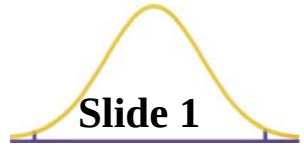
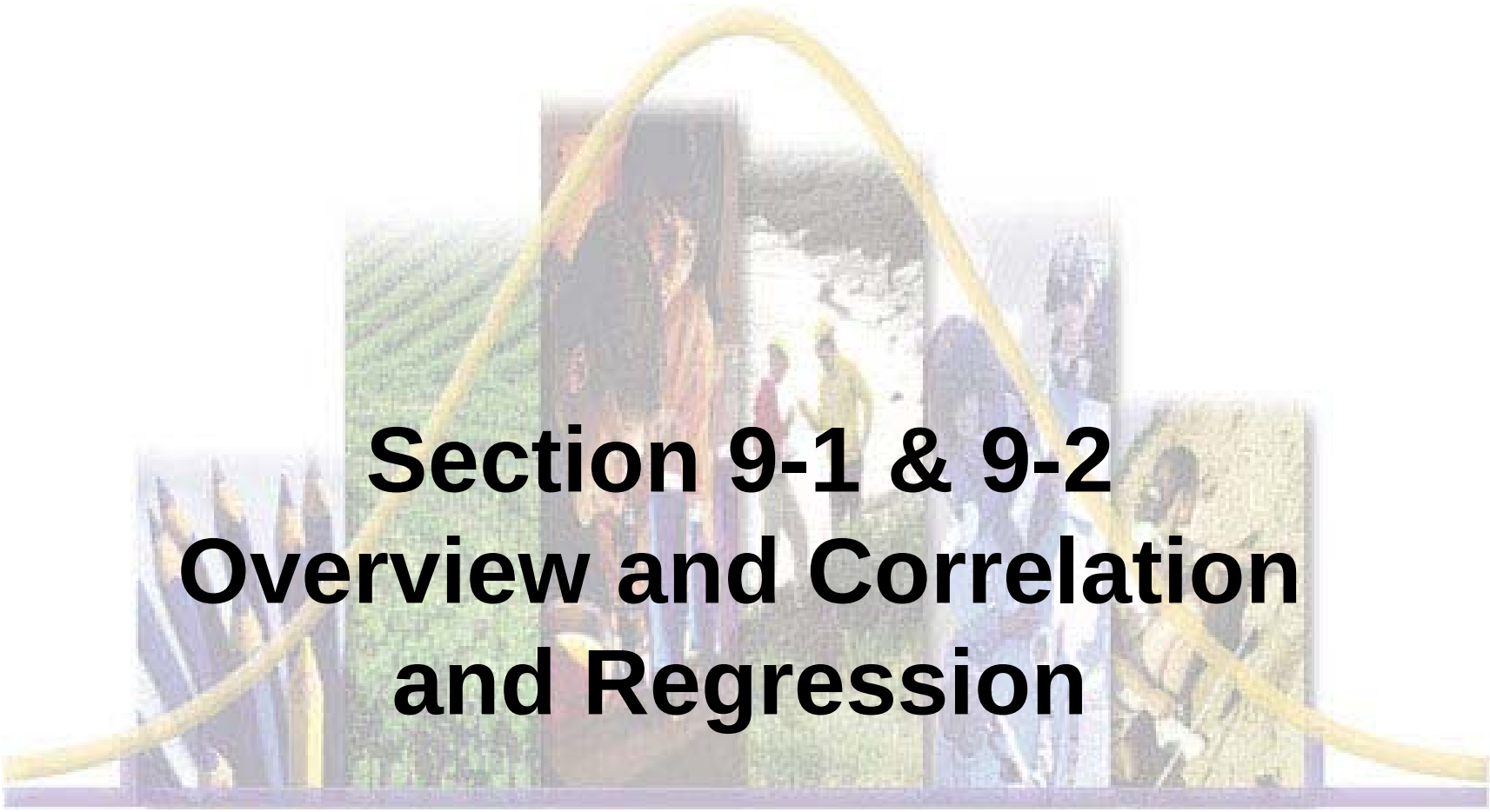


Chapter 9
Correlation and Regression



Section 9-1 & 9-2 Overview and Correlation and Regression

Created by Erin Hodgess, Houston, Texas

Overview

Paired Data (p.506)

- Is there a relationship?
- If so, what is the equation?
- Use that equation for prediction.

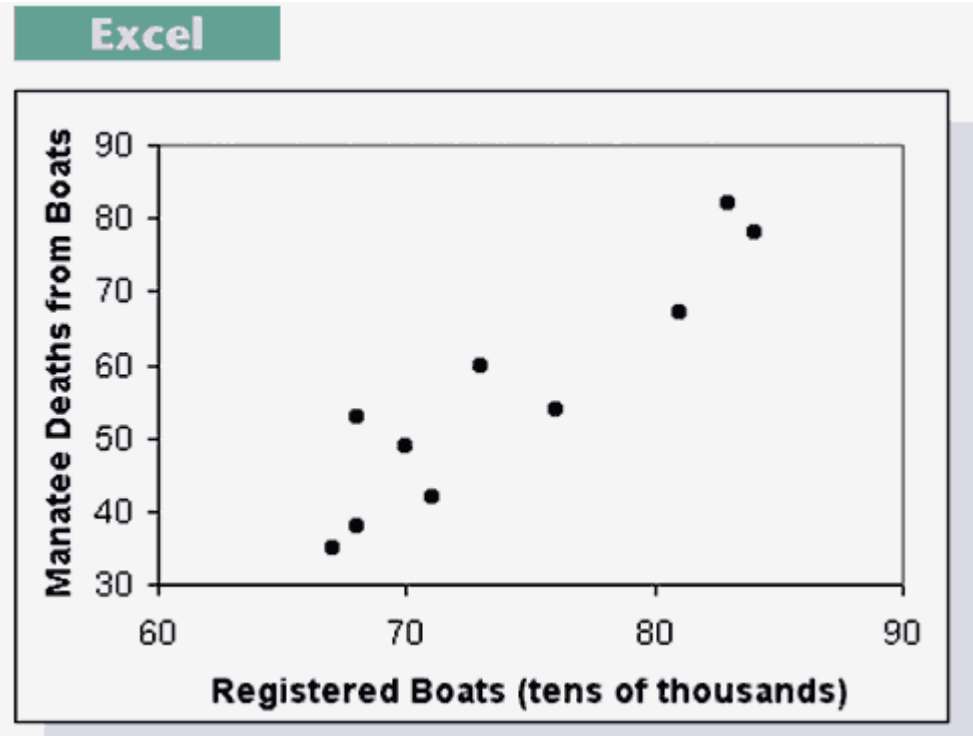
Definition

- A **correlation** exists between two variables when one of them is related to the other in some way.

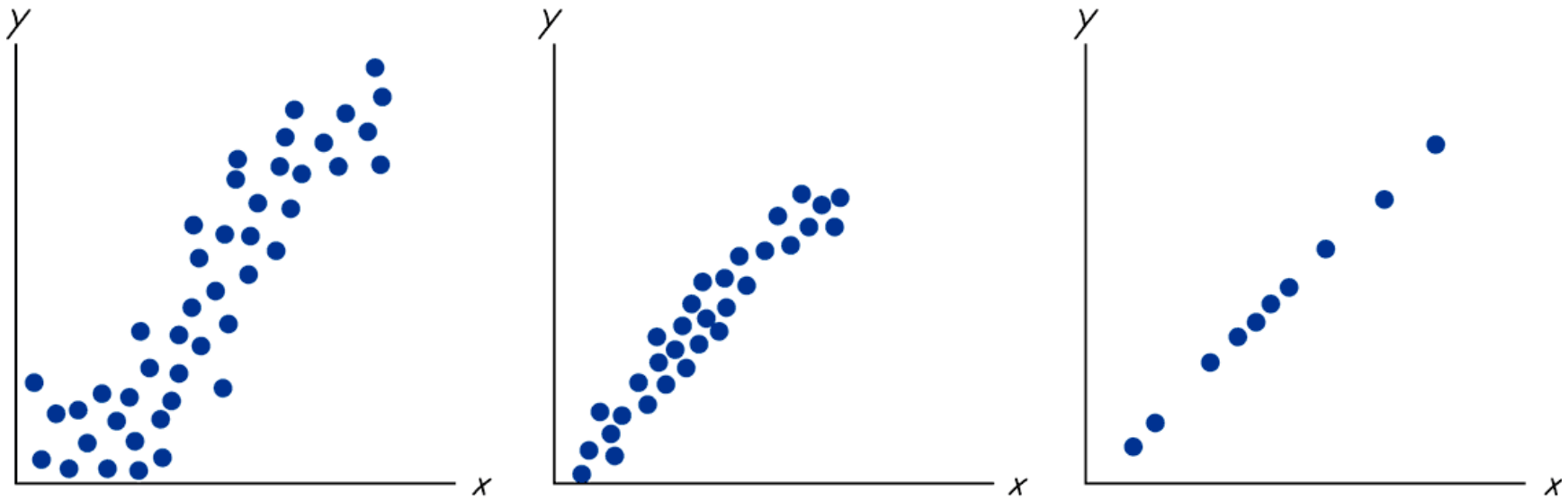
Definition

- **A Scatterplot (or scatter diagram) is a graph in which the paired (x, y) sample data are plotted with a horizontal x-axis and a vertical y-axis. Each individual (x, y) pair is plotted as a single point.**

Scatter Diagram of Paired Data (p.507)



Positive Linear Correlation (p.498)



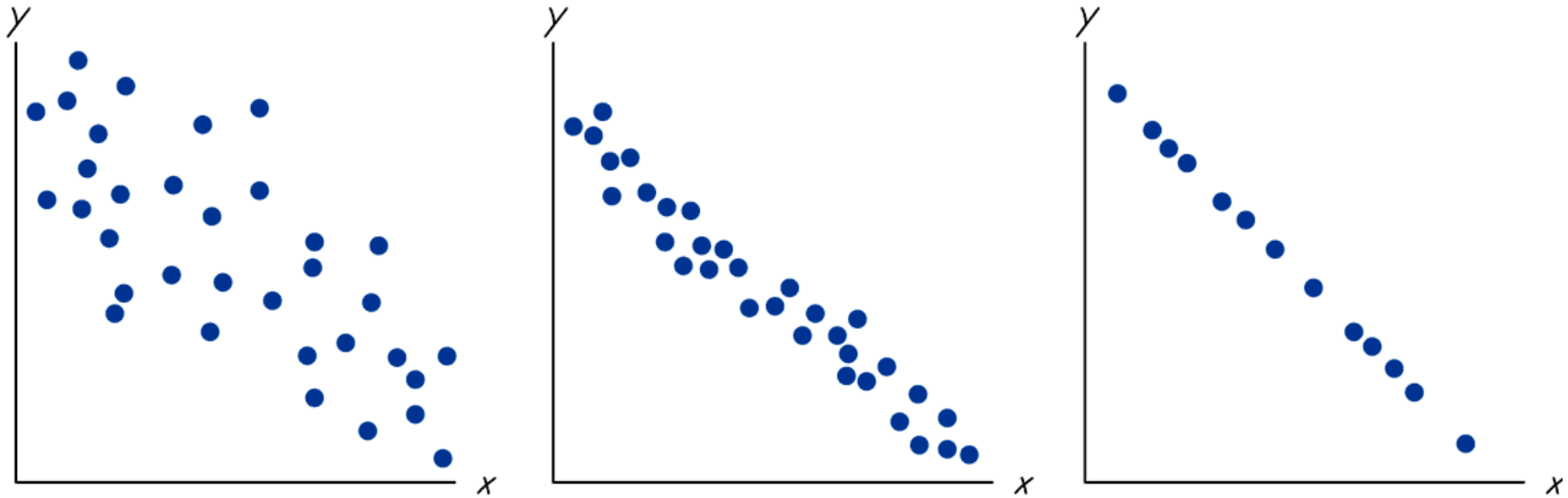
(a) Positive correlation between x and y

