SOC 4930/5050: Lab-14 - Regression Diagnostics

Christopher Prener, Ph.D.

November 27th, 2017

Directions

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

Pick Up Where You Left Off

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`). This is model 3 from the previous lab.

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 model 3 and the newly re-fit model 4. How do the betas differ (if at all)?