**MAS3903 Linear Models – Project**

This project is worth 10% of the overall mark for this course.

Before attempting this project, make sure you have read the associated paper, “Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity”. The paper aims to model the effect of the lecturers' beauty on their course evaluations. In this project we are only going to use a subset of the data, namely course evaluations where the class has at least forty-five students.

Your task is to determine which of the variables tenured, minority, age, female, students, lower, and beauty can be used to predict the course questionnaire results. To get the data, copy the file ‘Project Beauty.csv’ on the Blackboard site to your own directory. Then in R, check the directory is set to where you have put the file and then type:

url = "Project Beauty.csv"

project = read.table(url, header=T, sep=',')

**Description of the data:**

* courseevaluation - the questionnaire result.
* tenured - Does the lecturer have tenure? 1 = Yes.
* minority - Does the lecturer come from an ethnic minority (in the USA)? 1 = Yes.
* age
* female - 1 = Female.
* students - Number of students in the class.
* lower - 1 = course is foundation year.
* beauty - beauty score.

**Things to consider**

* Remember to answer the question, i.e. what variables should you keep in the model;
* Do you have to transform the data? What improves in your model if you do transform?
* How are your residuals?
* Does any of the data worry you - outliers, etc?
* Is your model any good? Comment on it.
* Do any interactions, quadratic terms, etc significantly improve the fit of the model
* This data is `real' and as such there isn't a perfect solution. In fact, there is more than one solution that I will accept. What I'm looking for is understanding of the problem!
* How do your results compare to the results in the paper, e.g. Table 3
* Describe your results. Don't just give me an equation. What do your results tell us?
* Part of the difficulty (and fun) of statistics is that we have to work outside our field of expertise. You should try to understand the general flow/theme of the paper, but not worry about every single detail.

**Comments on the project**

* The project will be marked out of 75:  
  55 marks for content, i.e. what statistics you do, 20 marks for style, i.e. introduction/conclusion/labels/spelling.
* The report should be in Word, using a 12point font, or LaTeX, using a comparable font.
* Do not bombard me with R output. Only place the relevant parts in the report, stick the rest in an Appendix if necessary.
* Don't leave it to the last minute.
* Remember to work in a group.
* Write the report for a reader that knows about statistics, i.e. understands about ANOVA, residuals, etc.
* However, you should also make the report concise.
* There is 12 page limit on your report (you can have appendices).
* Only include the residual plots for your final model.
* Explain how/why you have removed variables.
* At the end of the project I would like a short statement saying what each person contributed to the project, e.g. one or two sentences at most. Each person should sign this statement.
* Also state whether everyone should get equal marks for their project, i.e. overall did everyone do about the same level of work.
* **The project should be submitted to the General Office by 3.00pm on Friday 13th December 2019. Please attach a cover sheet for each member of the group to the front of the project. An electronic copy should be emailed to me before the deadline ready to run through Turnitin so I can check for plagiarism.**