

10.E: Correlation and Regression (Exercises)

These are homework exercises to accompany the Textmap created for "Introductory Statistics" by Shafer and Zhang.

10.1 Linear Relationships Between Variables

Basic

1. A line has equation $y = 0.5x + 2$.
 - a. Pick five distinct x -values, use the equation to compute the corresponding y -values, and plot the five points obtained.
 - b. Give the value of the slope of the line; give the value of the y -intercept.
2. A line has equation $y = x - 0.5$.
 - a. Pick five distinct x -values, use the equation to compute the corresponding y -values, and plot the five points obtained.
 - b. Give the value of the slope of the line; give the value of the y -intercept.
3. A line has equation $y = -2x + 4$.
 - a. Pick five distinct x -values, use the equation to compute the corresponding y -values, and plot the five points obtained.
 - b. Give the value of the slope of the line; give the value of the y -intercept.
4. A line has equation $y = -1.5x + 1$.
 - a. Pick five distinct x -values, use the equation to compute the corresponding y -values, and plot the five points obtained.
 - b. Give the value of the slope of the line; give the value of the y -intercept.
5. Based on the information given about a line, determine how y will change (increase, decrease, or stay the same) when x is increased, and explain. In some cases it might be impossible to tell from the information given.
 - a. The slope is positive.
 - b. The y -intercept is positive.
 - c. The slope is zero.
6. Based on the information given about a line, determine how y will change (increase, decrease, or stay the same) when x is increased, and explain. In some cases it might be impossible to tell from the information given.
 - a. The y -intercept is negative.
 - b. The y -intercept is zero.
 - c. The slope is negative.
7. A data set consists of eight (x, y) pairs of numbers:

$$\begin{array}{cccc} (0, 12) & (4, 16) & (8, 22) & (15, 28) \\ (2, 15) & (5, 14) & (13, 24) & (20, 30) \end{array} \quad (10.E.1)$$

 - a. Plot the data in a scatter diagram.
 - b. Based on the plot, explain whether the relationship between x and y appears to be deterministic or to involve randomness.
 - c. Based on the plot, explain whether the relationship between x and y appears to be linear or not linear.

8. A data set consists of ten (x, y) pairs of numbers:

$$\begin{array}{ccccc} (3, 20) & (6, 9) & (11, 0) & (14, 1) & (18, 9) \\ (5, 13) & (8, 4) & (12, 0) & (17, 6) & (20, 16) \end{array} \quad (10.E.2)$$

 - a. Plot the data in a scatter diagram.
 - b. Based on the plot, explain whether the relationship between x and y appears to be deterministic or to involve randomness.
 - c. Based on the plot, explain whether the relationship between x and y appears to be linear or not linear.

9. A data set consists of nine (x, y) pairs of numbers:

$$\begin{array}{ccccc} (8, 16) & (10, 4) & (12, 0) & (14, 4) & (16, 16) \\ (9, 9) & (11, 1) & (13, 1) & (15, 9) & \end{array} \quad (10.E.3)$$

 - a. Plot the data in a scatter diagram.
 - b. Based on the plot, explain whether the relationship between x and y appears to be deterministic or to involve randomness.
 - c. Based on the plot, explain whether the relationship between x and y appears to be linear or not linear.

10. A data set consists of five (x, y) pairs of numbers:

$$(0, 1) \quad (2, 5) \quad (3, 7) \quad (5, 11) \quad (8, 17) \quad (10.E.4)$$

- Plot the data in a scatter diagram.
- Based on the plot, explain whether the relationship between x and y appears to be deterministic or to involve randomness.
- Based on the plot, explain whether the relationship between x and y appears to be linear or not linear.

Applications

11. At $60^\circ F$ a particular blend of automotive gasoline weighs 6.17 lb/gal. The weight y of gasoline on a tank truck that is loaded with x gallons of gasoline is given by the linear equation

$$y = 6.17x \quad (10.E.5)$$

- Explain whether the relationship between the weight y and the amount x of gasoline is deterministic or contains an element of randomness.
 - Predict the weight of gasoline on a tank truck that has just been loaded with 6,750 gallons of gasoline.
12. The rate for renting a motor scooter for one day at a beach resort area is \$25 plus 30 cents for each mile the scooter is driven. The total cost y in dollars for renting a scooter and driving it x miles is

$$y = 0.30x + 25 \quad (10.E.6)$$

- Explain whether the relationship between the cost y of renting the scooter for a day and the distance x that the scooter is driven that day is deterministic or contains an element of randomness.
 - A person intends to rent a scooter one day for a trip to an attraction 17 miles away. Assuming that the total distance the scooter is driven is 34 miles, predict the cost of the rental.
13. The pricing schedule for labor on a service call by an elevator repair company is \$150 plus \$50 per hour on site.
- Write down the linear equation that relates the labor cost y to the number of hours x that the repairman is on site.
 - Calculate the labor cost for a service call that lasts 2.5 hours.
14. The cost of a telephone call made through a leased line service is 2.5 cents per minute.
- Write down the linear equation that relates the cost y (in cents) of a call to its length x .
 - Calculate the cost of a call that lasts 23 minutes.

Large Data Set Exercises

Large Data Sets not available

- Large Data Set 1 lists the SAT scores and GPAs of 1,000 students. Plot the scatter diagram with SAT score as the independent variable (x) and GPA as the dependent variable (y). Comment on the appearance and strength of any linear trend.
- Large Data Set 12 lists the golf scores on one round of golf for 75 golfers first using their own original clubs, then using clubs of a new, experimental design (after two months of familiarization with the new clubs). Plot the scatter diagram with golf score using the original clubs as the independent variable (x) and golf score using the new clubs as the dependent variable (y). Comment on the appearance and strength of any linear trend.
- Large Data Set 13 records the number of bidders and sales price of a particular type of antique grandfather clock at 60 auctions. Plot the scatter diagram with the number of bidders at the auction as the independent variable (x) and the sales price as the dependent variable (y). Comment on the appearance and strength of any linear trend.

Answers

- Answers vary.
 - Slope $m = 0.5$; y -intercept $b = 2$.
-
- Answers vary.
 - Slope $m = -2$; y -intercept $b = 4$.
-
-

- a. y increases.
 - b. Impossible to tell.
 - c. y does not change.
- 6.
- a. Scatter diagram needed.
 - b. Involves randomness.
 - c. Linear.
- 8.
- a. Scatter diagram needed.
 - b. Deterministic.
 - c. Not linear.
- 10.
- a. Deterministic.
 - b. 41,647.5pounds.
- 12.
- a. $y = 50x + 150$.
 - b. \$275.
- 14.
15. There appears to a hint of some positive correlation.
- 16.
17. There appears to be clear positive correlation.