(b) Strong positive correlation between x and y

(c) Perfect positive correlation between x and y

Figure 9-2 Scatter Plots

Negative Linear Correlation



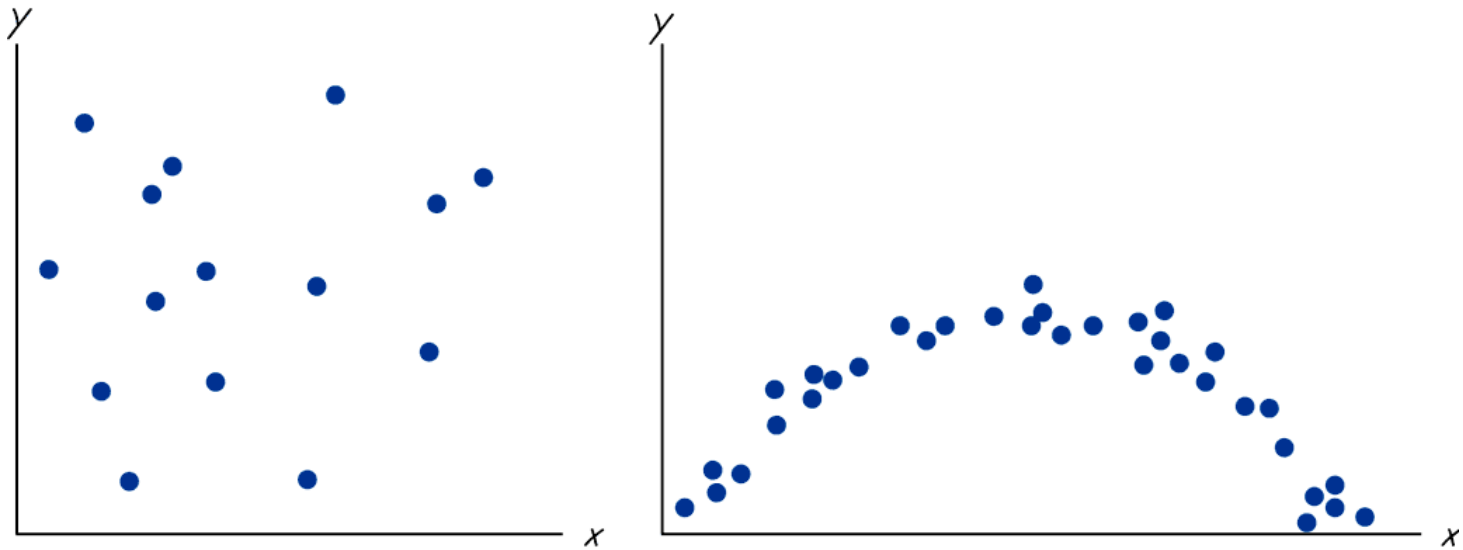
(d) Negative correlation between x and y

(e) Strong negative correlation between x and y

(f) Perfect negative correlation between x and y

Figure 9-2 Scatter Plots

No Linear Correlation



(g) No correlation between x and y

(h) Nonlinear relationship between x and y

Figure 9-2 Scatter Plots

Definition (p.509)

The linear correlation coefficient r measures strength of the linear relationship between paired x and y values in a sample.

Assumptions (p.507)

1. **The sample of paired data (x, y) is a random sample.**

2. **The pairs of (x, y) data have a bivariate normal distribution.**

Notation for the Linear Correlation Coefficient

n = number of pairs of data presented

Σ denotes the addition of the items indicated.

Σx denotes the sum of all x-values.

Σx^2 indicates that each x-value should be squared and then those squares added.

$(\Sigma x)^2$ indicates that the x-values should be added and the total then squared.

Σxy indicates that each x-value should be first multiplied by its corresponding y-value. After obtaining all such products, find their sum.

r represents linear correlation coefficient for a sample

ρ represents linear correlation coefficient for a population

Definition

The linear correlation coefficient r measures the strength of a linear relationship between the paired values in a sample.

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

Formula 9-1

Calculators can compute r

ρ (rho) is the linear correlation coefficient for all paired data in the population.

Rounding the Linear Correlation Coefficient *r*

- **Round to three decimal places so that it can be compared to critical values in Table A-6. (see p.510)**

- **Use calculator or computer if possible.**

Calculating *r*

Data

x	1	1	3	5
y	2	8	6	4

This data is from exercise #7 on p.521.

Calculating *r*

Table 9-2 Finding Statistics Used to Calculate *r*

	x	y	$x \cdot y$	x^2	y^2
	1	2	2	1	4
	1	8	8	1	64
	3	6	18	9	36
	5	4	20	25	16
Total	**10** ↑ Σx	**20** ↑ Σy	**48** ↑ Σxy	**36** ↑ Σx^2	**120** ↑ Σy^2

Calculating r

Data

x	1	1	3	5
y	2	8	6	4

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

$$r = \frac{4(48) - (10)(20)}{\sqrt{4(36) - (10)^2}\sqrt{4(120) - (20)^2}}$$

$$r = \frac{-8}{59.329} = -0.135$$

Interpreting the Linear Correlation Coefficient (p.511)

- If the absolute value of *r* exceeds the value in Table A - 6, conclude that there is a significant linear correlation.

- Otherwise, there is not sufficient evidence to support the conclusion of significant linear correlation.

Example: Boats and Manatees

Given the sample data in Table 9-1, find the value of the linear correlation coefficient r, then refer to Table A-6 to determine whether there is a significant linear correlation between the number of registered boats and the number of manatees killed by boats.

Using the same procedure previously illustrated, we find that $r = 0.922$.

Referring to Table A-6, we locate the row for which n=10. Using the critical value for α=5, we have 0.632. Because $r = 0.922$, its absolute value exceeds 0.632, so we conclude that there is a significant linear correlation between number of registered boats and number of manatee deaths from boats.

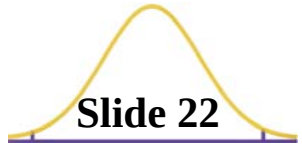
Properties of the Linear Correlation Coefficient *r*

1. $-1 \leq r \leq 1$ (see also p.512)
2. Value of *r* does not change if all values of either variable are converted to a different scale.
3. The *r* is not affected by the choice of *x* and *y*. interchange *x* and *y* and the value of *r* will not change.
4. *r* measures strength of a linear relationship.

Interpreting r: Explained Variation

The value of r^2 is the proportion of the variation in y that is explained by the linear relationship between x and y. (p.503 and p.533)

Example: Boats and Manatees



Using the boat/manatee data in Table 9-1, we have found that the value of the linear correlation coefficient $r = 0.922$. What proportion of the variation of the manatee deaths can be explained by the variation in the number of boat registrations?

With $r = 0.922$, we get $r^2 = 0.850$.

We conclude that 0.850 (or about 85%) of the variation in manatee deaths can be explained by the linear relationship between the number of boat registrations and the number of manatee deaths from boats. This implies that 15% of the variation of manatee deaths cannot be explained by the number of boat registrations.

Common Errors Involving Correlation (pp.503-504)

1. **Causation**: It is wrong to conclude that correlation implies causality.

2. **Averages**: Averages suppress individual variation and may inflate the correlation coefficient.

3. **Linearity**: There may be some relationship between x and y even when there is no significant linear correlation.

Common Errors Involving Correlation

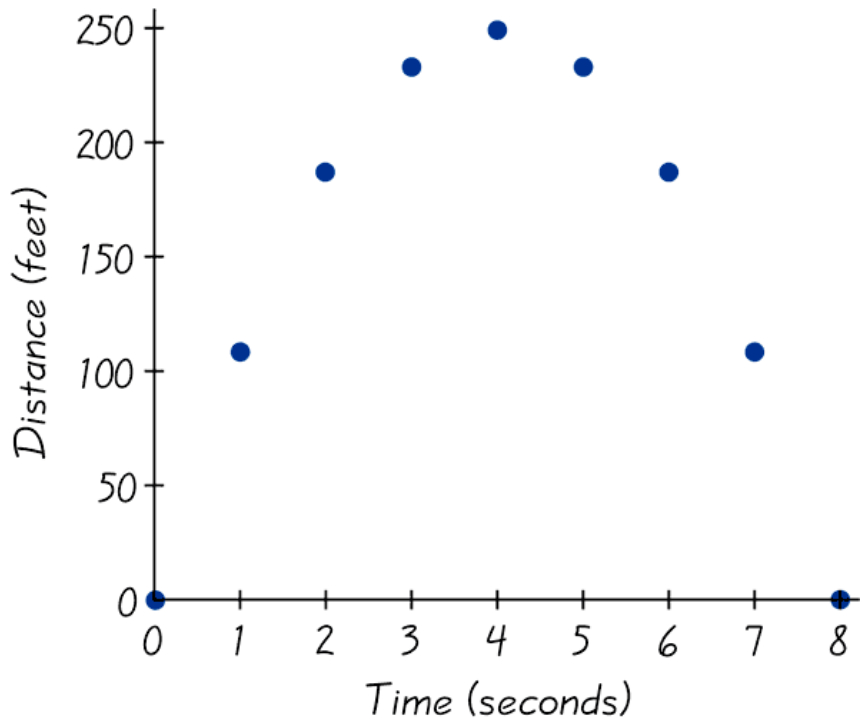


FIGURE 9-3

Scatterplot of Distance above Ground and Time for Object Thrown Upward

Formal Hypothesis Test (p.504)

