SOC 4015/5050: Lab-13 - Regression Diagnostics

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 3rd, 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-13`. 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!

Fit An Initial Model

1. Execute a multivariate regression model that shows how `displ` affects `fuelCost` controlling for characteristics of the engine (`cyl` and `gears`) and highway fuel efficiency (`hwy`).

Assess Model

2. Check model for non-linearity in the relationship between x variables and y.
3. Check the residuals for normality.
4. Check the residuals for homoskedastic errors.
5. Check the residuals for auto-correlation.
6. Check the independent variables for multi-collinearity.
7. Summarize your findings - do you have concerns about how the model is specified? Should variables be removed or added?

Check for Unusual Observations

8. Check for outliers.
9. Check for observations with high leverage values.
10. Check for observations with high influence values.
11. Summarize your findings - do you have concerns about unusual observations? Should observations be removed? If so, which ones?

Re-Fit Model

12. Re-fit the model based on your findings from the previous two sections.
13. Compare your initial model and the newly re-fit model. How do the betas differ (if at all)?