10.2 The Linear Correlation Coefficient

Basic

With the exception of the exercises at the end of Section 10.3, the first Basic exercise in each of the following sections through Section 10.7 uses the data from the first exercise here, the second Basic exercise uses the data from the second exercise here, and so on, and similarly for the Application exercises. Save your computations done on these exercises so that you do not need to repeat them later.

1. For the sample data

x	0	1	3	5	8
y	2	4	6	5	9

(10.E.7)

- a. Draw the scatter plot.
 - b. Based on the scatter plot, predict the sign of the linear correlation coefficient. Explain your answer.
 - c. Compute the linear correlation coefficient and compare its sign to your answer to part (b).
2. For the sample data

x	0	2	3	6	9
y	0	3	3	4	8

(10.E.8)

- a. Draw the scatter plot.
 - b. Based on the scatter plot, predict the sign of the linear correlation coefficient. Explain your answer.
 - c. Compute the linear correlation coefficient and compare its sign to your answer to part (b).
3. For the sample data

x	1	3	4	6	8
y	4	1	3	-1	0

(10.E.9)

- a. Draw the scatter plot.
- b. Based on the scatter plot, predict the sign of the linear correlation coefficient. Explain your answer.

c. Compute the linear correlation coefficient and compare its sign to your answer to part (b).

4. For the sample data

$$\begin{array}{c|ccccc} x & 1 & 2 & 4 & 7 & 9 \\ \hline y & 5 & 5 & 6 & -3 & 0 \end{array} \quad (10.E.10)$$

a. Draw the scatter plot.

b. Based on the scatter plot, predict the sign of the linear correlation coefficient. Explain your answer.

c. Compute the linear correlation coefficient and compare its sign to your answer to part (b).

5. For the sample data

$$\begin{array}{c|ccccc} x & 1 & 1 & 3 & 4 & 5 \\ \hline y & 2 & 1 & 5 & 3 & 4 \end{array} \quad (10.E.11)$$

a. Draw the scatter plot.

b. Based on the scatter plot, predict the sign of the linear correlation coefficient. Explain your answer.

c. Compute the linear correlation coefficient and compare its sign to your answer to part (b).

6. For the sample data

$$\begin{array}{c|ccccc} x & 1 & 3 & 5 & 5 & 8 \\ \hline y & 5 & -2 & 2 & -1 & -3 \end{array} \quad (10.E.12)$$

a. Draw the scatter plot.

b. Based on the scatter plot, predict the sign of the linear correlation coefficient. Explain your answer.

c. Compute the linear correlation coefficient and compare its sign to your answer to part (b).

7. Compute the linear correlation coefficient for the sample data summarized by the following information:

$$\begin{aligned} n &= 5 & \sum x &= 25 & \sum x^2 &= 165 \\ \sum y &= 24 & \sum y^2 &= 134 & \sum xy &= 144 \\ & & 1 \leq x \leq 9 & & & \end{aligned} \quad (10.E.13)$$

8. Compute the linear correlation coefficient for the sample data summarized by the following information:

$$\begin{aligned} n &= 5 & \sum x &= 31 & \sum x^2 &= 253 \\ \sum y &= 18 & \sum y^2 &= 90 & \sum xy &= 148 \\ & & 2 \leq x \leq 12 & & & \end{aligned} \quad (10.E.14)$$

9. Compute the linear correlation coefficient for the sample data summarized by the following information:

$$\begin{aligned} n &= 10 & \sum x &= 0 & \sum x^2 &= 60 \\ \sum y &= 24 & \sum y^2 &= 234 & \sum xy &= -87 \\ & & -4 \leq x \leq 4 & & & \end{aligned} \quad (10.E.15)$$

10. Compute the linear correlation coefficient for the sample data summarized by the following information:

$$\begin{aligned} n &= 10 & \sum x &= -3 & \sum x^2 &= 263 \\ \sum y &= 55 & \sum y^2 &= 917 & \sum xy &= -355 \\ & & -10 \leq x \leq 10 & & & \end{aligned} \quad (10.E.16)$$

Applications

11. The age x in months and vocabulary y were measured for six children, with the results shown in the table.

$$\begin{array}{c|cccccc} x & 13 & 14 & 15 & 16 & 16 & 18 \\ \hline y & 8 & 10 & 15 & 20 & 27 & 30 \end{array} \quad (10.E.17)$$

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

12. The curb weight x in hundreds of pounds and braking distance y in feet, at 50 miles per hour on dry pavement, were measured for five vehicles, with the results shown in the table.

x	25	27.5	32.5	35	45
y	105	125	140	140	150

(10.E.18)

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

13. The age x and resting heart rate y were measured for ten men, with the results shown in the table.

x	20	23	30	37	35
y	72	71	73	74	74
x	45	51	55	60	63
y	73	72	79	75	77

(10.E.19)

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

14. The wind speed x in miles per hour and wave height y in feet were measured under various conditions on an enclosed deep water sea, with the results shown in the table,

x	0	0	2	7	7
y	2.0	0.0	0.3	0.7	3.3
x	9	13	20	22	31
y	4.9	4.9	3.0	6.9	5.9

(10.E.20)

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

15. The advertising expenditure x and sales y in thousands of dollars for a small retail business in its first eight years in operation are shown in the table.

x	1.4	1.6	1.6	2.0
y	180	184	190	220
x	2.0	2.2	2.4	2.6
y	186	215	205	240

(10.E.21)

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

16. The height x at age 2 and y at age 20, both in inches, for ten women are tabulated in the table.

x	31.3	31.7	32.5	33.5	34.4
y	60.7	61.0	63.1	64.2	65.9
x	35.2	35.8	32.7	33.6	34.8
y	68.2	67.6	62.3	64.9	66.8

(10.E.22)

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

17. The course average x just before a final exam and the score y on the final exam were recorded for 15 randomly selected students in a large physics class, with the results shown in the table.

x	69.3	87.7	50.5	51.9	82.7
y	56	89	55	49	61

(10.E.23)

x	70.5	72.4	91.7	83.3	86.5
y	66	72	83	73	82

x	79.3	78.5	75.7	52.3	62.2
y	92	80	64	18	76

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

18. The table shows the acres x of corn planted and acres y of corn harvested, in millions of acres, in a particular country in ten successive years.

x	75.7	78.9	78.6	80.9	81.8
y	68.8	69.3	70.9	73.6	75.1

(10.E.24)

x	78.3	93.5	85.9	86.4	88.2
y	70.6	86.5	78.6	79.5	81.4

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

19. Fifty male subjects drank a measured amount x (in ounces) of a medication and the concentration y (in percent) in their blood of the active ingredient was measured 30 minutes later. The sample data are summarized by the following information.

