

Etude de cas

Modélisation d'une garantie d'un produit d'assurance auto

Etude à partir d'une base de données confidentielle de sinistres de
l'INSEE

Rapport réalisé par :

Rakotondrazaka Diamondra

Vitel Victor

Introduction

L'objectif est de modéliser la garantie d'un produit d'assurance automobile afin d'avoir une tarification segmentée grâce aux sinistres qu'il prend en charge au titre de ces garanties, à partir d'une base de données des sinistres.

La base de données que nous avons à notre disposition regroupe environ 934 000 individus assurés et 44 variables décrivant le profil des assurés (marque de voiture, nombre d'enfant, bonus/malus...). Les variables d'intérêts sont l'exposition (au risque de sinistres), le nombre et la charge de sinistres de bris de glace.

Nous effectuerons en premier lieu, une analyse univariée des critères explicatifs et un retraitement des données. Nous modéliserons la fréquence et du coût moyen des sinistres de chaque garantie via un modèle Linéaire Généralisé.

Ensuite, nous enrichirons la base de données avec des données externes, ce qui mènera à une seconde modélisation plus approfondie.

Enfin, l'étude se conclura après validation et comparaisons des modèles, une analyse de la prime pure et des résidus des modèles de chaque garantie.

Nous utiliserons pour cette modélisation le logiciel SAS.

Sommaire

Introduction.....	2
I Analyse univariée.....	3
1) Nettoyage de la table et codification de toutes les variables explicatives utilisées	3
2) Statistiques descriptives	6
II Modélisations	8
1) Première modélisation	8
2) Deuxième modélisation.....	9
III Validation et comparaison des modèles	10

I Analyse univariée

1) Nettoyage de la table et codification de toutes les variables explicatives utilisées

Cette étape a pour but d'adapter les variables afin de tester le pouvoir explicatif de chaque variable en vue de la modélisation. Nous allons catégoriser les variables quantitatives de la base de données afin de regrouper ou éventuellement supprimer les modalités qui ne contiennent pas assez d'exposition et de sinistre. Nous modifions également les variables qualitatives en renommant certaines classes et en effectuant des regroupements. Nous expliquons ci-dessous comment nous avons retravaillé les variables :

-Les marques de voiture sont regroupées par pays d'origine, les pays le moins représentés sont codés par la modalité « Autre ».

Nouvelle variable : CLASSEM

-On remplace les numéros et indices du CSP et du segment par les noms associés de la classe.

Nouvelles variables : CSPclasse, Segmentclasse

-Pour la variable ALIMENTATION, les modalités sans sinistres sont supprimées, celles avec le moins d'expositions et de sinistres sont regroupées dans la catégorie « AUTRE », la modalité IDs, étant remarquée comme une erreur de frappe, est corrigée en « IDS ».

Nouvelle variable : ALIMENTATION_p

-Les carrosseries sans sinistres sont supprimées et les autres sont regroupées par la première lettre de leur appellation (« B », « C » et celles restantes moins représentées : « Autre »).

Nouvelle variable : CARROSSERIE_p

-Les classes de réparation sans sinistres sont abandonnées et les regroupées en groupe de lettres suivant l'alphabet.

Nouvelle variable : CLASSE_REPARATON_p

-La modalité « P » de la variable GENRE est supprimée car elle est sans sinistres.

-Les variables CRM et ANNEE_05 portent toutes les deux sur le bonus-malus, donc une variable donnant des catégories de valeurs des coefficients de celui-ci est créée. Le coefficient vaudra donc « 0,5 » quand l'individu l'aura à cette valeur d'une durée en année supérieur à 0, (ANNEE_05 > 0), sinon les valeurs numériques CRM sont catégorisées (« 0,5 à 06 », « 0,6 à 0,7 » ...).

Nouvelle variable : COEFF_BONUS_MALUS

-Le nombre d'enfant est classé de la manière suivante : « 0 », « 1 », « 2 », « >=3 ».

Nouvelle variable : NB_ENFANT_p

-On prendra la variable GROUPE_SRA concernant la puissance car elle est mieux classée que la variable puissance. Cependant les premières et dernières modalités ayant moins de sinistres et d'expositions sont regroupées.

Nouvelle variable : GROUPE_SRA_p

-L'âge et le cout du pare-brise du véhicule sont catégorisés en un nombre de modalité restreint.

Nouvelles variables : AGE_VEHICULE_p, COUT_PARE_BRISE_p

-Les classes de prix sont groupées par groupes de lettres suivant l'alphabet.

Nouvelle variable : CLASSE_PRIX_p

-Concernant la variable CODE_POSTAL_GARAGE, on va plutôt s'intéresser aux régions des variables pour avoir moins de modalités à étudiées, donc on va placer dans une nouvelle variable DEPARTEMENT_GARAGE les 2 premiers chiffres du code postal, et ensuite associer à ces derniers, la région correspondante.

Nouvelle variable : REGION

-Les modalités de ENERGIE sont regroupées en 2 classes avec respectivement dans chacune une des 2 modalités concentrant le nombre de sinistres et d'exposition.

Nouvelle variable : ENERGIE_p

Liste des critères explicatifs obtenus :

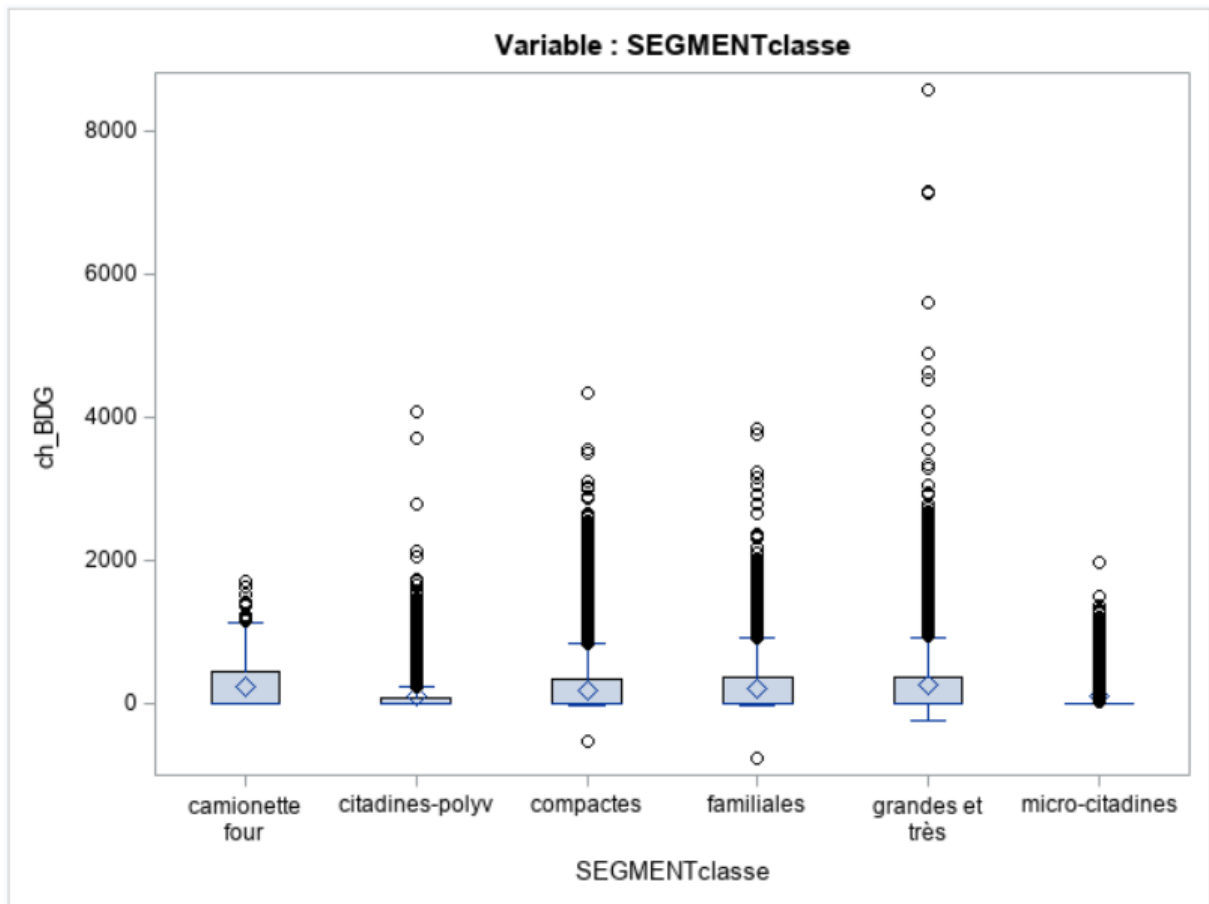
COEFF_BONUS_MALUS	CLASSEM	CSPclasse
SEGMENTclasse	GROUPE_SRA_p	REGION
ALIMENTATION_p	USAGE	MODE_VIE
AGE_VEHICULE_p	GARAGE	CARROSSERIE_p
classe_prix_p	CLASSE_REPARATION_p	NB_ENFANTS_p
COUT_PARE_BRISE_p	ENERGIE_p	GENRE
TYPE_LOGEMENT	STATUT_LOGEMENT	REGION

