Chapter 12
Simple Linear Regression

- Simple Linear Regression Model
- Least Squares Method
- Coefficient of Determination
- Model Assumptions
- Testing for Significance

Simple Linear Regression Model

- The equation that describes how *y* is related to *x* and an error term is called the regression model.
- The simple linear regression model is:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where:

β_0 and β_1 are called parameters of the model,

ε is a random variable called the error term.

Simple Linear Regression Equation

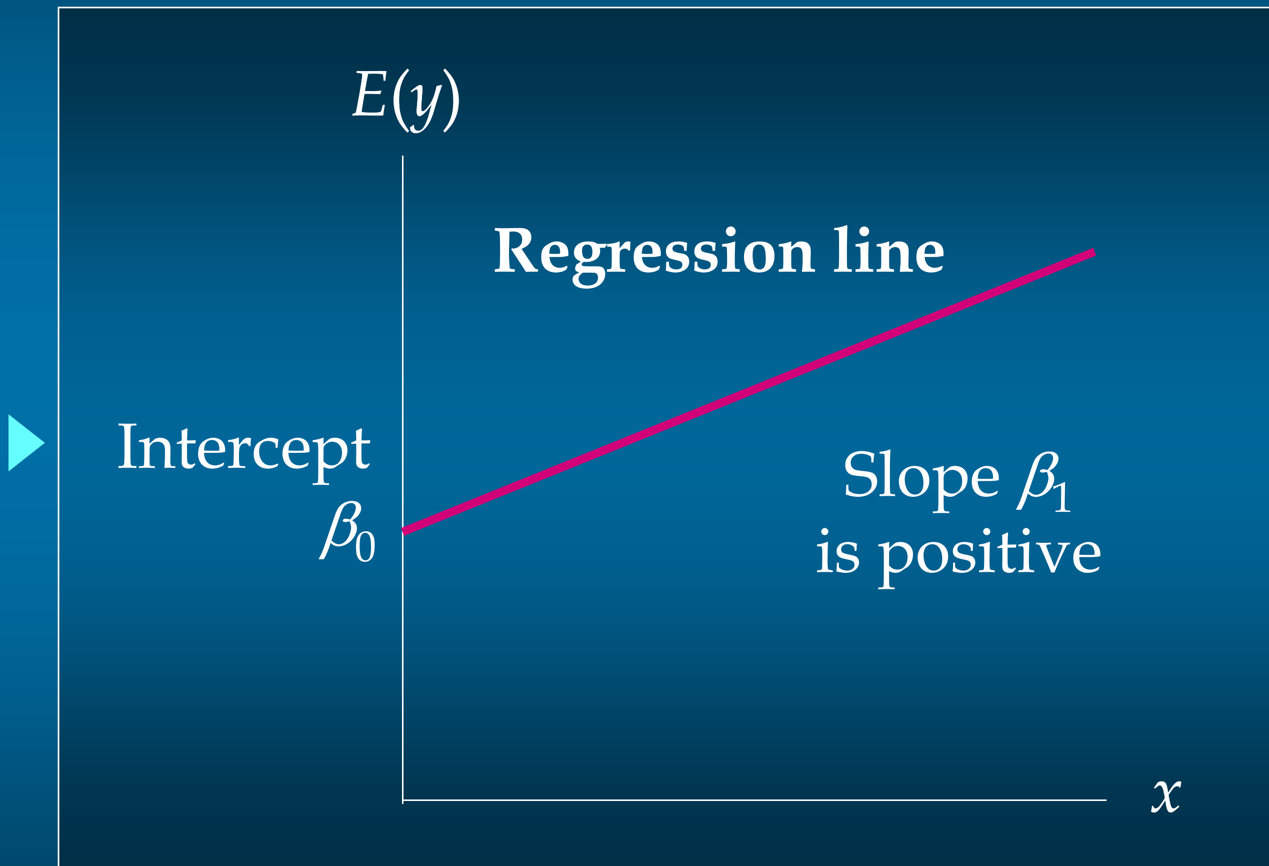
- The simple linear regression equation is:

$$E(y) = \beta_0 + \beta_1 x$$

- Graph of the regression equation is a straight line.
- β_0 is the y intercept of the regression line.
- β_1 is the slope of the regression line.
- $E(y)$ is the expected value of y for a given x value.