$$\begin{aligned}
 n &= 50 & \sum x &= 112.5 & \sum y &= 4.83 \\
 \sum xy &= 15.255 & 0 \leq x &\leq 4.5 \\
 \sum x^2 &= 356.25 & \sum y^2 &= 0.667
 \end{aligned}
 \tag{10.E.25}$$

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

20. In an effort to produce a formula for estimating the age of large free-standing oak trees non-invasively, the girth x (in inches) five feet off the ground of 15 such trees of known age y (in years) was measured. The sample data are summarized by the following information.

$$\begin{aligned}
 n &= 15 & \sum x &= 3368 & \sum y &= 6496 \\
 \sum xy &= 1,933,219 & 74 \leq x &\leq 395 \\
 \sum x^2 &= 917,780 & \sum y^2 &= 4,260,666
 \end{aligned}
 \tag{10.E.26}$$

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

21. Construction standards specify the strength of concrete 28 days after it is poured. For 30 samples of various types of concrete the strength x after 3 days and the strength y after 28 days (both in hundreds of pounds per square inch) were measured. The sample data are summarized by the following information.

$$\begin{aligned}
 n &= 30 & \sum x &= 501.6 & \sum y &= 1338.8 \\
 \sum xy &= 23,246.55 & 11 \leq x &\leq 22 \\
 \sum x^2 &= 8724.74 & \sum y^2 &= 61,980.14
 \end{aligned}
 \tag{10.E.27}$$

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

22. Power-generating facilities used forecasts of temperature to forecast energy demand. The average temperature x (degrees Fahrenheit) and the day's energy demand y (million watt-hours) were recorded on 40 randomly selected winter days in the

region served by a power company. The sample data are summarized by the following information.

$$\begin{aligned} n &= 40 & \sum x &= 2000 & \sum y &= 2969 \\ \sum xy &= 143,042 & 40 &\leq x \leq 60 \\ \sum x^2 &= 101,340 & \sum y^2 &= 243,027 \end{aligned} \quad (10.E.28)$$

Compute the linear correlation coefficient for these sample data and interpret its meaning in the context of the problem.

Additional Exercises

23. In each case state whether you expect the two variables x and y indicated to have positive, negative, or zero correlation.
 - a. the number x of pages in a book and the age y of the author
 - b. the number x of pages in a book and the age y of the intended reader
 - c. the weight x of an automobile and the fuel economy y in miles per gallon
 - d. the weight x of an automobile and the reading y on its odometer
 - e. the amount x of a sedative a person took an hour ago and the time y it takes him to respond to a stimulus
24. In each case state whether you expect the two variables x and y indicated to have positive, negative, or zero correlation.
 - a. the length x of time an emergency flare will burn and the length y of time the match used to light it burned
 - b. the average length x of time that calls to a retail call center are on hold one day and the number y of calls received that day
 - c. the length x of a regularly scheduled commercial flight between two cities and the headwind y encountered by the aircraft
 - d. the value x of a house and the its size y in square feet
 - e. the average temperature x on a winter day and the energy consumption y of the furnace
25. Changing the units of measurement on two variables x and y should not change the linear correlation coefficient. Moreover, most change of units amount to simply multiplying one unit by the other (for example, 1 foot = 12 inches). Multiply each x value in the table in Exercise 1 by two and compute the linear correlation coefficient for the new data set. Compare the new value of r to the one for the original data.
26. Refer to the previous exercise. Multiply each x value in the table in Exercise 2 by two, multiply each y value by three, and compute the linear correlation coefficient for the new data set. Compare the new value of r to the one for the original data.
27. Reversing the roles of x and y in the data set of Exercise 1 produces the data set

$$\begin{array}{c|ccccc} x & 2 & 4 & 6 & 5 & 9 \\ \hline y & 0 & 1 & 3 & 5 & 8 \end{array} \quad (10.E.29)$$

Compute the linear correlation coefficient of the new set of data and compare it to what you got in Exercise 1.

28. In the context of the previous problem, look at the formula for r and see if you can tell why what you observed there must be true for every data set.

Large Data Set Exercises

Large Data Sets not available

29. Large Data Set 1 lists the SAT scores and GPAs of 1,000 students. Compute the linear correlation coefficient r . Compare its value to your comments on the appearance and strength of any linear trend in the scatter diagram that you constructed in the first large data set problem for Section 10.1.
30. Large Data Set 12 lists the golf scores on one round of golf for 75 golfers first using their own original clubs, then using clubs of a new, experimental design (after two months of familiarization with the new clubs). Compute the linear correlation coefficient r . Compare its value to your comments on the appearance and strength of any linear trend in the scatter diagram that you constructed in the second large data set problem for Section 10.1.
31. Large Data Set 13 records the number of bidders and sales price of a particular type of antique grandfather clock at 60 auctions. Compute the linear correlation coefficient r . Compare its value to your comments on the appearance and strength of any linear trend in the scatter diagram that you constructed in the third large data set problem for Section 10.1.

Answers

1. $r = 0.921$
- 2.
3. $r = -0.794$
- 4.
5. $r = 0.707$
- 6.
7. 0.875
- 8.
9. -0.846
- 10.
11. 0.948
- 12.
13. 0.709
- 14.
15. 0.832
- 16.
17. 0.751
- 18.
19. 0.965
- 20.
21. 0.992
- 22.
23.
 - a. zero
 - b. positive
 - c. negative
 - d. zero
 - e. positive
- 24.
25. same value
- 26.
27. same value
- 28.
29. $r = 0.4601$
- 30.
31. $r = 0.9002$

10.3 Modelling Linear Relationships with Randomness Present

Basic

1. State the three assumptions that are the basis for the Simple Linear Regression Model.
2. The Simple Linear Regression Model is summarized by the equation

$$y = \beta_1 x + \beta_0 + \varepsilon \quad (10.E.30)$$

Identify the deterministic part and the random part.

3. Is the number β_1 in the equation $y = \beta_1 x + \beta_0$ a statistic or a population parameter? Explain.
4. Is the number σ in the Simple Linear Regression Model a statistic or a population parameter? Explain.
5. Describe what to look for in a scatter diagram in order to check that the assumptions of the Simple Linear Regression Model are true.
6. True or false: the assumptions of the Simple Linear Regression Model must hold exactly in order for the procedures and analysis developed in this chapter to be useful.

Answers

1.
 - i. The mean of y is linearly related to x .
 - ii. For each given x , y is a normal random variable with mean $\beta_1 x + \beta_0$ and a standard deviation σ .
 - iii. All the observations of y in the sample are independent.
- 2.
3. β_1 is a population parameter.
- 4.
5. A linear trend.

10.4 The Least Squares Regression Line

Basic

For the Basic and Application exercises in this section use the computations that were done for the exercises with the same number in Section 10.2.

