

Chapter 7 and 8: Introduction to Linear Regression

Review

The _____ measures the strength and direction of the **linear** relationship between two **quantitative** variables.

If there is enough evidence of an association we may want to find a line that is a good representation of the linear relationship between our two quantitative variables. Such a line is called _____.

We use linear regression to:

-
-

Recall that we use _____ to signify the explanatory variable and _____ to signify the response variable.

Regression

A _____ explains how the values of the response variable changes in relation to changes in the explanatory variable.

Equation

For a response variable y , and an explanatory variable x , the linear regression line between y and x is given by:

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

where,

- \hat{y} :
- b_0 :
- b_1 :

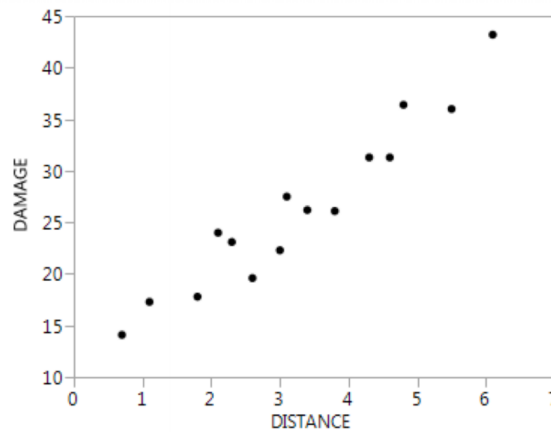
This is known as the slope-intercept equation of a line. For those less familiar with mathematics, the slope of a line represents _____. The intercept refers to _____. To calculate the slope and the intercept of a line we use the following formulas:

$$b_o = \bar{y} - b_1 * \bar{x} \quad b_1 = r * \frac{s_y}{s_x}$$

where,

- \bar{x} :
- \bar{y} :
- s_x :
- s_y :
- r :

Example To illustrate the processing of creating a linear regression let's look at some fire damage data. The explanatory variable in this is the distance between a fire and the nearest fire station in miles. The response variable is the damage in thousands of dollars caused by a fire. Data was collected from 15 homes in a major metropolitan area. From the scatterplot below, describe the form, strength, direction, and outliers of the data.



Do you think a linear regression would be appropriate in this case?

Summary Statistic	Value
Mean distance	3.28
Mean of damage	26.41
Standard deviation of distance	1.58
Standard deviation of damage	8.07
Correlation between distance and damage	0.961

Given the above summary statistics information, calculate the equation of the regression equation used to describe the relationship between distance from a fire station and fire damage.

y versus \hat{y}

- Correct linear equation:
- Incorrect linear equation:

It is important to include the "hat" in your linear equation!

- y :

—

—

- \hat{y} :

—

—

Interpretation

To interpret the results of our linear regressions into meaningful conclusions we can interpret the components of our equation in context. Recall:

$$\hat{y} = b_o + b_1 * x$$

To interpret b_1 , the slope:

- **Interpretation One:** For a one unit increase in the explanatory variable x , the *predicted* value of the response variable y will change by the amount of the slope.
- **Interpretation Two:** For a one unit increase in the explanatory variable x , the value of the response variable y will change by the amount of the slope, *on average*.

To interpret b_o , the intercept:

- **Interpretation:** When the value of the explanatory variable is equal to 0, the *predicted* value of the response variable is equal to the intercept.

Let's return to the example of our fire damage problem. Write the linear regression equation below and then interpret the results in the context of the problem.