# 10.3 Coefficient of Determination

$$\sum \left(y - \overline{y}\right)^2$$

# 10.3 Coefficient of Determination

- The **total variation** $\sum (y-\bar{y})^2$ is the sum of the squares of the vertical distances each point is from the mean.
- The total variation can be divided into two parts: that which is attributed to the relationship of *x* and *y*, and that which is due to chance.

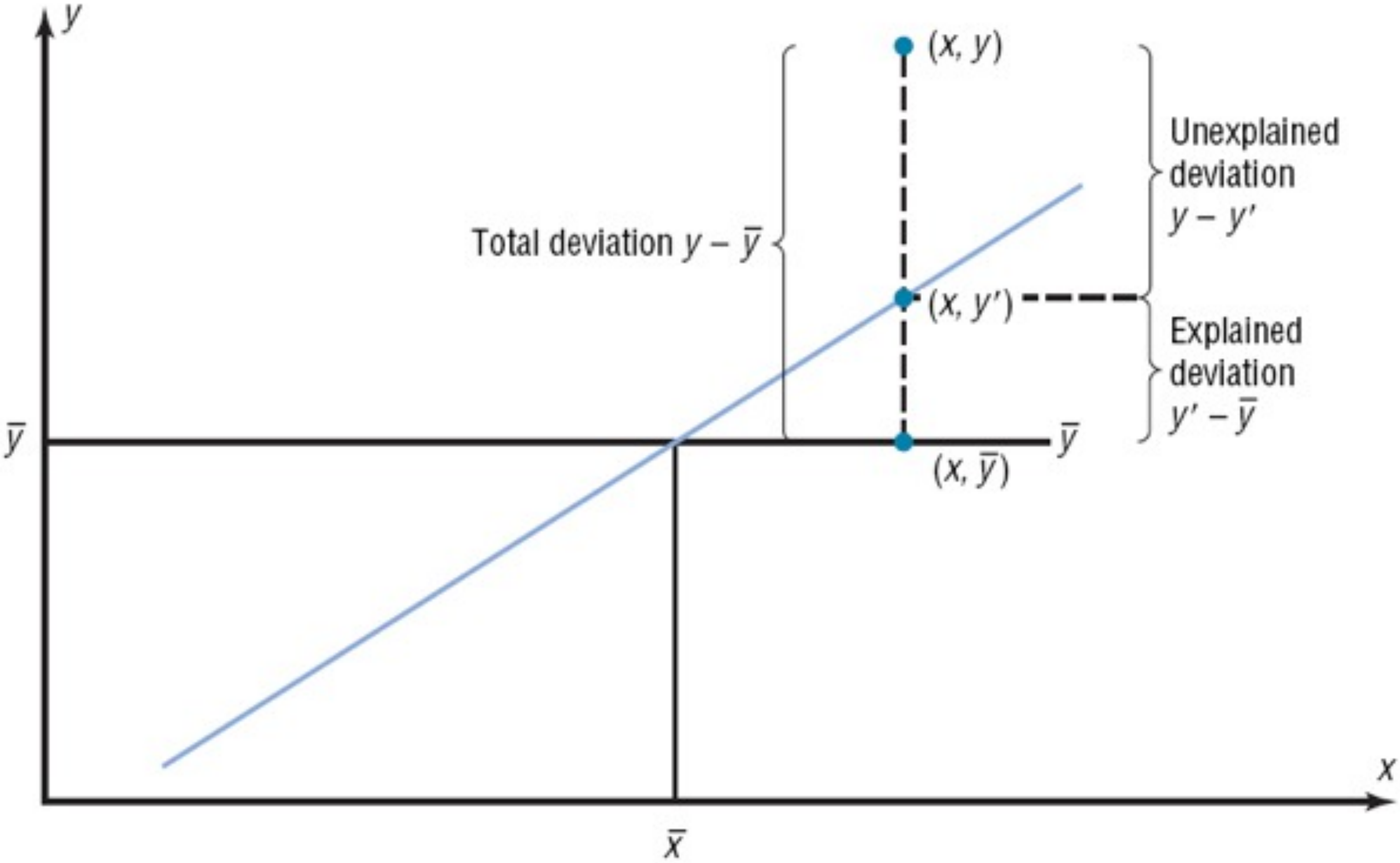
# Variation

$$\sum \left(y' - \bar{y}\right)^2$$

# Variation

- The variation obtained from the relationship (i.e., from the predicted $y'$ values) is $\sum(y'-\bar{y})^2$ and is called the **explained variation**.
- Variation due to chance, found by $\sum(y'-y)^2$, is called the **unexplained variation**. This variation cannot be attributed to the relationships.

