eldest child completes. When you make a scatterplot,

- a) the explanatory variable is years of education, and you expect to see a negative association.
- b) the explanatory variable is years of education, and you expect to see a positive association.
- c) the explanatory variable is parents' income, and you expect to see a positive association.
- d) the explanatory variable is parents' income, and you expect to see a negative association.

Parent income influences child's educational opportunities

Dependent variable is: Fish Activ
 No Selector
 R squared = 91.8% R squared (adjusted) = 89.2%
 s = 4.785 with 7 - 2 = 5 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	1152.95	1	1152.95	50.4
Residual	114.483	5	22.8967	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	148.617	10.71	13.9	0.0001
Temp	-3.21667	0.4533	-7.1	0.0009

19.

Scientists examined the activity level of a large number of fish at 7 different temperatures. Fish activity was rated on a scale of 0 (no activity) to 100 (maximal activity). The temperature was measured in degrees Celsius. A computer regression printout is provided. Approximately what percent of the variation in fish activity is explained by the linear model on temperature?

- a) 91
- b) 89
- c) 50.4
- d) 4.785

R² definition (r-squared)

Method	Male	Female
Firearms	13,950	2,641
Poison	3,148	2,469
Hanging	3,222	709
Other	1,457	690

20.

The two-way table categorizes suicides committed in a particular year by the gender of the victim and the method used.

Which of the following statements is consistent with the table?

- a) Men display a greater tendency to use firearms to commit suicide than do women.
- b) Females prefer poison over any other method of committing suicide.
- c) More women commit suicide than men.
- d) The correlation between method of suicide and sex of the victim is clearly positive.

higher proportion of men commit suicide by gun
 $\frac{13950}{\text{total males}} > \frac{2641}{\text{total females}}$