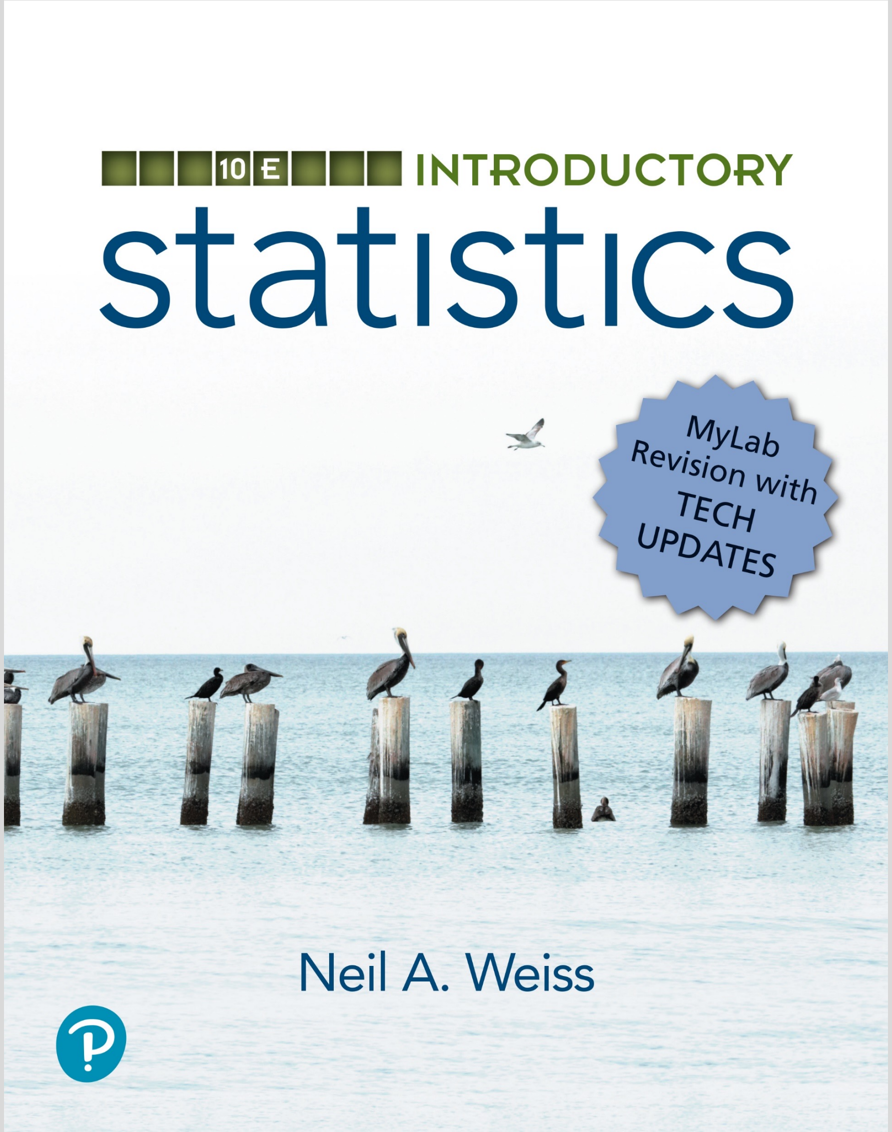
# Chapter 14

## Descriptive Methods in Regression and Correlation





Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Section 14.1

## Linear Equations with One Independent Variable


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Definition 14.1

**y-Intercept and Slope**

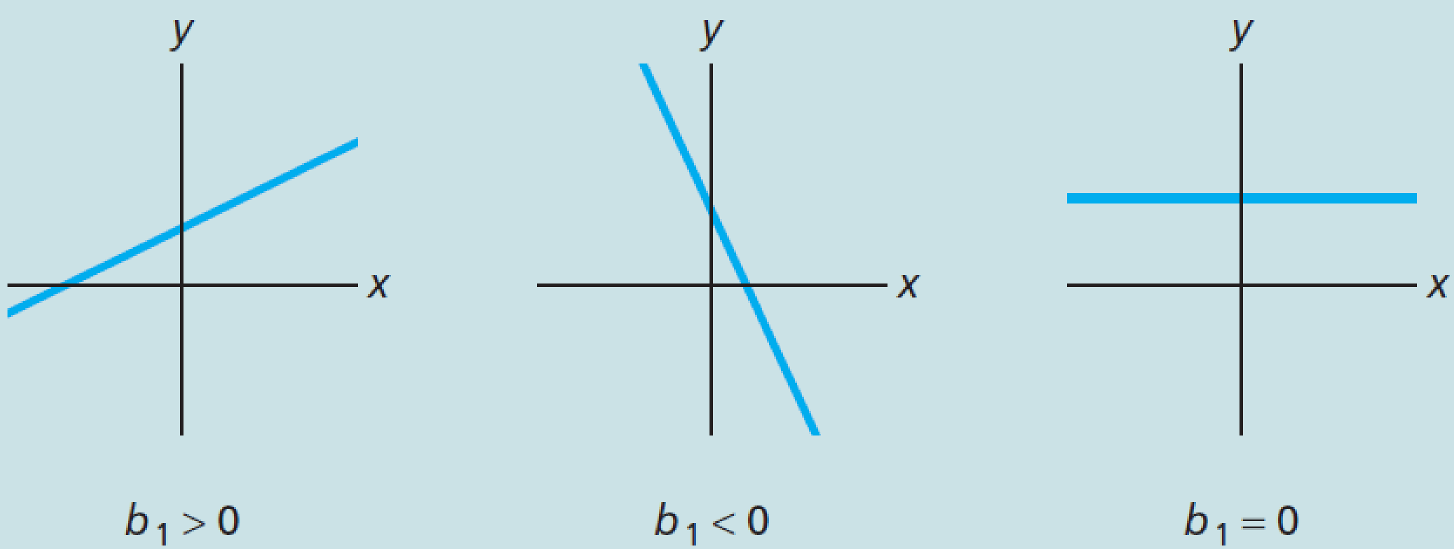
For a linear equation $y = b_0 + b_1 x$, the number $b_0$ is called the **y-intercept** and the number $b_1$ is called the **slope.**