Tableau 1 : Variables sélectionnées pour réaliser la modélisation

2) Statistiques descriptives

Une fois le recodage effectué, nous réalisons une analyse exploratoire des données. Nous nous intéressons à la distribution de l'exposition, des charges et nombres de bris de glaces selon les modalités des critères explicatifs.

Exemple :

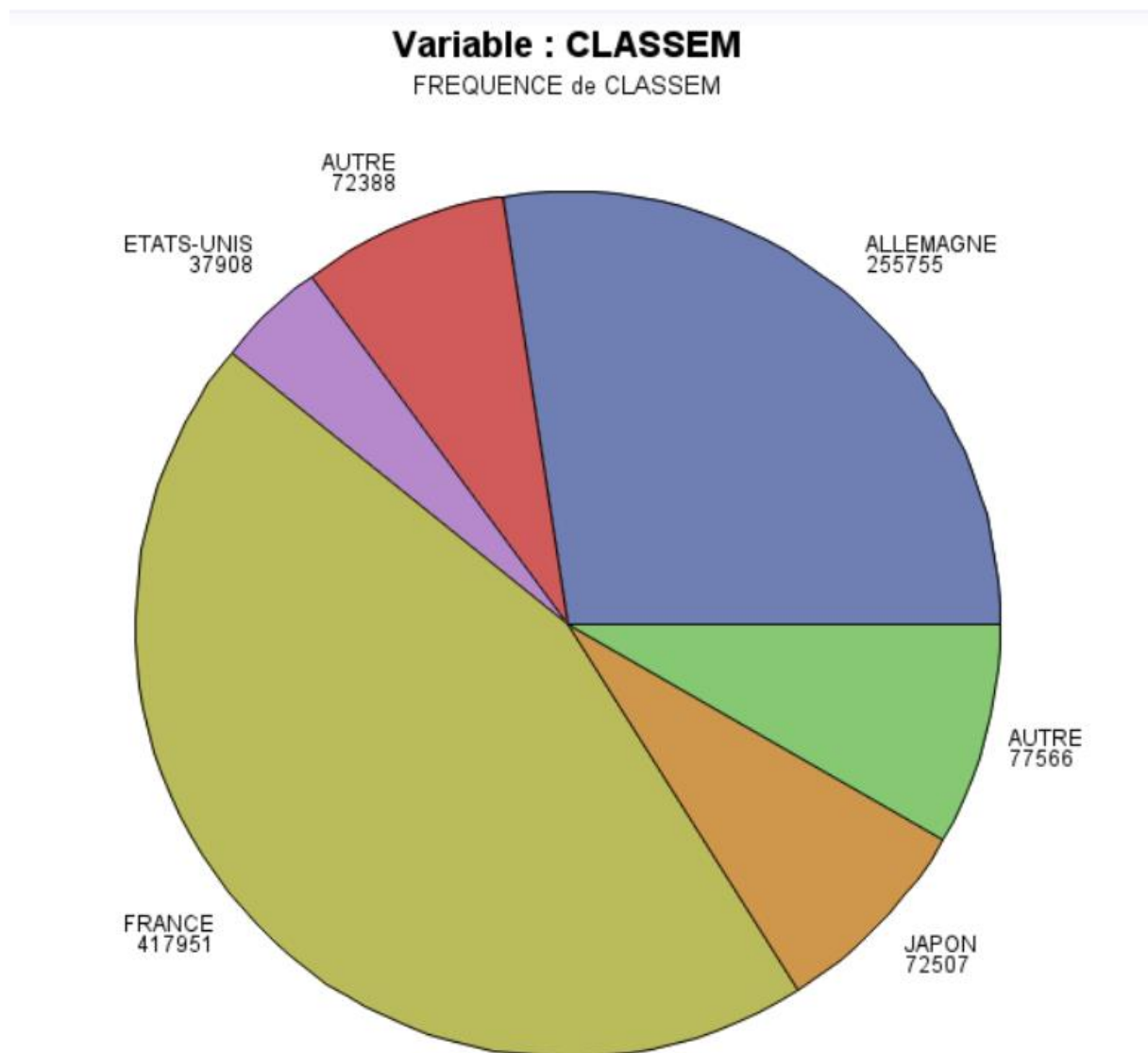


Graphique 1 : Diagramme en boîte de la charge de bris de glace en fonction du segment auquel appartient la voiture

La charge de sinistres de bris de glace semble moins élevée pour les citadines-polyvalentes et les micro-citadines, et équivalente pour les autres.

Nous pouvons ensuite étudier la répartition des modalités des critères explication pour voir quelles sont celles les plus présentes.

Exemple :



Graphique 2 : Diagramme circulaire des origines de marques de voiture

Il y a presque ici une majorité de marques d'origine française, la deuxième origine de marque la plus observée est allemande (environ un quart des observations).

II Modélisations

1) Première modélisation

À partir des critères explicatifs obtenus, nous allons effectuer deux modélisations linéaires généralisées, pour la fréquence des sinistres et leur coût moyen sur 60% des données de la table de données initiale.

Dans chaque modèle, nous retirerons une à une les variables non significatives ($p\text{-value} > 0.05$) de façon décroissante.

