Section 10-2

Regression

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 $\hat{y} = b_0 + b_1 x$, where b_0 is the y-intercept and b_1 is the slope.

REGRESSION EQUATION AND REGRESSION LINE

Given a collection of paired data, the **regression equation**

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

algebraically describes the ***relationship*** between the two variables.

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

ASSUMPTIONS

1. We are investigating only ***linear*** relationships.
2. The sample of paired data (x, y) is a random sample.
3. Visual examination of the scatterplot must confirm that the points approximate a straight-line pattern.
4. Any outliers must be removed if they are known to be errors. The effects of any outliers should be considered by calculating r with and without the outliers included.

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 regression line	$y = \beta_0 + \beta_1 x$	$\hat{y} = b_0 + b_1 x$

FORMULA FOR b_0 AND b_1

Slope: $b_1 = r \cdot \dfrac{s_y}{s_x}$

or $b_1 = \dfrac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

***y*-intercept**: $b_0 = \bar{y} - b_1 \bar{x}$

CALCULATING b_0 AND b_1 ON THE TI-83/84 CACLUATOR

The TI-83/84 calculators can compute these values. These values are given when we perform a **LinRegTTest**. The *y*-intercept (b_0) is given by the calculator as **a**; the slope (b_1) is given by the calculator as **b**.

FINDING THE REGRESSION EQUATION ON THE TI-83/84

1. Press **STAT** and arrow over to **CALC**.
2. Select **8:LinReg(a+bx)**.
3. Then enter **L1** and **L2** (or which ever lists you have your *x*- and *y*-values stored in).
4. Press **L1,L2**.
5. If you want to store your equation so that you can graph it, do Step 6. Otherwise, skip to Step 7.
6. Press „ **VARS**, arrow to **Y-VARS**, select **1:Function…**, and select **1:Y1**.
7. Press **ENTER**.

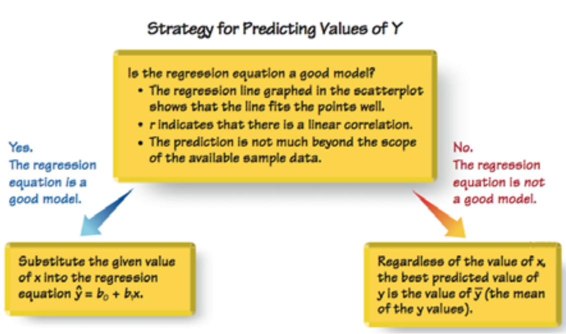
The regression line fits the sample points the best.

ROUND b_0 AND b_1

- Round to three significant digits.
- If you use the formulas, try not to round intermediate values.

USING THE REGRESSION EQUATION FOR PREDICTIONS

1. Use the regression equation for predictions only if the graph of the regression line on the scatter plot confirms that the regression line fits the points reasonable well.
2. Use the regression equation for predictions only if the linear correlation coefficient r indicates that there is a linear correlation between the data (as in Sect 10-2).
3. Use the regression equation for predictions only if the data do not go much beyond the scope of the available sample data.
4. If the regression equation does not appear to be useful for making predictions, the best predicted value of a variable is its sample mean.



FINDING A PREDICTION USING THE TI-83/84

Assuming there is a significant linear correlation and that you have stored the regression equation as Y1 in your calculator, you can find the prediction using your calculator as follows:

1. Go to **VARS** and arrow over to **Y-VARS**.
2. Select **1:Function**.
3. Select **1:Y1**.
4. Press **(**.
5. Enter the x-value for which you want the prediction.
6. Press **)**.
7. Press **Enter**.

DEFINITIONS

- **Marginal Change**: In working with two variables related by a linear regression equation, the marginal change in a variable is the amount that it changes when the other variable changes by exactly one unit. The slope b_1 in the regression equation represents the marginal change in y that occurs when the x changes by one unit.
- **Outlier**: An outlier is a point lying far away from the other data points.
- **Influential Points**: Paired sample data may include one or more influential points, which are points strongly affect the graph of the regression line.

RESIDUALS AND LEAST-SQUARES PROPERTY

- For a sample of paired (x, y) data, the **residual** is 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.
- A straight line satisfies the **least-squares 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

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