

SOC 4930/5050: Lab-13 - Multiple Regression

Christopher Prener, Ph.D.

November 20th, 2017

Directions

Please complete all steps below. Your your well formatted R Notebook (.Rmd file) and html output as well as your pdf with \LaTeX regression tables should be uploaded to your GitHub assignment repository by 4:15pm on Monday, November 27th, 2017. Use the auto17 data from testDriveR for both models.

Set-up

1. Construct a hypothesis and null hypothesis for the relationship between fuel cost (fuelCost) and engine size (represented by displ) that also accounts for characteristics of the engine (represented by cyl and gears) as well as the highway fuel efficiency (hwy).
2. Given the number of RHS variables and the sample size of the auto17 data, are we at risk of overfitting the model?

Regression Model 1: Main Effect Model

3. Construct a regression equation modeling how displ affects fuelCost using \LaTeX syntax.
4. Execute a bivariate regression model that shows how displ affects fuelCost.
5. In a sentence or two, provide an interpretation of the model's beta coefficient.
6. In a few sentences, summarize the model fit statistics.

Regression Model 2

7. Construct a regression equation modeling how displ affects fuelCost controlling for characteristics of the engine using \LaTeX syntax.

8. Execute a multivariate regression model that shows how `displ` affects `fuelCost` controlling for characteristics of the engine.
9. In a few sentences, provide an interpretation of the model's beta coefficients.
10. In a few sentences, summarize the model fit statistics.

Regression Model 3: Full Model

11. Construct a regression equation modeling how `displ` affects `fuelCost` controlling for characteristics of the engine and highway fuel efficiency using \LaTeX syntax.
12. Execute a multivariate regression model that shows how `displ` affects `fuelCost` controlling for characteristics of the engine and highway fuel efficiency.
13. In a paragraph, provide an interpretation of the model's beta coefficients.
14. In a few sentences, summarize the model fit statistics.

Communicating the Models' Results

15. Create a \LaTeX formatted regression table that includes all three models as well as their AIC and BIC summary statistics. Omit the un-adjusted R^2 value as well as the degrees of freedom values.
16. In a paragraph, compare how the beta values have changed between models.
17. In a paragraph, compare the models based on adjusted R^2 , AIC, and BIC summary statistics. Which model provides the best fit?