Nous obtenons les modèles suivants :

Statistique LR pour Analyse de Type 3			
Source	DDL	Khi-2	Pr > khi-2
CLASSEM	8	216.03	<.0001
CSPclasse	15	31.57	0.0074
SEGMENTclasse	5	35.14	<.0001
GROUPE_SRA_p	9	47.06	<.0001
REGION	12	42.45	<.0001
ALIMENTATION_p	3	13.60	0.0035
AGE_VEHICULE_p	5	49.90	<.0001
GARAGE	6	17.11	0.0089
classe_prix_p	5	65.35	<.0001
COUT_PARE_BRISE_p	4	215.91	<.0001
TYPE_LOGEMENT	2	23.22	<.0001

Tableau 2 : Modèle du coût moyen des sinistres

Statistique LR pour Analyse de Type 3			
Source	DDL	Khi-2	Pr > khi-2
COEFF_BONUS_MALUS	4	20.52	0.0004
CLASSEM	8	43.21	<.0001
CSPclasse	15	52.72	<.0001
SEGMENTclasse	5	81.56	<.0001
GROUPE_SRA_p	9	35.23	<.0001
REGION	12	586.07	<.0001
ALIMENTATION_p	3	8.01	0.0458
USAGE	3	56.71	<.0001
AGE_VEHICULE_p	5	224.87	<.0001
GARAGE	6	66.11	<.0001
CARROSSERIE_p	2	67.87	<.0001
ENERGIE_p	1	72.43	<.0001
GENRE	1	8.13	0.0044
STATUT_LOGEMENT	2	11.09	0.0039

Tableau 3 : Modèle du nombre des sinistres

2) Deuxième modélisation

Pour la deuxième modélisation, la base de données est enrichie avec des données externes (jointure avec les codes INSEE de communes). À partir de données d'une base de l'INSEE sur une trentaine d'indicateurs décrivant la population, les logements, les revenus, l'emploi et les établissements au niveau communal. Nous allons nous intéresser à la densité et au salaire médian par commune. On divise le nombre d'habitant de la commune par la superficie de la commune pour avoir la densité. Ces deux variables quantitatives sont ensuite catégorisées de la même manière que les autres.

On obtient les modèles suivants :

Statistique LR pour Analyse de Type 3			
Source	DDL	Khi-2	Pr > khi-2
CLASSEM	8	199.50	<.0001
CSPclasse	15	32.39	0.0057
SEGMENTclasse	5	28.32	<.0001
GROUPE_SRA_p	9	48.85	<.0001
REGION	12	42.37	<.0001
ALIMENTATION_p	3	12.12	0.0070
AGE_VEHICULE_p	5	48.89	<.0001
GARAGE	6	17.13	0.0088
classe_prix_p	5	68.26	<.0001
COUT_PARE_BRISE_p	4	209.32	<.0001
TYPE_LOGEMENT	2	13.67	0.0011
densite_p	3	8.26	0.0408

Tableau 4 : deuxième modèle sur coût moyen des sinistres

Statistique LR pour Analyse de Type 3			
Source	DDL	Khi-2	Pr > khi-2
COEFF_BONUS_MALUS	4	22.06	0.0002
CLASSEM	8	43.35	<.0001
CSPclasse	15	56.59	<.0001
SEGMENTclasse	5	87.13	<.0001
GROUPE_SRA_p	9	31.12	0.0003
REGION	12	577.20	<.0001
ALIMENTATION_p	3	9.44	0.0239
USAGE	3	47.74	<.0001
AGE_VEHICULE_p	5	222.74	<.0001
GARAGE	6	59.43	<.0001
CARROSSERIE_p	2	68.55	<.0001
ENERGIE_p	1	60.48	<.0001
GENRE	1	8.12	0.0044
STATUT_LOGEMENT	2	12.52	0.0019
densite_p	3	31.44	<.0001

Tableau 5 : Deuxième modèle sur le nombre de sinistres

Les deux modélisations gardent les mêmes variables, mais la deuxième retient en plus la densité comme variable significative (modélisation du coût moyen des sinistres).

III Validation et comparaison des modèles

À partir des coefficients obtenus pour chaque modalité des variables significatives après la deuxième modélisation linéaire généralisée, nous estimons la prime d'assurance de chaque individu.

On relève d'abord pour chaque individu, les valeurs des coefficients des modèles linéaires généralisés, des modalités.

On obtient ensuite la fréquence estimée (mod_freqsin) avec les coefficients de la modélisation du nombre de sinistres :

mod_freqsin = produit des exponentiels des coefficients des modalités du modèle

Ainsi que le coût moyen estimé des sinistres (mod_csin) avec les coefficients de la modélisation coût moyen des sinistres :

mod_csin = produit des exponentiels des coefficients des modalités du modèle

On en déduit une estimation de la prise en charge (cout) de sinistres :

cout = mod_freqsin * mod_csin

On a à présent une tarification segmentée (dépendant des modalités de chaque individu), de la charge de bris de glace.

Prime totale calculée par la modèle et charge totale :

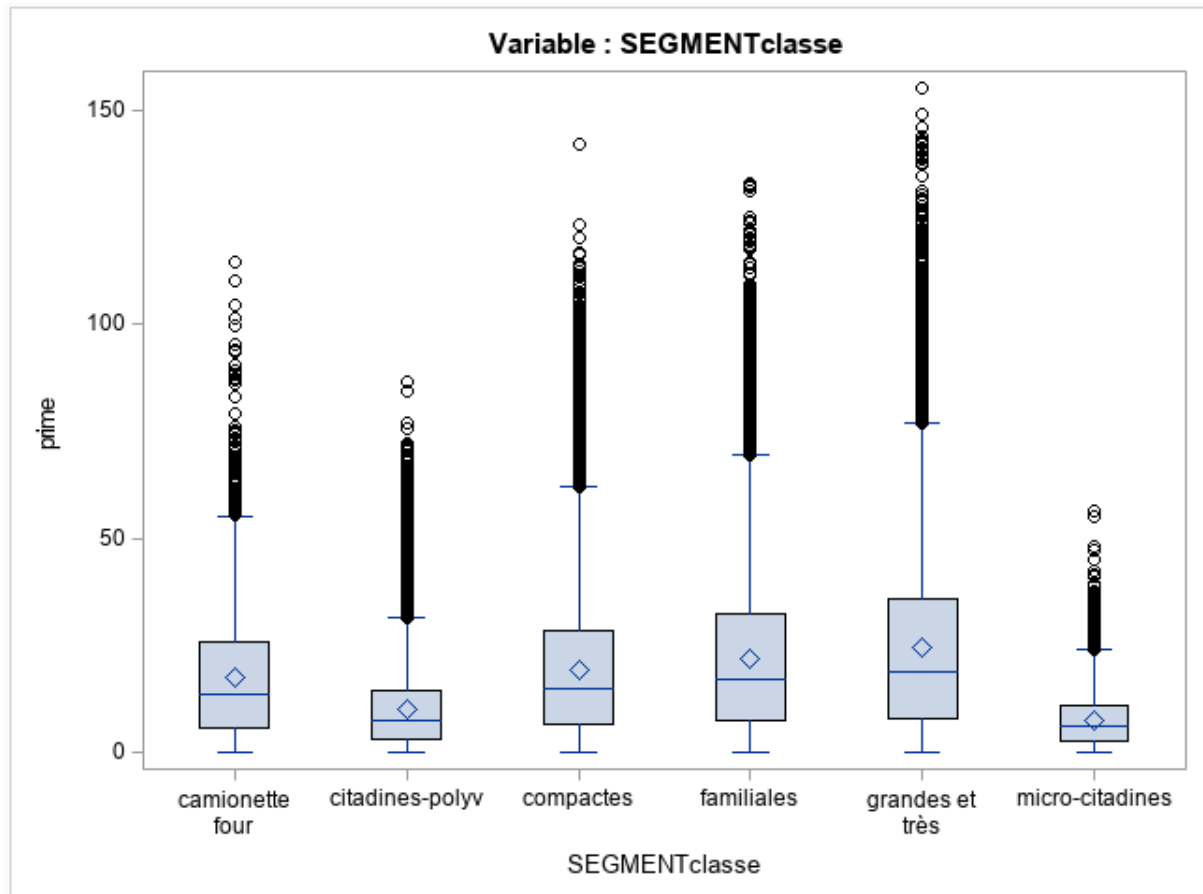
Elles sont calculées par la somme des coûts de la prise estimé de la prise en charge des sinistres multipliées par les expositions et par la somme des charges des sinistres de bris de glace.

