

Pricing Game

Data Science for Actuaries

July-November 2015

1 Database Description

Two datasets are available, online¹

<http://freakonometrics.free.fr/training.csv>

<http://freakonometrics.free.fr/pricing.csv>

Dataset `training` contains 100,000 unique insurance policies², for years 2009 and 2010, with the following variables

`PolNum`: policy number

`CalYear`: underwriting year

`Gender`: gender of the driver of the car

`Type`: type of car

`Category`: category of the car

`Occupation`: occupation of the driver

`Age`: age of the driver

`Group1`: group of the car

`Bonus`: Bonus Malus (French no claim discount)

`Poldur`: Age of the insurance contract

`Value`: Value of the car

`Adind`: Indicator for a 'material' cover

`SubGroup2`: Subregion of home of the driver

¹Those two datasets can be imported in R using `read.csv2()`. For other statistical language, be careful when importing the datasets: the first column (index) has no header.

²There are 100,000 unique policies, 21 of them are observed twice (i.e. over two years), so overall, there are 100,021 rows.

Group2: Region of home of the driver
Density: Population density
Expdays: Exposure (in days)
Numtppd: Number of third party material claims
Numtpbi: Number of third party bodily injury claims
Indtppd: Total cost of third party material claims
Indtpbi: Total cost of third party bodily injury claims

Dataset **pricing** contains the same variables, excepted those related to claims for 2011. It contains 36,311 insurance policies

2 Instructions

The goal is to compute a yearly third party premium (material + bodily injury) for the 36,311 contracts of the **pricing** dataset, for 2011. Are expected

- a dataset (a csv file), with two columns; the first one **PolNum** is the policy number and the second one is **Premium**, the proposed pure premium, with 36,611 rows
- the code for the pricing (in R, Python, etc.) and a brief summary of the methodology, including a brief description of the variable used.

Thanks for indicating a team name for each premium computed, and to send the csv file, the code, and the brief summary to PricingGame@institutdesactuares.com.

3 Rules

Once collected all the premium, we will work as a price aggrgator: each of the 36,311 drivers will chose the insured according to some criteria (that will be discussed), e.g.

- cheapest insurance company
- random selection among the cheapest insurance companies
- random matching for year 2010, and then choice based on retention probabilities (remains with the same insurance company if not too expensive).

The organisers have the right to normalize the premiums to avoid dumping strategies (normalization meaning that the sum of the premiums will be equal to the sum of the losses).

The target is to get the best combined ratio.

The results will be discussed during the 100% actuaires day, in Paris (and a details post will be published on <http://freakonometrics.hypotheses.org>).

For further information, PricingGame@institutdesactuaire.com or [@freakonometrics](#).