Dataset Exercises

Chapter 8 Logits

This set of questions changes the model slightly so that we can have a meaningful model with a dummy dependent variable. Instead of trying to 'explain' and predict car acceleration we will now look at the association between the variables we have been using and country of origin.

The equation will be:

USA =f { MPG,ACCEL, HORSE WEIGHT, U)

where USA is a dummy variable, which you should have created in the work for the last chapter.

1. Estimate an OLS version of the above model.

2. Interpret the coefficients and 't' ratios for the results.

3. Work out the predicted likelihood that a car will be American if it has the mean values of all four independent variables.

4. Retrieve the predicted values of USA from your equation [this involves doing the regression again and selecting /////{] in the options.

5. Look at the predicted values to see if any of them lie outside the unit [0-1] interval and are therefore 'unreal'. You can do this by getting the program to produce descriptive statistics for the Predicted Value as this will give max and min values which enables you to see if there are any >1 or <0. Alternately you could ask for a frequency distribution of the whole set of predicted values.

6. Round the predicted values to 1 if they are >0.5 and 0 if they are less than 0.5 and use this to produce a 2x2 contingency table.

7. Now you should produce a logit estimate of the same model. This can be done by going to the menus in this sequence: Analyze>Regression>Binary Logistic

8. Your output now has rather different test statistics. The 't statistic is not shown and the Wald statistic takes its place but you can work out the 't' by dividing the coefficient by the standard error. You can use the Sig in the penultimate column for two-tailed tests of the null of a zero coefficient.

9. Compare the signs of the coefficients and the results for statistical significance between the OLS and the Logit models.

10. The output gives a classification table for the number of cases correct versus actual cases based on rounding from the 0.5 level. You can use this to perform a Chi-square test as a measure of goodness of fit.

11. Discuss the following proposition "the work we have just done is pointless because it does not make sense to think of a car's country of origin as having been 'caused' by things like weight and miles per gallon. These do not make a car become Japanese or American."