Voici les valeurs trouvées :

prime	charge
6225078	6409860

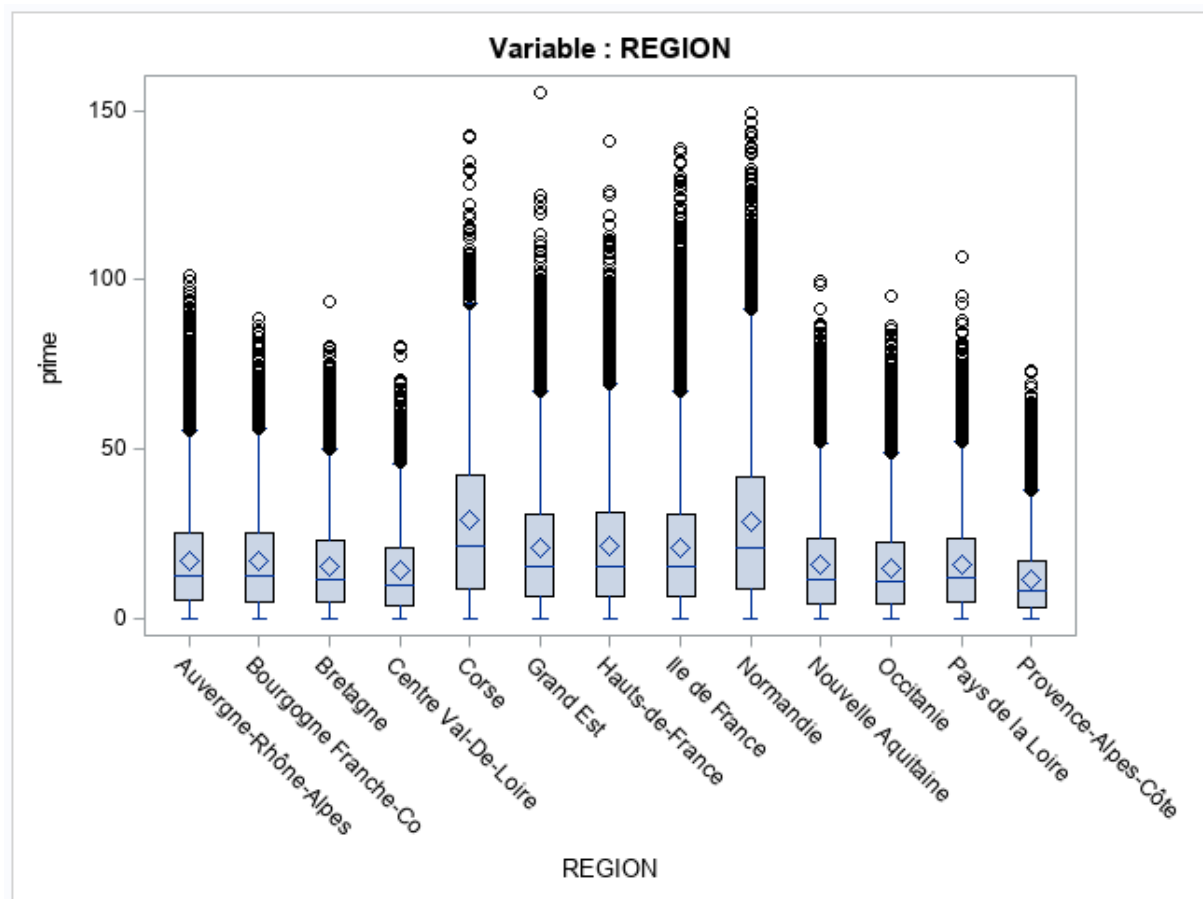
Une analyse exploratoire de la prime pure est réalisée pour chaque critères de tarification. On s'est donc intéressé à la distribution de celle-ci en fonction des modalités de ces critères.

Exemples :



Graphique 3 : distribution de la prime, en fonction du segment de la voiture.

Comme la charge de bris de glace, la prime versée à l'assurance pour la prise en charge de sinistres de bris de glace, est en moyenne plus faible pour les citadines-polyvalent et micro-citadines.



Graphique 4 : distribution de la prime, en fonction de la région.

Au vu des moyennes, médianes et troisième quartile, la prime est plus chère dans les régions Corse et Normandie.

Conclusion : on a obtenu un écart de 184 782 euros entre la prime et la charge totale du modèle. Cet écart plutôt grand n'est pas négligeable. Le modèle est donc perfectible en modifiant les codages des variables ou en rajoutant des variables via des données externes.

Références :

<https://www.insee.fr/fr/statistiques/1405599?geo=COM-50051> (données externes)