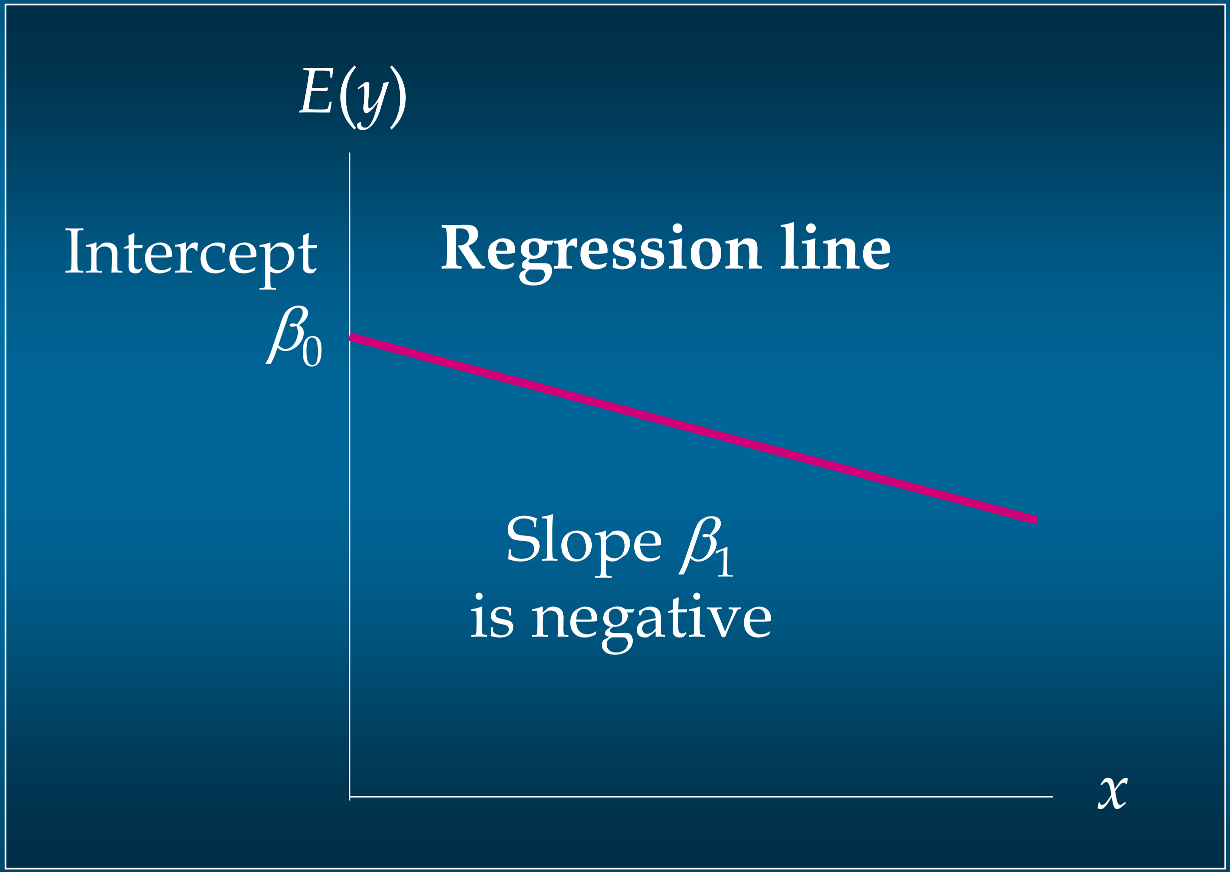
Simple Linear Regression Equation

- Positive Linear Relationship



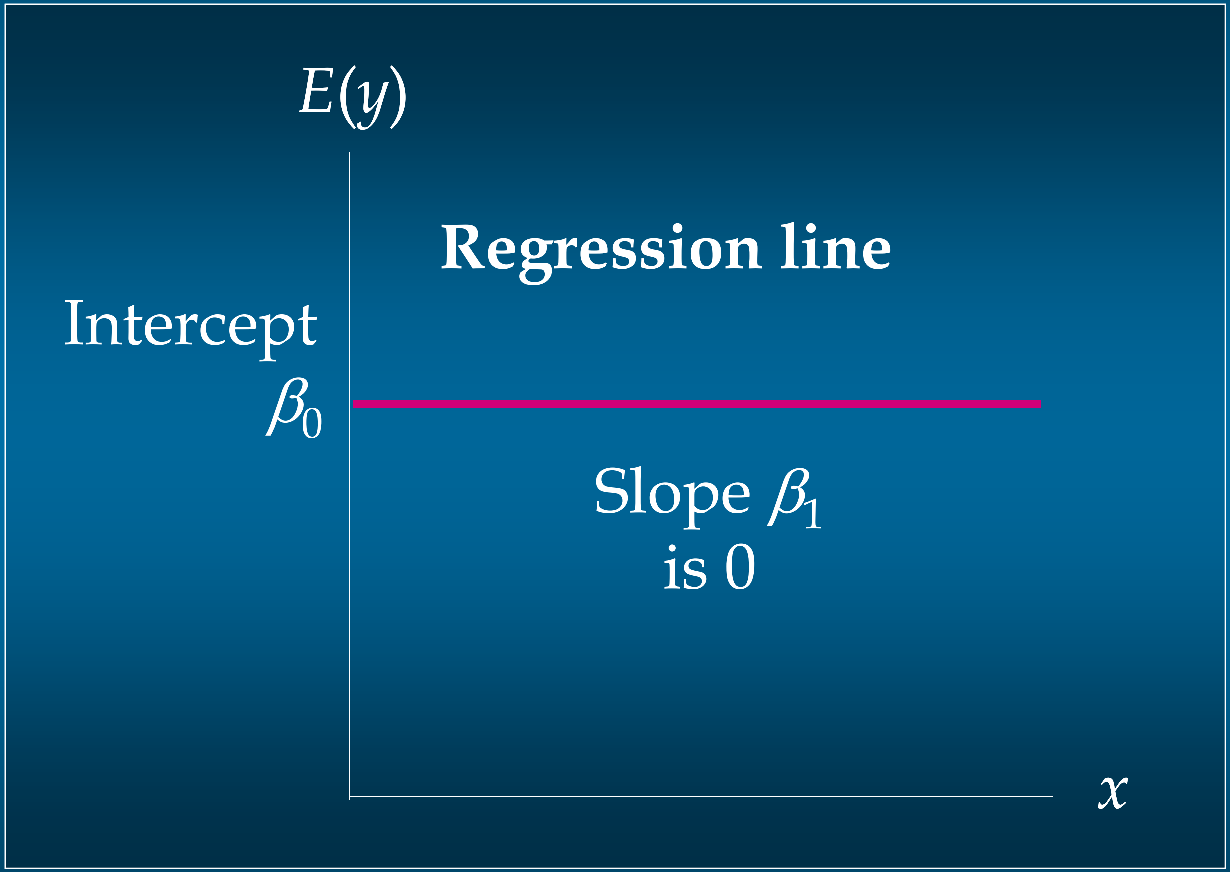
Simple Linear Regression Equation

- Negative Linear Relationship



Simple Linear Regression Equation

- No Relationship



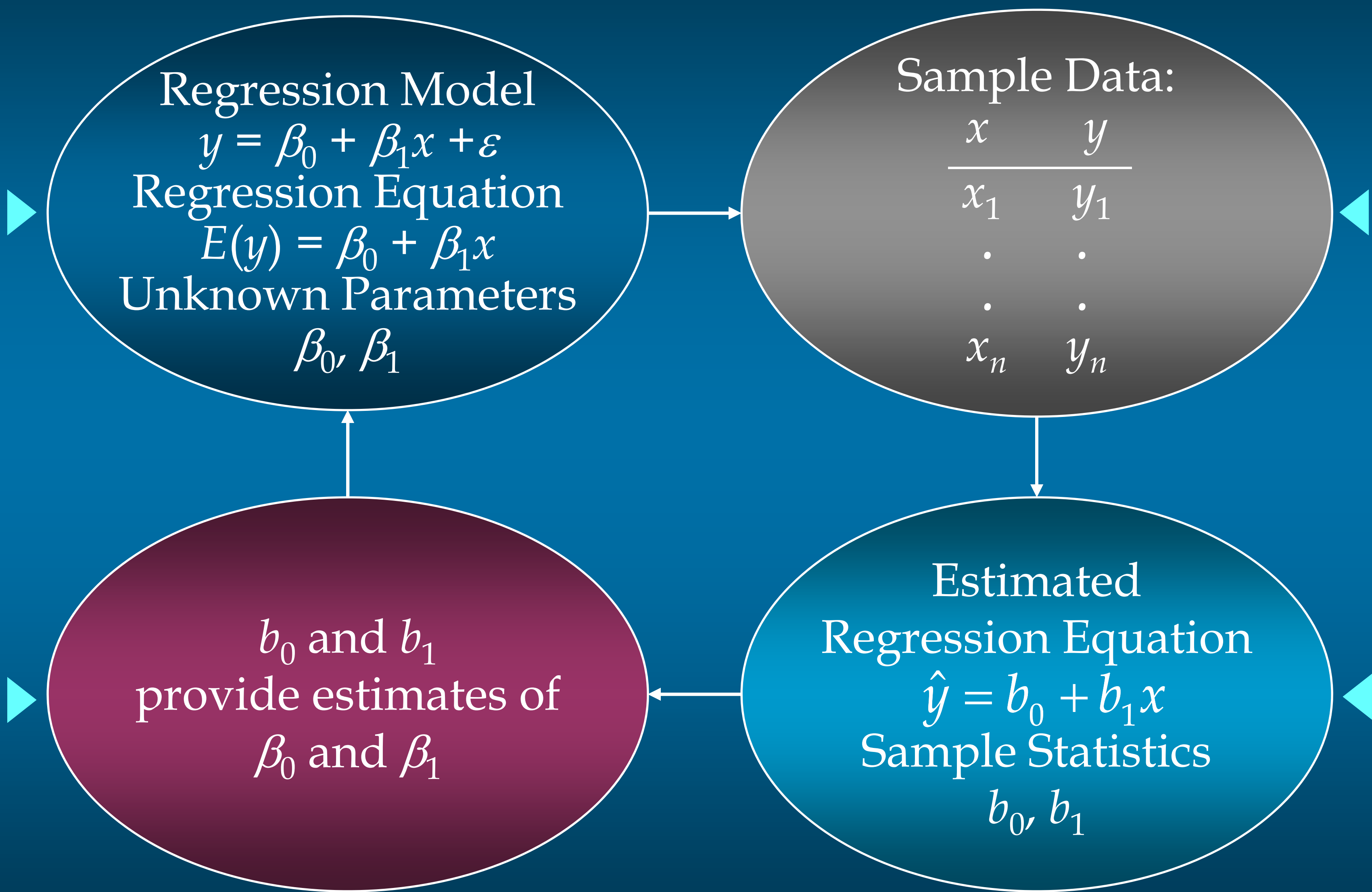
Estimated Simple Linear Regression Equation

- The estimated simple linear regression equation

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

  - The graph is called the estimated regression line.
  - b_0 is the y intercept of the line.
  - b_1 is the slope of the line.
