SOC 4015/5050: Lab-12 - Multiple Regression

Christopher Prener, Ph.D.

Fall 2018

Directions

Please complete all steps below. All work should be uploaded to your GitHub assignment repository by 4:15pm on Monday, November 19th, 2018. All data can be obtained from the `testDriveR` package's `auto17` data set.

Analysis Development

Using RStudio and your operating system's file manager, create an R Project in the *existing* directory in your assignments repository named `Lab-11`. Add a `README.md` file, notebook, and all necessary folders before beginning.[1]

[1] This initial section follows the project workflow that is available in the `lecture-03` repo!

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 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. Include the `.html` output in your `results/` directory; there is not a need to create a Word document for this lab.

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?