1. Compute the least squares regression line for the data in Exercise 1 of Section 10.2.
2. Compute the least squares regression line for the data in Exercise 2 of Section 10.2.
3. Compute the least squares regression line for the data in Exercise 3 of Section 10.2.
4. Compute the least squares regression line for the data in Exercise 4 of Section 10.2.
5. For the data in Exercise 5 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Compute the sum of the squared errors SSE using the definition $\sum (y - \hat{y})^2$.
 - c. Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \widehat{\beta}_1 SS_{xy}$.
6. For the data in Exercise 6 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Compute the sum of the squared errors SSE using the definition $\sum (y - \hat{y})^2$.
 - c. Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \widehat{\beta}_1 SS_{xy}$.
7. Compute the least squares regression line for the data in Exercise 7 of Section 10.2.
8. Compute the least squares regression line for the data in Exercise 8 of Section 10.2.
9. For the data in Exercise 9 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Can you compute the sum of the squared errors SSE using the definition $\sum (y - \hat{y})^2$? Explain.
 - c. Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \widehat{\beta}_1 SS_{xy}$.
10. For the data in Exercise 10 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Can you compute the sum of the squared errors SSE using the definition $\sum (y - \hat{y})^2$? Explain.
 - c. Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \widehat{\beta}_1 SS_{xy}$.

Applications

11. For the data in Exercise 11 of Section 10.2
 - a. Compute the least squares regression line.
 - b. On average, how many new words does a child from 13 to 18 months old learn each month? Explain.
 - c. Estimate the average vocabulary of all 16-month-old children.
12. For the data in Exercise 12 of Section 10.2
 - a. Compute the least squares regression line.
 - b. On average, how many additional feet are added to the braking distance for each additional 100 pounds of weight? Explain.
 - c. Estimate the average braking distance of all cars weighing 3,000 pounds.
13. For the data in Exercise 13 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Estimate the average resting heart rate of all 40-year-old men.

- c. Estimate the average resting heart rate of all newborn baby boys. Comment on the validity of the estimate.
14. For the data in Exercise 14 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Estimate the average wave height when the wind is blowing at 10 miles per hour.
 - c. Estimate the average wave height when there is no wind blowing. Comment on the validity of the estimate.
15. For the data in Exercise 15 of Section 10.2
 - a. Compute the least squares regression line.
 - b. On average, for each additional thousand dollars spent on advertising, how does revenue change? Explain.
 - c. Estimate the revenue if \$2,500 is spent on advertising next year.
16. For the data in Exercise 16 of Section 10.2
 - a. Compute the least squares regression line.
 - b. On average, for each additional inch of height of two-year-old girl, what is the change in the adult height? Explain.
 - c. Predict the adult height of a two-year-old girl who is 33 inches tall.
17. For the data in Exercise 17 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Compute SSE using the formula $SSE = SS_{yy} - \widehat{\beta}_1 SS_{xy}$.
 - c. Estimate the average final exam score of all students whose course average just before the exam is 85.
18. For the data in Exercise 18 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Compute SSE using the formula $SSE = SS_{yy} - \widehat{\beta}_1 SS_{xy}$.
 - c. Estimate the number of acres that would be harvested if 90 million acres of corn were planted.
19. For the data in Exercise 19 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Interpret the value of the slope of the least squares regression line in the context of the problem.
 - c. Estimate the average concentration of the active ingredient in the blood in men after consuming 1 ounce of the medication.
20. For the data in Exercise 20 of Section 10.2
 - a. Compute the least squares regression line.
 - b. Interpret the value of the slope of the least squares regression line in the context of the problem.
 - c. Estimate the age of an oak tree whose girth five feet off the ground is 92 inches.
21. For the data in Exercise 21 of Section 10.2
 - a. Compute the least squares regression line.
 - b. The 28-day strength of concrete used on a certain job must be at least 3,200 psi. If the 3-day strength is 1,300 psi, would we anticipate that the concrete will be sufficiently strong on the 28th day? Explain fully.
22. For the data in Exercise 22 of Section 10.2
 - a. Compute the least squares regression line.
 - b. If the power facility is called upon to provide more than 95 million watt-hours tomorrow then energy will have to be purchased from elsewhere at a premium. The forecast is for an average temperature of 42 degrees. Should the company plan on purchasing power at a premium?

Additional Exercises

23. Verify that no matter what the data are, the least squares regression line always passes through the point with coordinates (\bar{x}, \bar{y}) .
Hint: Find the predicted value of y when $x = \bar{x}$.
24. In Exercise 1 you computed the least squares regression line for the data in Exercise 1 of Section 10.2.
 - a. Reverse the roles of x and y and compute the least squares regression line for the new data set

$$\begin{array}{c|ccccc} x & 2 & 4 & 6 & 5 & 9 \\ \hline y & 0 & 1 & 3 & 5 & 8 \end{array} \quad (10.E.31)$$

- b. Interchanging x and y corresponds geometrically to reflecting the scatter plot in a 45-degree line. Reflecting the regression line for the original data the same way gives a line with the equation $\hat{y} = 1.346x - 3.600$. Is this the equation that you got in part (a)? Can you figure out why not? Hint: Think about how x and y are treated differently geometrically in the computation of the goodness of fit.
- c. Compute SSE for each line and see if they fit the same, or if one fits the data better than the other.

Large Data Set Exercises

Large Data Sets not available

25. Large Data Set 1 lists the SAT scores and GPAs of 1,000 students.
 - a. Compute the least squares regression line with SAT score as the independent variable (x) and GPA as the dependent variable (y).
 - b. Interpret the meaning of the slope $\hat{\beta}_1$ of regression line in the context of problem.
 - c. Compute SSE the measure of the goodness of fit of the regression line to the sample data.
 - d. Estimate the GPA of a student whose SAT score is 1350.
26. Large Data Set 12 lists the golf scores on one round of golf for 75 golfers first using their own original clubs, then using clubs of a new, experimental design (after two months of familiarization with the new clubs).
 - a. Compute the least squares regression line with scores using the original clubs as the independent variable (x) and scores using the new clubs as the dependent variable (y).
 - b. Interpret the meaning of the slope $\hat{\beta}_1$ of regression line in the context of problem.
 - c. Compute SSE the measure of the goodness of fit of the regression line to the sample data.
 - d. Estimate the score with the new clubs of a golfer whose score with the old clubs is 73.
27. Large Data Set 13 records the number of bidders and sales price of a particular type of antique grandfather clock at 60 auctions.
 - a. Compute the least squares regression line with the number of bidders present at the auction as the independent variable (x) and sales price as the dependent variable (y).
 - b. Interpret the meaning of the slope $\hat{\beta}_1$ of regression line in the context of problem.
 - c. Compute SSE the measure of the goodness of fit of the regression line to the sample data.
 - d. Estimate the sales price of a clock at an auction at which the number of bidders is seven.

