

ÉCOLE D'ACTUARIAT
UNIVERSITÉ LAVAL

Travail pratique 1

Guillaume MICHEL
Nathanaël PELCHAT
Mikael ROBERTSON
Olivier TURCOTTE

AUTOMNE 2018

1 Sommaire exécutif

Table des matières

1	Sommaire exécutif	2
2	Analyse des données	4
3	Modèle proposé	6
3.1	Équation	6
3.2	Variables qualitatives	6
3.3	Interactions	6
3.4	Interprétation	6
3.5	Statistiques	6
4	Analyse des résidus	6
5	Prévisions	6
6	Recommandations	6
	Annexes	6
A	Erreurs de données	6
B	Transformation	8
C	Sélection des variables	8

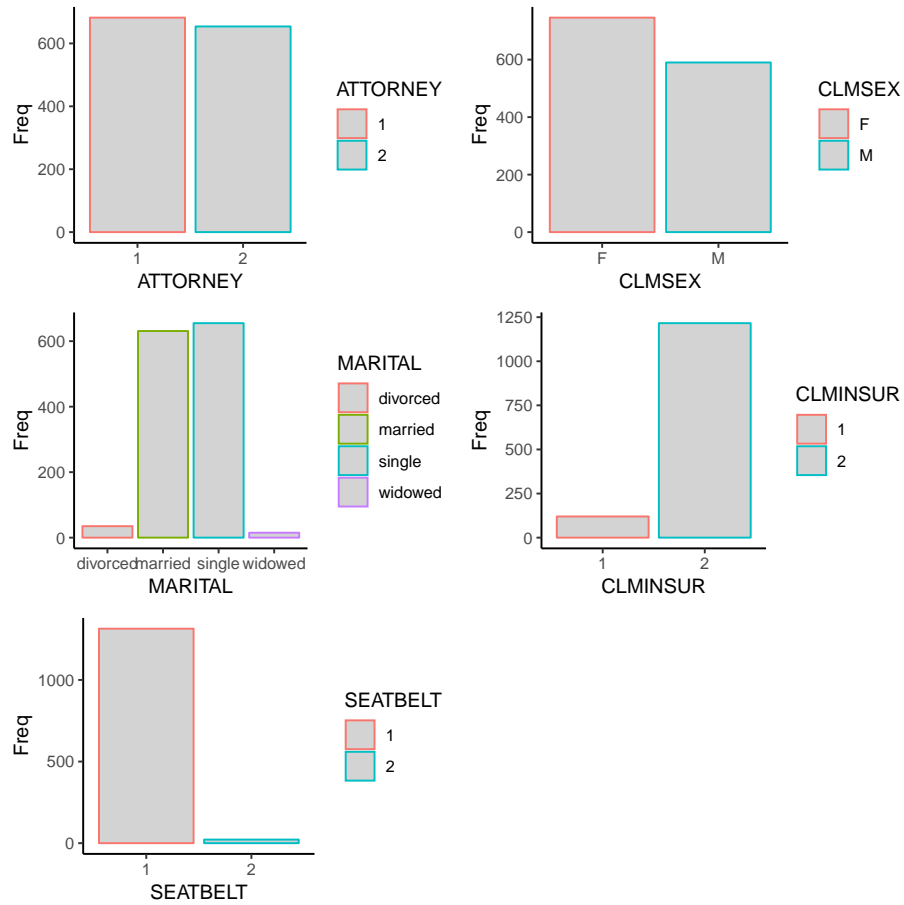
2 Analyse des données

Voici les variables disponibles afin d'effectuer un modèle prédictif de la perte économique :