# Variation



y
(x, y)
Unexplained
deviation
y − y′
Total deviation y − ȳ
(x, y′)
Explained
deviation
y′ − ȳ
ȳ
ȳ
(x, ȳ)
x
x̄

# Coefficient of Determiation

- The **coefficient of determination** is the ratio of the explained variation to the total variation.

# Coefficient of Determiation

- The **coefficient of determination** is the ratio of the explained variation to the total variation.
- The symbol for the coefficient of determination is $r^2$.

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

# Coefficient of Determiation

- The **coefficient of determination** is the ratio of the explained variation to the total variation.
- The symbol for the coefficient of determination is $r^2$.
- $$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

# Coefficient of Determiation

- The **coefficient of determination** is the ratio of the explained variation to the total variation.
- The symbol for the coefficient of determination is $r^2$.
- $$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

# Coefficient of Determiation

- The **coefficient of determination** is the ratio of the explained variation to the total variation.
- The symbol for the coefficient of determination is $r^2$.
- $$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$
- Another way to arrive at the value for $r^2$ is to square the correlation coefficient.

# Coefficient of Nondetermiation

- The **coefficient of nondetermination** is a measure of the unexplained variation.
- The formula for the coefficient of determination is $1.00 - r^2$.

# Standard Error of the Estimate

- The **standard error of estimate**, denoted by $s_{est}$ is the standard deviation of the observed *y* values about the predicted *y'* values. The formula for the standard error of estimate is:

$$s_{est} = \sqrt{\frac{\sum (y - y')^2}{n - 2}}$$

# Chapter 10
# Correlation and Regression

## Section 10-3

Example 10-12

Page #569

# Example 10-12: Copy Machine Costs

A researcher collects the following data and determines that there is a significant relationship between the age of a copy machine and its monthly maintenance cost. The regression equation is $y' = 55.57 + 8.13x$. Find the standard error of the estimate.

| Machine | Age $x$ (years) | Monthly cost $y$ |
|---|---|---|
| A | 1 | \$ 62 |
| B | 2 | 78 |
| C | 3 | 70 |
| D | 4 | 90 |
| E | 4 | 93 |
| F | 6 | 103 |

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | | | |
| B | 2 | 78 | | | |
| C | 3 | 70 | | | |
| D | 4 | 90 | | | |
| E | 4 | 93 | | | |
| F | 6 | 103 | | | |

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | | | |
| B | 2 | 78 | | | |
| C | 3 | 70 | | | |
| D | 4 | 90 | | | |
| E | 4 | 93 | | | |
| F | 6 | 103 | | | |

$$y' = 55.57 + 8.13x$$

$$y' = 55.57 + 8.13(1) = 63.70$$

$$y' = 55.57 + 8.13(2) = 71.83$$

$$y' = 55.57 + 8.13(3) = 79.96$$

$$y' = 55.57 + 8.13(4) = 88.09$$

$$y' = 55.57 + 8.13(6) = 104.35$$

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | | |
| B | 2 | 78 | 71.83 | | |
| C | 3 | 70 | 79.96 | | |
| D | 4 | 90 | 88.09 | | |
| E | 4 | 93 | 88.09 | | |
| F | 6 | 103 | 104.35 | | |

$$y' = 55.57 + 8.13x$$

$$y' = 55.57 + 8.13(1) = 63.70$$

$$y' = 55.57 + 8.13(2) = 71.83$$

$$y' = 55.57 + 8.13(3) = 79.96$$

$$y' = 55.57 + 8.13(4) = 88.09$$

$$y' = 55.57 + 8.13(6) = 104.35$$

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | | |
| B | 2 | 78 | 71.83 | | |
| C | 3 | 70 | 79.96 | | |
| D | 4 | 90 | 88.09 | | |
| E | 4 | 93 | 88.09 | | |
| F | 6 | 103 | 104.35 | | |

