Problem Set 8

Due date: 20 November

Table of contents

Please upload your completed assignment to the ELMs course site (under the assignments menu). Remember to include an annotated script file for all work with R and show your math for all other problems (if applicable, or necessary). Please also upload your completed assignment to the Github repository that you have shared with us. *We should be able to run your script with no errors.*

Total points: 30

Question 1

Points: 5

For the following regression equation, $\hat{Y} = 8.5 + 6x + \epsilon$, the standard error for β_0 is 2.5, the standard error for β_1 is 3.5, and the sample size is 2000. Find the t-statistic, 95% confidence interval, and p-value (using a two-tailed test) for β_1.

Is β_1 statistically significant at the 0.05-level with a two-tailed test? Why or why not?

t-statistic:

$t = \frac{6-0}{3.5} = 1.71$

The 95% CI:

Upper bound: 6 + 1.961 x 3.5 = 12.86

Lower bound: 6 -1.961 x 3.5 = -0.86

The 95% confidence interval for β_1 is approximately from -0.86 to 12.86. The interval suggests that, with 95% confidence, we can say the true value of β_1 lies somewhere within this range.

P-value:
With a t-statistic of 1.71 and 1998 degrees of freedom, we get a p-value of .087

1 is not statistically significant at the 0.05 level based on the t-statistic, p-value, and confidence interval, meaning we don't have enough evidence to say that β_1 is different from 0.

Question 2

Points: 5

Suppose you estimate an OLS regression and retrieve a R^2 value of 0.45. If the Total Sum of Squares (TSS) from that regression equals 4,700, what is the value for the Residual Sum of Squares (RSS)?

$RSS = 4700 \times (1 - 0.45) = 2585$

This tells us that the total variance in the DV, the models residuals account for 2585 units.

Question 3

Points: 5

Suppose you estimate a bivariate regression with a sample size of 102 and obtain a regression coefficient (β_1) of 5.0. What is the largest standard error that β_1 could have and still be statistically significant (i.e., reject the null hypothesis of no relationship) at the 0.05 level with a one-tailed test?

$SE(\beta 1\ ) = \frac{5.0}{1.66} = 3.01$

The largest standard error for β_1 that still allows for statistical significance at the 0.05 level about 3.01. This means if the actual SE of β_1 is less than or equal to 3.01, the coefficient is statistically significant at the 5% level, which will reject the null.

Question 4

Points: 5

Using the `states` dataset from the `poliscidata` package, produce a scatterplot of the variables `romney2012` and `hispanic10` (with `romney2012` as the dependent variable on the y-axis). Fit a regression line to the scatterplot. Describe the scatterplot and include a copy of it. Note any suspected outliers, if any (a visual inspection will suffice for this question).