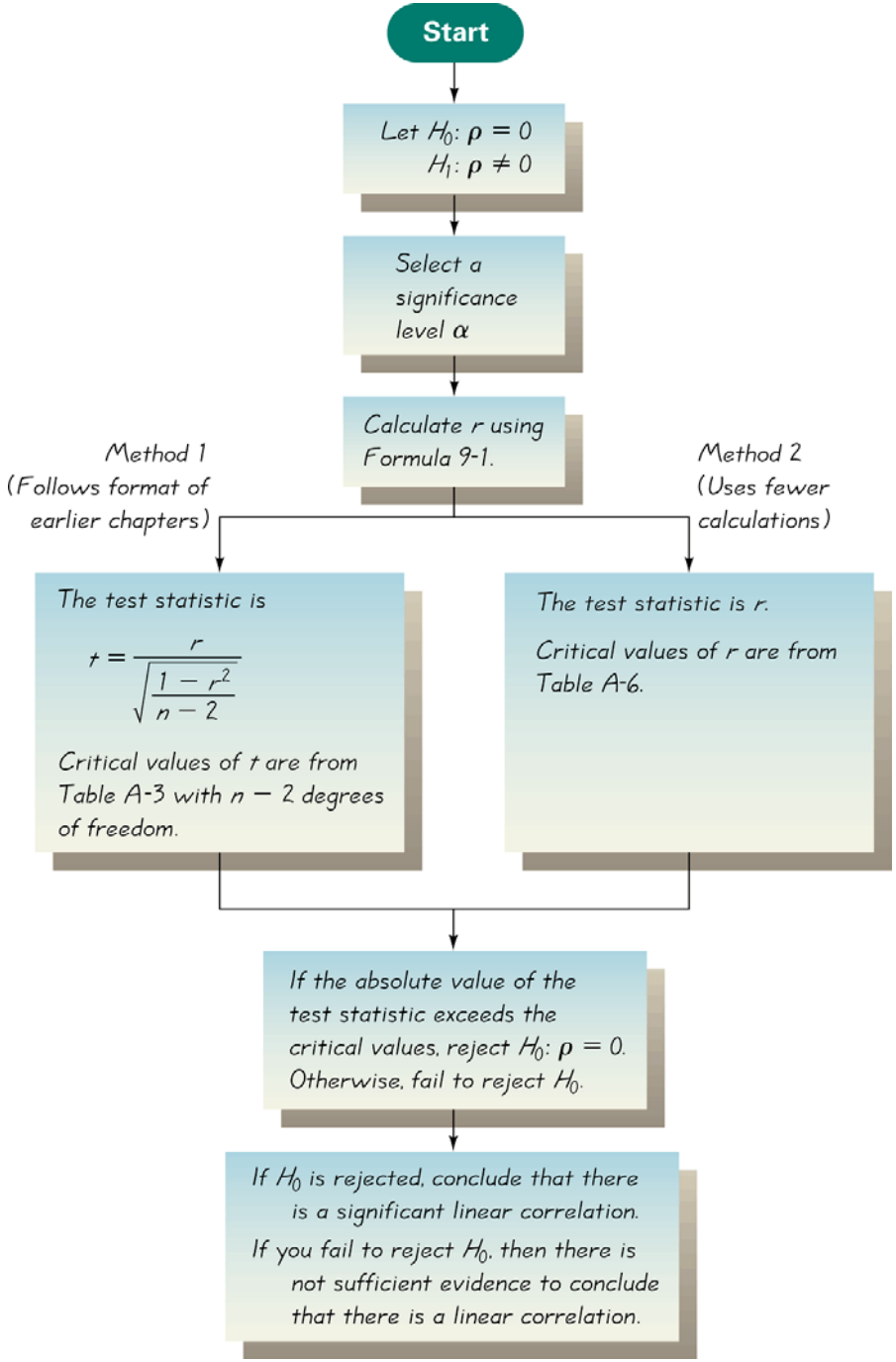
- We wish to determine whether there is a significant linear correlation between two variables.
- We present two methods.
- Both methods let H_0: $\rho = 0$

(no significant linear correlation)

H_1: $\rho \neq 0$

(significant linear correlation)

FIGURE 9-4 Testing for a Linear Correlation (p.505)



Method 1: Test Statistic is *t*

(follows format of earlier chapters)

Test statistic:

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

Critical values:

Use Table A-3 with degrees of freedom = $n - 2$

Method 2: Test Statistic is *r*

(uses fewer calculations)

- **Test statistic: *r***
- **Critical values: Refer to Table A-6**

(no degrees of freedom)

Example: Boats and Manatees

Using the boat/manatee data in Table 9-1, test the claim that there is a linear correlation between the number of registered boats and the number of manatee deaths from boats. Use Method 1.

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

$$t = \frac{0.922}{\sqrt{\frac{1 - 0.922^2}{10 - 2}}} = 6.735$$

Method 1: Test Statistic is *t*

(follows format of earlier chapters)

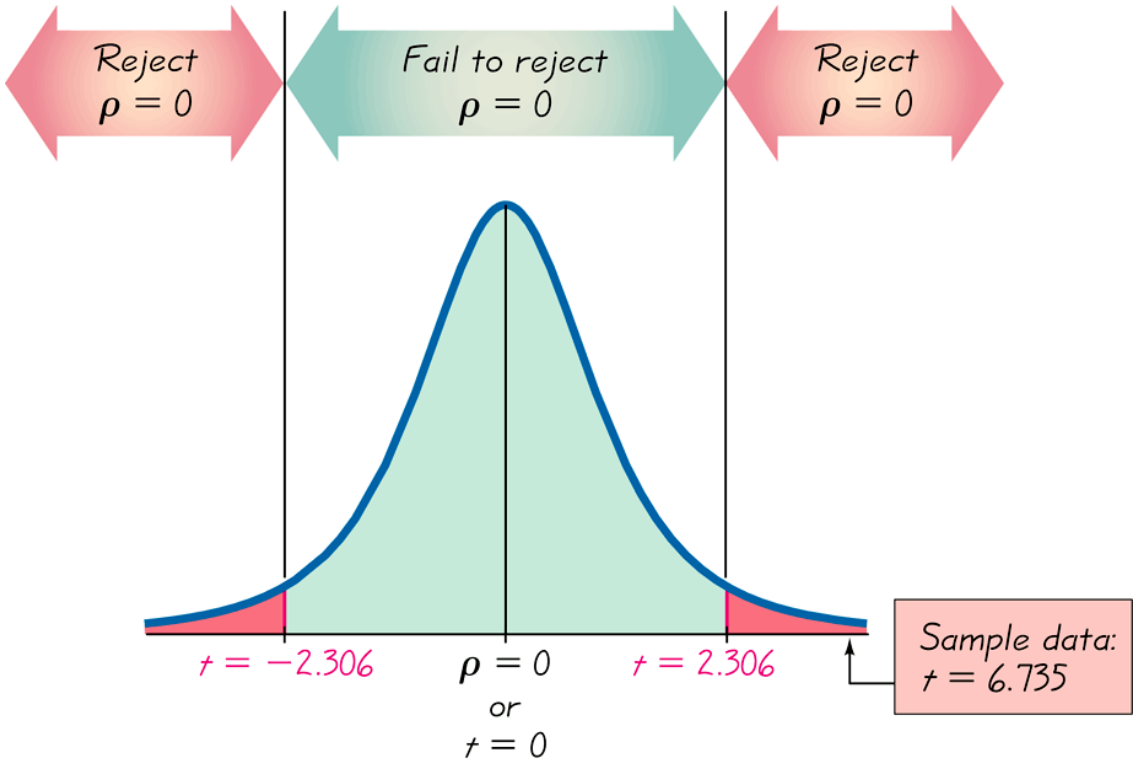
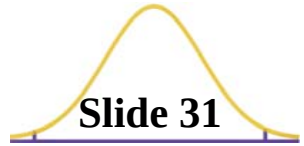


Figure 9-5 (p.516)

Example: Boats and Manatees



Using the boat/manatee data in Table 9-1, test the claim that there is a linear correlation between the number of registered boats and the number of manatee deaths from boats. Use Method 2.

The test statistic is $r = 0.922$. The critical values of $r = \pm 0.632$ are found in Table A-6 with $n = 10$ and $\alpha = 0.05$.

Method 2: Test Statistic is *r*

(uses fewer calculations)

- **Test statistic: *r***
- **Critical values: Refer to Table A-6**

(10 degrees of freedom)

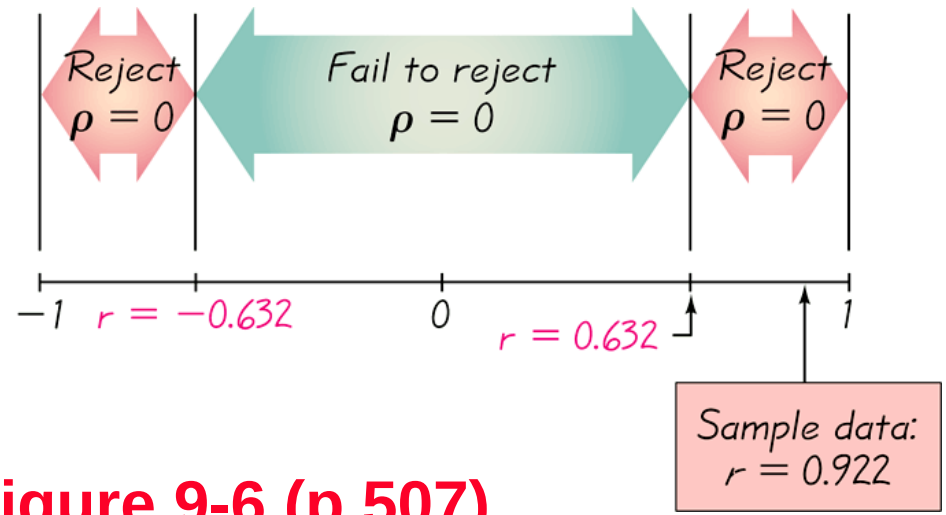


Figure 9-6 (p.507)

Example: Boats and Manatees

Using the boat/manatee data in Table 9-1, test the claim that there is a linear correlation between the number of registered boats and the number of manatee deaths from boats. Use both (a) Method 1 and (b) Method 2.

Using either of the two methods, we find that the absolute value of the test statistic does exceed the critical value (Method 1: 6.735 > 2.306. Method 2: 0.922 > 0.632); that is, the test statistic falls in the critical region.

We therefore reject the null hypothesis. There is sufficient evidence to support the claim of a linear correlation between the number of registered boats and the number of manatee deaths from boats.