Answers

1. $\hat{y} = 0.743x + 2.675$
- 2.
3. $\hat{y} = -0.610x + 4.082$
- 4.
5. $\hat{y} = 0.625x + 1.25$, $SSE = 5$
- 6.
7. $\hat{y} = 0.6x + 1.8$
- 8.
9. $\hat{y} = -1.45x + 2.4$, $SSE = 50.25$ (cannot use the definition to compute)
- 10.
11.
 - a. $\hat{y} = 4.848x - 56$
 - b. 4.8
 - c. 21.6
- 12.
13.
 - a. $\hat{y} = 0.114x + 69.222$
 - b. 73.8
 - c. 69.2, invalid extrapolation
- 14.
15.
 - a. $\hat{y} = 42.024x + 119.502$
 - b. increases by \$42,024

c. \$224,562

16.

17. a. $\hat{y} = 1.045x - 8.527$

b. 2151.93367

c. 80.3

18.

19. a. $\hat{y} = 0.043x + 0.001$

b. For each additional ounce of medication consumed blood concentration of the active ingredient increases by 0.043%

c. 0.044%

20.

21. a. $\hat{y} = 2.550x + 1.993$

b. Predicted 28-day strength is 3,514 psi; sufficiently strong

22.

23.

24.

25. a. $\hat{y} = 0.0016x + 0.022$

b. On average, every 100 point increase in SAT score adds 0.16 point to the GPA.

c. $SSE = 432.10$

d. $\hat{y} = 2.182$

26.

27. a. $\hat{y} = 116.62x + 6955.1$

b. On average, every 1 additional bidder at an auction raises the price by 116.62 dollars.

c. $SSE = 1850314.08$

d. $\hat{y} = 7771.44$

10.5 Statistical Inferences About β_1

Basic

For the Basic and Application exercises in this section use the computations that were done for the exercises with the same number in Section 10.2 and Section 10.4.

1. Construct the 95% confidence interval for the slope β_1 of the population regression line based on the sample data set of Exercise 1 of Section 10.2.
2. Construct the 90% confidence interval for the slope β_1 of the population regression line based on the sample data set of Exercise 2 of Section 10.2.
3. Construct the 90% confidence interval for the slope β_1 of the population regression line based on the sample data set of Exercise 3 of Section 10.2.
4. Construct the 99% confidence interval for the slope β_1 of the population regression line based on the sample data set of Exercise 4 of Section 10.2.
5. For the data in Exercise 5 of Section 10.2 test, at the 10% level of significance, whether x is useful for predicting y (that is, whether $\beta_1 \neq 0$).
6. For the data in Exercise 6 of Section 10.2 test, at the 5% level of significance, whether x is useful for predicting y (that is, whether $\beta_1 \neq 0$).
7. Construct the 90% confidence interval for the slope β_1 of the population regression line based on the sample data set of Exercise 7 of Section 10.2.
8. Construct the 95% confidence interval for the slope β_1 of the population regression line based on the sample data set of Exercise 8 of Section 10.2.
9. For the data in Exercise 9 of Section 10.2 test, at the 1% level of significance, whether x is useful for predicting y (that is, whether $\beta_1 \neq 0$).
10. For the data in Exercise 10 of Section 10.2 test, at the 1% level of significance, whether x is useful for predicting y (that is, whether $\beta_1 \neq 0$).

Applications

11. For the data in Exercise 11 of Section 10.2 construct a 90% confidence interval for the mean number of new words acquired per month by children between 13 and 18 months of age.
12. For the data in Exercise 12 of Section 10.2 construct a 90% confidence interval for the mean increased braking distance for each additional 100 pounds of vehicle weight.
13. For the data in Exercise 13 of Section 10.2 test, at the 10% level of significance, whether age is useful for predicting resting heart rate.
14. For the data in Exercise 14 of Section 10.2 test, at the 10% level of significance, whether wind speed is useful for predicting wave height.
15. For the situation described in Exercise 15 of Section 10.2
 - a. Construct the 95% confidence interval for the mean increase in revenue per additional thousand dollars spent on advertising.
 - b. An advertising agency tells the business owner that for every additional thousand dollars spent on advertising, revenue will increase by over \$25,000. Test this claim (which is the alternative hypothesis) at the 5% level of significance.
 - c. Perform the test of part (b) at the 10% level of significance.
 - d. Based on the results in (b) and (c), how believable is the ad agency's claim? (This is a subjective judgement.)
16. For the situation described in Exercise 16 of Section 10.2
 - a. Construct the 90% confidence interval for the mean increase in height per additional inch of length at age two.
 - b. It is claimed that for girls each additional inch of length at age two means more than an additional inch of height at maturity. Test this claim (which is the alternative hypothesis) at the 10% level of significance.
17. For the data in Exercise 17 of Section 10.2 test, at the 10% level of significance, whether course average before the final exam is useful for predicting the final exam grade.
18. For the situation described in Exercise 18 of Section 10.2, an agronomist claims that each additional million acres planted results in more than 750,000 additional acres harvested. Test this claim at the 1% level of significance.
19. For the data in Exercise 19 of Section 10.2 test, at the 1/10th of 1% level of significance, whether, ignoring all other facts such as age and body mass, the amount of the medication consumed is a useful predictor of blood concentration of the active ingredient.
20. For the data in Exercise 20 of Section 10.2 test, at the 1% level of significance, whether for each additional inch of girth the age of the tree increases by at least two and one-half years.
21. For the data in Exercise 21 of Section 10.2
 - a. Construct the 95% confidence interval for the mean increase in strength at 28 days for each additional hundred psi increase in strength at 3 days.
 - b. Test, at the 1/10th of 1% level of significance, whether the 3-day strength is useful for predicting 28-day strength.
22. For the situation described in Exercise 22 of Section 10.2
 - a. Construct the 99% confidence interval for the mean decrease in energy demand for each one-degree drop in temperature.
 - b. An engineer with the power company believes that for each one-degree increase in temperature, daily energy demand will decrease by more than 3.6 million watt-hours. Test this claim at the 1% level of significance.

Large Data Set Exercises

Large Data Sets not available

23. Large Data Set 1 lists the SAT scores and GPAs of 1,000 students.
 - a. Compute the 90% confidence interval for the slope β_1 of the population regression line with SAT score as the independent variable (x) and GPA as the dependent variable (y).
 - b. Test, at the 10% level of significance, the hypothesis that the slope of the population regression line is greater than 0.001, against the null hypothesis that it is exactly 0.001.
24. Large Data Set 12 lists the golf scores on one round of golf for 75 golfers first using their own original clubs, then using clubs of a new, experimental design (after two months of familiarization with the new clubs).
 - a. Compute the 95% confidence interval for the slope β_1 of the population regression line with scores using the original clubs as the independent variable (x) and scores using the new clubs as the dependent variable (y).
 - b. Test, at the 10% level of significance, the hypothesis that the slope of the population regression line is different from 1, against the null hypothesis that it is exactly 1.

