

This is “The Least Squares Regression Line”, section 10.4 from the book [Beginning Statistics](#) (v. 1.0). For details on it (including licensing), [click here](#).

For more information on the source of this book, or why it is available for free, please see [the project's home page](#). You can browse or download additional books there.

Has this book helped you? Consider passing it on:



Help Creative Commons

Creative Commons supports free culture from music to education. Their licenses helped make this book available to you.



Help a Public School

DonorsChoose.org helps people like you help teachers fund their classroom projects, from art supplies to books to calculators.



[Previous Section](#)



[Table of Contents](#)

[Next Section](#)



10.4 The Least Squares Regression Line

LEARNING OBJECTIVES

1. To learn how to measure how well a straight line fits a collection of data.
2. To learn how to construct the least squares regression line, the straight line that best fits a collection of data.
3. To learn the meaning of the slope of the least squares regression line.
4. To learn how to use the least squares regression line to estimate the response variable y in terms of the predictor variable x .

Goodness of Fit of a Straight Line to Data

Once the scatter diagram of the data has been drawn and the model assumptions described in the previous sections at least visually verified (and

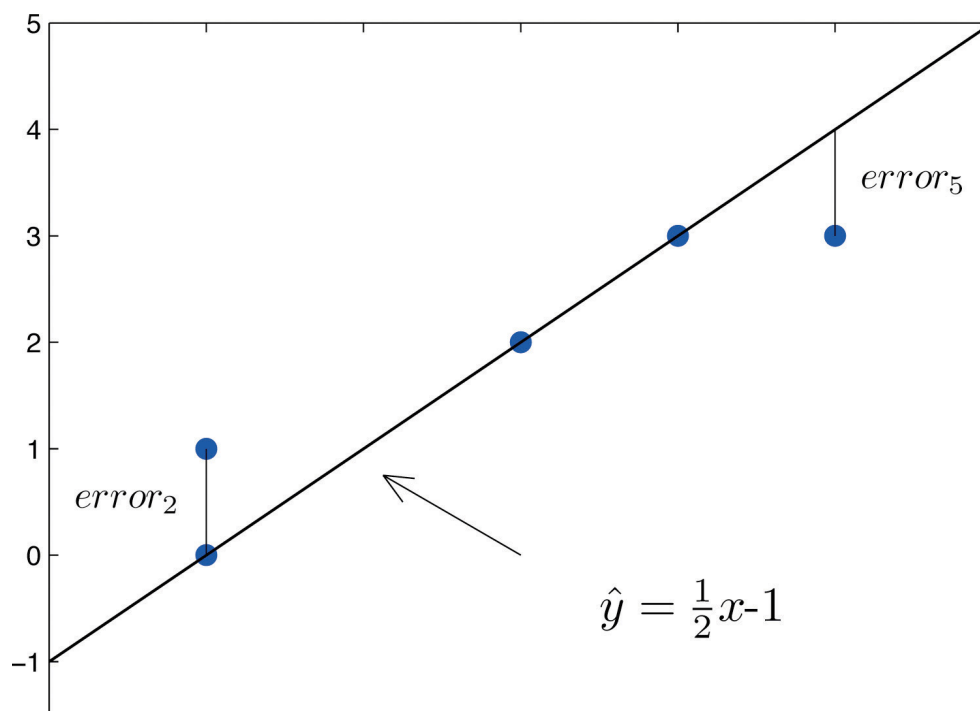
perhaps the correlation coefficient r computed to quantitatively verify the linear trend), the next step in the analysis is to find the straight line that best fits the data. We will explain how to measure how well a straight line fits a collection of points by examining how well the line $y = \frac{1}{2}x - 1$ fits the data set

x	2	2	6	8	10
y	0	1	2	3	3

(which will be used as a running example for the next three sections). We will write the equation of this line as $\hat{y} = \frac{1}{2}x - 1$ with an accent on the y to indicate that the y -values computed using this equation are not from the data. We will do this with all lines approximating data sets. The line $\hat{y} = \frac{1}{2}x - 1$ was selected as one that seems to fit the data reasonably well.

