

Describing the Relationship between Two Variables

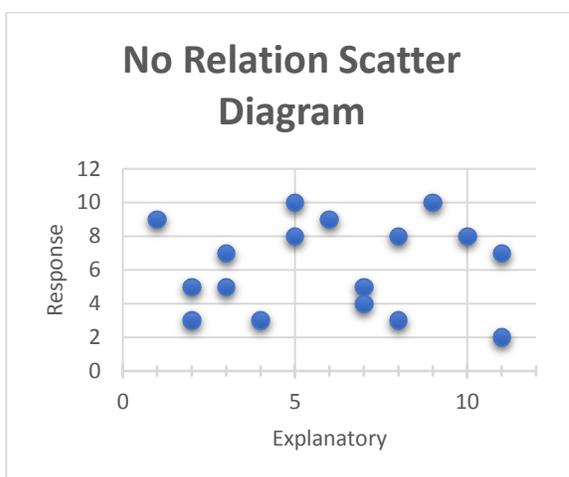
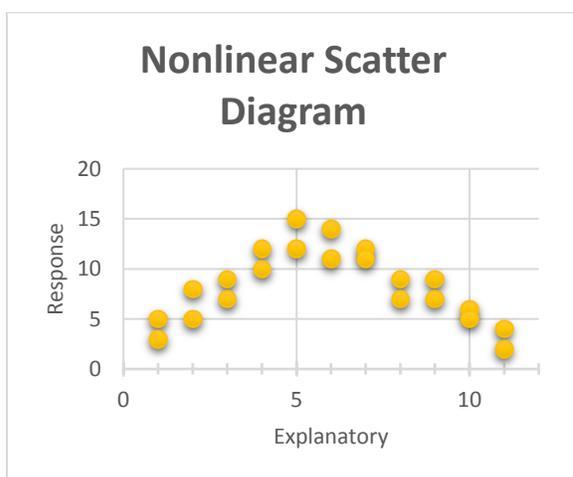
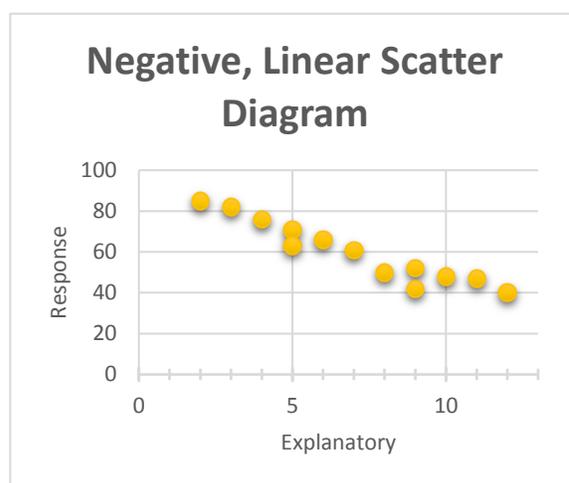
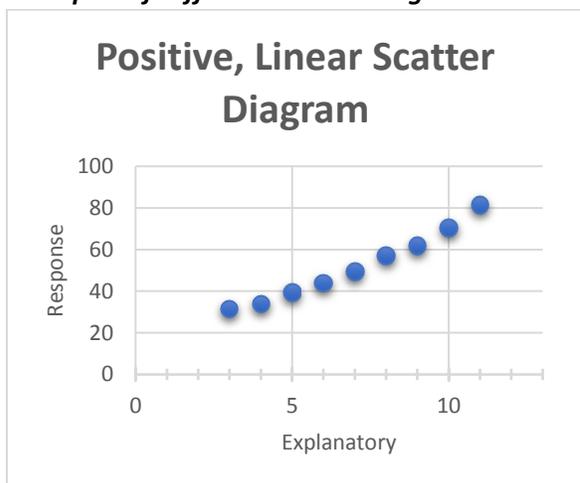
Key Definitions