Justification for *r* Formula

Formula 9-1 is developed from

$$r = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{(n-1) S_x S_y}$$

$(\bar{x}, \bar{y})$ **centroid of sample points**

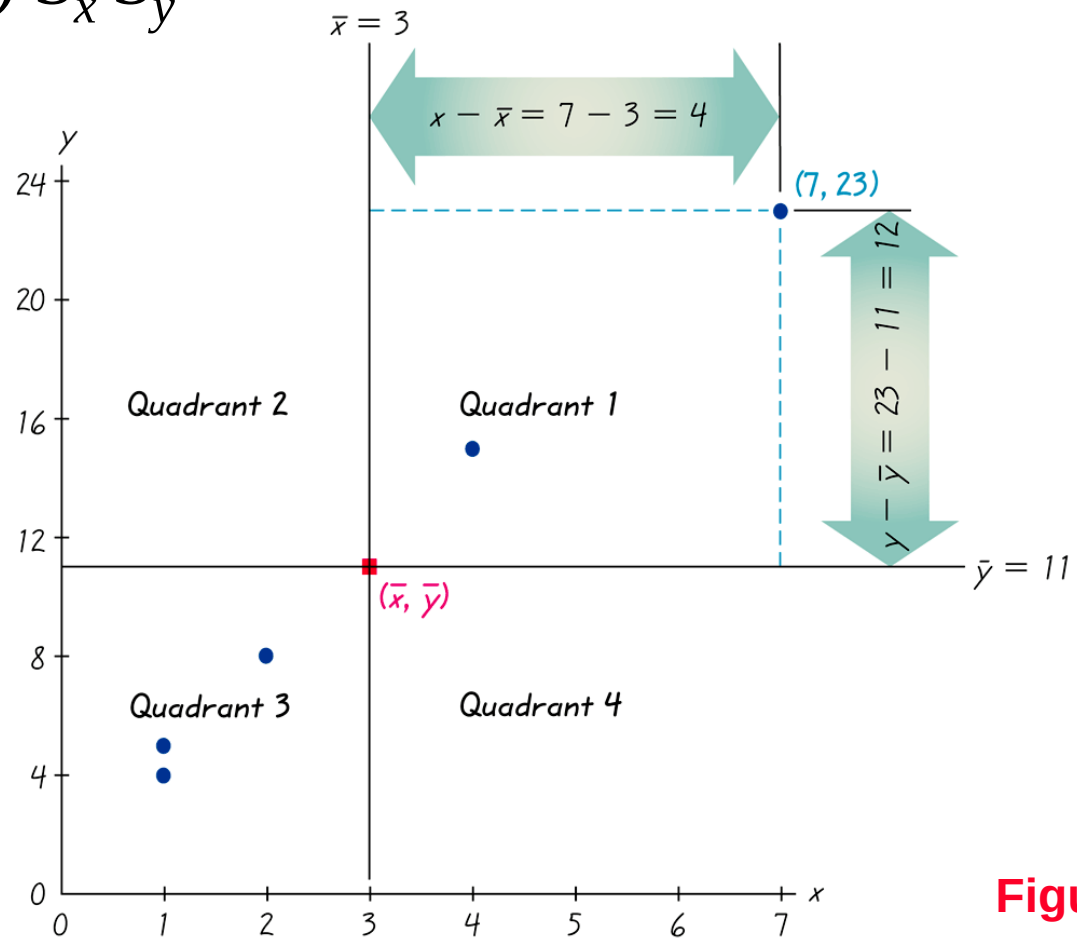
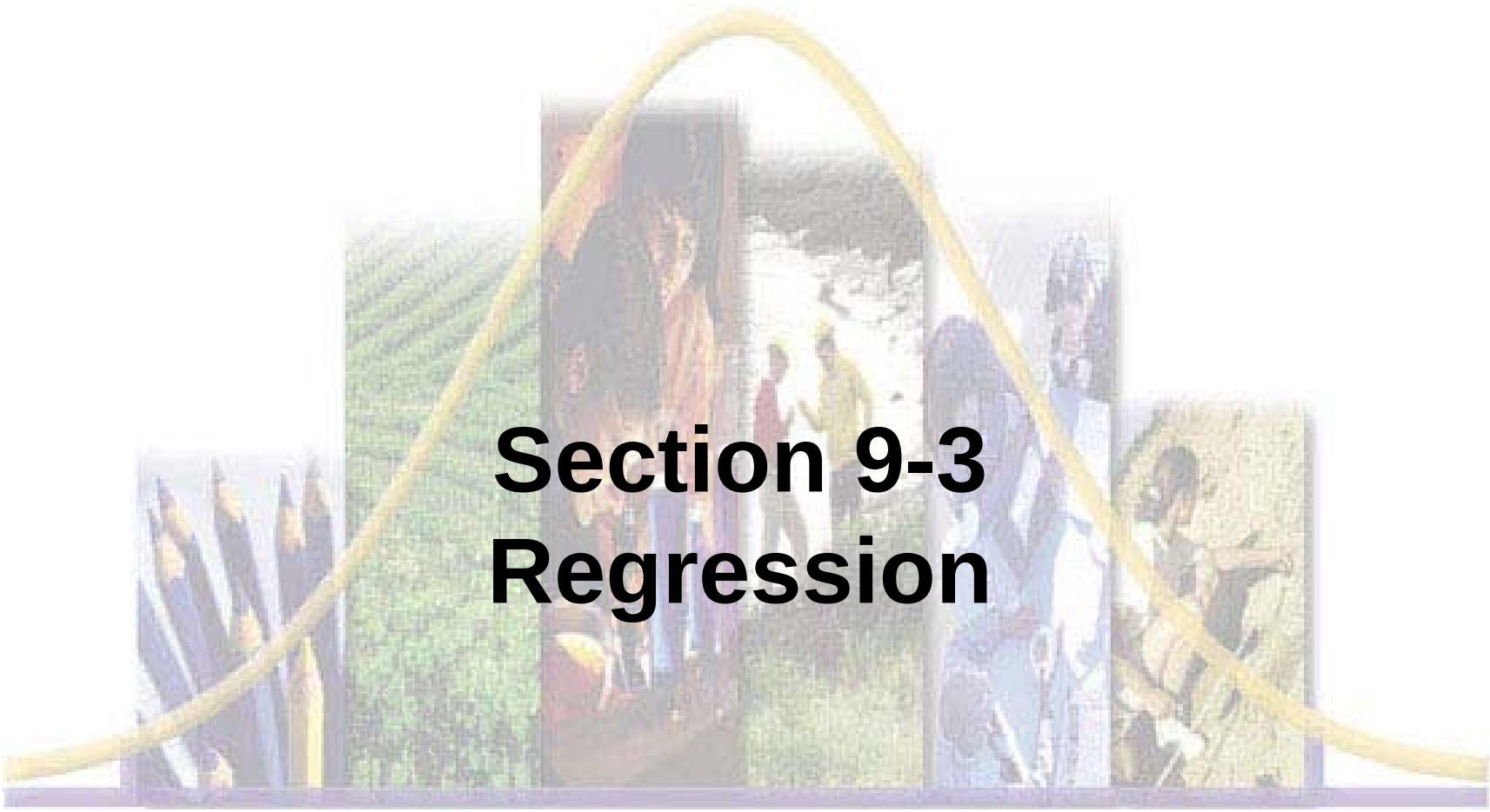


Figure 9-7



Section 9-3 Regression

Created by Erin Hodgess, Houston, Texas

Regression

Definition

- **Regression Equation**

  The regression equation expresses a relationship between x (called the independent variable, predictor variable or explanatory variable, and y (called the dependent variable or response variable.

  The typical equation of a straight line $y = mx + b$ is expressed in the form $\hat{y} = b_0 + b_1x$, where b_0 is the y-intercept and b_1 is the slope.

Assumptions

1. We are investigating only linear relationships.

2. For each x-value, y is a random variable having a normal (bell-shaped) distribution. All of these y distributions have the same variance. Also, for a given value of x, the distribution of y-values has a mean that lies on the regression line. (Results are not seriously affected if departures from normal distributions and equal variances are not too extreme.)

Regression

Definition

Regression Equation

Given a collection of paired data, the regression equation

$$\hat{y} = b_0 + b_1 x$$

algebraically describes the relationship between the two variables

Regression Line

The graph of the regression equation is called the regression line (or line of best fit, or least squares line).

Notation for Regression Equation

	Population Parameter	Sample Statistic
y-intercept of regression equation	β_0	b_0
Slope of regression equation	β_1	b_1
Equation of the regression line	$y = \beta_0 + \beta_1 x$	$\hat{y} = b_0 + b_1 x$

Formula for b_0 and b_1

Formula 9-2 $$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$ **(slope)**

Formula 9-3 $$b_0 = \bar{y} - b_1\bar{x}$$ **(*y*-intercept)**

calculators or computers can compute these values

If you find b_1 first, then

$$b_0 = \bar{y} - b_1\bar{x}$$

Formula 9-4

Can be used for Formula 9-2, where $\bar{y}$ is the mean of the y-values and $\bar{x}$ is the mean of the x values

The regression line fits the sample points best.

Rounding the y-intercept b_0 and the slope b_1

- Round to three significant digits.
- If you use the formulas 9-2 and 9-3, try not to round intermediate values. (see p.527)

Calculating the Regression Equation

Data

x	1	1	3	5
y	2	8	6	4

In Section 9-2, we used these values to find that the linear correlation coefficient of $r = -0.135$. Use this sample to find the regression equation.

Calculating the Regression Equation

Data

x	1	1	3	5
y	2	8	6	4

$n = 4$
$\Sigma x = 10$
$\Sigma y = 20$
$\Sigma x^2 = 36$
$\Sigma y^2 = 120$
$\Sigma xy = 48$

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b_1 = \frac{4(48) - (10)(20)}{4(36) - (10)^2}$$

$$b_1 = \frac{-8}{44} = -0.181818$$

Calculating the Regression Equation

Data

x	1	1	3	5
y	2	8	6	4

$n = 4$

$\Sigma x = 10$

$\Sigma y = 20$

$\Sigma x^2 = 36$

$\Sigma y^2 = 120$

$\Sigma xy = 48$

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$5 - (-0.181818)(2.5) = 5.45$$

Calculating the Regression Equation

Data

x	1	1	3	5
y	2	8	6	4

$n = 4$
$\Sigma x = 10$
$\Sigma y = 20$
$\Sigma x^2 = 36$
$\Sigma y^2 = 120$
$\Sigma xy = 48$

The estimated equation of the regression line is:

$$\hat{y} = 5.45 - 0.182x$$

Example: Boats and Manatees

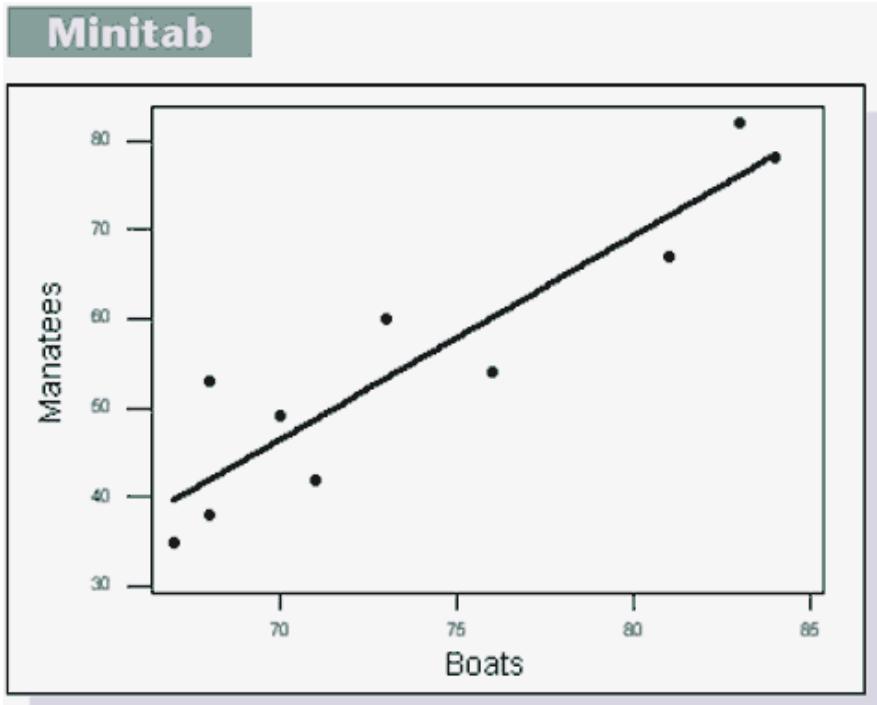
Given the sample data in Table 9-1, find the regression equation. (from pp.507-508)

Using the same procedure as in the previous example, we find that $b_1 = 2.27$ and $b_0 = -113$. Hence, the estimated regression equation is:

$$\hat{y} = -113 + 2.27x$$

Example: Boats and Manatees

Given the sample data in Table 9-1, find the regression equation.



Predictions

In predicting a value of y based on some given value of x ...