The idea for measuring the goodness of fit of a straight line to data is illustrated in [Figure 10.6 "Plot of the Five-Point Data and the Line"](#), in which the graph of the line $\hat{y} = \frac{1}{2}x - 1$ has been superimposed on the scatter plot for the sample data set.

Figure 10.6 Plot of the Five-Point Data and the Line $\hat{y} = \frac{1}{2}x - 1$





To each point in the data set there is associated an “error,” the positive or negative vertical distance from the point to the line: positive if the point is above the line and negative if it is below the line. The error can be computed as the actual y -value of the point minus the y -value \hat{y} that is “predicted” by inserting the x -value of the data point into the formula for the line:

$$\text{error at data point } (x, y) = (\text{true } y) - (\text{predicted } y) = y - \hat{y}$$

The computation of the error for each of the five points in the data set is shown in Table 10.1 "The Errors in Fitting Data with a Straight Line".

Table 10.1 The Errors in Fitting Data with a Straight Line

	x	y	$\hat{y} = \frac{1}{2}x - 1$	$y - \hat{y}$	$(y - \hat{y})^2$
	2	0	0	0	0
	2	1	0	1	1
	6	2	2	0	0
	8	3	3	0	0
	10	3	4	-1	1
Σ	-	-	-	0	2

A first thought for a measure of the goodness of fit of the line to the data would be simply to add the errors at every point, but the example shows that this cannot work well in general. The line does not fit the data perfectly (no line can), yet because of cancellation of positive and negative errors the sum of the errors (the fourth column of numbers) is zero. Instead goodness of fit is measured by the sum of the squares of the errors. Squaring eliminates the minus signs, so no cancellation can occur. For the data and line in Figure 10.6 "Plot of the Five-Point Data and the Line " the sum of the squared errors (the

last column of numbers) is 2. This number measures the goodness of fit of the line to the data.

Definition

The **goodness of fit** of a line $\hat{y} = mx + b$ to a set of n pairs (x, y) of numbers in a sample is the sum of the squared errors

$$\Sigma(y - \hat{y})^2$$

(n terms in the sum, one for each data pair).

The Least Squares Regression Line

Given any collection of pairs of numbers (except when all the x -values are the same) and the corresponding scatter diagram, there always exists exactly one straight line that fits the data better than any other, in the sense of minimizing the sum of the squared errors. It is called the *least squares regression line*. Moreover there are formulas for its slope and y -intercept.

Definition

Given a collection of pairs (x, y) of numbers (in which not all the x -values are the same), there is a line $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$ that best fits the data in the sense of minimizing the sum of the squared errors. It is called the **least squares regression line**. Its slope $\hat{\beta}_1$ and y -intercept $\hat{\beta}_0$ are computed using the formulas

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where

$$SS_{xx} = \sum x^2 - \frac{1}{n}(\sum x)^2, \quad SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y)$$

\bar{x} is the mean of all the x -values, \bar{y} is the mean of all the y -values, and n is the number of pairs in the data set.

The equation $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$ specifying the least squares regression line is called the **least squares regression equation**.

Remember from [Section 10.3 "Modelling Linear Relationships with Randomness Present"](#) that the line with the equation $y = \beta_1 x + \beta_0$ is called the population regression line. The numbers $\hat{\beta}_1$ and $\hat{\beta}_0$ are statistics that estimate the population parameters β_1 and β_0 .

We will compute the least squares regression line for the five-point data set, then for a more practical example that will be another running example for the introduction of new concepts in this and the next three sections.

EXAMPLE 2

Find the least squares regression line for the five-point data set

x	2	2	6	8	10
y	0	1	2	3	3

and verify that it fits the data better than the line $\hat{y} = \frac{1}{2}x - 1$ considered in [Section 10.4.1 "Goodness of Fit of a Straight Line to Data"](#).

