

- **Dependent and Independent Variables**

Dependent variables depend on the independent variable(s). With linear regression, we are attempting to predict values of the dependent variable. Dependent variables are usually designated Y and independent variables, X .

- **Correlation Coefficient -r**

The correlation coefficient measures the strength of the relationship between two variables. It can range from -1 to +1.

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{(n - 1)s_x s_y}$$

- **Regression Analysis**

Linear regression results in an equation used to predict a value for a dependent variable based on independent variable(s).

$$\hat{Y} = a + bX$$

- **Slope**

The slope determines how much Y increases with each increase in X .

$$b = r \frac{s_y}{s_x}$$

- **Intercept**

The value of Y when X is zero:

$$a = \bar{Y} - b\bar{X}$$

- **Standard Error of the Estimate**

The standard error of the estimate measures how widely the predicted values and actual values are dispersed on the regression line.

$$s_{y \cdot x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}}$$

- **Coefficient of Determination – r^2**

The coefficient of determination, r^2 , measures the proportion of variance in the dependent variable, Y , that can be explained by the independent variable, X .

- **Dummy Variables**

Dummy variables allow some qualitative variables, those that use a nominal scale to be used in regression.

- **The Multiple Regression Equation**

The multiple regression equation

$$\text{Predicted } Y = a + B_1(X_1) + B_2(X_2) + B_3(X_3) \dots$$

Chapter 13 Highlights – Correlation & Regression

- **Evaluating a Multiple Regression Output**

- **Correlation Coefficient**

Examine the correlation coefficient, r , to see if there is a relationship strong enough to continue.

- **The Global (F) Test**

The global test measures whether or not ALL the slopes of the independent variables are zero.

- **Evaluating Each Independent Variable**

The next step is to evaluate each variable to see if it should be included in the regression equation. We do this by testing the slope of each variable with a T test.

- **Creating a more accurate multiple regression equation**

Run another regression using only statistically significant independent variables.