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Key Fact 14.1

## Graphical Interpretation of Slope

The graph of the linear equation $y = b_0 + b_1 x$ slopes upward if $b_1 > 0$, slopes downward if $b_1 < 0$, and is horizontal if $b_1 = 0$, as shown in Fig. 14.6.





Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Section 14.2

## The Regression Equation


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Definition 14.2

**Scatterplot**

A **scatterplot** is a graph of data from two quantitative variables of a population. In a scatterplot, we use a horizontal axis for the observations of one variable and a vertical axis for the observations of the other variable. Each pair of observation is then plotted as a point.

 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Table 14.2

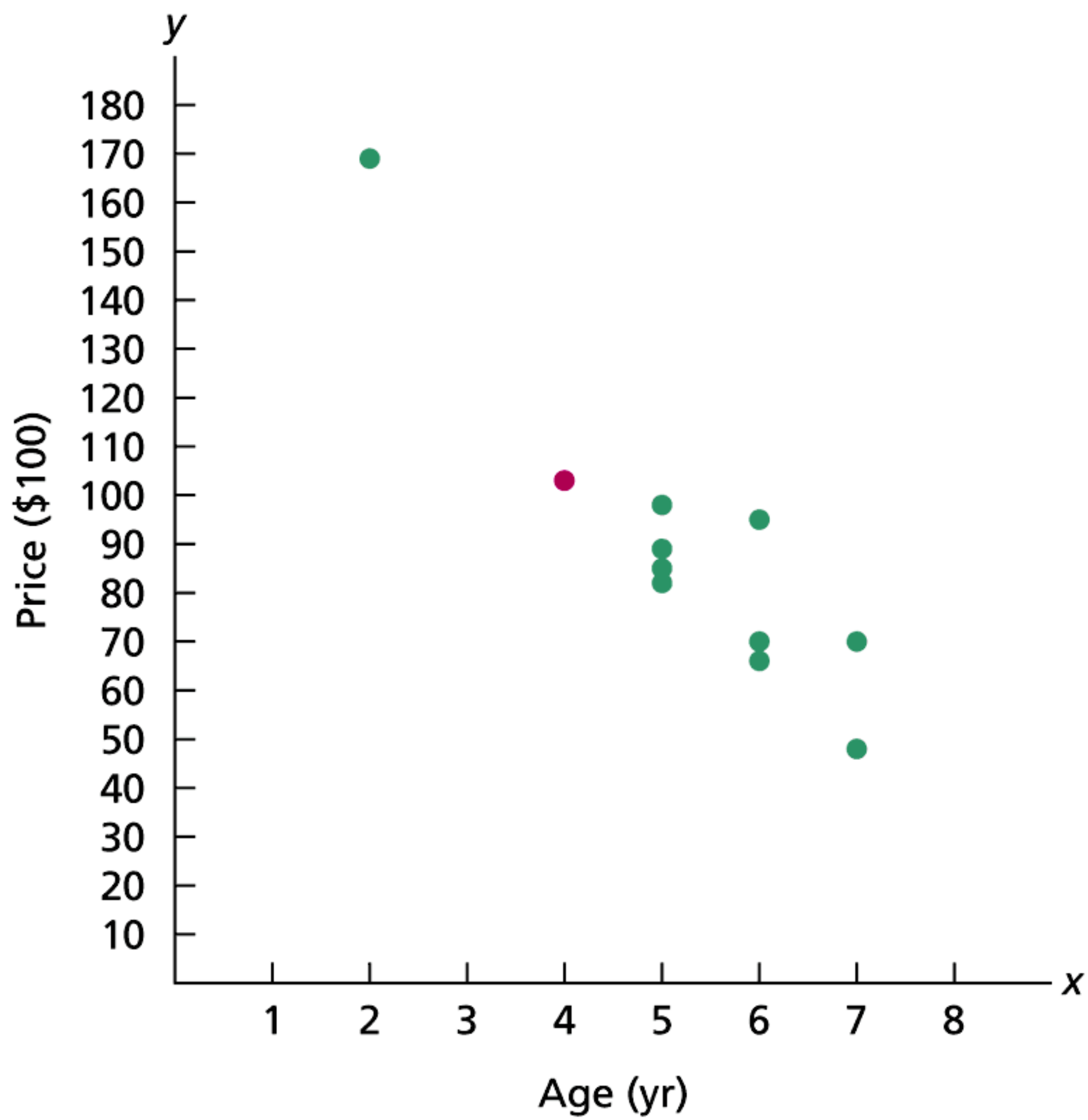
Age and price data for a sample of 11 Orions