Solution:

In actual practice computation of the regression line is done using a statistical computation package. In order to clarify the meaning of the formulas we display the computations in

tabular form.

	x	y	x ²	xy
	2	0	4	0
	2	1	4	2
	6	2	36	12
	8	3	64	24
	10	3	100	30
Σ	28	9	208	68

In the last line of the table we have the sum of the numbers in each column. Using them we compute:

$$SS_{xx} = \sum x^2 - \frac{1}{n}(\sum x)^2 = 208 - \frac{1}{5}(28)^2 = 51.2$$

$$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y) = 68 - \frac{1}{5}(28)(9) = 17.6$$

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{5} = 5.6$$

$$\bar{y} = \frac{\sum y}{n} = \frac{9}{5} = 1.8$$

so that

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{17.6}{51.2} = 0.34375 \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1.8 - (0.34375)(5.6) = -0.125$$

The least squares regression line for these data is

$$\hat{y} = 0.34375x - 0.125$$

The computations for measuring how well it fits the sample data are given in [Table 10.2 "The Errors in Fitting Data with the Least Squares Regression Line"](#). The sum of the squared errors is the sum of the numbers in the last column, which is 0.75. It is less than 2, the sum of the squared errors for the fit of the line $\hat{y} = \frac{1}{2}x - 1$ to this data set.

TABLE 10.2 THE ERRORS IN FITTING DATA WITH THE LEAST SQUARES REGRESSION LINE

x	y	$\hat{y} = 0.34375x - 0.125$	$y - \hat{y}$	$(y - \hat{y})^2$
2	0	0.5625	-0.5625	0.31640625
2	1	0.5625	0.4375	0.19140625
6	2	1.9375	0.0625	0.00390625
8	3	2.6250	0.3750	0.14062500
10	3	3.3125	-0.3125	0.09765625

EXAMPLE 3

Table 10.3 "Data on Age and Value of Used Automobiles of a Specific Make and Model" shows the age in years and the retail value in thousands of dollars of a random sample of ten automobiles of the same make and model.

- Construct the scatter diagram.
- Compute the linear correlation coefficient r . Interpret its value in the context of the problem.
- Compute the least squares regression line. Plot it on the scatter diagram.
- Interpret the meaning of the slope of the least squares regression line in the context of the problem.
- Suppose a four-year-old automobile of this make and model is selected at random. Use the regression equation to predict its retail value.
- Suppose a 20-year-old automobile of this make and model is selected at random. Use the regression equation to predict its retail value. Interpret the result.
- Comment on the validity of using the regression equation to predict the price of a brand new automobile of this make and model.

TABLE 10.3 DATA ON AGE AND VALUE OF USED

AUTOMOBILES OF A SPECIFIC MAKE AND MODEL

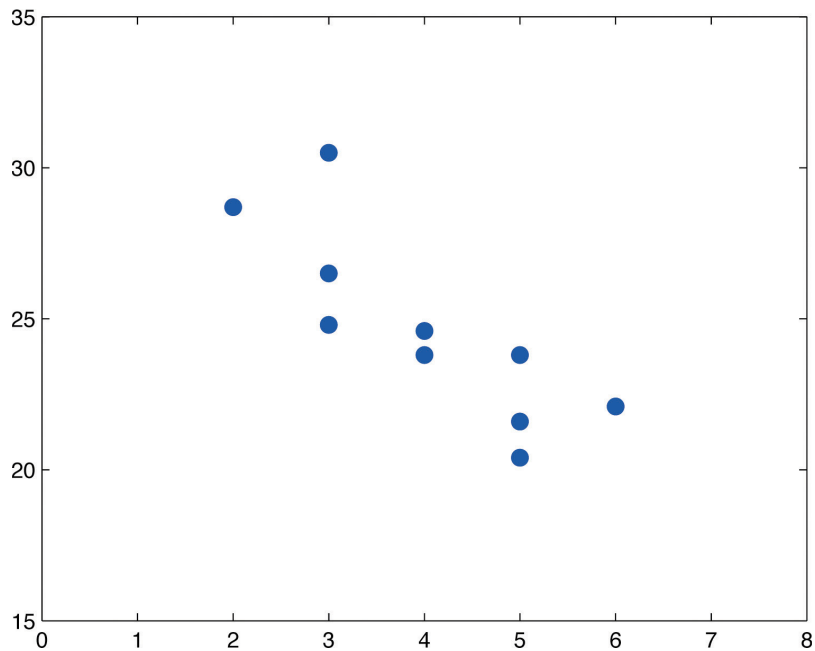
x	2	3	3	3	4	4	5	5	5	6
y	28.7	24.8	26.0	30.5	23.8	24.6	23.8	20.4	21.6	22.1

Solution:

- a. The scatter diagram is shown in Figure 10.7 "Scatter Diagram for Age and Value of Used Automobiles".

