

DEPARTMENT OF STATISTICS

Course STATS 330: Advanced Statistical Modelling

Tutorial Sheet 5: August 26, 2010

This tutorial is designed to give you practice in subset selection using APR and stepwise methods. In this tutorial we will again be using the **car data** used in Tutorial 3.

Task 1: Read in the data and create lm objects

If you haven't saved the data from the Tutorial 3, recreate the data and lm objects (i.e. the output from `lm`) as in Tutorial 3.

Task 2: Select a subset

Using the car data (use the final model `recip.no47.lm` derived in Tutorial 3) select a subset of variables. Use the all possible regressions method discussed in Lecture 14.

Task 3: Select a subset using backward elimination

As in Lecture 15, select a subset of variables for the car data using the “manual” version of BE (i.e. repeatedly using the function `lm` rather than `step`, with `direction="backward"`). Compare the result to that obtained with `step`. What do you conclude? What is different?

Note: when creating the lm objects use the form `data=cars.df[-47,]` (the subset form doesn't seem to work well with `step`)

Task 4: Select a subset using forward selection.

Use `step` to select a model using forward selection. Refer to Lecture 15 for the method. Examine the output so you understand what has happened at each step.

Task 5: Use stepwise regression

Use `step` with the option “`both`” to select a subset for the car data. How does this compare with the subset chosen in Tasks 3 and 4?

Task 6: All possible regressions

How do the stepwise methods compare with the all possible regressions method used in Task 2?

Task 7: Predictions

Calculate the predicted values for the model selected using all possible regressions, and the predicted values for the backward elimination procedure. How do these differ?