

LAB SESSION 8 – SIMPLE & MULTIPLE LINEAR REGRESSION

Analytics Primer

SIMPLE LINEAR REGRESSION

Complete Example

Example

- The director of admissions of a small college in the Midwest has hired you as an analyst to administer a newly designed entrance test. This test ranges from a score of 1-7. You administer the test to 213 students selected randomly from the new freshman class in a study to determine whether a student's GPA at the end of their freshman year can be predicted from the entrance test score. The sample's average GPA was 2.67 with a s.d. of 0.72. The average entrance test score was 4.81 with a s.d. of 0.69. The correlation between these two variables is 0.735. Use this information to answer the following questions...

Example

1. Create the sample regression line for predicting GPA from the entrance test score.

$$\hat{\beta}_1 = 0.735 \times \frac{0.72}{0.69} = 0.767 \quad \hat{\beta}_0 = 2.67 - 0.767(4.81) = -1.019$$

Example

2. What would a predicted GPA at the end of freshman year be for a student who scored a 6.1? How about a 2.9?

$$\hat{y}_{6.1} = -1.019 + 0.767(6.1) = 3.66 \qquad \hat{y}_{2.9} = 1.21$$

Example

3. What would the expected increase in GPA at the end of freshman year be for an increase of 1.5 points on the entrance test?

$$0.767(1.5) = 1.1505$$

Example

4. State the hypothesis for a test to determine whether the slope of the true regression line is equal to zero.

$$H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0$$

Example

5. Fill in the blanks for the following table:

Parameter	Estimate	Std. Error	t Value	P-value
Intercept	-1.019	0.238997	-4.264	< 0.001
Slope	0.767	0.049238	15.577	< 0.001

Example

6. Summarize the results of the hypothesis test of the slope. Enough evidence to say there is a relationship between entrance test score and GPA.

Example

7. What is the value and meaning of R^2 in this problem?

$$R^2 = r^2 = 0.735^2 = 0.54$$

54% of the variation in freshman year GPA is explained by the entrance test score.

MORE EXAMPLES!

More Examples

- At the last college football game of the season, a survey of attendants was conducted for a random sample of 95 people to asked them their age and how many games that they have attended this season. The average age was 29 with a standard deviation of 8.4. The average number of games attended was 4.2 with a standard deviation of 2.3. The correlation between these variables is 0.6.

Example

1. What is the slope for the simple linear regression line for using age to predict the number of games attended?

$$\hat{\beta}_1 = 0.6 \times \frac{2.3}{8.4} = 0.164$$

Example

2. What is the intercept for the simple linear regression line for using age to predict the number of games attended?

$$\hat{\beta}_0 = 4.2 - 0.164(29) = -0.556$$

Example

3. The following results were obtained from your regression above:

$$SSE = 14000$$

$$\Sigma(x - \bar{x})^2 = 12100$$

What is the test statistic for the test of the slope of the regression line?

$$t = \frac{0.164}{\left(\frac{\sqrt{\frac{14000}{95 - 2}}}{\sqrt{12100}} \right)} = 1.477$$

Example

4. What is the margin of error for a 95% confidence interval for the slope?

$$M.O.E. = 1.987 \times 0.111 = 0.22$$

Example

5. What is the proportion of variation in number of games attended that is explained by its linear relationship with age?

$$R^2 = r^2 = 0.6^2 = 0.36$$