

KUNGLIGA TEKNISKA HÖGSKOLAN

SF2930 REGRESSION ANALYSIS

Report II

Isac Karlsson
Ludvig Wärnberg Gerdin

Examiner
TATJANA PAVLENKO

March 9, 2020

Contents

1	Introduction	2
1.1	Background	2
1.2	Data	2
1.3	Problem description	2
1.3.1	Risk Differentiation and Grouping	2
1.3.2	Levelling	2
2	Methods	2
2.1	Grouping and Risk Differentiation	2
2.1.1	Levelling	3
3	Results	3
4	Conclusion	3

1 Introduction

1.1 Background

Most of the tractors in Sweden have a third party liability insurance, because they are required by law. In southern Europe a few large players have dominated the sales of tractor insurances. Our main task this project is to create our own tractor tariff on the form:

$$\text{Price} = \gamma_0 \prod_{k=1}^M \gamma_{k,i} \quad (1)$$

Here γ_0 corresponds to the base level and $\gamma_{k,i}$ are the risk factors and corresponds to each individual tractor.

1.2 Data

If P&C have provided us with data to train this price model, example given in the table below.

1.3 Problem description

1.3.1 Risk Differentiation and Grouping

Using GLM analysis we aim to make each group Risk homogeneous and that they contain enough data to get a stable GLM analysis, meanwhile handling imperfections in the dataset.

1.3.2 Levelling

Here we aim to calculate γ_0 such that the forecasted claim costs for each insurance are covered by the the price for each insurance, on a full year basis. We use a ratio between the estimated claim cost and the total premium of 90%. Lastly we calculate the base level γ_0 from the formula given in (1).

2 Methods

2.1 Grouping and Risk Differentiation

The criteria on which we based our groups were that

1. Each group should be risk homogeneous, and
2. Each group should have enough data to make the GLM estimates stable.

Greater emphasis were placed on fulfilling criteria 2) due to it being more concrete. In order to do that we considered cut-offs that placed a fairly equal shares of data in each risk group.

The resulting cut-offs and risk groups are found in section 3.

2.1.1 Levelling

Firstly, we subset the data rows (or customers) that had a `RiskYear` 2016. That way the GLM analysis were only conducted on the active customers. not those that weren't customers to If anymore. From the aggregated data, initialized as `glmdata2` in the provided script, we calculated the yearly claim-cost per tractor by dividing the summed duration in each duration by the corresponding summed claim cost in

In pseudo-code, the process was

```
for each
```

3 Results

4 Conclusion