| Car | Age (yr) $x$ | Price ($100) $y$ |
|---|---|---|
| 1 | 5 | 85 |
| 2 | 4 | 103 |
| 3 | 6 | 70 |
| 4 | 5 | 82 |
| 5 | 5 | 89 |
| 6 | 5 | 98 |
| 7 | 6 | 66 |
| 8 | 6 | 95 |
| 9 | 2 | 169 |
| 10 | 7 | 70 |
| 11 | 7 | 48 |


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Figure 14.7

Scatterplot for the age and price data of Orions from Table 14.2



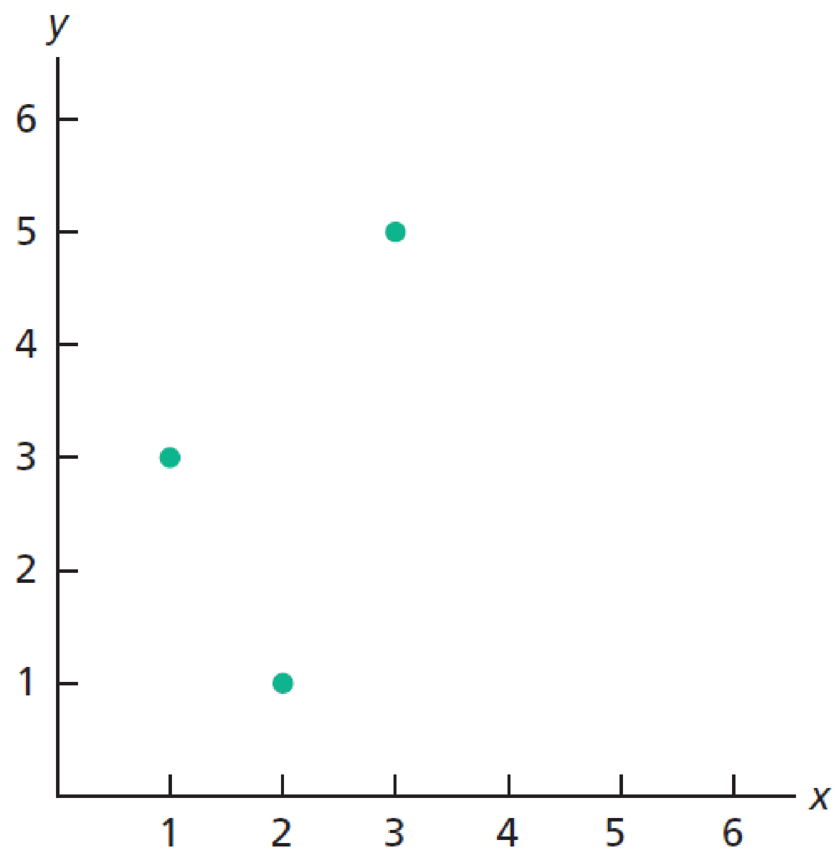

 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Table 14.3 & Figure 14.8

Three data points

| $x$ | $y$ |
|---|---|
| 1 | 3 |
| 2 | 1 |
| 3 | 5 |

Scatterplot for the data points in Table 14.3

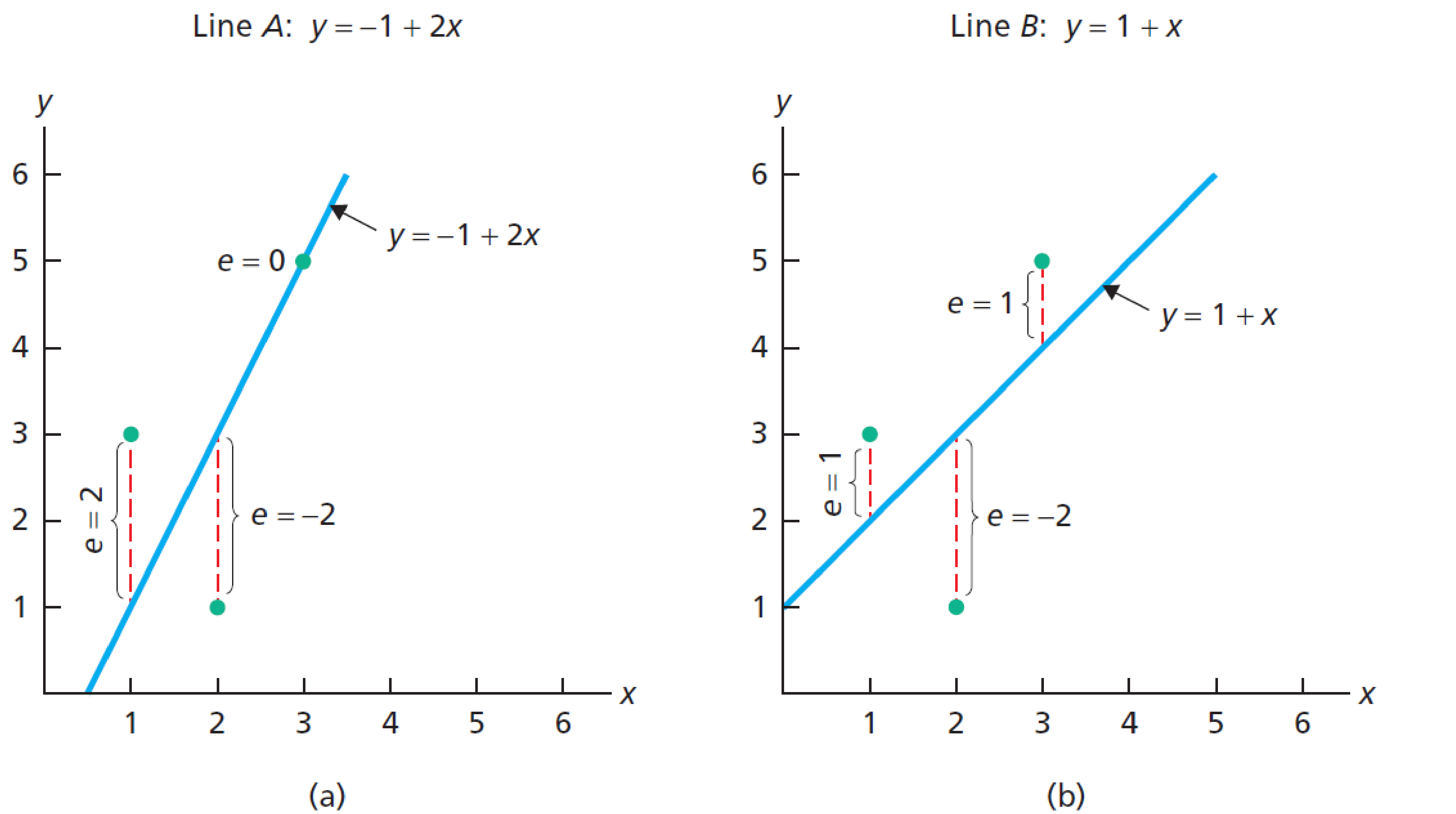




Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Figure 14.9

Two possible lines to fit the data points in Table 14.3




 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Table 14.4

Determining how well the data points in Table 14.3 are fit by (a) Line A and (b) Line B

Line $A$: $y = -1 + 2x$

| $x$ | $y$ | $\hat{y}$ | $e$ | $e^2$ |
|---|---|---|---|---|
| 1 | 3 | 1 | 2 | 4 |
| 2 | 1 | 3 | −2 | 4 |
| 3 | 5 | 5 | 0 | 0 |
| | | | | 8 |

(a)

Line $B$: $y = 1 + x$

| $x$ | $y$ | $\hat{y}$ | $e$ | $e^2$ |
|---|---|---|---|---|
| 1 | 3 | 2 | 1 | 1 |
| 2 | 1 | 3 | −2 | 4 |
| 3 | 5 | 4 | 1 | 1 |
| | | | | 6 |

(b)

 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Key Fact 14.2 & Definition 14.3

