Book

Add

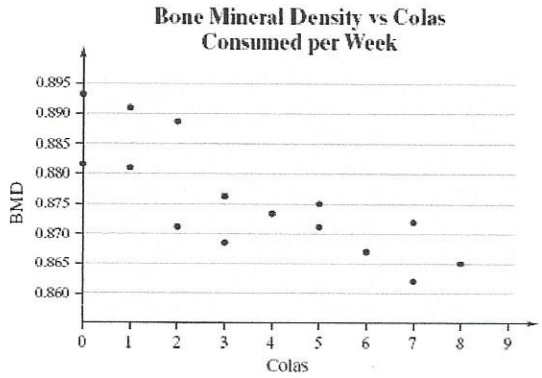
TABLE 4

Number of Colas per Week	Bone Mineral Density (g/cm^2)
0	0.893
0	0.882
1	0.891
1	0.881
2	0.888
2	0.871
3	0.868
3	0.876
4	0.873
5	0.875
5	0.871
6	0.867
7	0.862
7	0.872
8	0.865

Source: Based on data obtained from Katherine L. Tucker et. al., "Colas, but not other carbonated beverages, are associated with low bone mineral density in older women: The Framingham Osteoporosis Study." *American Journal of Clinical Nutrition* 2006, 84:936–942.

Now Work Problem 41

Figure 9



Solution In prospective cohort studies, data are collected on a group of subjects through questionnaires and surveys over time. Therefore, the data are observational. So the researchers cannot claim that increased cola consumption causes a decrease in bone mineral density.

In their article, the authors identified a number of lurking variables that could confound the results:

> Variables that could potentially confound the relation between cola consumption and bone mineral density ... included the following: age, body mass index, height, smoking, average daily intakes of alcohol, calcium, caffeine, total energy intake, physical activity, season of measurement, estrogen use, and menopause status.

The authors were careful to say that increased cola consumption is *associated* with lower bone mineral density because of potential lurking variables. They never stated that increased cola consumption *causes* lower bone mineral density.

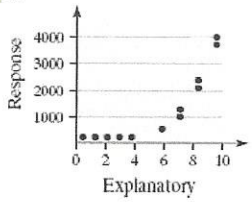
4.1 ASSESS YOUR UNDERSTANDING

VOCABULARY AND SKILL BUILDING

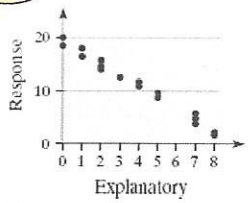
1. What is the difference between univariate data and bivariate data?
2. The _____ variable is the variable whose value can be explained by the value of the explanatory variable.
3. A _____ _____ is a graph that shows the relation between two quantitative variables.
4. What does it mean to say two variables are positively associated? Negatively associated?
5. If r = _____, then a perfect negative linear relation exists between the two quantitative variables.
6. *True or False*: If the linear correlation coefficient is close to 0, then the two variables have no relation.
7. A _____ variable is a variable that is related to both the explanatory and response variable.
8. *True or False*: Correlation implies causation.

In Problems 9–12, determine whether the scatter diagram indicates that a linear relation may exist between the two variables. If the relation is linear, determine whether it indicates a positive or negative association between the variables.

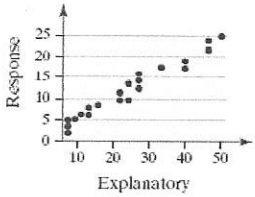
NW 9.



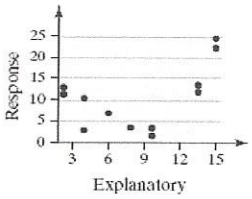
10.



11.



12.