Figure 10.7

Scatter Diagram for Age and Value of Used Automobiles



- a. We must first compute SS_{xx} , SS_{xy} , SS_{yy} , which means computing Σx , Σy , Σx^2 , Σy^2 , and Σxy . Using a computing device we obtain

$$\Sigma x = 40 \quad \Sigma y = 246.3 \quad \Sigma x^2 = 174 \quad \Sigma y^2 = 6154.15 \quad \Sigma xy = 956.5$$

Thus

$$SS_{xx} = \sum x^2 - \frac{1}{n}(\sum x)^2 = 174 - \frac{1}{10}(40)^2 = 14$$

$$SS_{xy} = \sum xy - \frac{1}{n}(\sum x)(\sum y) = 956.5 - \frac{1}{10}(40)(246.3) = -28.7$$

$$SS_{yy} = \sum y^2 - \frac{1}{n}(\sum y)^2 = 6154.15 - \frac{1}{10}(246.3)^2 = 87.781$$

so that

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} = \frac{-28.7}{\sqrt{(14)(87.781)}} = -0.819$$

The age and value of this make and model automobile are moderately strongly negatively correlated. As the age increases, the value of the automobile tends to decrease.

- b. Using the values of $\sum x$ and $\sum y$ computed in part (b),

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{10} = 4 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{246.3}{10} = 24.63$$

Thus using the values of SS_{xx} and SS_{xy} from part (b),

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-28.7}{14} = -2.05 \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 24.63 - (-2.05)(4) = 32.83$$

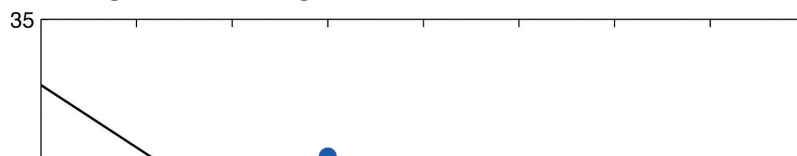
The equation $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$ of the least squares regression line for these sample data is

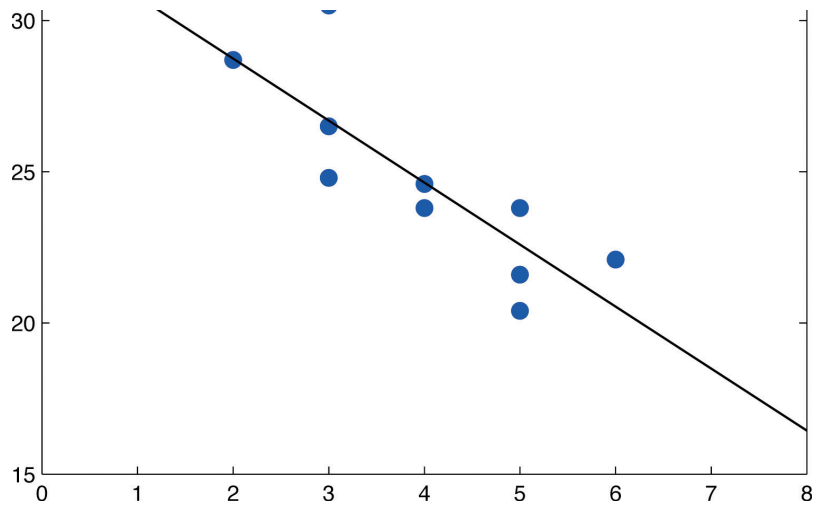
$$\hat{y} = -2.05x + 32.83$$

Figure 10.8 "Scatter Diagram and Regression Line for Age and Value of Used Automobiles" shows the scatter diagram with the graph of the least squares regression line superimposed.