1. **If there is not a significant linear correlation, the best predicted y-value is $\overline{y}$.**

2. **If there is a significant linear correlation, the best predicted y-value is found by substituting the x-value into the regression equation.**

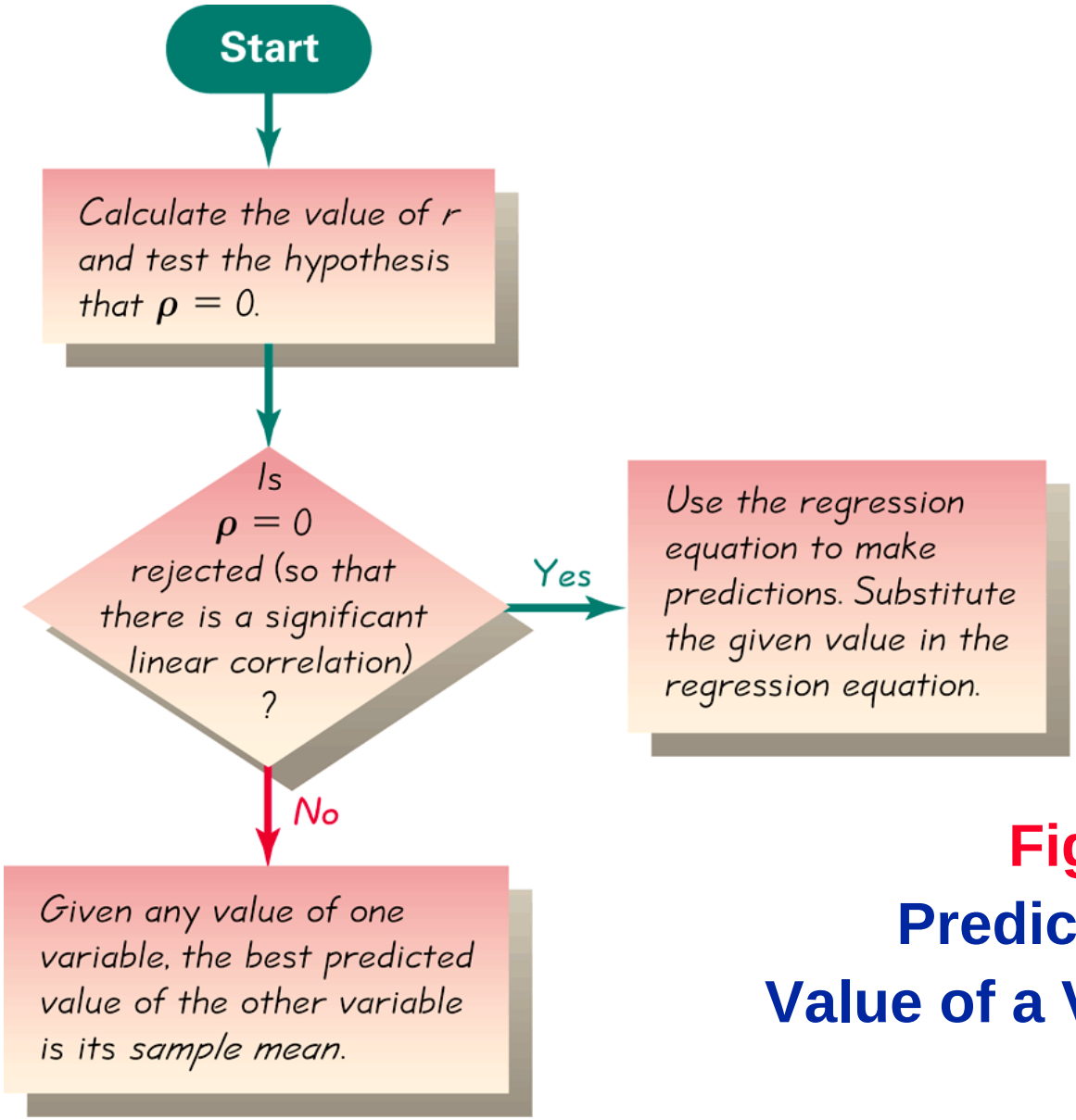


Figure 9-8
Predicting the Value of a Variable (p.522)

Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. Assume that in 2001 there were 850,000 registered boats. Because Table 9-1 lists the numbers of registered boats in tens of thousands, this means that for 2001 we have $x = 85$. Given that $x = 85$, find the best predicted value of y, the number of manatee deaths from boats.

Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. Given that $x = 85$, find the best predicted value of y, the number of manatee deaths from boats.

We must consider whether there is a linear correlation that justifies the use of that equation. We do have a significant linear correlation (with $r = 0.922$).

Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. Given that $x = 85$, find the best predicted value of y, the number of manatee deaths from boats.

$$\hat{y} = -113 + 2.27x$$

$$-113 + 2.27(85) = 80.0$$

The predicted number of manatee deaths is 80.0. The actual number of manatee deaths in 2001 was 82, so the predicted value of 80.0 is quite close.

Guidelines for Using The Regression Equation (p.523)

1. If there is no significant linear correlation, don't use the regression equation to make predictions.
2. When using the regression equation for predictions, stay within the scope of the available sample data.
3. A regression equation based on old data is not necessarily valid now.
4. Don't make predictions about a population that is different from the population from which the sample data was drawn.

Definitions

- **Marginal Change:** **The marginal change is the amount that a variable changes when the other variable changes by exactly one unit.**
- **Outlier:** **An outlier is a point lying far away from the other data points.**
- **Influential Points:** **An influential point strongly affects the graph of the regression line.**

Residuals and the Least-Squares Property

Definitions (p.525)

Residual

for a sample of paired (x, y) data, the difference ($y - \hat{y}$) between an observed sample y-value and the value of $\hat{y}$, which is the value of y that is predicted by using the regression equation.

Least-Squares Property

A straight line satisfies this property if the sum of the squares of the residuals is the smallest sum possible.

Residuals and the Least-Squares Property

x	1	2	4	5
y	4	24	8	32

$\hat{y} = 5 + 4x$

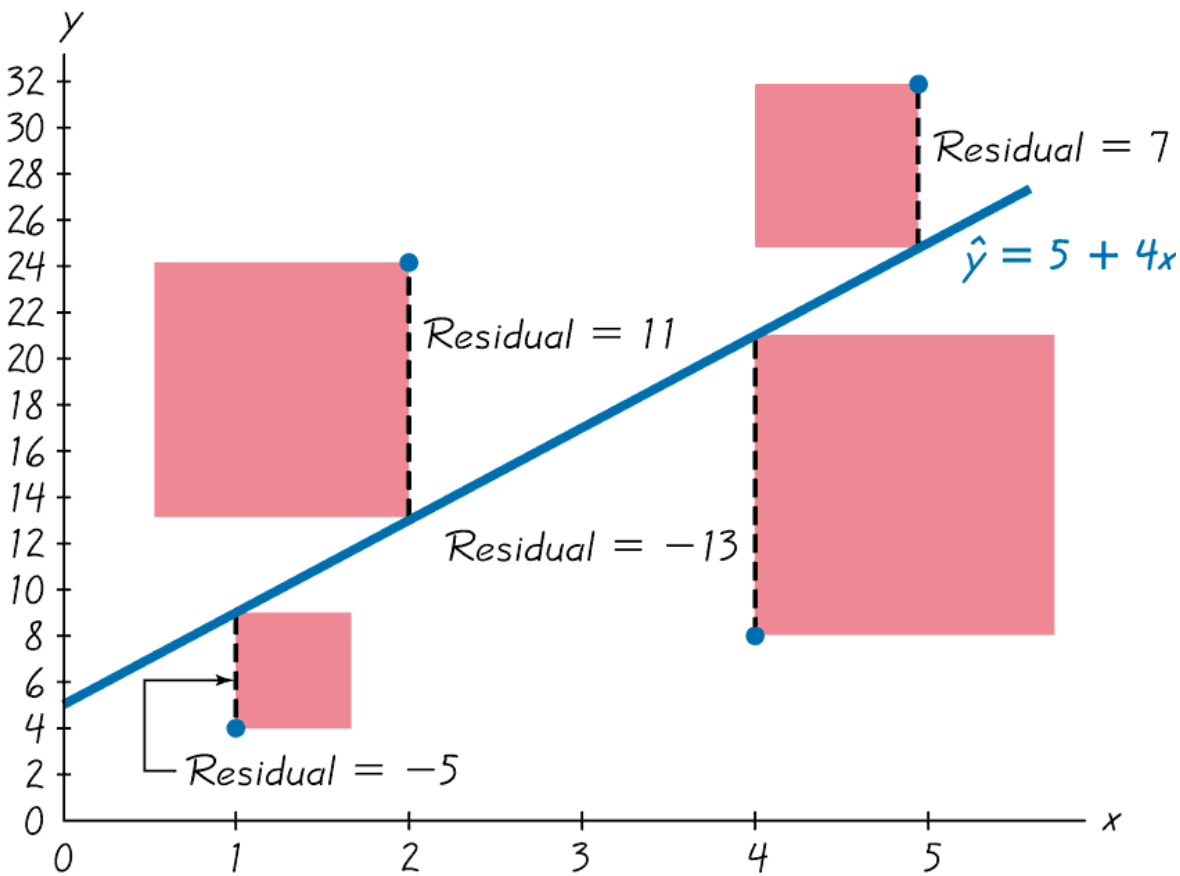
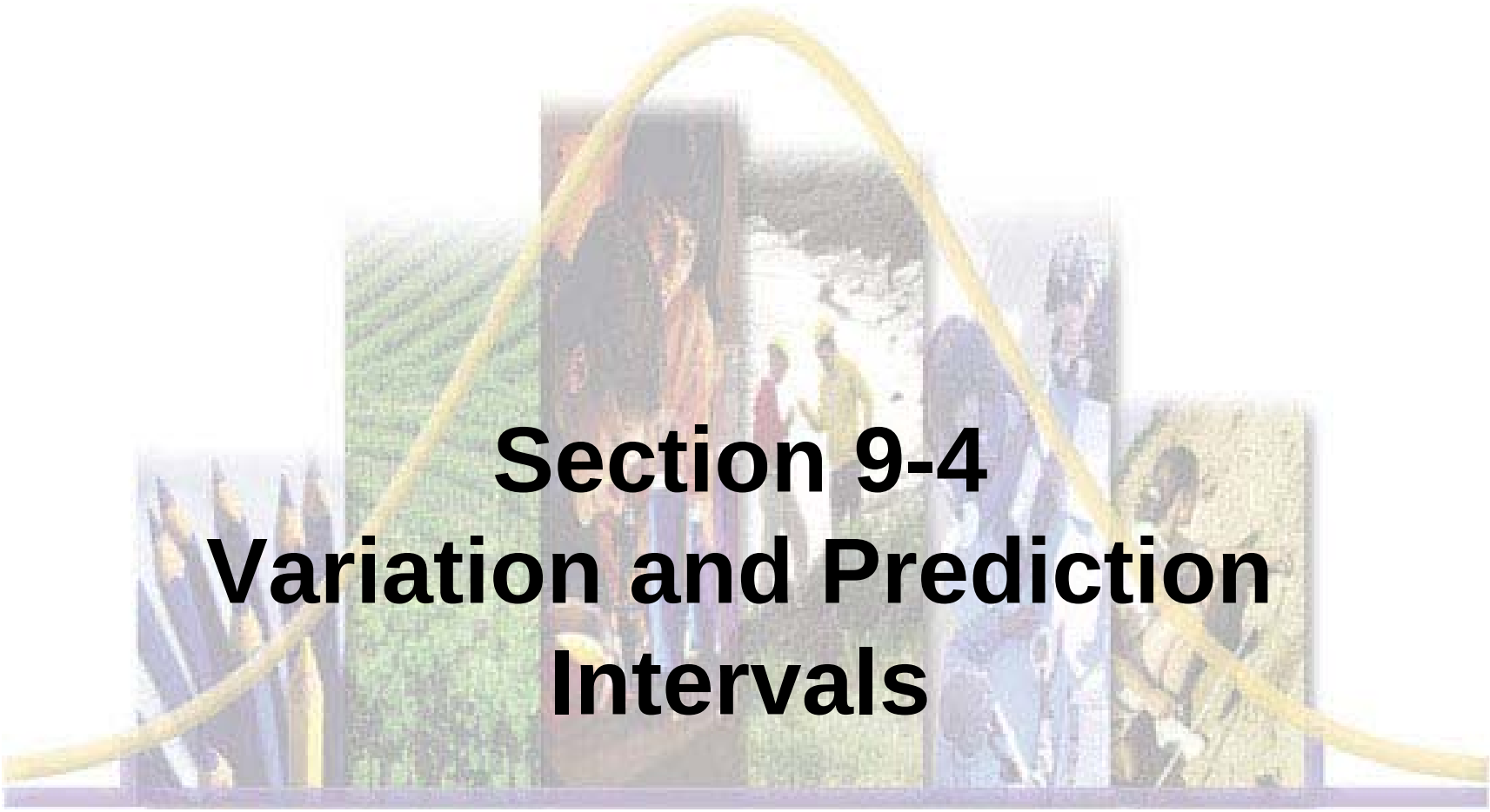


Figure 9-9 (p.525)



Section 9-4
Variation and Prediction Intervals

Created by Erin Hodgess, Houston, Texas

Definitions

We consider different types of variation that can be used for two major applications:

1. To determine the proportion of the variation in y that can be explained by the linear relationship between x and y.

2. To construct interval estimates of predicted y-values. Such intervals are called prediction intervals.

Definitions

Total Deviation The **total deviation** from the mean of the particular point (x, y) is the vertical distance $y - \overline{y}$, which is the distance between the point (x, y) and the horizontal line passing through the sample mean $\overline{y}$.

