Linear Relationships: Intercept and Slope

Learning Objective: In the special case of linear relationship, use the least squares regression line as a summary of the overall pattern, and use it to make predictions.

Like any other line, the equation of the least-squares regression line for summarizing the linear relationship between the response variable (Y) and the explanatory variable (X) has the form: $Y = a + bX$

All we need to do is calculate the intercept **a**, and the slope **b**, which is easily done if we know:

- $\overline{X}$—the mean of the explanatory variable's values
- S_X—the standard deviation of the explanatory variable's values
- $\overline{Y}$—the mean of the response variable's values
- S_Y—the standard deviation of the response variable's values
- r—the correlation coefficient

Given the five quantities above, the slope and intercept of the least squares regression line are found using the following formulas:

- $b = r\left(\frac{S_Y}{S_X}\right)$
- $a = \overline{Y} - b\overline{X}$

Comments

1. Note that since the formula for the intercept **a** depends on the value of the slope, **b**, you need to find **b** first.
2. The slope of the least squares regression line can be interpreted as the average change in the response variable when the explanatory variable increases by 1 unit.

Example: Age-Distance

Let's revisit our age-distance example, and find the **least-squares regression line**. The following output will be helpful in getting the 5 values we need:

Column	n	Mean	Std. Dev.	Std. Err.	Min	Q1	Median	Q3	Max
Age	30	51	21.776293	3.9757888	18	28	54	71	82
Distance	30	423	82.802216	15.117547	280	360	420	460	590

- Dependent Varialbe: Distance
- Independent Variable: Age
- R (correlation coefficient)= -0.7929
- The **slope** of the line is: $b = (-0.793) * \left(\frac{82.8}{21.78}\right) = -3$. This means that for every 1-unit increase of the explanatory variable, there is, on average, a 3-unit decrease in the response variable. The interpretation **in context** of the slope being -3 is, therefore: For every year a driver gets older, the maximum distance at which he/she can read a sign decreases, **on average**, by 3 feet.
- The **intercept** of the line is:

  a = 423 - (-3 * 51) = 576

  and therefore the **least squares regression line** for this example is:

  Distance = 576 + (- 3 * Age)

Here is the regression line plotted on the scatterplot:

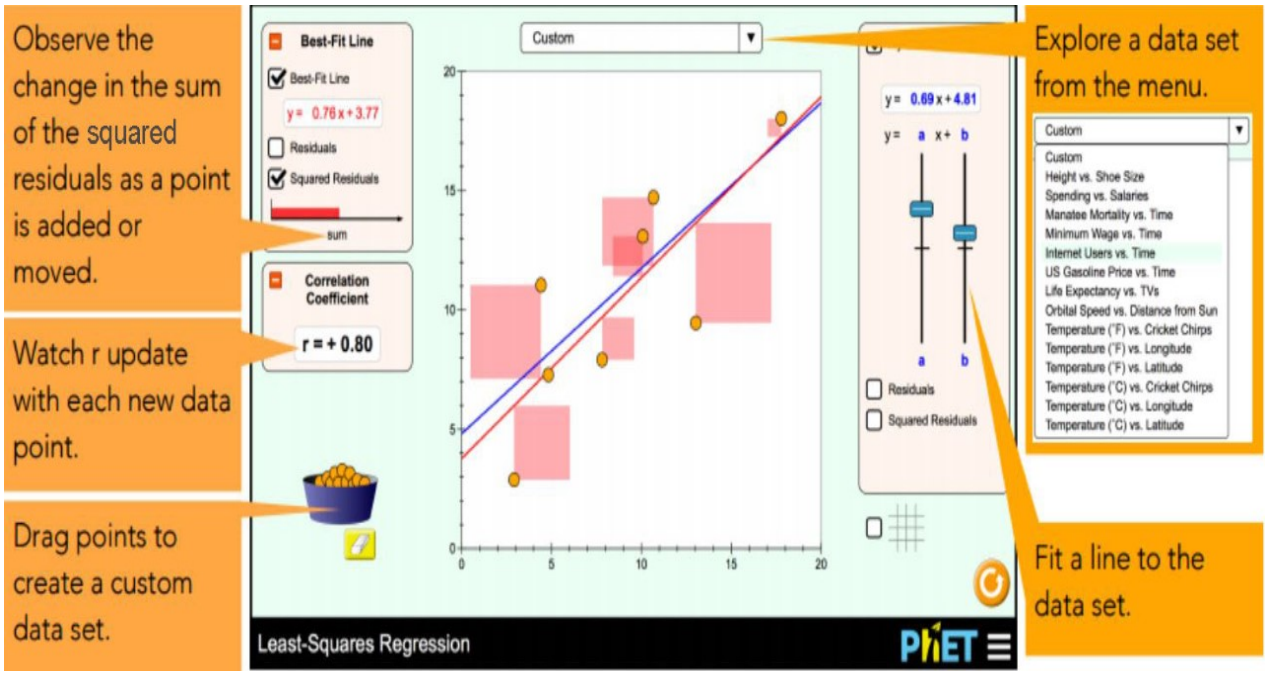
As we can see, the regression line fits the linear pattern of the data quite well.

Comment

As we mentioned before, hand-calculation is not the focus of this course. We wanted you to see one example in which the least squares regression line is calculated by hand, but in general we'll let a statistics package do that for us.

Interactive Simulation

Create your own scatter plot or use real-world data and try to fit a line to it! Explore how individual data points affect the correlation coefficient and best-fit line. Below is a graphic that shows how to use the simulation. **NOTE: If the simulation "freezes" while trying to use it, refresh the page to reset it.**



Learn By Doing

1/1 point (graded)

In the above simulation explore the correlation and best fit line for the "Life Expectancy vs. TVs" dataset. What is the intercept of the regression line?

- 0.04
- -0.04
- 0.61
- -0.61
- 69.6 ✔
- -69.6

Answer

Correct:

The regression line has the form: Y = intercept + (slope * X), which is sometimes written as Y = (slope * X) + intercept. Therefore, in y = -0.04X + 69.6 the intercept is 69.6.

Submit

Learn By Doing

1/1 point (graded)

In the above simulation explore the correlation and best fit line for the "Life Expectancy vs. TVs" dataset. Which of the following is the correct interpretation of the slope of the regression line?

- For each one unit increase in the average number of people per TV, the life expectancy is likely to change by (-0.04) + 69.6 = 69.2
- For each one unit increase in the average number of people per TV, the life expectancy is likely to increase by 0.04.
- For each one unit increase in the average number of people per TV, the life expectancy is likely to decrease by 0.04. ✔
- For each one unit decrease in the average number of people per TV, the life expectancy is likely to decrease by 0.04.

Answer

Correct:
The explanatory variable is the average number of people per TV. A negative slope of -0.04 means that for each 1-unit change in the explanatory variable, we would expect the response variable to change by 0.04 units. Since the slope is negative, the changes are in opposite directions. Thus, if the explanatory variable increases, the response variable decreases, and vice versa.

Submit