25. Large Data Set 13 records the number of bidders and sales price of a particular type of antique grandfather clock at 60 auctions.
- Compute the 95% confidence interval for the slope β_1 of the population regression line with the number of bidders present at the auction as the independent variable (x) and sales price as the dependent variable (y).
 - Test, at the 10% level of significance, the hypothesis that the average sales price increases by more than \$90 for each additional bidder at an auction, against the default that it increases by exactly \$90.

Answers

- 0.743 ± 0.578
-
- -0.610 ± 0.633
-
- $T = 1.732$, $\pm t_{0.05} = \pm 2.353$, do not reject H_0
-
- 0.6 ± 0.451
-
- $T = -4.481$, $\pm t_{0.005} = \pm 3.355$, reject H_0
-
- 4.8 ± 1.7 words
-
- $T = 2.843$, $\pm t_{0.05} = \pm 1.860$, reject H_0
-
- 42.024 ± 28.011 thousand dollars
 - $T = 1.487$, $\pm t_{0.05} = \pm 1.943$, do not reject H_0
 - $t_{0.10} = 1.440$, reject H_0
-
- $T = 4.096$, $\pm t_{0.05} = \pm 1.771$, reject H_0
-
- $T = 25.524$, $\pm t_{0.0005} = \pm 3.505$, reject H_0
-
- 2.550 ± 0.127 hundred psi
 - $T = 41.072$, $\pm t_{0.005} = \pm 3.674$, reject H_0
-
- $(0.0014, 0.0018)$
 - $H_0 : \beta_1 = 0.001$ vs $H_a : \beta_1 > 0.001$. Test Statistic: $Z = 6.1625$. Rejection Region: $[1.28, +\infty)$. Decision: Reject H_0
-
- $(101.789, 131.4435)$
 - $H_0 : \beta_1 = 90$ vs $H_a : \beta_1 > 90$. Test Statistic: $T = 3.5938$, $d.f. = 58$. Rejection Region: $[1.296, +\infty)$. Decision: Reject H_0

10.6 The Coefficient of Determination

Basic

For the Basic and Application exercises in this section use the computations that were done for the exercises with the same number in Section 10.2, Section 10.4, and Section 10.5.

- For the sample data set of Exercise 1 of Section 10.2 find the coefficient of determination using the formula $r^2 = \widehat{\beta}_1 SS_{xy} / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
- For the sample data set of Exercise 2 of Section 10.2 find the coefficient of determination using the formula $r^2 = \widehat{\beta}_1 SS_{xy} / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
- For the sample data set of Exercise 3 of Section 10.2 find the coefficient of determination using the formula $r^2 = \widehat{\beta}_1 SS_{xy} / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.

4. For the sample data set of Exercise 4 of Section 10.2 find the coefficient of determination using the formula $r^2 = \widehat{\beta}_1 SS_{xy} / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
5. For the sample data set of Exercise 5 of Section 10.2 find the coefficient of determination using the formula $r^2 = \widehat{\beta}_1 SS_{xy} / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
6. For the sample data set of Exercise 6 of Section 10.2 find the coefficient of determination using the formula $r^2 = \widehat{\beta}_1 SS_{xy} / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
7. For the sample data set of Exercise 7 of Section 10.2 find the coefficient of determination using the formula $r^2 = (SS_{yy} - SSE) / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
8. For the sample data set of Exercise 7 of Section 10.2 find the coefficient of determination using the formula $r^2 = (SS_{yy} - SSE) / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
9. For the sample data set of Exercise 7 of Section 10.2 find the coefficient of determination using the formula $r^2 = (SS_{yy} - SSE) / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.
10. For the sample data set of Exercise 7 of Section 10.2 find the coefficient of determination using the formula $r^2 = (SS_{yy} - SSE) / SS_{yy}$. Confirm your answer by squaring r as computed in that exercise.

Applications

11. For the data in Exercise 11 of Section 10.2 compute the coefficient of determination and interpret its value in the context of age and vocabulary.
12. For the data in Exercise 12 of Section 10.2" compute the coefficient of determination and interpret its value in the context of vehicle weight and braking distance.
13. For the data in Exercise 13 of Section 10.2 compute the coefficient of determination and interpret its value in the context of age and resting heart rate. In the age range of the data, does age seem to be a very important factor with regard to heart rate?
14. For the data in Exercise 14 of Section 10.2 compute the coefficient of determination and interpret its value in the context of wind speed and wave height. Does wind speed seem to be a very important factor with regard to wave height?
15. For the data in Exercise 15 of Section 10.2 find the proportion of the variability in revenue that is explained by level of advertising.
16. For the data in Exercise 16 of Section 10.2 find the proportion of the variability in adult height that is explained by the variation in length at age two.
17. For the data in Exercise 17 of Section 10.2 compute the coefficient of determination and interpret its value in the context of course average before the final exam and score on the final exam.
18. For the data in Exercise 18 of Section 10.2 compute the coefficient of determination and interpret its value in the context of acres planted and acres harvested.
19. For the data in Exercise 19 of Section 10.2 compute the coefficient of determination and interpret its value in the context of the amount of the medication consumed and blood concentration of the active ingredient.
20. For the data in Exercise 20 of Section 10.2 compute the coefficient of determination and interpret its value in the context of tree size and age.
21. For the data in Exercise 21 of Section 10.2 find the proportion of the variability in 28-day strength of concrete that is accounted for by variation in 3-day strength.
22. For the data in Exercise 22 of Section 10.2 find the proportion of the variability in energy demand that is accounted for by variation in average temperature.

Large Data Set Exercises

Large Data Sets not available

23. Large Data Set 1 lists the SAT scores and GPAs of 1, 000 students. Compute the coefficient of determination and interpret its value in the context of SAT scores and GPAs.
24. Large Data Set 12 lists the golf scores on one round of golf for 75 golfers first using their own original clubs, then using clubs of a new, experimental design (after two months of familiarization with the new clubs). Compute the coefficient of determination and interpret its value in the context of golf scores with the two kinds of golf clubs.
25. Large Data Set 13 records the number of bidders and sales price of a particular type of antique grandfather clock at 60 auctions. Compute the coefficient of determination and interpret its value in the context of the number of bidders at an auction and the price of this type of antique grandfather clock.