Explained Deviation is the vertical distance $\hat{y} - \overline{y}$, which is the distance between the predicted y-value and the horizontal line passing through the sample mean $\overline{y}$.

Definitions

Unexplained Deviation is the vertical distance $y - \hat{y}$, which is the vertical distance between the point (x, y) and the regression line. (The distance $y - \hat{y}$ is also called a *residual*, as defined in Section 9-3.).

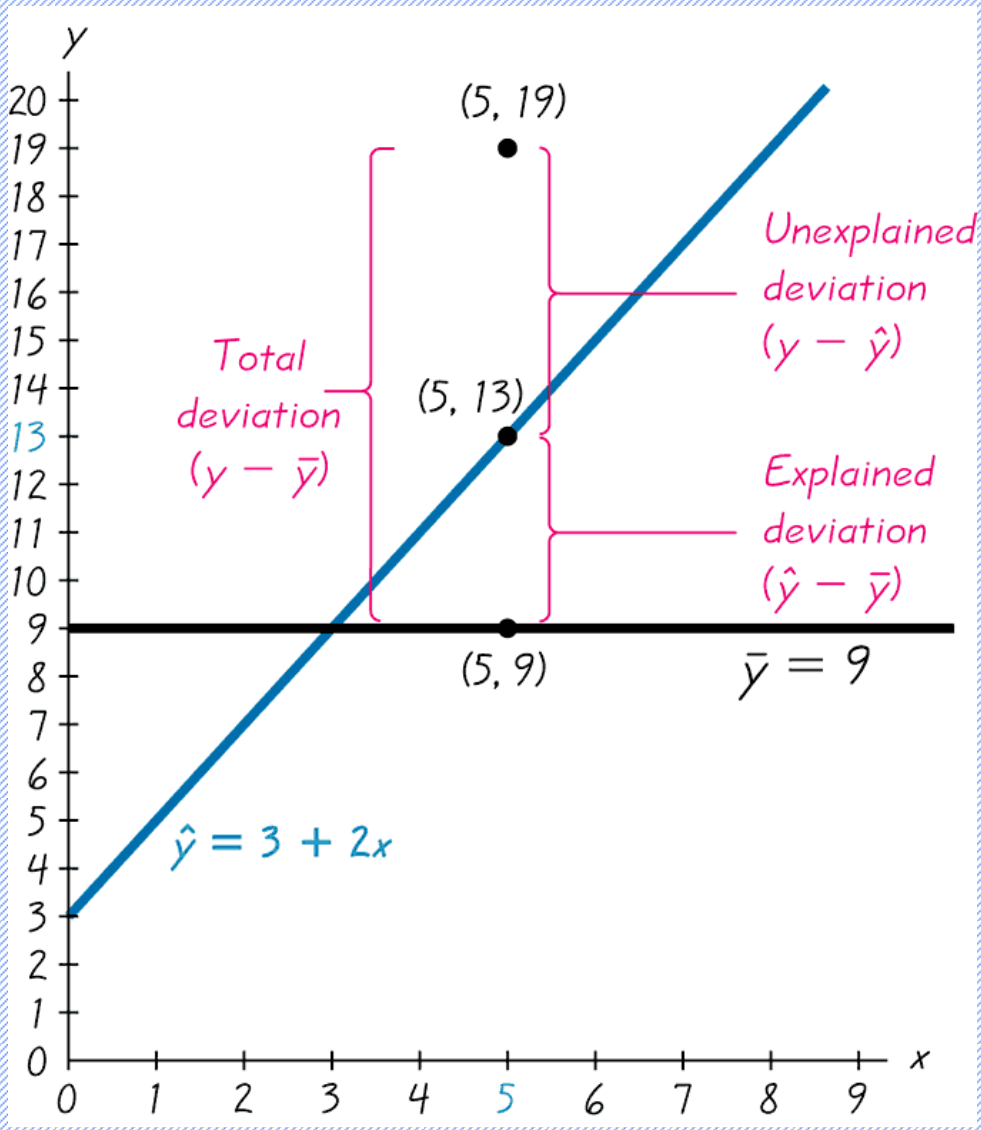


Figure 9-10 Unexplained, Explained, and Total Deviation

(total deviation) = (explained deviation) + (unexplained deviation)

$$(y - \bar{y}) = (\hat{y} - \bar{y}) + (y - \hat{y})$$

(total variation) = (explained variation) + (unexplained variation)

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

Formula 9-4

Definition

Coefficient of determination

the amount of the variation in y that is explained by the regression line

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

or

simply square *r*

(determined by Formula 9-1, section 9-2)

Prediction Intervals

Definition

The standard error of estimate is a measure of the differences (or distances) between the observed sample y values and the predicted values $\hat{y}$ that are obtained using the regression equation.

Standard Error of Estimate

$$s_e = \sqrt{\frac{\Sigma (y - \hat{y})^2}{n - 2}}$$

or

$$s_e = \sqrt{\frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n - 2}}$$

Formula 9-5

Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. Find the standard error of estimate s_e for the boat/manatee data.

$n = 10$
$\Sigma y^2 = 33456$
$\Sigma y = 558$
$\Sigma xy = 42214$
$b_0 = -112.70989$
$b_1 = 2.27408$

$$s_e = \sqrt{\frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n - 2}}$$

$$s_e = \sqrt{\frac{33456 - (-112.70989)(558) - (2.27408)(42414)}{10 - 2}}$$

Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. Find the standard error of estimate s_e for the boat/manatee data.

$n = 10$

$\Sigma y^2 = 33456$

$\Sigma y = 558$

$\Sigma xy = 42214$

$b_0 = -112.70989$

$b_1 = 2.27408$

$s_e = 6.61234 = 6.61$

Prediction Interval for an Individual y

$$\hat{y} - E < y < \hat{y} + E$$

where

$$E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$$

x_0 represents the given value of x

$t_{\alpha/2}$ has $n - 2$ degrees of freedom

Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. We have also found that when $x = 85$, the predicted number of manatee deaths is 80.0. Construct a 95% prediction interval given that $x = 85$.

$$E = t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$$

$$E = (2.306)(6.6123) \sqrt{1 + \frac{1}{10} + \frac{10(85-74)^2}{10(55289) - (741)^2}}$$

$$E = 18.1$$

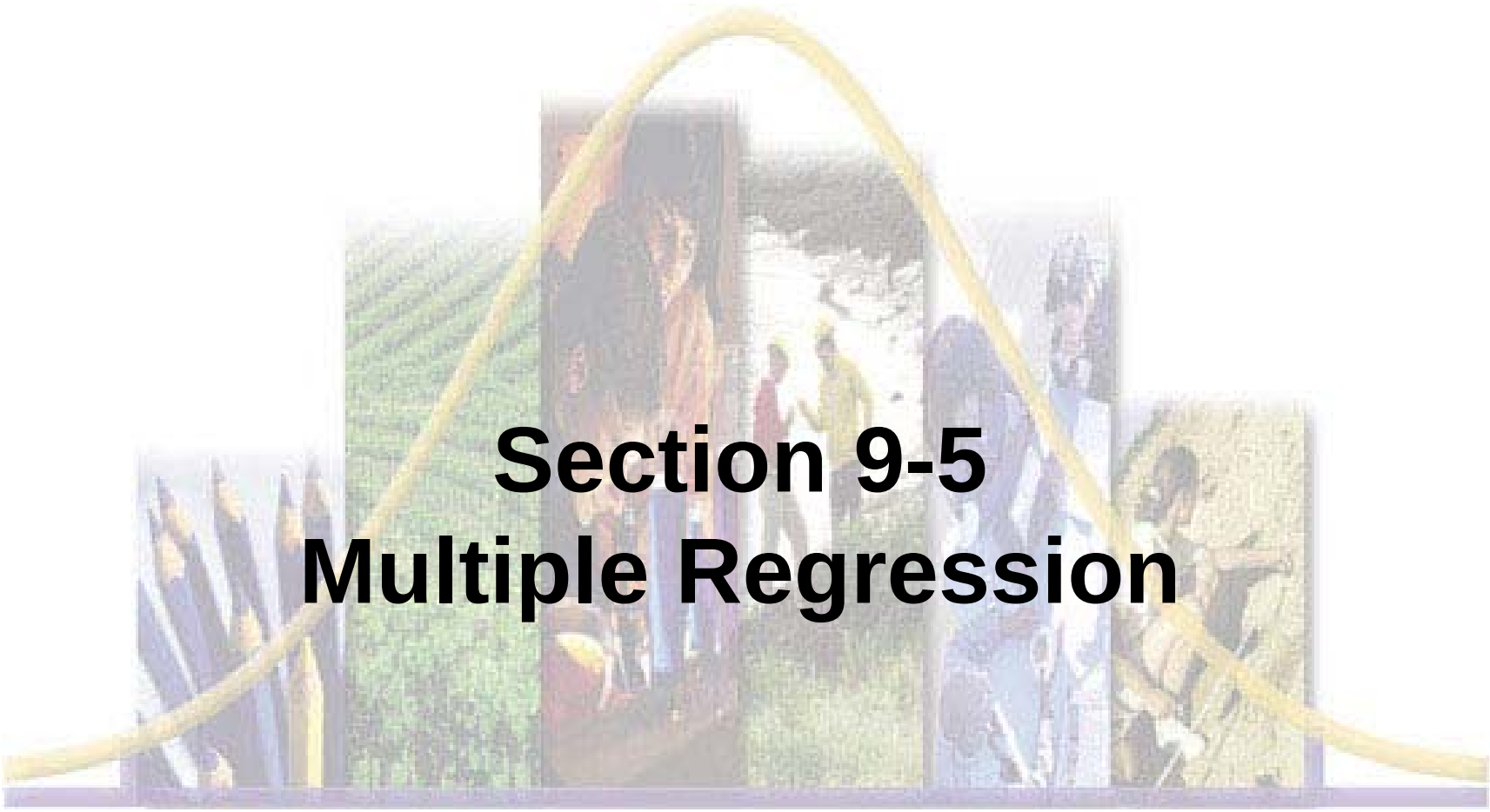
Example: Boats and Manatees

Given the sample data in Table 9-1, we found that the regression equation is $\hat{y} = -113 + 2.27x$. We have also found that when $x = 85$, the predicted number of manatee deaths is 80.0. Construct a 95% prediction interval given that $x = 85$.

$$\hat{y} - E < y < \hat{y} + E$$

$$80.6 - 18.1 < y < 80.6 + 18.1$$

$$62.5 < y < 98.7$$



Section 9-5
Multiple Regression

Created by Erin Hodgess, Houston, Texas

Multiple Regression

Definition

Multiple Regression Equation

A linear relationship between a dependent variable y and two or more independent variables ($x_1, x_2, x_3 \ldots, x_k$)

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k$$

Notation

$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_k x_k$

(General form of the estimated multiple regression equation)

n = **sample size**

k = **number of independent variables**

$\hat{y}$ = **predicted value of the dependent variable** y

$x_1, x_2, x_3 \ldots, x_k$ **are the independent variables**

Notation

β_0 = the y-intercept, or the value of y when all of the predictor variables are 0

b_0 = estimate of β_0 based on the sample data

$\beta_1, \beta_2, \beta_3 \ldots, \beta_k$ are the coefficients of the independent variables $x_1, x_2, x_3 \ldots, x_k$

$b_1, b_2, b_3 \ldots, b_k$ are the sample estimates of the coefficients $\beta_1, \beta_2, \beta_3 \ldots, \beta_k$

Assumption

Use a statistical software package such as

- **STATDISK**
- **Minitab**
- **Excel**

Example: Bears

For reasons of safety, a study of bears involved the collection of various measurements that were taken after the bears were anesthetized. Using the data in Table 9-3, find the multiple regression equation in which the dependent variable is weight and the independent variables are head length and total overall length.