  - $\hat{y}$ is the estimated value of y for a given x value.

Estimation Process



Least Squares Method

- Least Squares Criterion

$$\min \sum (y_i - \hat{y}_i)^2$$

where:

y_i = observed value of the dependent variable for the *i*th observation

$\hat{y}_i$ = estimated value of the dependent variable for the *i*th observation

Least Squares Method

- Slope for the Estimated Regression Equation

$$b_1 = \frac{\sum x_i y_i - \frac{\left(\sum x_i \sum y_i\right)}{n}}{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}$$

Least Squares Method

- y-Intercept for the Estimated Regression Equation

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

where:

x_i = value of independent variable for ith observation

y_i = value of dependent variable for ith observation

$\bar{x}$ = mean value for independent variable

$\bar{y}$ = mean value for dependent variable

n = total number of observations

Simple Linear Regression

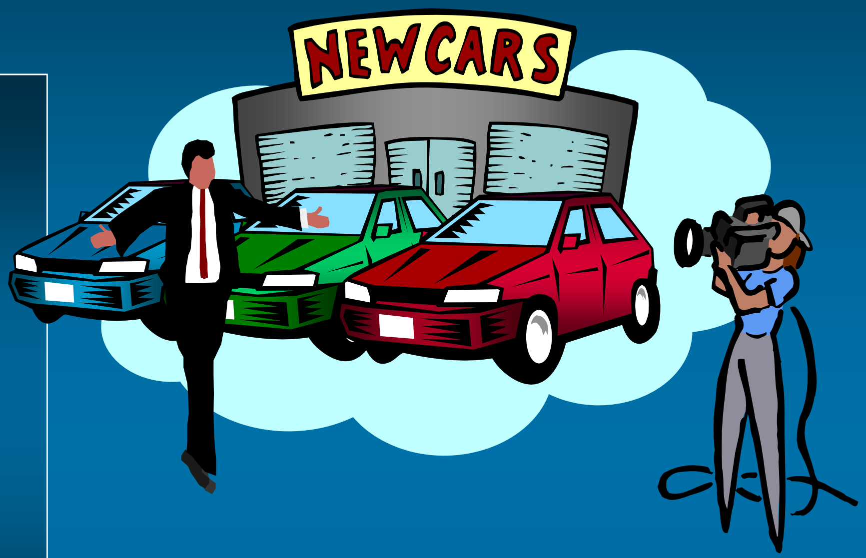
- Example: Reed Auto Sales

    - Reed Auto periodically has a special week-long sale. As part of the advertising campaign Reed runs one or more television commercials during the weekend preceding the sale. Data from a sample of 5 previous sales are shown on the next slide.

