

PUSASQF206
Application of IT – Basics of R

Time: 2 hours
Total Marks: 60 marks

HealthStar is a health insurance company in the country of Actuarial. You are an analyst at the company and are asked to analyse the claims experience of the past 6 months since the inception of the company.

The claims department has provided the data of the claims in a file called *Claims_Experience.csv* along with the details regarding what each column means :

LOCATION : The location of residence of the insured

JOB : Profession of the insured

SEX : Biological gender of the insured

AGE : Age of the Insured

CLAIM : Amount of the claim paid by the insurer.

- (a) Import the file in R and print the number of rows it contains to the console. (1)
- (b) Fit a linear regression model to the data with CLAIM as the response variable and AGE as the explanatory variable.
Provide your interpretation of the model by commenting on the p – value of the F test. (4)
- (c) Generate all the diagnostic plots of the model in (b) and comment on whether the underlying assumptions of the linear regression model are fulfilled. (6)
Your manager has asked you to update the methods and instead fit a GLM model by still assuming that the CLAIM are from a Normal distribution.
- (d) Fit an equivalent model to the model in (b) (2)
- (e) Add the main effects of the other variables to the model in (d) and comment on the effect of each variable to the claim amount. (4)
- (f) Another suggestion mentions that the log of claim amounts is a better fit for the normal distribution than the claim amounts directly. Change the link function of the model with all explanatory variables remaining the same. Compare the model with that in (e) (4)
- (g) By performing forward selection and AIC as a criteria, determine which is the best possible model.