MATH 204 - ASSIGNMENT 2

Please Hand in Assignment in the Lecture on Monday 17th March.

FEV (*forced expiratory volume*) is an index of lung function that measures the volume of air expelled after one second of constant effort. The data set to be studied contains measurements of FEV on 654 children ages 6-22 who were seen in the Childhood Respiratory Disease Study in 1980 in East Boston, Massachusetts.

The following variables are recorded:

- **ID**: Child identification number.
- **Age**: Age in years
- **FEV**: FEV in litres
- **Height**: Height in inches
- **Sex**: 1-Female or 2-Male
- **Smoker**: 1-Nonsmoker or 2-Smoker.

The data may be downloaded

- in plain text format from

  http://www.math.mcgill.ca/∼dstephens/204/Data/FEV.txt

- in SPSS format from

  http://www.math.mcgill.ca/∼dstephens/204/Data/FEV.sav

Research question: Which is the most appropriate model for explaining the variation in response variable **FEV** ? Use the regression analysis methods covered in lectures to find the most appropriate model.

20 Marks

Some issues to consider when performing the analysis are:

- The most appropriate model for explaining the variation in FEV is the **simplest** one that fits the data adequately, and that cannot be improved by introducing or removing predictors.
- Stepwise selection is a good method to find the best model. However, you should still look at **residual plots** to verify that the model that is ultimately selected gives a good fit.
- Two of the predictor variables are **continuous covariates**, two are **factor predictors**; you should handle the different types of predictors appropriately.
- It might be necessary to **transform** the response variable to meet the required assumptions (of constant variance/Normality). *An .avi file demonstrating how to perform transformations in SPSS can be found on the course website.*
- A scatterplot of **FEV** versus **Height** reveals that it may be necessary to fit **polynomial** terms in some of the variables.
- A scatterplot of **Age** versus **Height** reveals an interesting (if unsurprising) relationship.

Please limit your solution to no more than SIX sides of paper (8.5in x 11in or 215.9mm x 279.4mm). You may use SPSS, and the output generated, provided that you write comments pointing out the key results.