<https://documentation.sas.com/>

Annexe du code SAS :

```
libname dir 'C:\Users\Ascensio\Downloads\Etude_de_cas';

/*création de la table de données */

data work.table_modif1;
set TMP1.assurance;
run;

proc sql;          /*calculs du nb total d'expo, de sinistre, charge de bruit
de glace */

select sum(expo) as expo, sum(nb_sinbdg) as nb_sinbdg, sum(ch_BDG) as
ch_BDG, sum(nb_sinbdg) / sum(expo) as freq, sum(ch_BDG) / sum(nb_sinbdg) as
CM from table_modif1;

quit;

proc sql;          /*calculs du nombre d'exposition, de charge de sinistres de
bris degace et de nombre de sinistres dans les modalités d'un
critère explicatif (C) donné*/

select &var, sum(expo) as expo, sum(nb_sinbdg) as nb_sinbdg, sum(ch_BDG) as
ch_BDG,
sum(nb_sinbdg) / sum(expo) as freq, sum(ch_BDG) / sum(nb_sinbdg) as CM from
table_modif1 group by /*(C)*/;

quit;

/*propriétés de variables quantitatives et corrélations */

proc means data = TMP1.assurance;
var expo ch_BDG nb_sinbdg AGE_VEHICULE COUT_PARE_BRISE NB_ENFANTS;
run;

proc corr data = TMP1.assurance;
var expo ch_BDG nb_sinbdg AGE_VEHICULE COUT_PARE_BRISE NB_ENFANTS;
run;

/*recodage des variables */

data table_modif1 ;
set TMP1.assurance;

/*définir le pays de la marque de la voiture*/

length CLASSEM $20. ;
if MARQUE ="ALFA ROMEO" or MARQUE ="BERTONE" or MARQUE ="FIAT"
or MARQUE ="IVECO " or MARQUE ="LANCIA" or MARQUE ="PIAGGIO" then CLASSEM =
"ITALIE";
```

```

        else if MARQUE = "ALPINE" or MARQUE = "RENAULT" or MARQUE
="CITROEN" or MARQUE = "DS" or MARQUE = "GRANDIN" or MARQUE = "MATRA" or
MARQUE = "MEGA" or MARQUE = "PEUGEOT" or MARQUE = "RENAULT" or MARQUE
="TALBOT" or MARQUE = "Z_FICTIF" then CLASSEM = "FRANCE";
        else if MARQUE = "AUDI" or MARQUE = "BMW" or MARQUE = "MCC" or
MARQUE = "MERCEDES" or MARQUE = "OPEL" or MARQUE = "PORSCH" or MARQUE
="SMART" or MARQUE = "VOLKSWAGEN" then CLASSEM = "ALLEMAGNE";
        else if MARQUE = "AUSTIN" or MARQUE = "BEDFORD" or MARQUE
="DAIMLER" or MARQUE = "JAGUAR" or MARQUE = "LAND ROVER" or MARQUE = "LDV" or
MARQUE = "LOTUS" or MARQUE = "MG " or MARQUE = "MINI" or MARQUE = "ROVER" or
MARQUE = "TRIUMPH" then CLASSEM = "GRANDE-BRETAGNE";
        else if MARQUE = "BUICK" or MARQUE = "CADILLAC" or MARQUE
="CHEVROLET" or MARQUE = "CHRYSLER" or MARQUE = "DODGE" or MARQUE = "FORD"
or MARQUE = "JEEP" or MARQUE = "PONTIAC" or MARQUE = "TESLA" then CLASSEM =
"ETATS-UNIS";
        else if MARQUE = "DACIA " or MARQUE = "LADA" or MARQUE = "SAAB" or
MARQUE = "SKODA" or MARQUE = "TATA" or MARQUE = "VOLVO" then CLASSEM = "AUTRE";
        else if MARQUE = "DAEWOO " or MARQUE = "HYUNDAI " or MARQUE = "KIA
" or MARQUE = "SSANGYONG" then CLASSEM = "COREE DU SUD" ;
        else if MARQUE = "EBRO " or MARQUE = "SANTANA" or MARQUE = "SEAT"
then CLASSEM = "ESPAGNE";
        else if MARQUE = "DAIHATSU" or MARQUE = "HONDA " or
MARQUE = "INFINITI " or MARQUE = "ISUZU " or MARQUE = "LEXUS" or MARQUE = "MAZDA"
or MARQUE = "MITSUBISHI" or MARQUE = "NISSAN" or MARQUE = "SUBARU" or
MARQUE = "SUZUKI" or MARQUE = "TOYOTA" then CLASSEM = "JAPON";
        /*else if MARQUE ~= . ; */

/*CSP*/

if CSP = 10 then CSPclasse = "Agriculteur/Pêcheur";
else if CSP = 21 then CSPclasse = "Artisan";
else if CSP = 22 then CSPclasse = "Commerçant";
else if CSP = 23 then CSPclasse = "Chef d'entreprise";
else if CSP = 31 then CSPclasse = "Profession libérale";
else if CSP = 32 then CSPclasse = "Fonctionnaire et assimilé:
cadre";
        else if CSP = 35 then CSPclasse = "Professionnel du spectacle";
        else if CSP = 36 then CSPclasse = "Salarié du privé: cadre";
        else if CSP = 41 then CSPclasse = "Fonctionnaire et assimilé:
technicien";
        else if CSP = 42 then CSPclasse = "Fonctionnaire de
l'enseignement";
        else if CSP = 47 then CSPclasse = "Salarié du privé:
technicien";
        else if CSP = 50 then CSPclasse = "Salarié du privé:
employé/ouvrier";
        else if CSP = 51 then CSPclasse = "Fonctionnaire et assimilé:
employé/ouvrier";
        else if CSP = 53 then CSPclasse = "Policier/Militaire";
        else if CSP = 70 then CSPclasse = "Retraité";
        else if CSP = 82 then CSPclasse = "Sans activité
professionnelle";
        else if CSP = 84 then CSPclasse = "Etudiant";
        else if CSP = 85 then CSPclasse = "Sans activité et en recherche
d'emploi";
        else if CSP = 97 then CSPclasse = "Forain";
        else if CSP = 99 then CSPclasse = "V.R.P / Taxi";
        /*else if CSP ~= . ; */

```

```

/*Changement du format de la date */

format DATE_PREM_CIRCULATION_d ddmmyy10.;
DATE_PREM_CIRCULATION_d = DATEPART(DATE_PREM_CIRCULATION);

/* SEGMENT*/

if SEGMENT = "A" then SEGMENTclasse = "micro-citadines";
else if SEGMENT = "B" then SEGMENTclasse = "citadines-
polyvalentes";
else if SEGMENT = "H" then SEGMENTclasse = "grandes et très
grandes berlines";
else if SEGMENT = "K1" or SEGMENT = "K2" then SEGMENTclasse =
"camionnette fourgonnette";
else if SEGMENT = "M1" then SEGMENTclasse = "compactes";
else if SEGMENT = "M2" then SEGMENTclasse = "familiales";
/*else if SEGMENT ~= . ; */

/*ENERGIE */

if ENERGIE = "AU" or ENERGIE = "EL" or ENERGIE = "ES" then
ENERGIE_p="AU, EL, ES";
if ENERGIE = "GO" or ENERGIE = "GP" then ENERGIE_p="GO, GP";

/* Carrosserie */

length CARROSSERIE_p $10.;
if CARROSSERIE = "CPE" or CARROSSERIE = "LIM" or CARROSSERIE =
"TTS" or CARROSSERIE = "BER" then DELETE;
if CARROSSERIE = "BE+" or CARROSSERIE = "BE3" or CARROSSERIE
="BE4" or CARROSSERIE="BE5" or CARROSSERIE="BK+" or CARROSSERIE="BK3" or
CARROSSERIE="BK5" then CARROSSERIE_p = "B";
else if CARROSSERIE = "CAB" or CARROSSERIE = "CPC" or
CARROSSERIE = "CTF" then CARROSSERIE_p = "C";
else CARROSSERIE_p = "Autre";

/* CLASSE REPARATION */
length CLASSE_REPARATION_p $5. ;
if CLASSE_REPARATION = "F" or CLASSE_REPARATION = "H"
or CLASSE_REPARATION = "HC" or CLASSE_REPARATION = "I" then
CLASSE_REPARATION_p = "F à I";
if CLASSE_REPARATION = "J" or CLASSE_REPARATION = "K"
or CLASSE_REPARATION = "L" or CLASSE_REPARATION = "M" then
CLASSE_REPARATION_p = "J à M";
if CLASSE_REPARATION = "N" or CLASSE_REPARATION = "O"
or CLASSE_REPARATION = "P" or CLASSE_REPARATION = "Q" then
CLASSE_REPARATION_p = "N à Q";
if CLASSE_REPARATION = "R" or CLASSE_REPARATION = "S"
or CLASSE_REPARATION = "T" or CLASSE_REPARATION = "U" or
CLASSE_REPARATION = "V" then CLASSE_REPARATION_p = "R à V";

/*GENRE */

if GENRE = "P" then DELETE ;

/*Date d'acquisition */