## Least-Squares Criterion

The **least-squares criterion** is that the line that best fits a set of data points is the one having the smallest possible sum of squared errors.

## Regression Line and Regression Equation

**Regression line:** The line that best fits a set of data points according to the least-squares criterion.

**Regression equation:** The equation of the regression line.

ALWAYS LEARNING Copyright © 2020, 2016, 2012 Pearson Education, Inc. PEARSON

# Definition 14.4

## Notation Used in Regression and Correlation

For a set of $n$ data points, the defining and computing formulas for $\boldsymbol{S_{xx}}$, $\boldsymbol{S_{xy}}$, and $\boldsymbol{S_{yy}}$ are as follows.

| Quantity | Defining formula | Computing formula |
|---|---|---|
| $S_{xx}$ | $\Sigma(x_i - \bar{x})^2$ | $\Sigma x_i^2 - (\Sigma x_i)^2/n$ |
| $S_{xy}$ | $\Sigma(x_i - \bar{x})(y_i - \bar{y})$ | $\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)/n$ |
| $S_{yy}$ | $\Sigma(y_i - \bar{y})^2$ | $\Sigma y_i^2 - (\Sigma y_i)^2/n$ |

 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Formula 14.1

## Regression Equation

The regression equation for a set of $n$ data points is $\hat{y} = b_0 + b_1 x$, where

$$b_1 = \frac{S_{xy}}{S_{xx}} \quad \text{and} \quad b_0 = \bar{y} - b_1\bar{x} = \frac{1}{n}(\Sigma y_i - b_1 \Sigma x_i).$$

These two equations give the slope and $y$-intercept of the regression line, respectively.