Example: Bears

Table 9-3 Data from Anesthetized Male Bears

Variable	Minitab Column	Name	Sample Data							
y	C1	WEIGHT	80	344	416	348	262	360	332	34
x_2	C2	AGE	19	55	81	115	56	51	68	8
x_3	C3	HEADLEN	11.0	16.5	15.5	17.0	15.0	13.5	16.0	9.0
x_4	C4	HEADWDTH	5.5	9.0	8.0	10.0	7.5	8.0	9.0	4.5
x_5	C5	NECK	16.0	28.0	31.0	31.5	26.5	27.0	29.0	13.0
x_6	C6	LENGTH	53.0	67.5	72.0	72.0	73.5	68.5	73.0	37.0
x_7	C7	CHEST	26	45	54	49	41	49	44	19

Example: Bears

Minitab

The regression equation is

WEIGHT = −374 + 18.8 HEADLEN + 5.87 LENGTH ← ① Multiple regression equation

Predictor	Coef	Stdev	t-ratio	p
Constant	−374.3	134.1	−2.79	0.038
HEADLEN	18.82	23.15	0.81	0.453
LENGTH	5.875	5.065	1.16	0.299

s = 68.56 R-sq = 82.8% R-sq(adj) = 75.9%

↖ $R^2 = 0.828$ ↖ ② Adjusted $R^2 = 0.759$

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	2	113142	56571	12.03	0.012
Error	5	23506	4701		↑
Total	7	136648			

③ Overall significance of multiple regression equation

Example: Bears

The regression equation is:

$$\text{WEIGHT} = -374 + 18.8\,\text{HEADLEN} + 5.87\,\text{LENGTH}$$

$$y = -374 + 18.8x_3 + 5.87x_6$$

Adjusted R^2

Definitions

- The **multiple coefficient of determination** is a measure of how well the multiple regression equation fits the sample data.
- The **Adjusted coefficient of determination** R^2 is modified to account for the number of variables and the sample size.

Adjusted R^2

$$\text{Adjusted } R^2 = 1 - \frac{(n-1)}{[n-(k+1)]}(1-R^2)$$

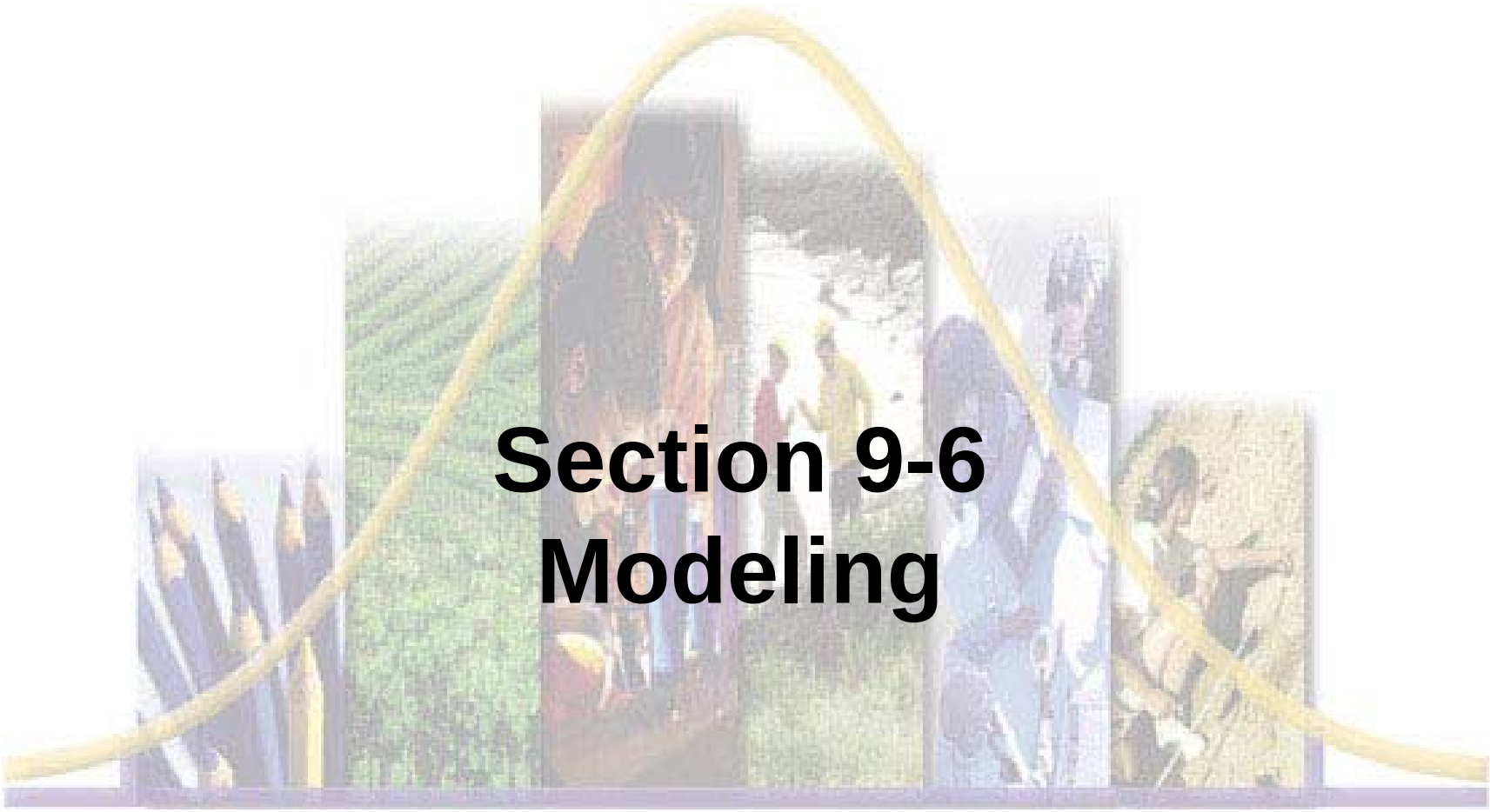
Formula 9-6

where n = sample size

k = number of independent (x) variables

Finding the Best Multiple Regression Equation

1. Use common sense and practical considerations to include or exclude variables.
2. Instead of including almost every available variable, include relatively few independent (x) variables, weeding out independent variables that don't have an effect on the dependent variable.
3. Select an equation having a value of adjusted R^2 with this property: If an additional independent variable is included, the value of adjusted R^2 does not increase by a substantial amount.
4. For a given number of independent (x) variables, select the equation with the largest value of adjusted R^2.
5. Select an equation having overall significance, as determined by the P-value in the computer display.



Section 9-6 Modeling

Created by Erin Hodgess, Houston, Texas

Definition

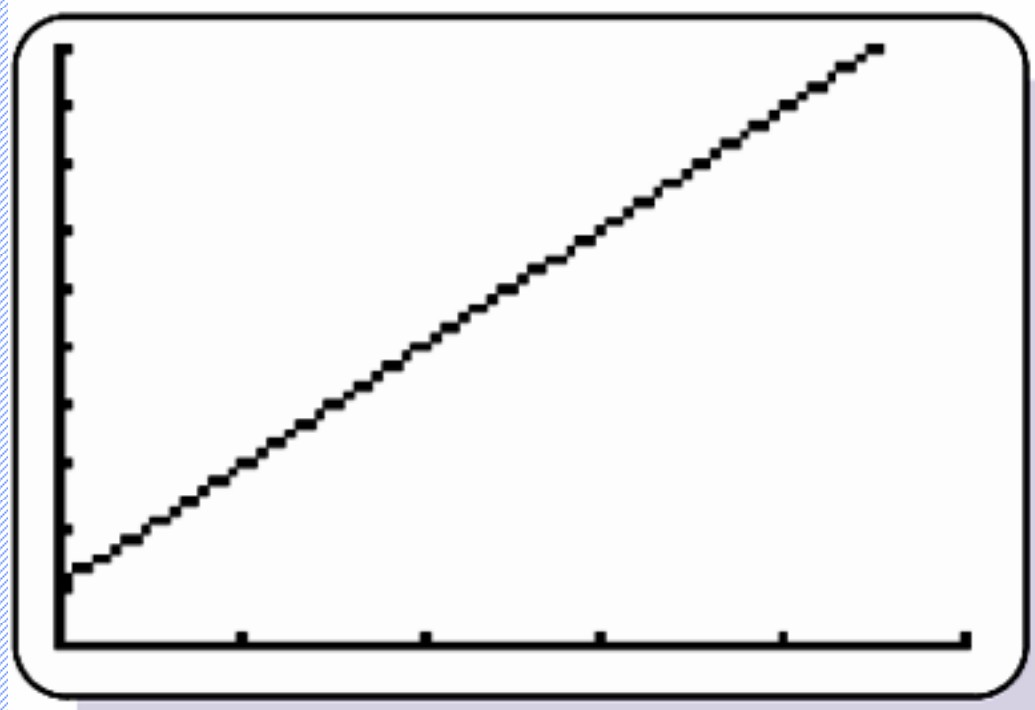
Mathematical Model

A mathematical model is a mathematical function that 'fits' or describes real-world data.

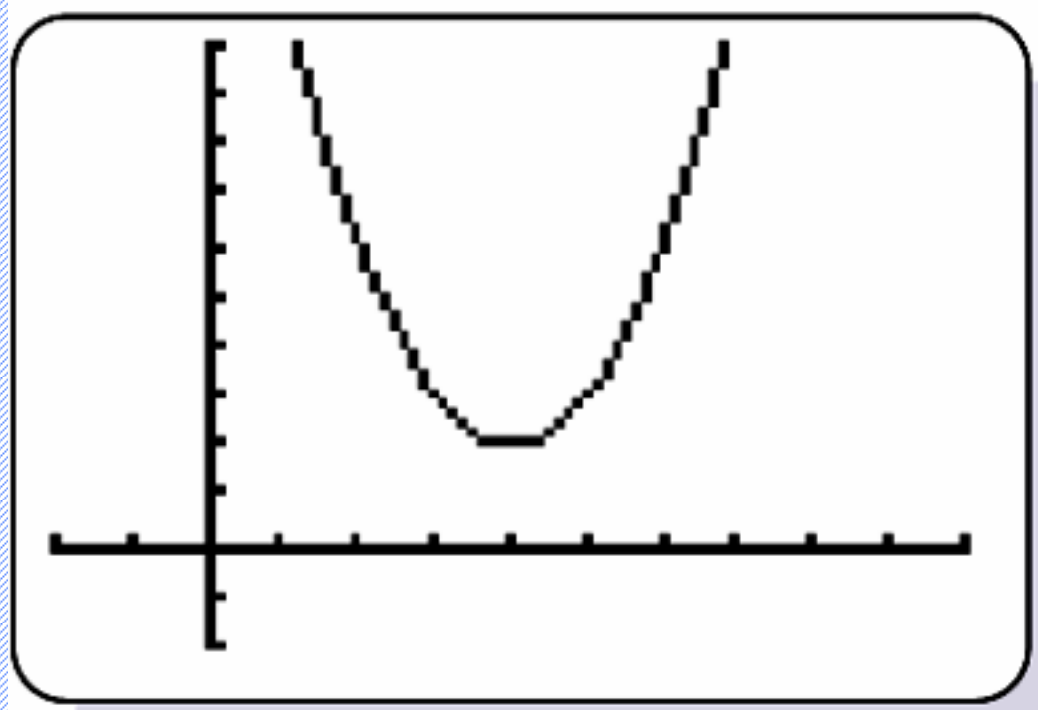
TI-83 Generic Models

- **Linear:** $y = a + bx$
- **Quadratic:** $y = ax^2 + bx + c$
- **Logarithmic:** $y = a + b \ln x$
- **Exponential:** $y = ab^x$
- **Power:** $y = ax^b$
- **Logistic:** $y = \dfrac{c}{1 + ae^{-bx}}$