```

```

format DATE_ACQUISITION_d ddmmyy10.;
DATE_ACQUISITION_d = DATEPART(DATE_ACQUISITION);

/*CRM + ANNEE_05 */

length COEFF_BONUS_MALUS $10.;
if ANNEE_05 > 0 then COEFF_BONUS_MALUS = "0.5";

if CRM <= 0.6 then COEFF_BONUS_MALUS = "0.5 à 0.6";
if 0.6<CRM and CRM <= 0.7 then COEFF_BONUS_MALUS = "0.6 à 0.7";
if 0.7<CRM and CRM<= 0.8 then COEFF_BONUS_MALUS = "0.7 à 0.8";
if 0.8<CRM and CRM<= 0.9 then COEFF_BONUS_MALUS = "0.8 à 0.9";
if 0.9<CRM then COEFF_BONUS_MALUS = ">0.9";

/*NB_ENFANTS*/

length NB_ENFANTS_p $3.;
if NB_ENFANTS <= 0 then NB_ENFANTS_p = "0" ;
if NB_ENFANTS = 1 then NB_ENFANTS_p = "1" ;
if NB_ENFANTS = 2 then NB_ENFANTS_p = "2" ;
if NB_ENFANTS >= 3 then NB_ENFANTS_p = ">=3" ;

/*GROUPE_SRA*/

if GROUPE_SRA < 29 then GROUPE_SRA_p = "20 à 27";
if 36<GROUPE_SRA then GROUPE_SRA_p = "37 à 46";
if GROUPE_SRA = 29 then GROUPE_SRA_p = "29";
if GROUPE_SRA = 30 then GROUPE_SRA_p = "30";
if GROUPE_SRA = 31 then GROUPE_SRA_p = "31";
if GROUPE_SRA = 32 then GROUPE_SRA_p = "32";
if GROUPE_SRA = 33 then GROUPE_SRA_p = "33";
if GROUPE_SRA = 34 then GROUPE_SRA_p = "34";
if GROUPE_SRA = 35 then GROUPE_SRA_p = "35";
if GROUPE_SRA = 36 then GROUPE_SRA_p = "36";

/* GARAGE */

if GARAGE = -1 or GARAGE = 0 then DELETE;

/* Age du véhicule */

length AGE_VEHICULE_p $10. ;
if AGE_VEHICULE < 4 then AGE_VEHICULE_p = "<4 ans";
if AGE_VEHICULE >=4 and AGE_VEHICULE< 8 then AGE_VEHICULE_p =
"4 à 8 ans";
if AGE_VEHICULE >=8 and AGE_VEHICULE< 12 then AGE_VEHICULE_p =
"8 à 12 ans";
if AGE_VEHICULE >=12 and AGE_VEHICULE < 16 then AGE_VEHICULE_p =
"12 à 16 ans";
if AGE_VEHICULE >=16 and AGE_VEHICULE< 20 then AGE_VEHICULE_p =
"16 à 20 ans";
if AGE_VEHICULE >=20 then AGE_VEHICULE_p = ">20 ans";

/* Cout du pare brise */

```



```

length COUT_PARE_BRISE_p $20. ;
if COUT_PARE_BRISE < 200 then COUT_PARE_BRISE_p = "<200";
if COUT_PARE_BRISE >=200 and COUT_PARE_BRISE< 400 then
COUT_PARE_BRISE_p = "200 à 400";
if COUT_PARE_BRISE >=400 and COUT_PARE_BRISE< 600 then
COUT_PARE_BRISE_p = "400 à 600";
if COUT_PARE_BRISE >=600 and COUT_PARE_BRISE< 800 then
COUT_PARE_BRISE_p = "600 à 800";
if COUT_PARE_BRISE >=800 then COUT_PARE_BRISE_p = ">800 ";

/*Alimentation */

length ALIMENTATION_p $5. ;
if ALIMENTATION = "GNV" or ALIMENTATION = "ONJ" then DELETE;
if ALIMENTATION = "CAR" or ALIMENTATION = "CTU" or ALIMENTATION
= "ELC" or ALIMENTATION = "ELP" or ALIMENTATION = "ELS" or ALIMENTATION =
"GPL" or ALIMENTATION = "HES" or ALIMENTATION = "HGO" or ALIMENTATION =
"IND" then ALIMENTATION_p ="Autre";
if ALIMENTATION = "INS" then ALIMENTATION_p = "INS";
if ALIMENTATION = "IDS" or ALIMENTATION = "IDs" then
ALIMENTATION_p = "IDS";
if ALIMENTATION = "INJ" then ALIMENTATION_p = "INJ";

/*Classe prix */
length classe_prix_p $5. ;
if classe_prix = "A" or classe_prix = "B" or classe_prix = "C"
or classe_prix = "D" then classe_prix_p ="A à D";
if classe_prix = "E" or classe_prix = "F" or classe_prix = "G"
or classe_prix = "H" or classe_prix = "HC" then classe_prix_p ="E à H";
if classe_prix = "I" or classe_prix = "J" or classe_prix = "K"
or classe_prix = "L" then classe_prix_p ="I à L";
if classe_prix = "M" or classe_prix = "N" or classe_prix = "O"
or classe_prix = "P" then classe_prix_p ="M à P";
if classe_prix = "Q" or classe_prix = "R" or classe_prix = "S"
or classe_prix = "T" then classe_prix_p ="Q à T";
if classe_prix = "U" or classe_prix = "V" then classe_prix_p
="U à V";

/* DEPARTEMENT */

DEPARTEMENT_GARAGE = substr(CODE_POSTAL_GARAGE, 1, 2);

/*Régions*/

if DEPARTEMENT_GARAGE= "03" or DEPARTEMENT_GARAGE="63" or
DEPARTEMENT_GARAGE="15" or DEPARTEMENT_GARAGE="43" or
DEPARTEMENT_GARAGE="42" or DEPARTEMENT_GARAGE="69" or
DEPARTEMENT_GARAGE="07" or DEPARTEMENT_GARAGE="26" or
DEPARTEMENT_GARAGE="38" or DEPARTEMENT_GARAGE="01" or
DEPARTEMENT_GARAGE="74" or DEPARTEMENT_GARAGE="73" then REGION ="Auvergne-
Rhône-Alpes";
if DEPARTEMENT_GARAGE="84" or DEPARTEMENT_GARAGE="13" or
DEPARTEMENT_GARAGE="83" or DEPARTEMENT_GARAGE="06" or
DEPARTEMENT_GARAGE="04" or DEPARTEMENT_GARAGE="05" then REGION = "Provence-
Alpes-Côte d'Azur";
if DEPARTEMENT_GARAGE="30" or DEPARTEMENT_GARAGE="34" or
DEPARTEMENT_GARAGE="11" or DEPARTEMENT_GARAGE="66" or