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Table 14.6

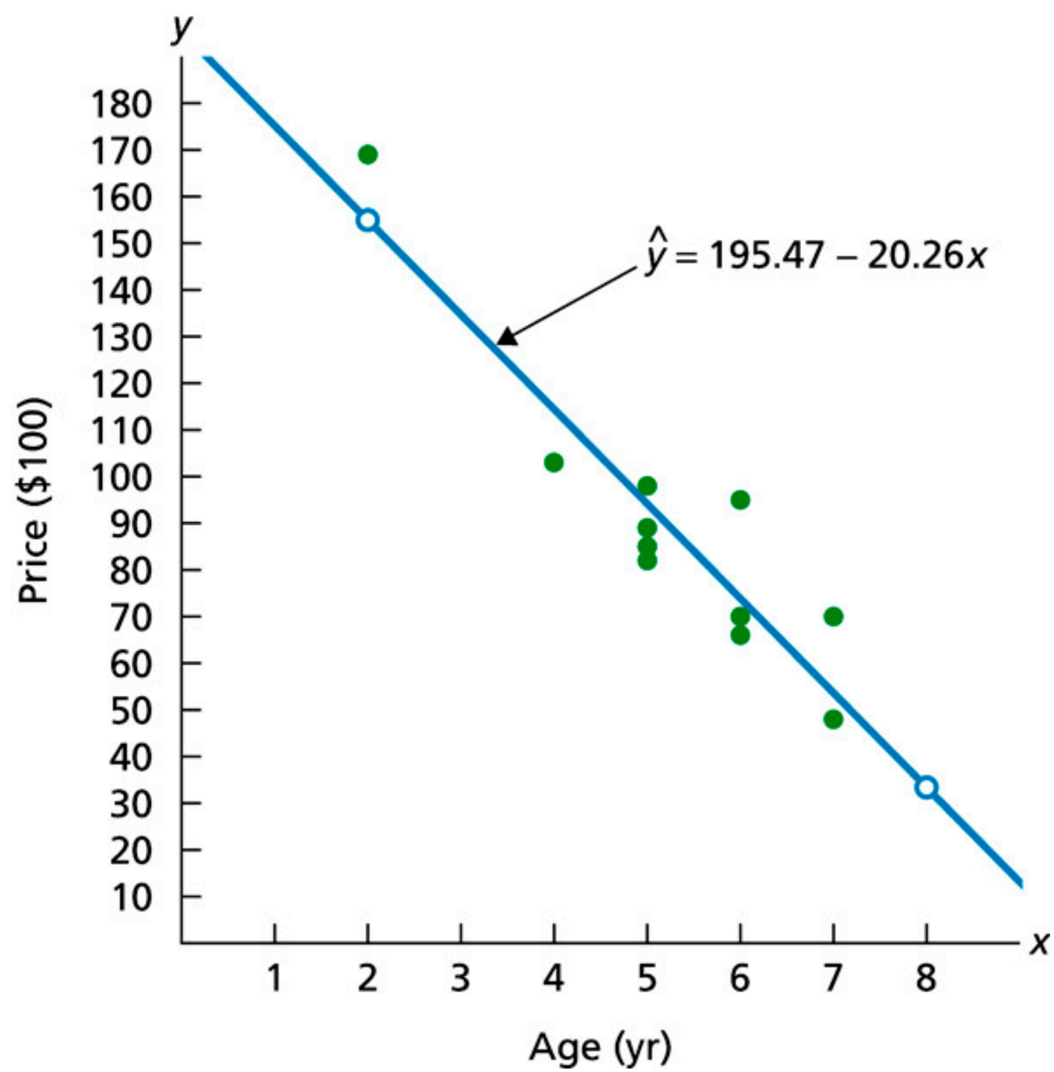
Table for computing the regression equation for the Orion data

| Age (yr) $x$ | Price ($100) $y$ | $xy$ | $x^2$ |
|---|---|---|---|
| 5 | 85 | 425 | 25 |
| 4 | 103 | 412 | 16 |
| 6 | 70 | 420 | 36 |
| 5 | 82 | 410 | 25 |
| 5 | 89 | 445 | 25 |
| 5 | 98 | 490 | 25 |
| 6 | 66 | 396 | 36 |
| 6 | 95 | 570 | 36 |
| 2 | 169 | 338 | 4 |
| 7 | 70 | 490 | 49 |
| 7 | 48 | 336 | 49 |
| 58 | 975 | 4732 | 326 |


