

100% Actuaire: exercice de pricing

Formation Data Science pour l'Actuariat

Juillet-Novembre 2015

1 Description des Bases de Données

Deux jeux de données sont disponibles en ligne¹

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

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

La base `training` est constituée de 100,000 polices différentes², pour les années 2009 et 2010, et est composée des variables suivantes

PolNum: numéro de police

CalYear: année calendaire de souscription

Gender: genre du conducteur

Type: type de véhicule

Category: catégorie du véhicule

Occupation: profession

Age: âge du conducteur

Group1: groupe du véhicule

Bonus: Bonus Malus

Poldur: Ancienneté du contrat

Value: Valeur du véhicule

Adind: Indicateur d'une garantie dommages

SubGroup2: Sous-région d'habitation

¹Les bases se lisent directement en R à l'aide de la fonction `read.csv2()`. Pour tout autre logiciel, il faudra faire attention lors de l'importation: la première colonne (indice) n'a pas d'intitulé, ce qui peut causer un décalage des noms des variables lors de l'importation.

²Il y a 100,000 polices distinctes, dont 21 présentes les deux années, soit en tout 100,021 observations en tout.

- Group2:** Région d'habitation
- Density:** Densité de population
- Expdays:** Exposition (en jours)
- Numtppd:** Nombre de sinistres RC matériels
- Numtpbi:** Nombre de sinistres RC corporels
- Indtppd:** Coût total des sinistres RC matériels
- Indtpbi:** Coût total des sinistres corporels

La base `pricing` est composée des mêmes variables, sauf celles relatives aux sinistres. Elle est composée de 36,311 polices pour l'année 2011.

2 Instructions

L'objectif est de proposer une prime annuelle RC global (RC matériels et RC corporels) pour les 36,311 contrats de la base `pricing`, pour l'année 2011. Les attendus sont

- une base sous forme de fichier `csv`, contenant deux variables `PolNum` et `Prime`, et 36,611 lignes (les 36,311 polices de la base `pricing`)
- le code utilisé pour le `pricing` (en R, Python, etc.) et un (rapide) descriptif de la méthodologie utilisée, ainsi que des variables retenues.

Merci d'indiquer un nom d'équipe pour chacun des tarifs proposés et d'envoyer le tout à l'adresse arthur.charpentier@gmail.com.

3 Fonctionnement du Jeu

Une fois collectées toutes les primes, nous créerons alors un agrégateur de prix, et chacun des 36,311 assurés choira l'assureur suivant quelques règles qui seront discutées lors de la journée, par exemple

- choix de l'assureur le moins cher
- choix de l'assureur au hasard parmi les moins chers (avec des probabilités fonctions des primes, avec une probabilité plus forte d'aller chez le moins cher)
- répartition aléatoire des entre les assureurs pour l'année 2010 puis choix de l'assureur en tenant compte de la distance entre la prime la plus basse offerte par le marché en 2011, et la prime offerte par son 'ancien' assureur.

Les organisateurs s'autorisent à normaliser les primes proposées par un participant afin d'éviter toute stratégie de *dumping*. La normalisation se fera de manière proportionnelle, de telle sorte que la somme des primes égale la somme des coûts de sinistres.

Le but du jeu n'est pas d'avoir la plus grosse part de marché, mais le meilleur ratio combiné.

Le comportement du marché sera discuté lors de la journée 100% actuaires.

Pour davantage d'information, par courriel arthur.charpentier@gmail.com ou par Twitter [@freakonometrics](https://twitter.com/freakonometrics).