```

```

DEPARTEMENT_GARAGE="09" or DEPARTEMENT_GARAGE="81" or
DEPARTEMENT_GARAGE="12" or DEPARTEMENT_GARAGE="46" or
DEPARTEMENT_GARAGE="82" or DEPARTEMENT_GARAGE="32" or
DEPARTEMENT_GARAGE="65" or DEPARTEMENT_GARAGE="31" then REGION =
"Occitanie";
    if DEPARTEMENT_GARAGE="64" or DEPARTEMENT_GARAGE="40" or
DEPARTEMENT_GARAGE="47" or DEPARTEMENT_GARAGE="33" or
DEPARTEMENT_GARAGE="24" or DEPARTEMENT_GARAGE="19" or
DEPARTEMENT_GARAGE="23" or DEPARTEMENT_GARAGE="87" or
DEPARTEMENT_GARAGE="16" or DEPARTEMENT_GARAGE="17" or
DEPARTEMENT_GARAGE="79" or DEPARTEMENT_GARAGE="86" then REGION = "Nouvelle
Aquitaine";
    if DEPARTEMENT_GARAGE="85" or DEPARTEMENT_GARAGE="44" or
DEPARTEMENT_GARAGE="49" or DEPARTEMENT_GARAGE="72" or
DEPARTEMENT_GARAGE="53" then REGION = "Pays de la Loire";
    if DEPARTEMENT_GARAGE="22" or DEPARTEMENT_GARAGE="29" or
DEPARTEMENT_GARAGE="35" or DEPARTEMENT_GARAGE="56" then REGION =
"Bretagne";
    if DEPARTEMENT_GARAGE="50" or DEPARTEMENT_GARAGE="14" or
DEPARTEMENT_GARAGE="61" or DEPARTEMENT_GARAGE="27" or
DEPARTEMENT_GARAGE="76" then REGION = "Normandie";
    if DEPARTEMENT_GARAGE="62" or DEPARTEMENT_GARAGE="80" or
DEPARTEMENT_GARAGE="60" or DEPARTEMENT_GARAGE="02" or
DEPARTEMENT_GARAGE="59" then REGION = "Hauts-de-France";
    if DEPARTEMENT_GARAGE="78" or DEPARTEMENT_GARAGE="77" or
DEPARTEMENT_GARAGE="75" or DEPARTEMENT_GARAGE="91" or
DEPARTEMENT_GARAGE="92" or DEPARTEMENT_GARAGE="93" or
DEPARTEMENT_GARAGE="94" or DEPARTEMENT_GARAGE="95" then REGION = "Ile de
France";
    if DEPARTEMENT_GARAGE="67" or DEPARTEMENT_GARAGE="68" or
DEPARTEMENT_GARAGE="88" or DEPARTEMENT_GARAGE="54" or
DEPARTEMENT_GARAGE="57" or DEPARTEMENT_GARAGE="55" or
DEPARTEMENT_GARAGE="52" or DEPARTEMENT_GARAGE="10" or
DEPARTEMENT_GARAGE="51" or DEPARTEMENT_GARAGE="08" then REGION = "Grand
Est";
    if DEPARTEMENT_GARAGE="25" or DEPARTEMENT_GARAGE="90" or
DEPARTEMENT_GARAGE="39" or DEPARTEMENT_GARAGE="70" or
DEPARTEMENT_GARAGE="21" or DEPARTEMENT_GARAGE="71" or
DEPARTEMENT_GARAGE="58" or DEPARTEMENT_GARAGE="89" then REGION = "Bourgogne
Franche-Comté";
    if DEPARTEMENT_GARAGE="45" or DEPARTEMENT_GARAGE="18" or
DEPARTEMENT_GARAGE="36" or DEPARTEMENT_GARAGE="41" or
DEPARTEMENT_GARAGE="28" or DEPARTEMENT_GARAGE="37" then REGION = "Centre
Val-De-Loire";
    if DEPARTEMENT_GARAGE="28" or DEPARTEMENT_GARAGE="2A" then
REGION = "Corse";

run;
/*Statistiques descriptives : boxplots de l'expo du nombre et charge de
bris de glace en fonction de chaque critère explicatif*/

%let var =
CLASSEM
CSPclasse
SEGMENTclasse
COEFF_BONUS_MALUS
NB_ENFANTS_p
GROUPE_SRA_p
USAGE
MODE_VIE
GARAGE

```

```

ALIMENTATION_p
classe_prix_p
CARROSSERIE_p
CLASSE_REPARATION_p
COUT_PARE_BRISE_p
ENERGIE_p
GENRE
TYPE_LOGEMENT
STATUT_LOGEMENT
AGE_VEHICULE_p
REGION
;

%macro boxplots;
    %do i = 1 %to 20;
        %let variable = %scan( &var,&i);
        proc sgplot data = table_modif1;
            title "Variable : &variable";
            vbox expo /category = &variable; /*ch_BDG nb_sinbdg */
            run;
        %end;
        title;
    %mend boxplots;

%boxplots;
proc freq data = table_modif1;
tables CARROSSERIE_p;
run;

/* Séparation de la table en 2 parties (60% et 40% des données) */
DATA base;
SET joint1;
alea = ranuni(41);
RUN;

DATA Base_Apprentissage;
SET base;
WHERE alea < 0.6;
RUN;

DATA Base_Test;
SET base;
WHERE alea > 0.6;
RUN;

/*Création du log pour la variable expo */
DATA base_Apprentissage;
SET base_Apprentissage;
ln = log(expo);
RUN;

```

```

/*Modélisation de la fréquence des sinistres*/
PROC GENMOD DATA = base_Apprentissage;
  CLASS
    COEFF_BONUS_MALUS
    CLASSEM
    CSPclasse
    SEGMENTclasse
    GROUPE_SRA_p
    REGION
    ALIMENTATION_p
    USAGE
    /*MODE_VIE*/
    AGE_VEHICULE_p
    GARAGE
    CARROSSERIE_p
    /*classe_prix_p*/
    /*CLASSE_REPARATION_p*/
    /*NB_ENFANTS_p*/
    /*COUT_PARE_BRISE_p*/
    ENERGIE_p
    GENRE
    /*TYPE_LOGEMENT*/
    STATUT_LOGEMENT
    REGION
  ;
  MODEL nb_sinbdg =
    COEFF_BONUS_MALUS
    CLASSEM
    CSPclasse
    SEGMENTclasse
    GROUPE_SRA_p
    REGION
    ALIMENTATION_p
    USAGE
    /*MODE_VIE*/
    AGE_VEHICULE_p
    GARAGE
    CARROSSERIE_p
    /*classe_prix_p*/
    /*CLASSE_REPARATION_p*/
    /*NB_ENFANTS_p*/
    /*COUT_PARE_BRISE_p*/
    ENERGIE_p
    GENRE
    /*TYPE_LOGEMENT*/
    STATUT_LOGEMENT
    REGION
    / DIST = poisson LINK = log OFFSET = ln TYPE3;
RUN;

```

```

/*Modélisation du coût des sinistres*/
PROC GENMOD DATA = base_Apprentissage;
  CLASS /*COEFF_BONUS_MALUS*/
    CLASSEM
    CSPclasse
    SEGMENTclasse
    GROUPE_SRA_p
    REGION

```

```

ALIMENTATION_p
/*USAGE*/
/*MODE_VIE*/
AGE_VEHICULE_p
GARAGE
/*CARROSSERIE_p*/
classe_prix_p
/*CLASSE_REPARATION_p*/
/*NB_ENFANTS_p*/
COUT_PARE_BRISE_p
/*ENERGIE_p*/
/*GENRE*/
TYPE_LOGEMENT
/*STATUT_LOGEMENT*/
REGION
;
MODEL ch_BDG=
/*COEFF_BONUS_MALUS*/
CLASSEM
CSPclasse
SEGMENTclasse
GROUPE_SRA_p
REGION
ALIMENTATION_p
/*USAGE*/
/*MODE_VIE*/
AGE_VEHICULE_p
GARAGE
/*CARROSSERIE_p*/
classe_prix_p
/*CLASSE_REPARATION_p*/
/*NB_ENFANTS_p*/
COUT_PARE_BRISE_p
/*ENERGIE_p*/
/*GENRE*/
TYPE_LOGEMENT
/*STATUT_LOGEMENT*/
REGION

/ DIST = gamma LINK = log TYPE3 ;
WEIGHT nb_sinbdg;
RUN;

/*Table contenant les données externes */

data work.Table2;
infile 'C:\Users\Ascensio\Downloads\Etude_de_cas\base_cc_comparateur.csv'
dsd misover dlm=';' firstobs=2;
input CODE_INSEE_DOMICILE$ POP SUPERFICIE DENSITE SALAIRE_MEDIAN;

length densite_p $10.;
if densite < 100 then densite_p = "0 à 100";
if densite => 100 and densite < 200 then densite_p = "100 à 200";
if densite => 200 and densite < 300 then densite_p = "200 à 300";
if densite => 300 then densite_p = ">300";

