MTH5120 Statistical Modelling 1

Assessed Coursework 1

This coursework counts for 15% of your overall module mark.

To complete this coursework you will need to use your own data set which you should have uploaded to QM Plus in week 2 as well as completing a short questionnaire about your data set. The coursework question below asks you to fit a simple linear regression model to your own data in R programming and then to assess that model and comment on your results. The marks for this coursework are all for correct coding in R and interpreting those modelling results in found in R. There are no marks for the quality of the data or how well a linear regression model describes your chosen data.

Please complete the coursework in a MS Word document. You should copy your R code and R output into your Word document along with your own typed comments. R code and output can be copied from R-Studio and pasted into Word without needing to use screen shots. Once in Word if you change the font back to `Courier New` your code and output in Word will format the same way as the R-Studio original. Your MS Word document should then be uploaded to the submission point in QM Plus.

Please ensure that you submit your own work on your own data set. You should be aware of QMUL policies on collusion and plagiarism.

The deadline for submission of this coursework is 5pm UK time on Thursday 22nd February (week 5). Late submissions or submissions by email will not be accepted. If you do not submit by the deadline and do not have an accepted Extenuating Circumstances claim, you will score zero for this coursework.

Question

(a) Load the data set that you submitted previously to QM Plus into R and assign the data to an explanatory variable and a response variable. [4]

(b) Explain briefly (in less than 50 words) why you chose this data set. [3]

(c) Construct a simple linear regression model using your data and display a summary of the model results. Copy your R code and output into your Word document and then write down the values of the least squares estimates of the two regression parameters. [4]

(d) Write in one sentence an interpretation of the model in (c) and its two parameters. [2]

(e) How well does the model in (c) explain your data? Your answer to this part should be typed into your Word document in no more than 500 words. You should use the methods covered in the module lectures and IT labs for assessing a simple linear regression model and then make your own conclusions. Note that the marks for this part will be awarded for the quality of your analysis and for conclusions made from evidence generated in R and not for how well the model explains the data. Where you use output and plots from R to support your conclusions these should be copied into your Word document along with the R code used to generate them. [12]

[Total 25]