SF2930 Regression analysis VT2017

Project 2

The project should be done in groups of **two**.

- A computer written[1], self-containing, report of the subjects presented below should be sent by email to `flrios@math.kth.se` no later than 6th of March. The subject of the email should be "SF2930 Project 2: Full Name 1, Full Name 2"
- In addition to this, hand in your resulting factors on the form given by the template TariffFactors.xlsx found on the course site, no later than March 2nd by email to `flrios@math.kth.se`. The subject of the email should be "SF2930 Project 2: Full Name 1, Full Name 2"

Introduction

A tractor is a vehicle designed to deliver a high torque at slow speeds, mostly used in agriculture or construction. In Sweden, most of these vehicles are registred in Vägtrafikregistret, and are therefore required by law to have a third part liability insurance. Many tractor owners complement this legally required insurance with an insurance covering vehicle damage to their own tractor. If P&C seeks your help to price this insurance as risk correct as possible, i.e. you are going to make your own tractor tariff on the form

$$price = \gamma_0 \prod_{k=1}^{M} \gamma_{k,i} \tag{1}$$

where γ_0 is the base level and $\gamma_{k,i}, k = 1, ..., M$ are the risk factors corresponding to variable number k and variable group number i. For example, let $k = 1$ be VehicleAge and for one particular tractor the age is 3 years old. Then, according to below table, $\gamma_1 = 0.95$.

VehicleAge group i	Risk factor $\gamma_{1,i}$
1: Age <= 1	1.00
2: Age = 2	0.98
3: Age = 3	0.95
4: Age = 4	0.90
5: Age >= 5	0.85

[1] Preferably using LaTeX

The project

1. Risk differentiation

What describes the risk? Use a multiplicative GLM model to model both claims frequency and claim severity separately. Use the same variables and variable groups in both models for simplicity, and propose the final risk factor $\gamma_{k,i}$, where risk = claim frequency $\cdot$ claim severity.

Present and explain your choice of risk arguments, grouping of data and risk factors. How does this comply with Likelihood Ratio Test and different measures for goodness of fit discussed in this course? Compare all tests and measures.

The template GLM.R contains a structure for the GLM-analysis which you can use.

2. Levelling

Having found the risk factors $\gamma_{k,i}$, determine the base level γ_0. Note that γ_0 is estimated already by the GLM-program, but this base level corresponds to the total claims cost of the analysis data, not hte policies that are active today. This is done in the following two steps:

a. Estimate the expected claim cost for 2015. Assume that If P&C has a ratio target between the estimated claim cost and the total premium of 90% – what should the total sum of tractors' premium be to accommodate this target?

b. Determine the base level γ_0 using the results found in a), combined with the factors $\gamma_{k,i}$ found from the risk differentiation.

All tariffs will compete against each other – so make sure to describe the risk as good as you can!

Dataset

In the file Tractors.csv you will find claims history on vehicle damage for all tractors insured by If P&C during 2004-2014. The file also contains information regarding the insured tractor and the company that owns the tractor in the following manner:

RiskYear	VehicleAge	Weight	Climate	ActivityCode	Duration	NoOfClaims	ClaimCost
2010	009	3830	North	Construction	0.63	1	627 099
2008	001	400	South	Missing	0.59	1	253 850
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

where *RiskYear* states the year of the claim and insurance period, *VehicleAge* and *Weight* denotes the age and weight of the tractor respectively, *Climate* denotes the geographical location in Sweden where the tractor is used and *ActivityCode* is the activity code registred on the company that owns the tractor. For every tractor there is also

information regarding *Duration* - namely the share of the risk year the tractor was insured - and the number of claims and claim cost tied to that insurance period, denoted by *NoOfClaims* and *ClaimCost*.