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | -1.70 | |
| B | 2 | 78 | 71.83 | 6.17 | |
| C | 3 | 70 | 79.96 | -9.96 | |
| D | 4 | 90 | 88.09 | 1.91 | |
| E | 4 | 93 | 88.09 | 4.91 | |
| F | 6 | 103 | 104.35 | -1.35 | |

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | -1.70 | 2.89 |
| B | 2 | 78 | 71.83 | 6.17 | 38.0689 |
| C | 3 | 70 | 79.96 | -9.96 | 99.2016 |
| D | 4 | 90 | 88.09 | 1.91 | 3.6481 |
| E | 4 | 93 | 88.09 | 4.91 | 24.1081 |
| F | 6 | 103 | 104.35 | -1.35 | 1.8225 |

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | -1.70 | 2.89 |
| B | 2 | 78 | 71.83 | 6.17 | 38.0689 |
| C | 3 | 70 | 79.96 | -9.96 | 99.2016 |
| D | 4 | 90 | 88.09 | 1.91 | 3.6481 |
| E | 4 | 93 | 88.09 | 4.91 | 24.1081 |
| F | 6 | 103 | 104.35 | -1.35 | 1.8225 |
| | | | | | 169.7392 |

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | -1.70 | 2.89 |
| B | 2 | 78 | 71.83 | 6.17 | 38.0689 |
| C | 3 | 70 | 79.96 | -9.96 | 99.2016 |
| D | 4 | 90 | 88.09 | 1.91 | 3.6481 |
| E | 4 | 93 | 88.09 | 4.91 | 24.1081 |
| F | 6 | 103 | 104.35 | -1.35 | 1.8225 |
| | | | | | 169.7392 |

$$s_{est} = \sqrt{\frac{\sum (y - y')^2}{n - 2}}$$

# Example 10-12: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $y'$ | $y - y'$ | $(y - y')^2$ |
|---|---|---|---|---|---|
| A | 1 | 62 | 63.70 | -1.70 | 2.89 |
| B | 2 | 78 | 71.83 | 6.17 | 38.0689 |
| C | 3 | 70 | 79.96 | -9.96 | 99.2016 |
| D | 4 | 90 | 88.09 | 1.91 | 3.6481 |
| E | 4 | 93 | 88.09 | 4.91 | 24.1081 |
| F | 6 | 103 | 104.35 | -1.35 | 1.8225 |
| | | | | | 169.7392 |

$$s_{est} = \sqrt{\frac{\sum (y - y')^2}{n-2}}$$

$$s_{est} = \sqrt{\frac{169.7392}{4}} = 6.51$$

# Chapter 10
# Correlation and Regression

## Section 10-3

Example 10-13

Page #570

# Example 10-13: Copy Machine Costs

$$s_{est} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n-2}}$$

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | | |
| B | 2 | 78 | | |
| C | 3 | 70 | | |
| D | 4 | 90 | | |
| E | 4 | 93 | | |
| F | 6 | 103 | | |

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | |
| B | 2 | 78 | 156 | |
| C | 3 | 70 | 210 | |
| D | 4 | 90 | 360 | |
| E | 4 | 93 | 372 | |
| F | 6 | 103 | 618 | |

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | 3,844 |
| B | 2 | 78 | 156 | 6,084 |
| C | 3 | 70 | 210 | 4,900 |
| D | 4 | 90 | 360 | 8,100 |
| E | 4 | 93 | 372 | 8,649 |
| F | 6 | 103 | 618 | 10,609 |

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | 3,844 |
| B | 2 | 78 | 156 | 6,084 |
| C | 3 | 70 | 210 | 4,900 |
| D | 4 | 90 | 360 | 8,100 |
| E | 4 | 93 | 372 | 8,649 |
| F | 6 | 103 | 618 | 10,609 |
| | | 496 | | |