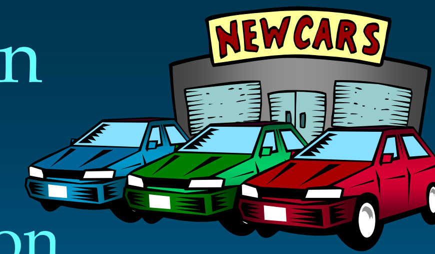
Simple Linear Regression

- Example: Reed Auto Sales

Number of TV Ads	Number of Cars Sold
1	14
3	24
2	18
1	17
3	27



Estimated Regression Equation



- q Slope for the Estimated Regression Equation

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{20}{4} = 5$$

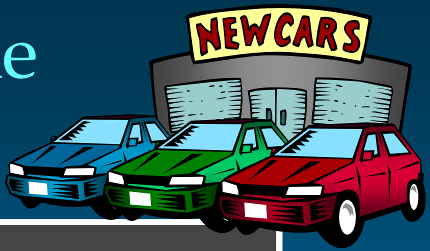
- q *y*-Intercept for the Estimated Regression Equation

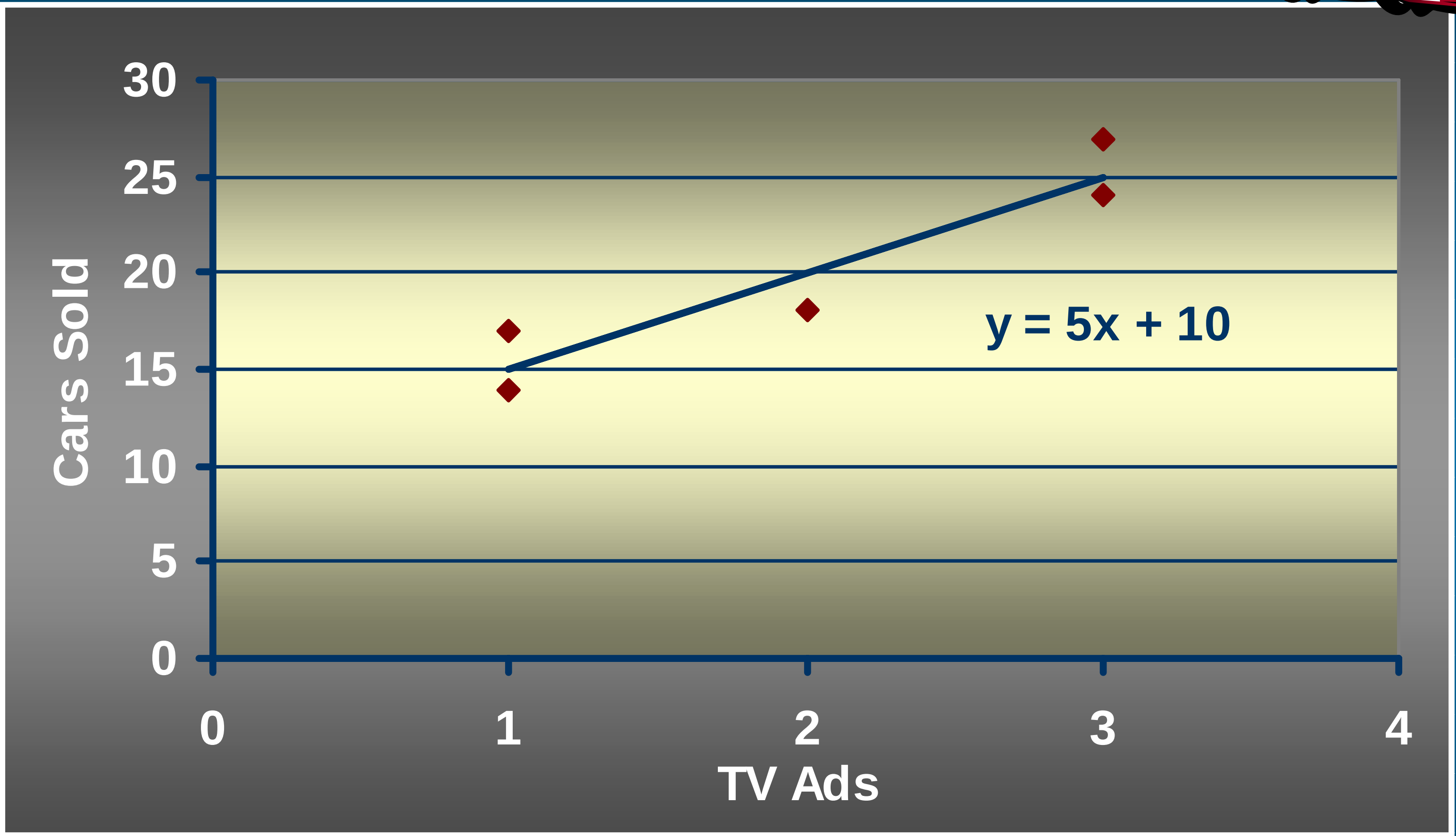
$$b_0 = \bar{y} - b_1\bar{x} = 20 - 5(2) = 10$$