Answers

1. 0.848
- 2.
3. 0.631
- 4.
5. 0.5
- 6.
7. 0.766
- 8.
9. 0.715
- 10.
11. 0.898; about 90% of the variability in vocabulary is explained by age
- 12.
13. 0.503; about 50% of the variability in heart rate is explained by age. Age is a significant but not dominant factor in explaining heart rate.
- 14.
15. The proportion is $r^2 = 0.692$
- 16.
17. 0.563; about 56% of the variability in final exam scores is explained by course average before the final exam
- 18.
19. 0.931; about 93% of the variability in the blood concentration of the active ingredient is explained by the amount of the medication consumed
- 20.
21. The proportion is $r^2 = 0.984$
- 22.
23. $r^2 = 21.17\%$
- 24.
25. $r^2 = 81.04\%$

10.7 Estimation and Prediction

Basic

For the Basic and Application exercises in this section use the computations that were done for the exercises with the same number in previous sections.

1. For the sample data set of Exercise 1 of Section 10.2
 - a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 4$.
 - b. Construct the 90% confidence interval for that mean value.
2. For the sample data set of Exercise 2 of Section 10.2
 - a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 4$.
 - b. Construct the 90% confidence interval for that mean value.
3. For the sample data set of Exercise 3 of Section 10.2
 - a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 7$.
 - b. Construct the 95% confidence interval for that mean value.
4. For the sample data set of Exercise 4 of Section 10.2
 - a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 2$.
 - b. Construct the 80% confidence interval for that mean value.
5. For the sample data set of Exercise 5 of Section 10.2
 - a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 1$.
 - b. Construct the 80% confidence interval for that mean value.
6. For the sample data set of Exercise 6 of Section 10.2

- a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 5$.
 - b. Construct the 95% confidence interval for that mean value.
7. For the sample data set of Exercise 7 of Section 10.2
- a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 6$.
 - b. Construct the 99% confidence interval for that mean value.
 - c. Is it valid to make the same estimates for $x = 12$? Explain.
8. For the sample data set of Exercise 8 of Section 10.2
- a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 12$.
 - b. Construct the 80% confidence interval for that mean value.
 - c. Is it valid to make the same estimates for $x = 0$? Explain.
9. For the sample data set of Exercise 9 of Section 10.2
- a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 0$.
 - b. Construct the 90% confidence interval for that mean value.
 - c. Is it valid to make the same estimates for $x = -1$? Explain.
10. For the sample data set of Exercise 9 of Section 10.2
- a. Give a point estimate for the mean value of y in the sub-population determined by the condition $x = 8$.
 - b. Construct the 95% confidence interval for that mean value.
 - c. Is it valid to make the same estimates for $x = 0$? Explain.

Applications

11. For the data in Exercise 11 of Section 10.2
- a. Give a point estimate for the average number of words in the vocabulary of 18-month-old children.
 - b. Construct the 95% confidence interval for that mean value.
 - c. Construct the 95% confidence interval for that mean value.
12. For the data in Exercise 12 of Section 10.2
- a. Give a point estimate for the average braking distance of automobiles that weigh 3,250 pounds.
 - b. Construct the 80% confidence interval for that mean value.
 - c. Is it valid to make the same estimates for 5,000-pound automobiles? Explain.
13. For the data in Exercise 13 of Section 10.2
- a. Give a point estimate for the resting heart rate of a man who is 35 years old.
 - b. One of the men in the sample is 35 years old, but his resting heart rate is not what you computed in part (a). Explain why this is not a contradiction.
 - c. Construct the 90% confidence interval for the mean resting heart rate of all 35-year-old men.
14. For the data in Exercise 14 of Section 10.2
- a. Give a point estimate for the wave height when the wind speed is 13 miles per hour.
 - b. One of the wind speeds in the sample is 13 miles per hour, but the height of waves that day is not what you computed in part (a). Explain why this is not a contradiction.
 - c. Construct the 95% confidence interval for the mean wave height on days when the wind speed is 13 miles per hour.
15. For the data in Exercise 15 of Section 10.2
- a. The business owner intends to spend \$2,500 on advertising next year. Give an estimate of next year's revenue based on this fact.
 - b. Construct the 90% prediction interval for next year's revenue, based on the intent to spend \$2,500 on advertising.
16. For the data in Exercise 16 of Section 10.2
- a. A two-year-old girl is 32.3 inches long. Predict her adult height.
 - b. Construct the 95% prediction interval for the girl's adult height.
17. For the data in Exercise 17 of Section 10.2
- a. Lodovico has a 78.6 average in his physics class just before the final. Give a point estimate of what his final exam grade will be.
 - b. Explain whether an interval estimate for this problem is a confidence interval or a prediction interval.

- c. Based on your answer to (b), construct an interval estimate for Lodovico's final exam grade at the 90% level of confidence.
18. For the data in Exercise 18 of Section 10.2
- This year 86.2 million acres of corn were planted. Give a point estimate of the number of acres that will be harvested this year.
 - Explain whether an interval estimate for this problem is a confidence interval or a prediction interval.
 - Based on your answer to (b), construct an interval estimate for the number of acres that will be harvested this year, at the 99% level of confidence.
19. For the data in Exercise 19 of Section 10.2
- Give a point estimate for the blood concentration of the active ingredient of this medication in a man who has consumed 1.5 ounces of the medication just recently.
 - Gratiano just consumed 1.5 ounces of this medication 30 minutes ago. Construct a 95% prediction interval for the concentration of the active ingredient in his blood right now.
20. For the data in Exercise 20 of Section 10.2
- You measure the girth of a free-standing oak tree five feet off the ground and obtain the value 127 inches. How old do you estimate the tree to be?
 - Construct a 90% prediction interval for the age of this tree.
21. For the data in Exercise 21 of Section 10.2
- A test cylinder of concrete three days old fails at 1,750 psi. Predict what the 28-day strength of the concrete will be.
 - Construct a 99% prediction interval for the 28-day strength of this concrete.
 - Based on your answer to (b), what would be the minimum 28-day strength you could expect this concrete to exhibit?
22. For the data in Exercise 22 of Section 10.2
- Tomorrow's average temperature is forecast to be 53 degrees. Estimate the energy demand tomorrow.
 - Construct a 99% prediction interval for the energy demand tomorrow.
 - Based on your answer to (b), what would be the minimum demand you could expect?

