SOC 5050: Lab 15

Christopher Prener, Ph.D.

November 28th, 2016

Directions

Please complete all steps below. Your final work by hand, do-file, log-file, plots, and markdown file with answers should be uploaded to your GitHub assignment repository by 4:20pm on Monday, December 5th, 2016.

Fit a Series of Regression Models

1. Construct a hypothesis and null hypothesis for the relationship between cost (`price`) and place of manufacture (`foreign`).

2. Construct a regression equation modeling how `foreign` affect `price` on a separate piece of paper. Include controls for the size of the vehicle (`displacement`), the vehicle's engine (`gear_ratio`), the engine's efficiency (`mpg`), and repair record (`rep78`). Scan this and turn it in with your repository.

3. Execute a bivariate regression model that shows how `foreign` affects `price`. For this model, do not use robust standard errors. Fully interpret the results of this model.

4. Generate information criterion output for this model.

5. Execute a multivariate regression model that shows how `foreign` affects `price`, controlling for characteristics of a vehicle's size and efficiency. Fully interpret the results of this model.

6. Generate information criterion output for this model.

7. Execute a multivariate regression model that shows how `foreign` affects `price`, controlling for characteristics of a vehicle's size and efficiency as well as its repair record. Fully interpret the results of this model.

8. Generate information criterion output for this model.

9. Compare the model fit information for all three models - which model is the best fit and why?

10. Compare how beta values changed across all three models.

Test Regression Assumptions

11. Check for outliers and influential observations.

12. Re-specify the model so that problematic observations are dropped.

13. Compare the fit this fourth model to your third model that contained all observations.

14. Compare how beta values from this fourth model to your third model that contained all observations.

15. Consider how the number of observations have changed. Should we be worried about the model now being overfit?

16. Test the fourth model's residuals - are they normally distributed?

17. Test the fourth model's residuals - are they heteroskedastic?

18. Test the fourth model for multicollinearity.

19. Re-specify the fourth model with robust standard errors - do any of the p-values change?

Document Details

Document produced by Christopher Prener, Ph.D. for the Saint Louis University course SOC 5050 - QUANTITATIVE ANALYSIS: APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository README.md file for additional details.