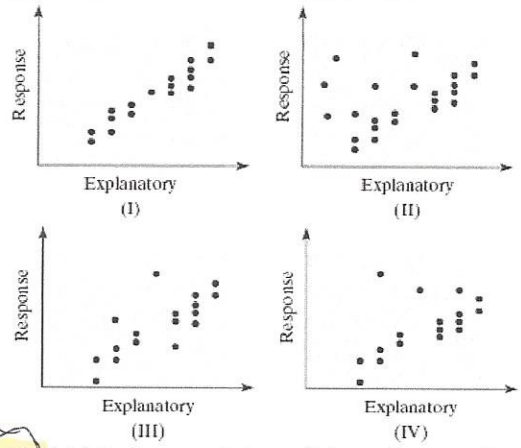


Book

Add

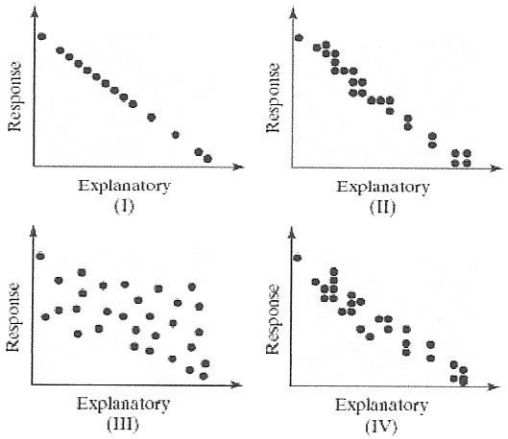
NW 13. Match the linear correlation coefficient to the scatter diagrams. The scales on the x- and y-axes are the same for each diagram.

(a) $r = 0.787$ **(b)** $r = 0.523$
(c) $r = 0.053$ **(d)** $r = 0.946$

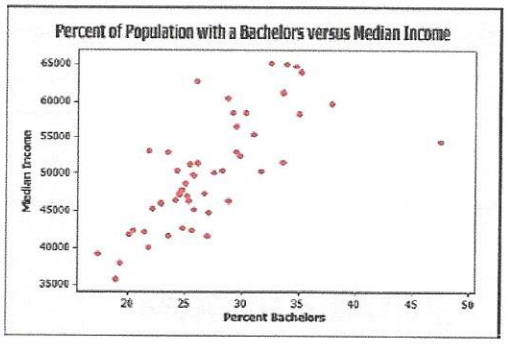


14. Match the linear correlation coefficient to the scatter diagram. The scales on the x- and y-axes are the same for each diagram.

(a) $r = -0.969$ **(b)** $r = -0.049$
(c) $r = -1$ **(d)** $r = -0.992$



15. Does Education Pay? The scatter diagram drawn in MINITAB shows the relation between the percentage of the population of a state plus Washington, DC, that has at least a bachelor's degree and the median income (in dollars) of the state for 2009. *Source:* U.S. Census Bureau

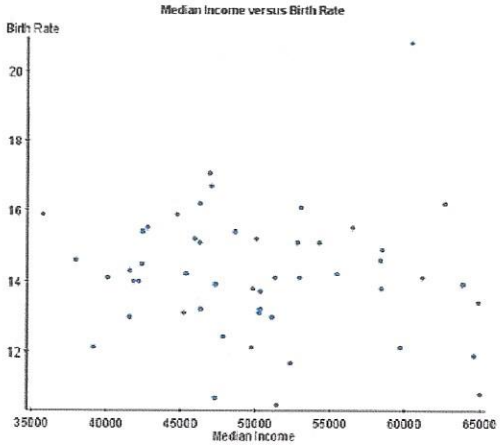


(a) Describe any relation that exists between level of education and median income.

(b) One observation appears to stick out from the rest. Which one? This particular observation is for Washington, DC. Can you think of any reasons why Washington, DC, might have a high percentage of residents with a bachelor's degree, but a lower than expected median income?

(c) The correlation coefficient between the percentage of the population with a bachelor's degree and median income with Washington, DC, included in the data set is 0.718. The correlation coefficient without Washington, DC, included is 0.802. What property does this illustrate about the linear correlation coefficient?

16. Relation between Income and Birthrate? The following scatter diagram drawn in StatCrunch shows the relation between median income (in dollars) in a state and birthrate (births per 1000 women 15 to 44 years old).



(a) Describe any relation that exists between median income and birthrate.

(b) One observation sticks out from the rest. Which one? This particular observation is for the state of Utah. Are there any explanations for this result?

(c) The correlation coefficient between median income and birthrate is −0.070. What does this imply about the relation between median income and birthrate?

