Project description

(due April 24, Monday)

Overview

In the semester project, you will have the opportunity to apply the models we learn in class, to an actuarial problem. Your team may have up to four students. You are encouraged to work in a team, but solo work is also fine if you prefer that.

Instructions

Imagine you are a team of actuaries working for an actuarial consulting group (you may come up with a name for your group). One day a client approaches you and asks you to construct a rating engine. The client provides you a dataset of insurance claim frequencies and severities. You should choose *one* of the following datasets:

1. *The Swedish motor insurance dataset*: A dataset compiled by the Swedish Committee on the Analysis of Risk Premium in Motor Insurance. More details are available on the PDF description file posted on D2L.

2. *The 2019 SoA Student Research Case Study Challenge data*: A dataset of claim frequencies and severities for multiple lines for an imaginary automobile insurance company called Safelife. More details about the context of this dataset is available on the original case study challenge handout: Click here.

3. *The LGPIF data*: A dataset of local government property insurance claims experience from the Wisconsin Office of the Commissioner of Insurance. This is a real dataset of government property insurance covering structures owned by Wisconsin local government entities. It covers structures like, for example, the Mitchell Park Dome, etc.

The datasets are posted on D2L. Once you choose your data, follow the instructions below.

- Install R (`https://www.r-project.org`) and RStudio (`https://rstudio.com`).
- Read the data into RStudio, and create a training sample and a validation sample. Fit your models using the training sample, and test them using the validation sample.
- Create some figures and tables of summary statistics of the data.
- Use a frequency-severity model pair to determine the rate for each policyholder. The values for the parameters of your model should be obtained using maximum likelihood estimation (MLE). Your engine should charge a higher rate for policies with higher exposure. Justify your model.
- Report the summary statistics, model parameters, standard errors, and prediction results for the models you have attempted, in the format of an actuarial report.

Grading criteria

1. Formatting (30%): Your report should be in the format of an actuarial report. This means it should be neatly presented for your client, with an executive summary including a description of the actuarial problem, a section on the assumptions you make, details of your analysis, and your actuarial recommendation.

2. Data summary (30%): Your report should give the reader a good idea of what the dataset looks like. You can use creative figures and summary tables for this part.

3. Analysis (40%): Your analysis should properly apply MLE using the training sample, report the estimates and standard errors for your model parameters, and validate your model using the validation sample. The report should recommend a solution to the problem defined in your executive summary. You must justify your model using the diagnostic methods, which we will cover in class.

Resources & advice

1. The 2019 SoA Student Research Case Study Challenge has examples of good actuarial reports. Click: https://www.soa.org/research/opportunities/2019-student-case-study/ and click one of the winning submissions for that year.

2. The R code, which you write for homework 3 and 4 will be a good place to start. Example MLE code will be posted on D2L and discussed during class.

3. Start early by working on the summary statistics first. This will help you get familiar with your data, and give you some time until we learn the model diagnostic methods later during the course. A good way to create figures in R is by using libraries such as ggplot.

4. Discuss about your progress with other students, and the instructor!

For honors credit

If you are working on an honors project, then here are some ideas for you:

1. Compare two (or more) models and determine which model fits the data better,
2. or calculate insurance rates under imaginary coverage modifications,
3. or you may have your own idea, which is also fine. ☺