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | 3,844 |
| B | 2 | 78 | 156 | 6,084 |
| C | 3 | 70 | 210 | 4,900 |
| D | 4 | 90 | 360 | 8,100 |
| E | 4 | 93 | 372 | 8,649 |
| F | 6 | 103 | 618 | 10,609 |
| | | 496 | 1778 | |

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | 3,844 |
| B | 2 | 78 | 156 | 6,084 |
| C | 3 | 70 | 210 | 4,900 |
| D | 4 | 90 | 360 | 8,100 |
| E | 4 | 93 | 372 | 8,649 |
| F | 6 | 103 | 618 | 10,609 |
| | | 496 | 1778 | 42,186 |

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | 3,844 |
| B | 2 | 78 | 156 | 6,084 |
| C | 3 | 70 | 210 | 4,900 |
| D | 4 | 90 | 360 | 8,100 |
| E | 4 | 93 | 372 | 8,649 |
| F | 6 | 103 | 618 | 10,609 |
| | | 496 | 1778 | 42,186 |

$$s_{est} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n-2}}$$

# Example 10-13: Copy Machine Costs

| Machine | Age $x$ (years) | Monthly cost, $y$ | $xy$ | $y^2$ |
|---|---|---|---|---|
| A | 1 | 62 | 62 | 3,844 |
| B | 2 | 78 | 156 | 6,084 |
| C | 3 | 70 | 210 | 4,900 |
| D | 4 | 90 | 360 | 8,100 |
| E | 4 | 93 | 372 | 8,649 |
| F | 6 | 103 | 618 | 10,609 |
| | | 496 | 1778 | 42,186 |

$$s_{est} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n-2}}$$

$$s_{est} = \sqrt{\frac{42,186 - 55.57(496) - 8.13(1778)}{4}} = 6.48$$

# Formula for the Prediction Interval about a Value $y'$

# Formula for the Prediction Interval about a Value $y'$

$$y' - t_{\alpha/2} s_{est} \sqrt{1 + \frac{1}{n} + \frac{n\left(x - \bar{X}\right)^2}{n\sum x^2 - \left(\sum x\right)^2}} < y$$

$$< y' + t_{\alpha/2} s_{est} \sqrt{1 + \frac{1}{n} + \frac{n\left(x - \bar{X}\right)^2}{n\sum x^2 - \left(\sum x\right)^2}}$$

with d.f. $= n - 2$

# Chapter 10
# Correlation and Regression

## Section 10-3

Example 10-14

Page #571

# Example 10-14: Copy Machine Costs

For the data in Example 10–12, find the 95% prediction interval for the monthly maintenance cost of a machine that is 3 years old.

**Step 1:** Find

**Step 2:** Find $y'$ *for* $x$ = 3.

**Step 3:** Find $s_{est}$.

(as shown in Example 10-13)

# Example 10-14: Copy Machine Costs

For the data in Example 10–12, find the 95% prediction interval for the monthly maintenance cost of a machine that is 3 years old.

**Step 1:** Find $\sum x, \sum x^2, \text{ and } \overline{X}.$

**Step 2:** Find $y'$ *for* $x$ = 3.

**Step 3:** Find $s_{est}$.

(as shown in Example 10-13)

# Example 10-14: Copy Machine Costs

For the data in Example 10–12, find the 95% prediction interval for the monthly maintenance cost of a machine that is 3 years old.

**Step 1:** Find $\sum x, \sum x^2, \text{ and } \bar{X}.$

$$\sum x = 20 \quad \sum x^2 = 82 \quad \bar{X} = \frac{20}{6} = 3.3$$

**Step 2:** Find $y'$ *for* $x$ = 3*.*

**Step 3:** Find $s_{est}$.

(as shown in Example 10-13)

# Example 10-14: Copy Machine Costs

For the data in Example 10–12, find the 95% prediction interval for the monthly maintenance cost of a machine that is 3 years old.

