UNIVERSITY OF EAST ANGLIA
School of Computing Sciences

COURSEWORK ASSIGNMENT

Module	Applied Statistics
ASSIGNMENT TITLE	Course Work 2
DATE AND TIME TO HAND IN	Wednesday 7th March 2018
ASSIGNMENT VALUE	25 %
SET BY	Dr Nikolaos Korfiatis

Aims

To help assimilate material presented in the lectures by providing appropriate exercises.

Learning outcomes:

To improve understanding of the material by working individually on problems based on the material.

Assessment criteria:

1. Accuracy of answers.
2. Understanding of subject displayed.
3. Clarity of explanations of working.
4. Quality of reporting.

As always credit is given for a persuasive argument.

Description of assignment:

Solution of problems related to material in lectures.

1. Present solutions to all problems.
2. Giving full working that demonstrates the steps required to obtain the correct solution and the important principles used.
3. The logic of the solution is important

Handing in procedure:

Please submit your piece of coursework in the following way:

1. Attach this coursework assignment sheet to your piece of coursework. **This should be a word document reporting your results as per the lecture notes. Do not exceed the 10 pages with minimum font size 11pt.**

2. In addition to your report submit a printed `R` file with all the regression commands.

Plagiarism

Plagiarism is the copying or close paraphrasing of published or unpublished work, including the work of another student; without due acknowledgement. Plagiarism is regarded a serious offence by the University, and all cases will be investigated. Possible consequences of plagiarism include deduction of marks and disciplinary action, as detailed by UEA's Policy on Plagiarism and Collusion.

If you have any questions drop by my office (Thomas Paine Study Centre 0.15) or contact me at N.Korfiatis@uea.ac.uk

To explore which patient demographic characteristics and organisational factors influence the length of stay (LOS) among medical admissions, we record data for several hospitals. The data obtained are as follows (SPSS data set length.sav must be read in R):

Variable	description
id	id
age	Patients' average age
risk	Risk of infection in the hospital
rcr	Ratio of the number of infected to the number of uninfected (×100)
xray	Ratio of the number of xrays to the number of patients (×100)
beds	Average number of beds
aff	University Affiliation (0: No, 1: Yes)
avdaily	Average number of admissions per day
nurses	Average number of nurses
fac	Percentage of surgeries
slength	LOS

1. Determine the best regression model with the backward elimination and with the forward selection. Explain what happens in every step. Compare the models and select the best regression model. *[marks 20]*
2. Check the adequacy of this model. *[marks 10]*
3. If the assumptions are violated transform the dependent variable to correct the model and determine the best regression model. *[marks 20]*
4. Report any outlier or influential value in this model. *[marks 10]*
5. In this model decide whether the addition of the interaction term between age and aff contributes substantially to the prediction of the length of stay. **Use all the taught tests and methods**. *[marks 10]*
6. Adding the above interaction term which are the predicted regression models for
   (a) University affiliation
   (b) Non-University affiliation.
   [marks 10]
7. Interpret the regression coefficients in the best regression model with the interaction term. *[marks 15]*
8. Interpret the multiple R^2 in the best regression model with the interaction term. *[marks 5]*