Large Data Set Exercises

Large Data Sets not available

23. Large Data Set 1 lists the SAT scores and GPAs of 1,000 students.
- Give a point estimate of the mean GPA of all students who score 1350 on the SAT.
 - Construct a 90% confidence interval for the mean GPA of all students who score 1350 on the SAT.
24. Large Data Set 12 lists the golf scores on one round of golf for 75 golfers first using their own original clubs, then using clubs of a new, experimental design (after two months of familiarization with the new clubs).
- Thurio averages 72 strokes per round with his own clubs. Give a point estimate for his score on one round if he switches to the new clubs.
 - Explain whether an interval estimate for this problem is a confidence interval or a prediction interval.
 - Based on your answer to (b), construct an interval estimate for Thurio's score on one round if he switches to the new clubs, at 90% confidence.
25. Large Data Set 13 records the number of bidders and sales price of a particular type of antique grandfather clock at 60 auctions.
- There are seven likely bidders at the Verona auction today. Give a point estimate for the price of such a clock at today's auction.
 - Explain whether an interval estimate for this problem is a confidence interval or a prediction interval.
 - Based on your answer to (b), construct an interval estimate for the likely sale price of such a clock at today's sale, at 95% confidence.

Answers

- 5.647
 - 5.647 ± 1.253
-

3. a. -0.188
b. -0.188 ± 3.041
- 4.
5. a. 1.875
b. 1.875 ± 1.423
- 6.
7. a. 5.4
b. 5.4 ± 3.355
c. invalid (extrapolation)
- 8.
9. a. 2.4
b. 2.4 ± 1.474
c. valid (-1 is in the range of the x -values in the data set)
- 10.
11. a. 31.3 words
b. 31.3 ± 7.1 words
c. not valid, since two years is 24 months, hence this is extrapolation
- 12.
13. a. 73.2 beats/min
b. The man's heart rate is not the predicted average for all men his age.
c. 73.2 ± 1.2 beats/min
- 14.
15. a. $\$224,562$
b. $\$224,562 \pm \$28,699$
- 16.
17. a. 74
b. Prediction (one person, not an average for all who have average 78.6 before the final exam)
c. 74 ± 24
- 18.
19. a. 0.066%
b. $0.066 \pm 0.034\%$
- 20.
21. a. $4,656$ psi
b. $4,656 \pm 321$ psi
c. $4,656 - 321 = 4,335$ psi
- 22.
23. a. 2.19
b. $(2.1421, 2.2316)$
- 24.
25. a. 7771.39
b. A prediction interval.
c. $(7410.41, 8132.38)$

10.8 A Complete Example

Basic

The exercises in this section are unrelated to those in previous sections.

1. The data give the amount x of silicofluoride in the water (mg/L) and the amount y of lead in the bloodstream ($\mu\text{g/dL}$) of ten children in various communities with and without municipal water. Perform a complete analysis of the data, in analogy with the discussion in this section (that is, make a scatter plot, do preliminary computations, find the least squares regression line, find SSE , s_e and r , and so on). In the hypothesis test use as the alternative hypothesis $\beta_1 > 0$, and test at the 5% level of significance. Use confidence level 95% for the confidence interval for β_1 . Construct 95% confidence and predictions intervals at $x_p = 2$ at the end.

x	0.0	0.0	1.1	1.4	1.6
y	0.3	0.1	4.7	3.2	5.1
x	1.7	2.0	2.0	2.2	2.2
y	7.0	5.0	6.1	8.6	9.5

(10.E.32)

2. The table gives the weight x (thousands of pounds) and available heat energy y (million BTU) of a standard cord of various species of wood typically used for heating. Perform a complete analysis of the data, in analogy with the discussion in this section (that is, make a scatter plot, do preliminary computations, find the least squares regression line, find SSE , s_e and r , and so on). In the hypothesis test use as the alternative hypothesis β_1 , and test at the 5% level of significance. Use confidence level 95% for the confidence interval for β_1 . Construct 95% confidence and predictions intervals at $x_p = 5$ at the end.

x	3.37	3.50	4.29	4.00	4.64
y	23.6	17.5	20.1	21.6	28.1
x	4.99	4.94	5.48	3.26	4.16
y	25.3	27.0	30.7	18.9	20.7

(10.E.33)

Large Data Set Exercises

Large Data Sets not available

3. Large Data Sets 3 and 3A list the shoe sizes and heights of 174 customers entering a shoe store. The gender of the customer is not indicated in Large Data Set 3. However, men's and women's shoes are not measured on the same scale; for example, a size 8 shoe for men is not the same size as a size 8 shoe for women. Thus it would not be meaningful to apply regression analysis to Large Data Set 3. Nevertheless, compute the scatter diagrams, with shoe size as the independent variable (x) and height as the dependent variable (y), for (i) just the data on men, (ii) just the data on women, and (iii) the full mixed data set with both men and women. Does the third, invalid scatter diagram look markedly different from the other two?
4. Separate out from Large Data Set 3A just the data on men and do a complete analysis, with shoe size as the independent variable (x) and height as the dependent variable (y). Use $\alpha = 0.05$ and $x_p = 10$ whenever appropriate.
5. Separate out from Large Data Set 3A just the data on women and do a complete analysis, with shoe size as the independent variable (x) and height as the dependent variable (y). Use $\alpha = 0.05$ and $x_p = 10$ whenever appropriate.

Answers

1.

$$\begin{aligned} \sum x &= 14.2, \sum y = 49.6, \sum xy = 91.73, \sum x^2 = 26.3, \sum y^2 = 333.86 \\ SS_{xx} &= 6.136, SS_{xy} = 21.298, SS_{yy} = 87.844 \\ \bar{x} &= 1.42, \bar{y} = 4.96 \\ \widehat{\beta}_1 &= 3.47, \widehat{\beta}_0 = 0.03 \\ SSE &= 13.92 \\ s_e &= 1.32 \\ r &= 0.9174, r^2 = 0.8416 \\ df &= 8, T = 6.518 \end{aligned}$$
(10.E.34)

The 95% confidence interval for β_1 is: (2.24, 4.70)

At $x_p = 2$ the 95% confidence interval for $E(y)$ is (5.77, 8.17)

At $x_p = 2$ the 95% confidence interval for y is (3.73, 10.21)

- 2.

3. The positively correlated trend seems less profound than that in each of the previous plots.
- 4.
5. The regression line: $\hat{y} = 3.3426x + 138.7692$ Coefficient of Correlation: $r = 0.9431$. Coefficient of Determination: $r^2 = 0.8894$. $SSE = 283.2473$. $s_e = 1.9305$. A 95% confidence interval for β_1 : $(3.0733, 3.6120)$ Test Statistic for $H_0 : \beta_1 = 0 : T = 24.7209$. At $x_p = 10$, $\hat{y} = 172.1956$; a 95% confidence interval for the mean value of y is: $(171.5577, 172.8335)$ and a 95% prediction interval for an individual value of y is: $(168.2974, 176.0938)$

This page titled [10.E: Correlation and Regression \(Exercises\)](#) is shared under a [CC BY-NC-SA 3.0](#) license and was authored, remixed, and/or curated by [Anonymous](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [10.E: Correlation and Regression \(Exercises\)](#) has no license indicated.