Copyright © 2020, 2016, 2012 Pearson Education, Inc. PEARSON

# Figure 14.11

Regression line and data points for Orion data





Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Definition 14.5

**Response Variable and Predictor Variable**

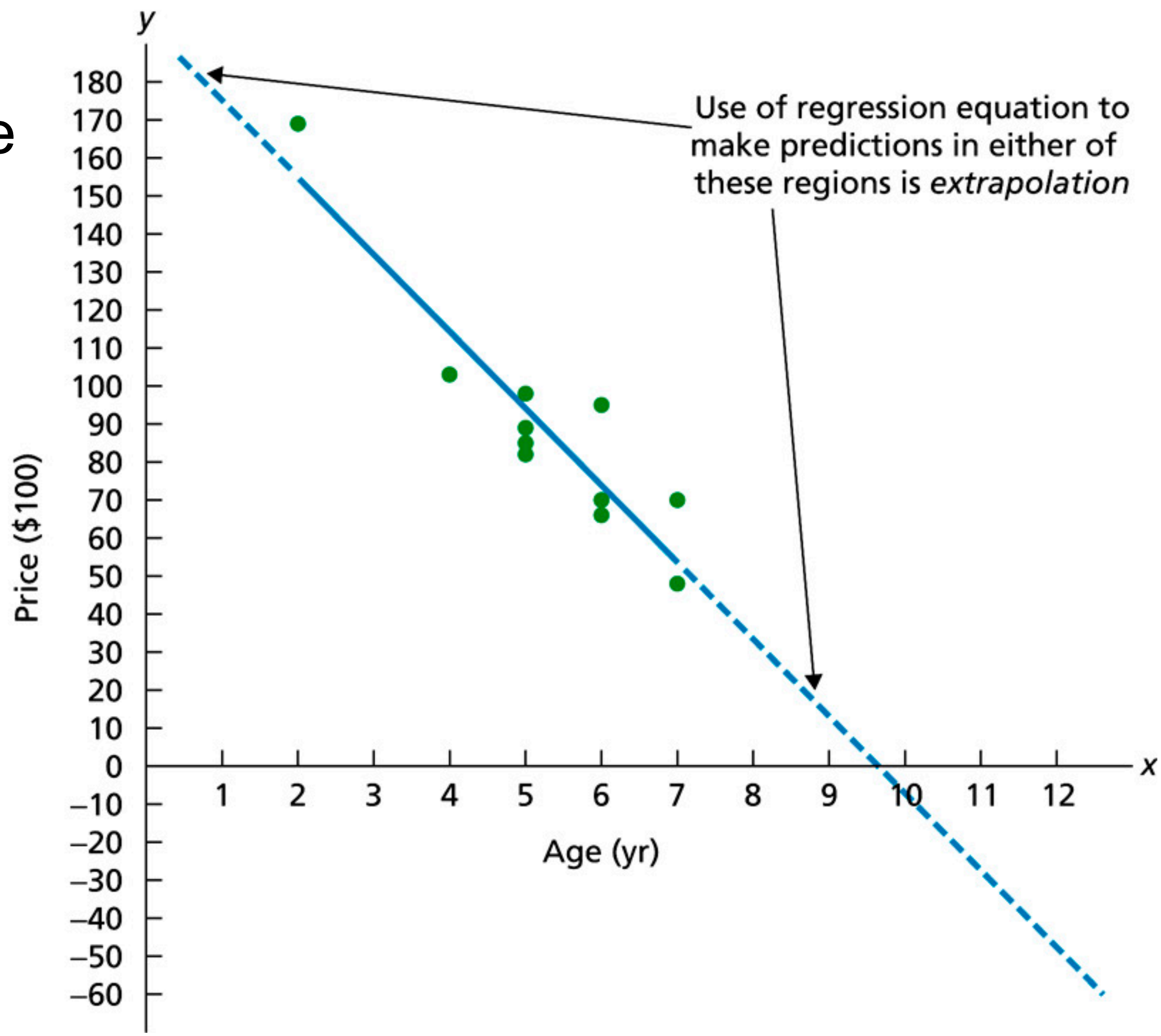
**Response variable**: The variable to be measured or observed.