Figure 10.8

Scatter Diagram and Regression Line for Age and Value of Used Automobiles





- a. The slope -2.05 means that for each unit increase in x (additional year of age) the average value of this make and model vehicle decreases by about 2.05 units (about \$2,050).
- b. Since we know nothing about the automobile other than its age, we assume that it is of about average value and use the average value of all four-year-old vehicles of this make and model as our estimate. The average value is simply the value of \hat{y} obtained when the number 4 is inserted for x in the least squares regression equation:

$$\hat{y} = -2.05(4) + 32.83 = 24.63$$

which corresponds to \$24,630.

- c. Now we insert $x = 20$ into the least squares regression equation, to obtain

$$\hat{y} = -2.05(20) + 32.83 = -8.17$$

which corresponds to $-\$8,170$. Something is wrong here, since a negative makes no sense. The error arose from applying the regression equation to a value of x not in the range of x -values in the original data, from two to six years.

Applying the regression equation $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$ to a value of x outside the range of x -values in the data set is called *extrapolation*. It is an invalid use of the

regression equation and should be avoided.

- d. The price of a brand new vehicle of this make and model is the value of the automobile at age 0. If the value $x = 0$ is inserted into the regression equation the result is always $\hat{\beta}_0$, the y-intercept, in this case 32.83, which corresponds to \$32,830. But this is a case of extrapolation, just as part (f) was, hence this result is invalid, although not obviously so. In the context of the problem, since automobiles tend to lose value much more quickly immediately after they are purchased than they do after they are several years old, the number \$32,830 is probably an underestimate of the price of a new automobile of this make and model.

For emphasis we highlight the points raised by parts (f) and (g) of the example.

Definition

*The process of using the least squares regression equation to estimate the value of y at a value of x that does not lie in the range of the x -values in the data set that was used to form the regression line is called **extrapolation**. It is an invalid use of the regression equation that can lead to errors, hence should be avoided.*

The Sum of the Squared Errors SSE

In general, in order to measure the goodness of fit of a line to a set of data, we must compute the predicted y -value \hat{y} at every point in the data set, compute each error, square it, and then add up all the squares. In the case of the least squares regression line, however, the line that best fits the data, the sum of the squared errors can be computed directly from the data using the following formula.

The sum of the squared errors for the least squares regression line is denoted by SSE . It can be computed using the formula

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$$

EXAMPLE 4

Find the sum of the squared errors SSE for the least squares regression line for the five-point data set

x	2	2	6	8	10
y	0	1	2	3	3

Do so in two ways:

- a. using the definition $\Sigma(y - \hat{y})^2$;
- b. using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.

Solution:

- a. The least squares regression line was computed in [Note 10.18 "Example 2"](#) and is $\hat{y} = 0.34375x - 0.125$. SSE was found at the end of that example using the definition $\Sigma(y - \hat{y})^2$. The computations were tabulated in [Table 10.2 "The Errors in Fitting Data with the Least Squares Regression Line"](#). SSE is the sum of the numbers in the last column, which is 0.75.
- b. The numbers SS_{xy} and $\hat{\beta}_1$ were already computed in [Note 10.18 "Example 2"](#) in the process of finding the least squares regression line. So was the number $\Sigma y = 9$. We must compute SS_{yy} . To do so it is necessary to first compute $\Sigma y^2 = 0 + 1^2 + 2^2 + 3^2 + 3^2 = 23$. Then

$$SS_{yy} = \Sigma y^2 - \frac{1}{n} (\Sigma y)^2 = 23 - \frac{1}{5} (9)^2 = 6.8$$

so that

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 6.8 - (0.34375)(17.6) = 0.75$$

EXAMPLE 5

Find the sum of the squared errors SSE for the least squares regression line for the data set, presented in Table 10.3 "Data on Age and Value of Used Automobiles of a Specific Make and Model", on age and values of used vehicles in Note 10.19 "Example 3".

Solution:

From Note 10.19 "Example 3" we already know that

$$SS_{xy} = -28.7, \quad \hat{\beta}_1 = -2.05, \quad \text{and} \quad \Sigma y = 246.3$$

To compute SS_{yy} we first compute

$$\Sigma y^2 = 28.7^2 + 24.8^2 + 26.0^2 + 30.5^2 + 23.8^2 + 24.6^2 + 23.8^2 + 20.4^2 + 21.6^2 + 22.1^2$$

Then

$$SS_{yy} = \Sigma y^2 - \frac{1}{n} (\Sigma y)^2 = 6154.15 - \frac{1}{10} (246.3)^2 = 87.781$$

Therefore

$$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 87.781 - (-2.05)(-28.7) = 28.946$$

KEY TAKEAWAYS

- How well a straight line fits a data set is measured by the sum of the squared errors.
- The least squares regression line is the line that best fits the data. Its slope and

y-intercept are computed from the data using formulas.