In Problems 17–20, (a) draw a scatter diagram of the data, (b) by hand, compute the correlation coefficient, and (c) determine whether there is a linear relation between x and y.

17.

x	2	4	6	6	7
y	4	8	10	13	20

18.

x	2	3	5	6	6
y	10	9	7	4	2

Book

19.

x	2	6	6	7	9
y	8	7	6	9	5

20.

x	0	5	7	8	9
y	3	8	6	9	4

21. Name the Relation, Part I For each of the following statements, explain whether you think the variables will have positive correlation, negative correlation, or no correlation. Support your opinion.

(a) Number of children in the household under the age of 3 and expenditures on diapers
(b) Interest rates on car loans and number of cars sold
(c) Number of hours per week on the treadmill and cholesterol level
(d) Price of a Big Mac and number of McDonald's French fries sold in a week
(e) Shoe size and IQ

22. Name the Relation, Part II For each of the following statements, explain whether you think the variables will have positive correlation, negative correlation, or no correlation. Support your opinion.

(a) Number of cigarettes smoked by a pregnant woman each week and birth weight of her baby
(b) Years of education and annual salary
(c) Number of doctors on staff at a hospital and number of administrators on staff
(d) Head circumference and IQ
(e) Number of movie goers and movie ticket price

APPLYING THE CONCEPTS

23. The TIMMS Exam The Trends in International Mathematics and Science (TIMMS) is a mathematics and science achievement exam given internationally. On each exam, students are asked to respond to a variety of background questions. For the 41 nations that participated in TIMMS, the correlation between the percentage of items answered in the background questionnaire (used as a proxy for student task persistence) and mean score on the exam was 0.79. Does this suggest there is a linear relation between student task persistence and achievement score? Write a sentence that explains what this result might mean.

24. The TIMMS Exam Part II (See Problem 23) For the 41 nations that participated in TIMMS, the correlation between the percentage of students who skipped class at least once in the past month and the mean score on the exam was −0.52. Does this suggest there is a linear relation between attendance and achievement score? Write a sentence that explains what this result might mean.

NW **25. An Unhealthy Commute** The Gallup Organization regularly surveys adult Americans regarding their commute time to work. In addition, they administer a Well-Being Survey. According to the Gallup Organization, "The Gallup-Healthways Well-Being Index Composite Score is comprised of six sub-indices: Life Evaluation, Emotional Health, Physical Health, Healthy Behavior, Work Environment and Basic Access." A complete description of the index can be found at http://www.well-beingindex.com/. The data in the following table are based on the results of the survey, which represent commute time to work (in minutes) and well-being index score.

Add

Commute Time (in minutes)	Gallup-Healthways Well-Being Index Composite Score
5	69.2
15	68.3
25	67.5
35	67.1
50	66.4
72	66.1
105	63.9

Source: The Gallup Organization

(a) Which variable do you believe is likely the explanatory variable and which is the response variable?
(b) Draw a scatter diagram of the data.
(c) Determine the linear correlation coefficient between commute time and well-being index score.
(d) Does a linear relation exist between the commute time and well-being index score?

26. Credit Scores Your Fair Isaacs Corporation (FICO) credit score is used to determine your creditworthiness. It is used to help determine whether you qualify for a mortgage or credit and is even used to determine insurance rates. FICO scores have a range of 300 to 850, with a higher score indicating a better credit history. The given data represent the interest rate (in percent) a bank would offer on a 36-month auto loan for various FICO scores.

Credit Score	Interest Rate (percent)
545	18.982
595	17.967
640	12.218
675	8.612
705	6.680
750	5.150

Source: www.myfico.com

(a) Which variable do you believe is likely the explanatory variable and which is the response variable?
(b) Draw a scatter diagram of the data.
(c) Determine the linear correlation coefficient between FICO score and interest rate on a 36-month auto loan.
(d) Does a linear relation exist between the FICO score and interest rate?

27. Height versus Head Circumference A pediatrician wants to determine the relation that may exist between a child's height and head circumference. She randomly selects eleven 3-year-old children from her practice, measures their heights and head circumference, and obtains the data shown in the table on the next page.