**Predictor variable:** A variable used to predict or explain the values of the response variable.

 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Figure 14.12

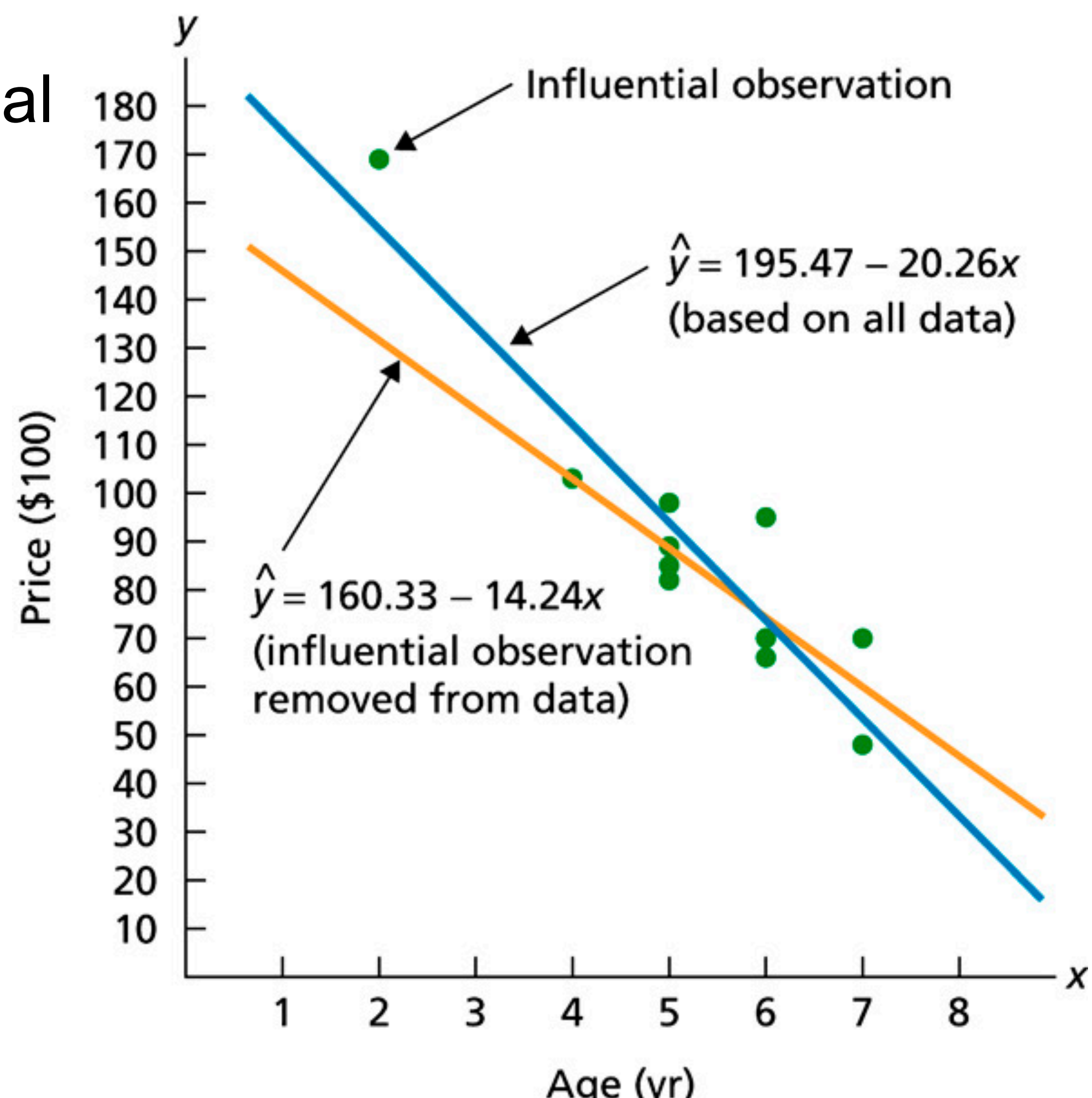
Extrapolation in the Orion example




 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Figure 14.13

Regression lines with and without the influential observation removed




 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Key Fact 14.3

**Criterion for Finding a Regression Line**

Before finding a regression line for a set of data points, draw a scatterplot. If the data points do not appear to be scattered about a line, do not determine a regression line.


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Section 14.3

## The Coefficient of Determination


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Definition 14.6

## Sums of Squares in Regression

**Total sum of squares, *SST*:** The total variation in the observed values of the response variable: $SST = \Sigma(y_i - \bar{y})^2$.

**Regression sum of squares, *SSR*:** The variation in the observed values of the response variable explained by the regression: $SSR = \Sigma(\hat{y}_i - \bar{y})^2$.

**Error sum of squares, *SSE*:** The variation in the observed values of the response variable not explained by the regression: $SSE = \Sigma(y_i - \hat{y}_i)^2$.


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Definition 14.7

## Coefficient of Determination

The **coefficient of determination, $r^2$,** is the proportion of variation in the observed values of the response variable explained by the regression. Thus,

$$r^2 = \frac{SSR}{SST}.$$

 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Table 14.7

Table for finding the three sums of squares

| $x$ | $y$ | $\hat{y}$ | $y - \bar{y}$ | $(y - \bar{y})^2$ | $\hat{y} - \bar{y}$ | $(\hat{y} - \bar{y})^2$ | $y - \hat{y}$ | $(y - \hat{y})^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 0 | 0 | −1 | 1 | 1 | 1 |
| 2 | 1 | 3 | −2 | 4 | 0 | 0 | −2 | 4 |
| 3 | 5 | 4 | 2 | 4 | 1 | 1 | 1 | 1 |
| | | | | 8 | | 2 | | 6 |


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Key Fact 14.4

**Regression Identity**

The total sum of squares equals the regression sum of squares plus the error sum of squares: $SST = SSR + SSE$.


Copyright © 2020, 2016, 2012 Pearson Education, Inc. PEARSON

# Formula 14.2

## Computing Formulas for the Sums of Squares

The computing formulas for the three sums of squares are

$$SST = \Sigma y_i^2 - (\Sigma y_i)^2/n, \quad SSR = \frac{[\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)/n]^2}{\Sigma x_i^2 - (\Sigma x_i)^2/n},$$

and $SSE = SST - SSR$.


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Table 14.8

Table for finding *SST* and *SSR* for the Orion data by using the computing formulas

| Age (yr) $x$ | Price ($100) $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 5 | 85 | 425 | 25 | 7,225 |
| 4 | 103 | 412 | 16 | 10,609 |
| 6 | 70 | 420 | 36 | 4,900 |
| 5 | 82 | 410 | 25 | 6,724 |
| 5 | 89 | 445 | 25 | 7,921 |
| 5 | 98 | 490 | 25 | 9,604 |
| 6 | 66 | 396 | 36 | 4,356 |
| 6 | 95 | 570 | 36 | 9,025 |
| 2 | 169 | 338 | 4 | 28,561 |
| 7 | 70 | 490 | 49 | 4,900 |
| 7 | 48 | 336 | 49 | 2,304 |
| 58 | 975 | 4732 | 326 | 96,129 |


Copyright © 2020, 2016, 2012 Pearson Education, Inc. PEARSON

# Section 14.4

## Linear Correlation

ALWAYS LEARNING Copyright © 2020, 2016, 2012 Pearson Education, Inc. PEARSON 

# Definition 14.8 & Formula 14.3

## Linear Correlation Coefficient

For a set of $n$ data points, the **linear correlation coefficient, $r$,** is defined by

$$r = \frac{\frac{1}{n-1}\Sigma(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y},$$

where $s_x$ and $s_y$ denote the sample standard deviations of the $x$-values and $y$-values, respectively.