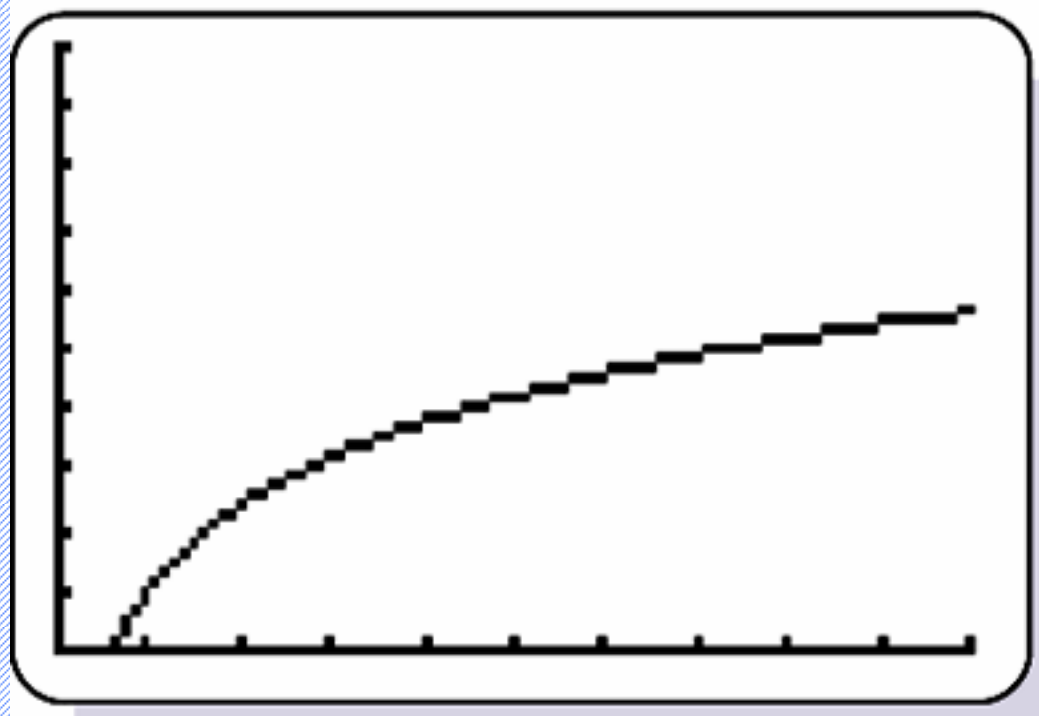
Linear: $y = 1 + 2x$



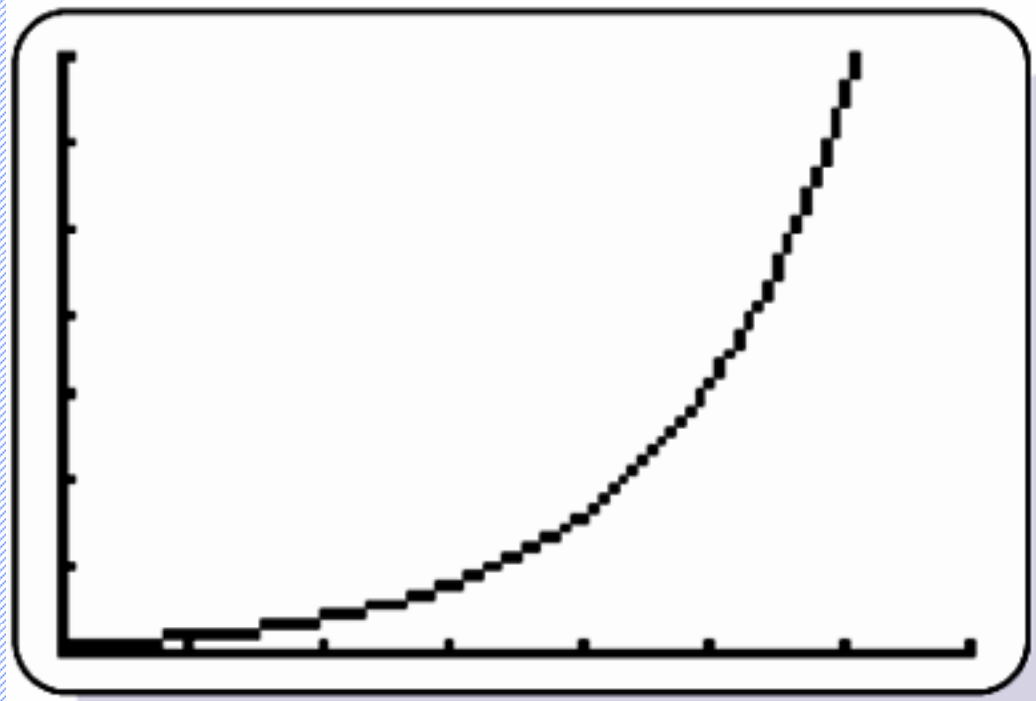
Quadratic: $y = 2x^2 - 8x + 9$



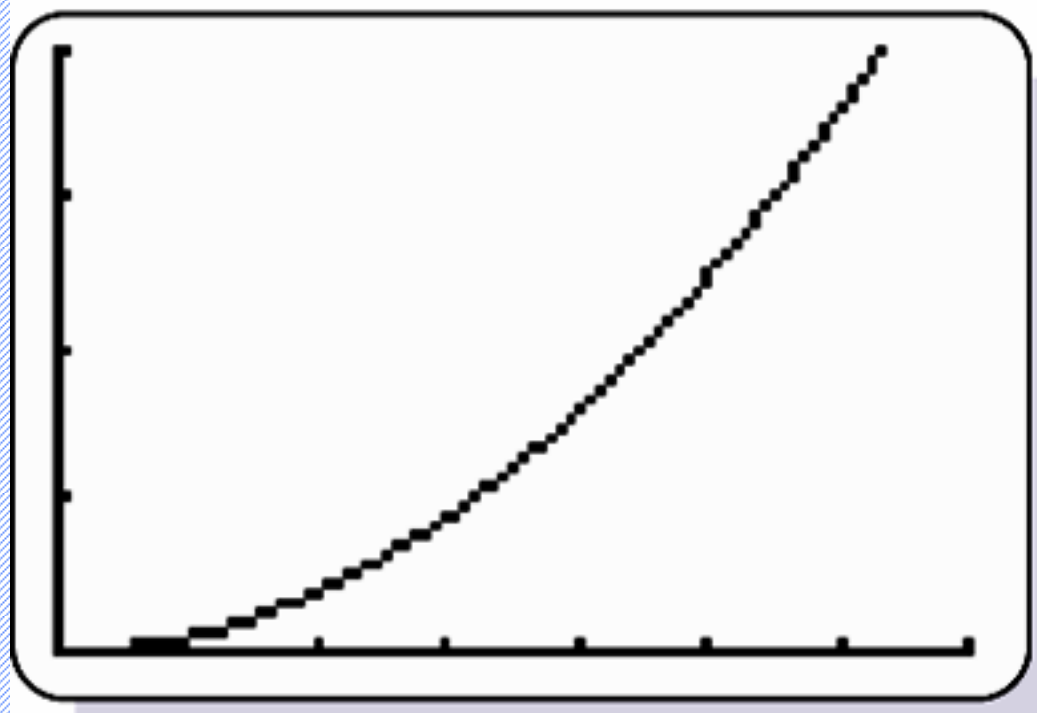
Logarithmic: $y = 1 + 2\ln x$



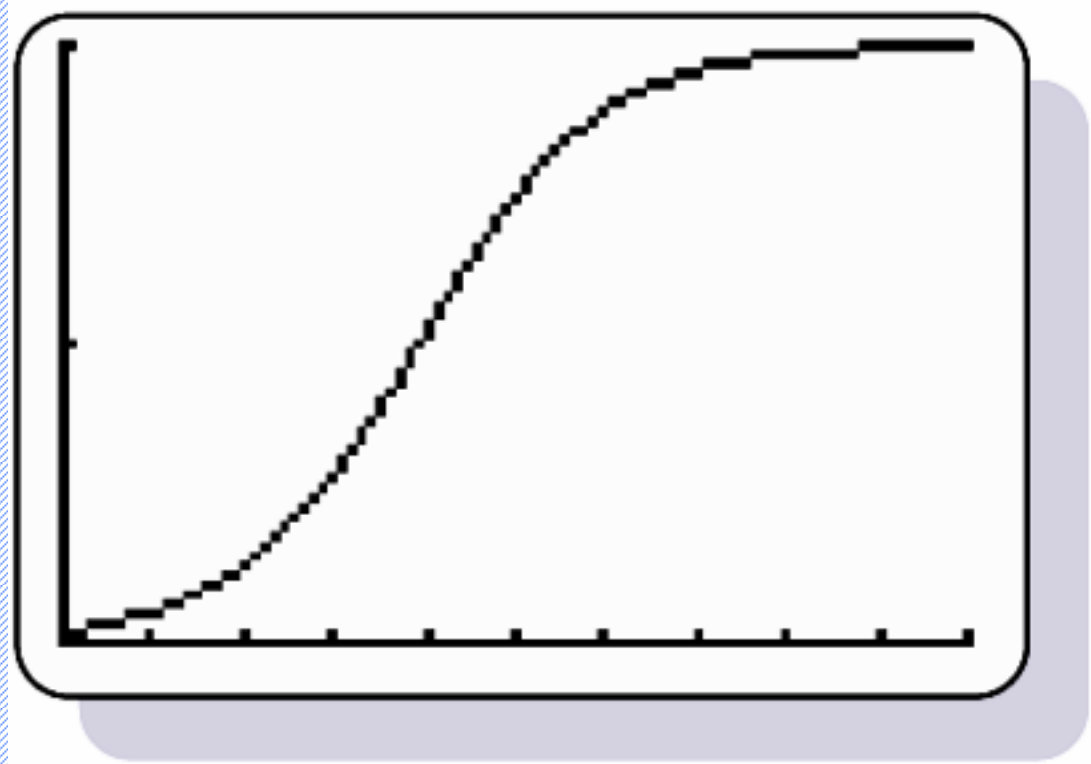
Exponential: $y = 2^x$



Power: $y = x^2$



Logistic: $y = \dfrac{2}{1 + 50e^{-x}}$



Development of a Good Mathematics Model

- **Look for a Pattern in the Graph:** Examine the graph of the plotted points and compare the basic pattern to the known generic graphs.
- **Find and Compare Values of R^2:** Select functions that result in larger values of R^2, because such larger values correspond to functions that better fit the observed points.
- **Think:** Use common sense. Don't use a model that lead to predicted values known to be totally unrealistic.