- **Scatter Diagram:** A graph made to show the relationship between two different variables (each pair of x 's and y 's) measured from the same equation.
- **Linear Relationship:** A linear relationship will have all the points close together and no curves, dips, etc. in the graph. It will be an almost straight line up or down.
- **Nonlinear Relationship:** A nonlinear relationship will still have the points close together, but will have a curve or dip.
- **No Relationship:** Having no relationship in a scatter diagram means that the data doesn't have a particular pattern to it. The data is dispersed.
- **Positively Associated:** This is when you have one value increase, the other value will increase as well. The line on your diagram will start low and increase upwards. This is only for equations who have a linear relationship.
- **Negatively Associated:** This is when you have one value decrease, the other will decrease as well. The line on your diagram will start high and decrease downwards. This is only for equations who have a linear relationship.
- **Linear Correlation Coefficient:** A calculation that shows us the strength of the linear relation and the direction of the linear relation.
- **Lurking Variable:** This is something that shows a linear relation between the variables, but is not actually correlated.
- **Residual:** A residual is the space in between the observed y and the predicted y which is also known as the error.
- **Slope:** The rate that the line is increasing or decreasing at.
- **Y-Intercept:** The point on the line where $x = 0$.
- **Coefficient of Determination:** The percentage of total variation in the response variable that is explained in the least-squares regression line.
- **Total Variation:** The deviation between the observed and mean values.
- **Explained Variation:** The deviation between the predicted and mean values.
- **Unexplained Variation:** The deviation between the observed and predicted values.

Scatter Diagram, Relations, and Association

- **How a Scatter Diagram is Set Up:** Your scatter diagram is like a usual graph from algebra where we graph x and y values, but there is no line connecting each point. On the graph, the x -axis is called the explanatory variables and the y -axis is called the response variables. Plot each point on the graph to create the graph.

- **Determining Meaning from the Scatter Diagram:** After each point is plotted on the graph, you are able to determine if the equation has a linear relation, nonlinear relation, or no relation. You can also determine whether the equation is positively associated or negatively associated.
- **Examples of Different Scatter Diagrams:**



Linear Correlation Coefficient

How to Find the Correlation Coefficient: The following is the formula given to us on how to find the correlation coefficient:

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n - 1}$$

This formula takes a very long time to do by hand. Therefore, we use technology to help us find the answer. We do this using Excel. The following are step by step instructions:

1. First, input your data. Make sure you x-values and y-values are in separate columns.
2. In a blank cell, type in =CORREL(*click and drag down to highlight the cells of the x-values, click and drag down to highlight the cells of the y-values*) [The text italicized is instructions on what to input]. Press enter and the correlation coefficient answer will replace what you wrote.

- **Example of Finding the Correlation Coefficient:**

You are given the following data points to find the correlation coefficient:

x-values	y-values	Correlation Coefficient:
82	103	=CORREL(A2:A11, B2:B11)
77	101	0.900702096
90	121	
71	100	
62	95	
68	98	
74	100	
84	118	
94	116	
88	114	

- **How to Determine if the Correlation Coefficient Shows a Linear Relation:** Using your textbook, go to the table in the back of your book that gives us critical values. Match up your sample size with the “n” on the chart. The number next to your “n” is the critical value. If the critical value is less than the absolute value of the correlation coefficient, we have a linear relation between the two variables. If it is not, there is not a linear relation.
- **How to Interpret the Linear Correlation Coefficient:** The correlation coefficient is only between -1 and 1. If your correlation coefficient is towards -1, that means you have a negative, linear relation. If your correlation coefficient is towards 1, that means you have a positive, linear relation. If your correlation coefficient is towards 0, that means you have no relation.
- **Example on How to Determine and Interpret the Correlation Coefficient:** Using the following data points, we need to interpret details of the equation from the correlation coefficient. Going into our textbook, we look for the critical value first which is 0.632 since our sample size is 10. From there we must see if the absolute value of our correlation coefficient is greater than the critical value:

$$|0.90| > .632$$

Since our correlation coefficient is greater than the critical value, it is a strong linear relationship. Since our correlation coefficient is positive without the absolute values, it has a positive associativity.

