

C528: Predictive Analytics

Homework 1

Due July 2, 11:59pm

Your submission should consist of a neatly formatted PDF report and Stata do files that document your commands.

In this assignment, you will use the dataset *cars89.dta*. This dataset collects data on 107 car models in 1989.

1. Report the sample means and standard deviations of the variables price, weight, horsepower, seating, and engine displacement.
2. Test the hypothesis that the population mean car price is \$19,000, considering 90%, 95%, and 99% degrees of support. Go through each step of the process and explain your reasoning and method. Be sure to include the statement of the null and alternative hypotheses, calculation of the t statistic, and your conclusion.
3. Create a new variable in the dataset indicating whether or not (from the American perspective) a car is imported. Estimate the average treatment effect of a car being imported on price.
4. Use the difference in means test to determine the statistical significance of the average treatment effect of a car being imported. Determine at which degrees of support we can conclude the effect is significant. Be sure to include the statement of the null and alternative hypotheses, calculation of the t statistic, and your conclusion.

5. Estimate the equation

$$price = \beta_0 + \beta_1 weight + \beta_2 horsepower + \varepsilon$$

and report the results.

6. Continue using the model from part 5. Test the hypothesis that an increase of one horsepower corresponds with a price increase of \$70. Go through each step of the process and explain your reasoning and method. Be sure to include the statement of the null and alternative hypotheses, calculation of the t statistic, and your conclusion.
7. Estimate the equation

$$price = \beta_0 + \beta_1 weight + \beta_2 horsepower + \beta_3 displacement + \beta_4 seating + \beta_5 imported + \varepsilon$$

and report the results. The variable *imported* is your dummy variable from part 3. Briefly explain the interpretation of the parameter estimates.

8. Use your estimated model to predict the price of a Mercedes Benz 300E, which has a displacement of 180.8 cubic inches, horsepower of 178, weight of 3131 lb., and five seats.
9. Which additional assumptions would need to be true to use the model in part 8 to make active predictions? Do you think these assumptions are reasonable here? Why or why not?