Class Exercise 1: Claims Data

Use the Claims data set.

Select only the nonzero claims.

Randomly sample 1000 observations from that data set using your date of birth as a seed.

Fit different distributions to the claim size variable and choose the most appropriate one.

Build a model containing the best selection of variables, validate the model and interpret the final model.

Present your analysis in the form of a short report.

Example of code for sampling:

set.seed(20)

mysample<-sample(1:67856,replace=F,size=1000)

mydata<-car[mysample,]

attach(mydata)

1. Remember, first get rid of observations with zero claims, then take your random sample of 1000 observations.

2. You may find that doing your data exploration on the logged transformed response makes for clearer displays. but then use the untransformed response in your modelling but with a log link specified.

3. You should follow the steps I described in the Introductory lectures: Data Exploration, Model Building, Model Validation, Presentation and Interpretation of your final model.

4. Summarise approach and results concisely and clearly in your 4 page report. Attach R program and additional computer output ( but not too much) . Try not to copy computer output straight into report.

Hand in Date 29th July 12 noon at Statistical Sciences Reception