Tableau 1 – Description des variables

Variables	Type	Description
CASENUM	Valeur entière	Numéro d'identification de la réclamation
ATTORNEY	Variable indicatrice	Indique si le réclamant est représenté par un avocat
CLMSEX	Variable indicatrice	Indique le sexe du réclamant
MARITAL	Variable polytomique	Indique le statut marital du réclamant
CLMINSUR	Variable polytomique	Indique si le réclamant est assuré
SEATBELT	Variable polytomique	Indique si le réclamant portait une ceinture de sécurité
CLMAGE	Valeur entière	Âge du réclamant
LOSS	Valeur continue	Perte économique totale du réclamant en milliers de dollars

Ces variables sont en majorité qualitative. Une analyse de fréquences de celles-ci permet d'avoir un meilleur ressenti quant à leurs interaction avec la variable exogène LOSS :



##	CLMAGE	LOSS
##	Min. : 0.00	Min. : 0.005
##	1st Qu.: 21.00	1st Qu.: 0.640
##	Median : 33.00	Median : 2.331
##	Mean : 33.01	Mean : 5.965
##	3rd Qu.: 41.00	3rd Qu.: 3.998
##	Max. : 610.00	Max. : 1067.697

3 Modèle proposé

3.1 Équation

3.2 Variables qualitatives

3.3 Interactions

3.4 Interprétation

3.5 Statistiques

4 Analyse des résidus

5 Prévisions

6 Recommendations

Annexes

A Erreurs de données

La base de données originelles utilisé dans la création du modèle a dû subir quelques modifications afin d'être utilisable. Voici les quelques erreurs répertoriés ainsi que les techniques utilisés pour les rectifier :

1. Fautes d'ortographe

(a) MARITAL

Cette colonne contient à l'origine plusieurs faute de frappe des états maritaux. Afin d'unifier le tout, il a fallut substituer les états dans ces quatres variables distinctes : *divorced*, *widowed*, *married*, *single*.

(b) CLMSEX

Cette colonne supposé contenir les états **F** ou **M** contient à l'origine quelques états *male*. Afin d'unifier le tout, ces états ont été substituer en **M**.

2. Données aberrantes

(a) LOSS

Cette colonne contient une valeur très extrême de *1067.697*. En analysant le boxplot ci-dessous, on voit bien que la valeur est très énorme à comparer au reste. Ainsi, la ligne contenant cette valeur a été retiré afin de préserver au mieux les données jugés adéquates.