The Least-Squares Regression Line

- **Finding the Linear Equation:** We select two points from our data. From there, we find the slope for our equation using the slope formula:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Note that y_2 and x_2 are the larger values. From there, we plug it into the point slope formula:

$$y - y_1 = m(x - x_1)$$

Make sure when using the point-slope formula use plug in only one of the points that you used to get the slope. We want to have the “y” and “x” remain unknown variable for the equation. Solve the equation for “y” to get your equation.

- **Example of Finding the Linear Equation:**

Using the data points given above in the correlation coefficient example, we can find the linear equation. First using the points (74, 100) and (68, 98), we will find the slope:

$$m = \frac{100 - 98}{74 - 68} = \frac{2}{6} = \frac{1}{3} = 0.33$$

Next, we plug the slope and one of the data points into the slope-point formula to find the equation:

$$y - 98 = 0.33(x - 68)$$

$$y - 98 = 0.33x - 22.44$$

$$y = 0.33x - 22.44 + 98$$

$$y = 0.33x + 75.56$$

- **Least Squares Method:** A residual is the space in-between the observed y and the predicted y . The least squares method tries to make this distance and error as small as possible. To do this, we need to have the observed y (the linear equation) and the predicted y (the least-squares regression line).
- **How to Find the Least-Squares Regression Line:** To find the least-squares regression line, you need to find the slope and y -intercept first. We do not find it the same way we find the linear equation's slope and y -intercept. We do this by first finding the slope (we use the symbol b_1). To find the slope, you need to have the correlation coefficient, the standard deviation of the y -values, and the standard deviation of the x -values. The following is the formula for the slope:

$$b_1 = r \cdot \frac{s_y}{s_x}$$

After finding the slope, you can now find the y -intercept. To find the y -intercept, you need to have the slope, mean of the x -values, and the mean of the y -values. The following is the formula:

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

From there, you have all the information needed to put your information into the least-squares regression line formula:

$$\hat{y} = b_1x + b_0$$

- **Example of Finding the Least-Squares Regression Line:**

Using the data values from the example on the correlation coefficient, we know $r = .90$, $\bar{x} = 79$, $\bar{y} = 106.6$, $s_x = 10.35$, and $s_y = 9.55$. First, we find the slope:

$$b_1 = r \cdot \frac{s_y}{s_x} = 0.9 \cdot \frac{9.55}{10.35} = 0.9(0.92) = 0.83$$

Now that we know the slope, we can find the y -intercept:

$$b_0 = \bar{y} - b_1\bar{x} = 106.6 - 0.83(79) = 106.6 - 65.57 = 41.03$$

Now, we can put the information into our least-squares regression line formula:

$$\hat{y} = b_1x + b_0$$

$$\hat{y} = .083x + 41.03$$

The Coefficient of Determination

- **How to Find the Coefficient of Determination:** To find the coefficient of determination for a least-squares regression line, you take your linear correlation coefficient and square it. Therefore, the formula is the following:

$$R^2 = r^2$$

- **Example of How to Find the Coefficient of Determination:**

Using the data from the linear correlation coefficient example, we know that $r = .90$. Now we just plug it into the formula:

$$R^2 = (0.90)^2 = .81$$

Since the coefficient of determination is a percentage, we just turn our decimal into a percentage by multiplying it by 100%. Therefore, our answer is 81%.

Symbol Guide

Chapter Title Symbols		
Term	Symbol	Use
Sample Mean	\bar{x}	To identify the sample mean
Sample Standard Deviation	s	To identify the sample standard deviation
Sample Size	n	To identify the sample size
Sum	Σ	To identify when we add up everything
Linear Correlation Coefficient	r	To identify the correlation coefficient
Linear Line Slope	m	To identify the slope of a linear line
Y-Intercept	b	To identify the y-intercept of a linear line
Least-Squares Regression Line Slope	b_1	To identify the least-squares regression line slope
Least-Squares Regression Line Y-Intercept	b_0	To identify the least-squares regression line y-intercept
Predicted y	\hat{y}	To identify the predicted y of a least-squares regression line
The Coefficient of Determination	R^2	To identify the coefficient of determination