Assignment 2, due 5pm Thursday 1 April

- Your assignment needs to show the R code you used, and your well discussed answers to the questions (in Word, or some other word processor). There is no one right answer, this assignment is about you exploring data and model fitting.
- Submit your assignments on Learn.

Multiple Linear Regression

Use the dataset CarData.txt to investigate the relationship between the price of cars and a range of possible factors that might influence it. The factors recorded are as follows:

Price: suggested retail price of the used car still in excellent condition.
Mileage: miles the car has been driven
Make: manufacturer of the car
Model: model of the car – ignore this for now
Type: body type such as sedan, coupe, etc.
Cylinder: number of cylinders in the engine
Litre: a more specific measure of engine size
Doors: number of doors
Cruise: indicator variable representing whether the car has cruise control (1 = cruise)
Sound: indicator variable representing whether the car has upgraded speakers (1 = upgraded)
Leather: indicator variable representing whether the car has leather seats (1 = leather)

1. Start by exploring the data. Use R and create summary statistics and plots for each variable (ignore the variable for the model of car).
2. Use suitable graphs and plots to explore the relationship between variables.
3. Next, fit a linear model starting with all the main effects. Reduce your model to the most parsimonious final model. Make sure you look at the residuals.
4. Try some different model fitting methods such as starting with a minimal model and adding terms to it (forward selection), or using the step-both-ways option in R. Compare your final models with your backwards selection model. You can also try some different ways to make decisions, such as compare your final, best model from using AIC, BIC, or, if you have done STAT202, Mallows Cp.
5. Discuss what your final, best model means – what effects car price and in what way?

STAT463 students: include one or two sensible higher order two-way interactions.