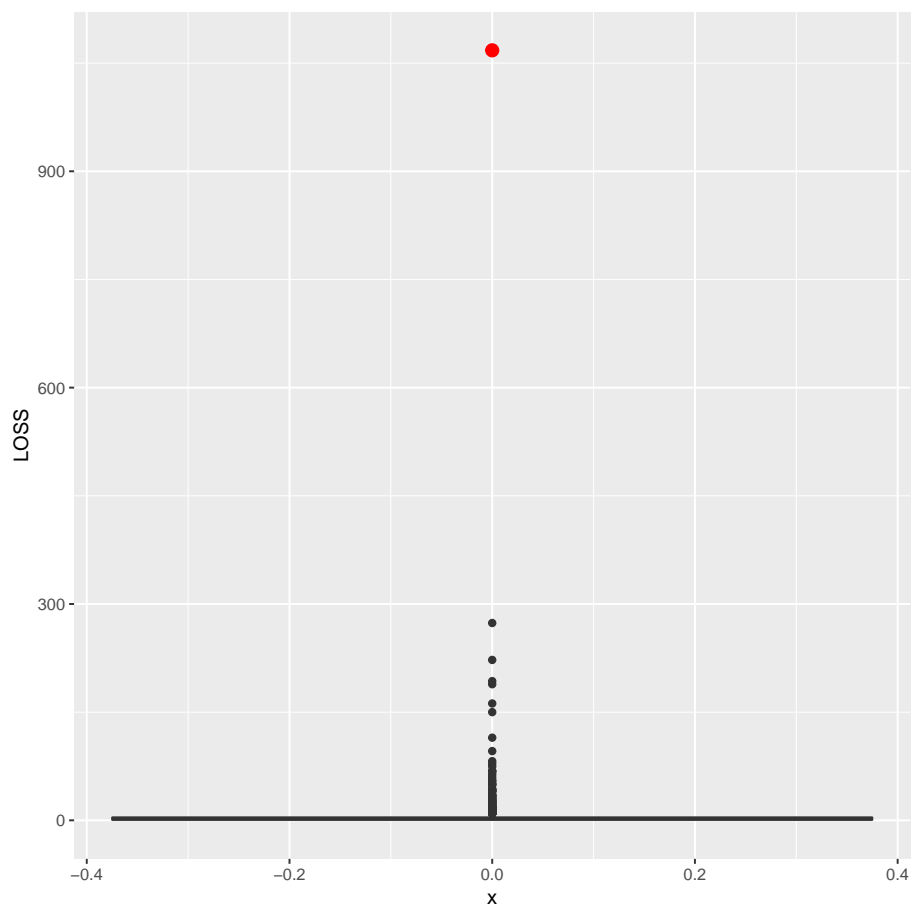
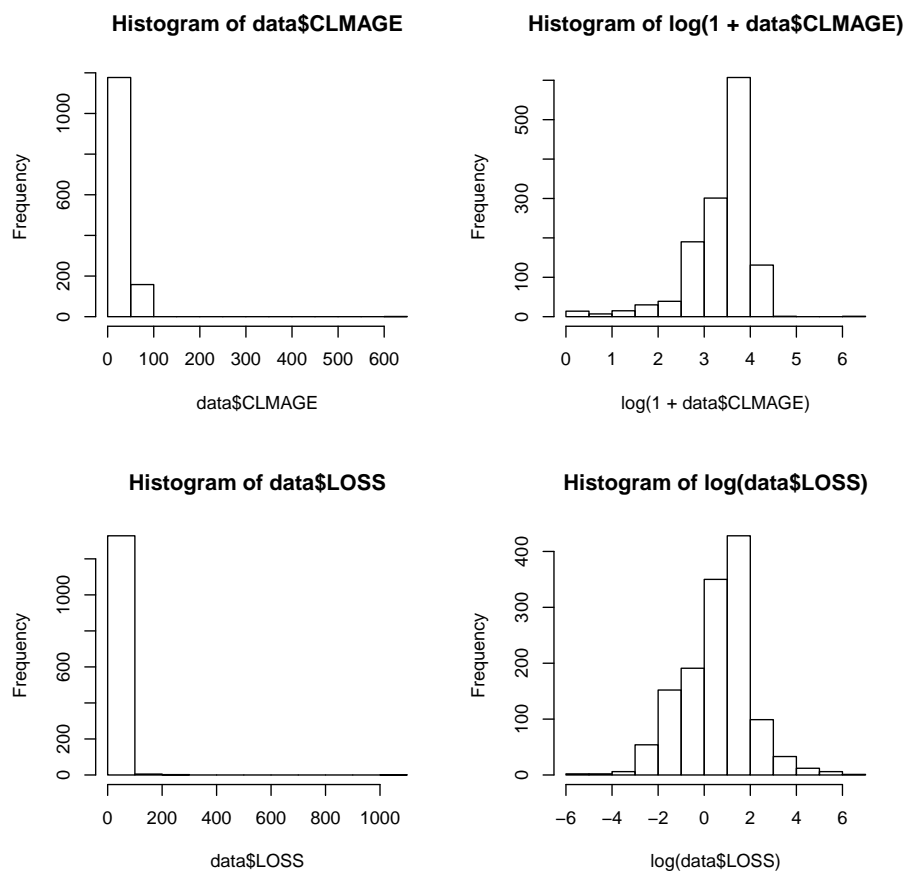


FIGURE 1 – Boxplot de LOSS

(b) CLMAGE

Cette colonne contient une valeur de *610*, valeur impossible selon la description de la variable, soit l'âge du réclamant. Afin de ne pas fausser les résultat, la ligne contenant cette valeur a été retiré.

B Transformation



C Sélection des variables