## Computing Formula for a Linear Correlation Coefficient

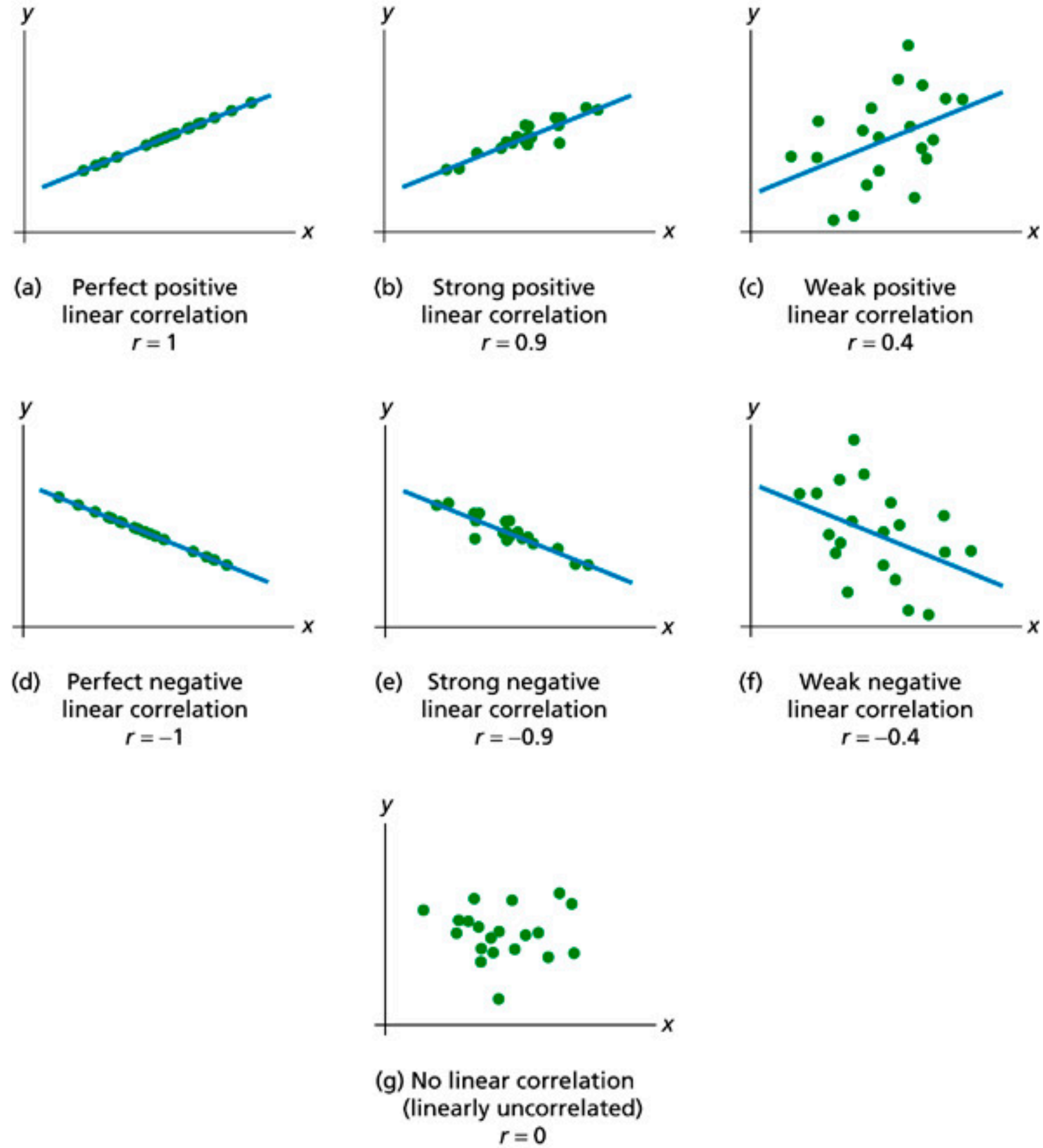
The computing formula for a linear correlation coefficient is

$$r = \frac{\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)/n}{\sqrt{[\Sigma x_i^2 - (\Sigma x_i)^2/n][\Sigma y_i^2 - (\Sigma y_i)^2/n]}}.$$


Copyright © 2020, 2016, 2012 Pearson Education, Inc.

# Figure 14.18

Various degrees of linear correlation




 Copyright © 2020, 2016, 2012 Pearson Education, Inc. 

# Key Fact 14.5

**Relationship between the Correlation Coefficient and the Coefficient of Determination**

The coefficient of determination equals the square of the linear correlation coefficient.


Copyright © 2020, 2016, 2012 Pearson Education, Inc.