LAB SESSION 7 – CORRELATION & SLR

Analytics Primer

SIMPLE LINEAR REGRESSION

Example

- An analyst for the State of North Carolina has collected data on salaries and years of education with the intention of finding out how education affects the salary an individual makes monthly (in dollars). The correlation between the two variables is 0.327. The simple linear regression line between these two variables is the following:

$$\hat{y} = 66.271 + 66.054x$$

Example

1. Interpret the values of the coefficients.
   - The average salary of people with no education is $66.27/month.
   - The average increase in salary for every additional year of education is $66.05/month.

Example

2. How much salary per month do you expect someone with only a high school education to earn monthly (13 years of education)?

$$\hat{y} = 66.271 + 66.05(13) = \$924.97$$

Example

3. How much more salary can a person earn monthly if they go to school 4 more years after high school?

$$66.054(4) = \$264.22$$

Example

4. Calculate and interpret R^2.

$$R^2 = r^2 = 0.327^2 = 0.107$$

- 10.7% of the variation in monthly salary can be explained by the relationship with years of education.

MORE EXAMPLES!

More Examples

- You have a sample of data where the mean value of y is 100.7 with a standard deviation of 15.4. Your sample also has a mean value of x as 2.8 with a standard deviation of 4.8. The correlation between these two variables is 0.52.

Example

1. What is the slope of the regression line between these two variables?

$$\hat{\beta}_1 = 0.52 \times \frac{15.4}{4.8} = 1.67$$

Example

2. What is the intercept of the regression line between these two variables?

$$\hat{\beta}_0 = 100.7 - (1.67 \times 2.8) = 96.02$$

Example

3. What is the proportion of the variation in the y variable explained by the linear relationship with the x variable?

$$R^2 = r^2 = 0.52^2 = 0.27$$

Example

4. What would the predicted value of y be if the x variable took a value of 10?

$$\hat{y} = 96.02 + 1.67(10) = 112.72$$