- The slope $\hat{\beta}_1$ of the least squares regression line estimates the size and direction of the mean change in the dependent variable y when the independent variable x is increased by one unit.
- The sum of the squared errors SSE of the least squares regression line can be computed using a formula, without having to compute all the individual errors.

EXERCISES

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 "The Linear Correlation Coefficient".

1. Compute the least squares regression line for the data in Exercise 1 of Section 10.2 "The Linear Correlation Coefficient".
2. Compute the least squares regression line for the data in Exercise 2 of Section 10.2 "The Linear Correlation Coefficient".
3. Compute the least squares regression line for the data in Exercise 3 of Section 10.2 "The Linear Correlation Coefficient".
4. Compute the least squares regression line for the data in Exercise 4 of Section 10.2 "The Linear Correlation Coefficient".
5. For the data in Exercise 5 of Section 10.2 "The Linear Correlation Coefficient"
 - a. Compute the least squares regression line.
 - b. Compute the sum of the squared errors SSE using the definition $\Sigma(y - \hat{y})^2$.
 - c. Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.
6. For the data in Exercise 6 of Section 10.2 "The Linear Correlation Coefficient"
 - a. Compute the least squares regression line.
 - b. Compute the sum of the squared errors SSE using the definition $\Sigma(y - \hat{y})^2$.

- c. Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.
7. Compute the least squares regression line for the data in Exercise 7 of Section 10.2 "The Linear Correlation Coefficient".
8. Compute the least squares regression line for the data in Exercise 8 of Section 10.2 "The Linear Correlation Coefficient".
9. For the data in Exercise 9 of Section 10.2 "The Linear Correlation Coefficient"
- Compute the least squares regression line.
 - Can you compute the sum of the squared errors SSE using the definition $\sum(y - \hat{y})^2$? Explain.
 - Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.
10. For the data in Exercise 10 of Section 10.2 "The Linear Correlation Coefficient"
- Compute the least squares regression line.
 - Can you compute the sum of the squared errors SSE using the definition $\sum(y - \hat{y})^2$? Explain.
 - Compute the sum of the squared errors SSE using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.

APPLICATIONS

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

13. For the data in Exercise 13 of Section 10.2 "The Linear Correlation Coefficient"

- 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 "The Linear Correlation Coefficient"

- 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 "The Linear Correlation Coefficient"

- 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 "The Linear Correlation Coefficient"

- 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 "The Linear Correlation Coefficient"

- a. Compute the least squares regression line.
- b. Compute SSE using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.
- 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 "The Linear Correlation Coefficient"

- a. Compute the least squares regression line.
- b. Compute SSE using the formula $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$.

- 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 "The Linear Correlation Coefficient"
- 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 "The Linear Correlation Coefficient"
- 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 "The Linear Correlation Coefficient"
- 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 "The Linear Correlation Coefficient"
- 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 "The Linear Correlation Coefficient".

- a. Reverse the roles of x and y and compute the least squares regression line for the new data set

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

- 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

25. Large Data Set 1 lists the SAT scores and GPAs of 1,000 students.

<http://www.flatworldknowledge.com/sites/all/files/data1.xls>

- 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).

<http://www.flatworldknowledge.com/sites/all/files/data12.xls>

- 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.

<http://www.flatworldknowledge.com/sites/all/files/data13.xls>

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

- b. increases by \$42,024,
c. \$224,562
17. a. $\hat{y} = 1.045x - 8.527$,
b. 2151.93367,
c. 80.3
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%
21. a. $\hat{y} = 2.550x + 1.993$,
b. Predicted 28-day strength is 3,514 psi; sufficiently strong
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. $\hat{y} = 2.182$
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. $\hat{y} = 7771.44$



[Previous Section](#)



[Table of Contents](#)

[Next Section](#)