**Step 1:** Find $\sum x, \sum x^2$, and $\bar{X}$.

$$\sum x = 20 \quad \sum x^2 = 82 \quad \bar{X} = \frac{20}{6} = 3.3$$

**Step 2:** Find $y'$ *for* $x$ = 3.

$$y' = 55.57 + 8.13(3) = 79.96$$

**Step 3:** Find $s_{est}$.

(as shown in Example 10-13)

# Example 10-14: Copy Machine Costs

For the data in Example 10–12, find the 95% prediction interval for the monthly maintenance cost of a machine that is 3 years old.

**Step 1:** Find $\sum x, \sum x^2, \text{ and } \overline{X}.$

$$\sum x = 20 \quad \sum x^2 = 82 \quad \overline{X} = \frac{20}{6} = 3.3$$

**Step 2:** Find $y'$ *for* $x$ = 3*.*

$$y' = 55.57 + 8.13\left(3\right) = 79.96$$

# Example 10-14: Copy Machine Costs

For the data in Example 10–12, find the 95% prediction interval for the monthly maintenance cost of a machine that is 3 years old.

**Step 1:** Find $\sum x, \sum x^2, \text{ and } \bar{X}.$

$$\sum x = 20 \quad \sum x^2 = 82 \quad \bar{X} = \frac{20}{6} = 3.3$$

**Step 2:** Find $y'$ *for* $x$ = 3.

$$y' = 55.57 + 8.13(3) = 79.96$$

**Step 3:** Find $s_{est}$.

$s_{est} = 6.48$ (as shown in Example 10-13)

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

$$y' - t_{\alpha/2} s_{est} \sqrt{1 + \frac{1}{n} + \frac{n\left(x - \overline{X}\right)^2}{n\sum x^2 - \left(\sum x\right)^2}} < y$$

$$< y' + t_{\alpha/2} s_{est} \sqrt{1 + \frac{1}{n} + \frac{n\left(x - \overline{X}\right)^2}{n\sum x^2 - \left(\sum x\right)^2}}$$

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

$$y' - t_{\alpha/2} s_{est} \sqrt{1 + \frac{1}{n} + \frac{n\left(x - \bar{X}\right)^2}{n\sum x^2 - \left(\sum x\right)^2}} < y$$

$$< y' + t_{\alpha/2} s_{est} \sqrt{1 + \frac{1}{n} + \frac{n\left(x - \bar{X}\right)^2}{n\sum x^2 - \left(\sum x\right)^2}}$$

$$79.96 - (2.776)(6.48)\sqrt{1 + \frac{1}{6} + \frac{6(3 - 3.3)^2}{6(82) - (20)^2}} < y$$

$$< 79.96 + (2.776)(6.48)\sqrt{1 + \frac{1}{6} + \frac{6(3 - 3.3)^2}{6(82) - (20)^2}}$$

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

$$79.96-(2.776)(6.48)\sqrt{1+\frac{1}{6}+\frac{6(3-3.3)^2}{6(82)-(20)^2}} < y$$

$$< 79.96+(2.776)(6.48)\sqrt{1+\frac{1}{6}+\frac{6(3-3.3)^2}{6(82)-(20)^2}}$$

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

$$79.96-(2.776)(6.48)\sqrt{1+\frac{1}{6}+\frac{6(3-3.3)^2}{6(82)-(20)^2}}<y$$

$$<79.96+(2.776)(6.48)\sqrt{1+\frac{1}{6}+\frac{6(3-3.3)^2}{6(82)-(20)^2}}$$

$$79.96-19.43<y<79.96+19.43$$

$$60.53<y<99.39$$

# Example 10-14: Copy Machine Costs

**Step 4:** Substitute in the formula and solve.

$$79.96-(2.776)(6.48)\sqrt{1+\frac{1}{6}+\frac{6(3-3.3)^2}{6(82)-(20)^2}}<y$$

$$<79.96+(2.776)(6.48)\sqrt{1+\frac{1}{6}+\frac{6(3-3.3)^2}{6(82)-(20)^2}}$$

$$79.96-19.43<y<79.96+19.43$$

$$60.53<y<99.39$$

Hence, you can be 95% confident that the interval $60.53 < y < 99.39$ contains the actual value of *y*.