- q Estimated Regression Equation

$$\hat{y} = 10 + 5x$$

Scatter Diagram and Trend Line





Coefficient of Determination

- Relationship Among SST, SSR, SSE

  - SST = SSR + SSE

$$\sum(y_i-\overline{y})^2=\sum(\hat{y}_i-\overline{y})^2+\sum(y_i-\hat{y}_i)^2$$

where:

SST = total sum of squares
SSR = sum of squares due to regression
SSE = sum of squares due to error

Coefficient of Determination

- The <u>coefficient of determination</u> is:

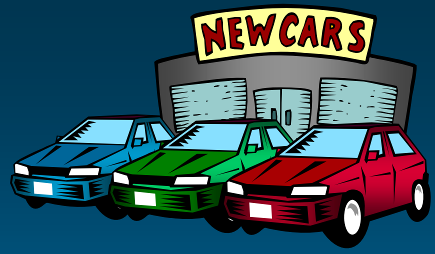
  - $r^2 = SSR/SST$

where:

SSR = sum of squares due to regression

SST = total sum of squares

Coefficient of Determination



- r^2 = SSR/SST = 100/114 = .8772
- The regression relationship is very strong; 88% of the variability in the number of cars sold can be explained by the linear relationship between the number of TV ads and the number of cars sold.

Sample Correlation Coefficient

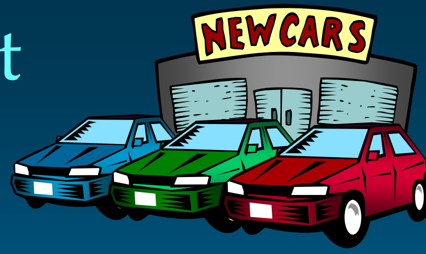
$$r_{xy} = (\text{sign of } b_1)\sqrt{\text{Coefficient of Determination}}$$

$$r_{xy} = (\text{sign of } b_1)\sqrt{r^2}$$

where:

b_1 = the slope of the estimated regression equation $\hat{y} = b_0 + b_1 x$

Sample Correlation Coefficient



- $$r_{xy} = (\text{sign of } b_1)\sqrt{r^2}$$
- The sign of b_1 in the equation $\hat{y} = 10 + 5x$ is "+".
- $r_{xy} = +\sqrt{.8772}$
- $r_{xy} = +.9366$

Assumptions About the Error Term ε

1. The error ε is a random variable with mean of zero.
2. The variance of ε, denoted by σ^2, is the same for all values of the independent variable.
3. The values of ε are independent.
4. The error ε is a normally distributed random variable.

Testing for Significance

- To test for a significant regression relationship, we must conduct a hypothesis test to determine whether the value of β_1 is zero.
- Two tests are commonly used:

  t Test and *F* Test

- Both the *t* test and *F* test require an estimate of σ^2, the variance of ε in the regression model.

Testing for Significance

- An Estimate of σ
  - The mean square error (MSE) provides the estimate of σ^2, and the notation s^2 is also used.

$$s^2 = \text{MSE} = \text{SSE}/(n-2)$$

  - where:

$$\text{SSE} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b_1 x_i)^2$$

Testing for Significance

- An Estimate of σ
  - To estimate σ we take the square root of σ^2.
  - The resulting s is called the standard error of the estimate.

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{\text{SSE}}{n-2}}$$