```

```

length salaire_median_p $10.;
if salaire_median >= 10000 and salaire_median <16000 then salaire_median_p=
"12000 à 14000";

if salaire_median >= 16000 and salaire_median <18000 then salaire_median_p=
"16000 à 18000";
if salaire_median >= 18000 and salaire_median <20000 then salaire_median_p=
"18000 à 20000";
if salaire_median > 20000 then salaire_median_p= "plus de 20000";
run;

proc sort data = Table2;
by CODE_INSEE_DOMICILE;
run;

/*jointure tablemodif1 et Table 2 */
proc sort data = table_modif1;
by CODE_INSEE_DOMICILE;
run;

data joint1;
merge table_modif1 (in=a) Table2;
by CODE_INSEE_DOMICILE;
if a then output;
run;

/*Modélisation du coût des sinistres + variables ajoutées*/
PROC GENMOD DATA = base_Apprentissage;
CLASS /*COEFF_BONUS_MALUS*/
CLASSEM
CSPclasse
SEGMENTclasse
GROUPE_SRA_p
REGION
ALIMENTATION_p
/*USAGE*/
/*MODE_VIE*/
AGE_VEHICULE_p
GARAGE
/*CARROSSERIE_p*/
classe_prix_p
/*CLASSE_REPARATION_p*/
/*NB_ENFANTS_p*/
COUT_PARE_BRISE_p
/*ENERGIE_p*/
/*GENRE*/
TYPE_LOGEMENT
/*STATUT_LOGEMENT*/
REGION
densite_p
/*salaire_median_p*/
;
MODEL ch_BDG=
/*COEFF_BONUS_MALUS*/
CLASSEM
CSPclasse
SEGMENTclasse
GROUPE_SRA_p

```

```

REGION
ALIMENTATION_p
/*USAGE*/
/*MODE_VIE*/
AGE_VEHICULE_p
GARAGE
/*CARROSSERIE_p*/
classe_prix_p
/*CLASSE_REPARATION_p*/
/*NB_ENFANTS_p*/
COUT_PARE_BRISE_p
/*ENERGIE_p*/
/*GENRE*/
TYPE_LOGEMENT
/*STATUT_LOGEMENT*/
REGION
densite_p
/*salaire_median_p*/
/ DIST = gamma LINK = log TYPE3 ;
WEIGHT nb_sinbdg;
RUN;

/*Modélisation de la fréquence des sinistres + variables ajoutées*/

PROC GENMOD DATA = base_Apprentissage;
CLASS
COEFF_BONUS_MALUS
CLASSEM
CSPclasse
SEGMENTclasse
GROUPE_SRA_p
REGION
ALIMENTATION_p
USAGE
/*MODE_VIE*/
AGE_VEHICULE_p
GARAGE
CARROSSERIE_p
/*classe_prix_p*/
/*CLASSE_REPARATION_p*/
/*NB_ENFANTS_p*/
/*COUT_PARE_BRISE_p*/
ENERGIE_p
GENRE
/*TYPE_LOGEMENT*/
STATUT_LOGEMENT
REGION
densite_p
/*salaire_median_p*/
;
MODEL nb_sinbdg =
COEFF_BONUS_MALUS
CLASSEM
CSPclasse
SEGMENTclasse
GROUPE_SRA_p
REGION

```

```

ALIMENTATION_p
USAGE
/*MODE_VIE*/
AGE_VEHICULE_p
GARAGE
CARROSSERIE_p
/*classe_prix_p*/
/*CLASSE_REPARATION_p*/
/*NB_ENFANTS_p*/
/*COUT_PARE_BRISE_p*/
ENERGIE_p
GENRE
/*TYPE_LOGEMENT*/
STATUT_LOGEMENT
REGION
densite_p
/*salaire_median_p*/
/ DIST = poisson LINK = log OFFSET = ln TYPE3;
RUN;

/* Modèle final (sur la table contenant 40% des données)*/

data MODEL_FINAL;
    set base_Test;

    Intercept = -3.4462;

    if COEFF_BONUS_MALUS = "0.5 à 0.6" then estimation_BONUS = -
0.1638;
    if COEFF_BONUS_MALUS = "0.6 à 0.7" then estimation_BONUS = -
0.0537;
    if COEFF_BONUS_MALUS = "0.7 à 0.8" then estimation_BONUS = -
0.0667;
    if COEFF_BONUS_MALUS = "0.8 à 0.9" then estimation_BONUS = -
0.1121;
    if COEFF_BONUS_MALUS = ">0.9" then estimation_BONUS = 0;

    if CLASSEM = "ALLEMAGNE" then estimation_CLASSEM = 0.0615;
    if CLASSEM = "AUTRE" then estimation_CLASSEM = -0.0813;
    if CLASSEM = "COREE DU SUD" then estimation_CLASSEM = -0.0903;
    if CLASSEM = "ESPAGNE" then estimation_CLASSEM = -0.0846;
    if CLASSEM = "ETATS-UNIS" then estimation_CLASSEM = -0.0481;
    if CLASSEM = "ITALIE" then estimation_CLASSEM = -0.0448;
    if CLASSEM = "FRANCE" then estimation_CLASSEM = 0.0377;
    if CLASSEM = "GRANDE-BRETAGNE" then estimation_CLASSEM =
0.2274;
    if CLASSEM = "JAPON" then estimation_CLASSEM = 0;

    if CSPclasse = "Agriculteur/Pêcheur" then estimation_CSPclasse
= -0.1587;
    if CSPclasse = "Artisan" then estimation_CSPclasse = -0.0081;
    if CSPclasse = "Chef d'entreprise" then estimation_CSPclasse =
0.1233;
    if CSPclasse = "Commerçant" then estimation_CSPclasse = -
0.0462;
    if CSPclasse = "Etudiant" then estimation_CSPclasse = 0.4911;
    if CSPclasse = "Fonctionnaire de l'" then estimation_CSPclasse
= 0.0718;
    if CSPclasse = "Fonctionnaire et as" then estimation_CSPclasse
= -0.0076;

```



```

    if CSPclasse = "Policier/Militaire" then estimation_CSPclasse =
-0.0215;
    if CSPclasse = "Profession libérale" then estimation_CSPclasse
= 0.0549;
    if CSPclasse = "Professionnel du sp" then estimation_CSPclasse
= 0.1016;
    if CSPclasse = "Retraité" then estimation_CSPclasse = -0.1406;
    if CSPclasse = "Salarié du privé: c" then estimation_CSPclasse
= 0.0116;
    if CSPclasse = "Salarié du privé: e" then estimation_CSPclasse
= -0.0207;
    if CSPclasse = "Salarié du privé: t" then estimation_CSPclasse
= -0.0854;
    if CSPclasse = "Sans activité et en" then estimation_CSPclasse
= 0.0586;
    if CSPclasse = "Sans activité profe" then estimation_CSPclasse
= 0;

    if SEGMENTclasse = "camionette four" then
estimation_SEGMENTclasse = 0.3612;
    if SEGMENTclasse = "citadines-polyv" then
estimation_SEGMENTclasse = 0.0345;
    if SEGMENTclasse = "compactes" then estimation_SEGMENTclasse =
0.2625;
    if SEGMENTclasse = "familiales" then estimation_SEGMENTclasse =
0.2109;
    if SEGMENTclasse = "grandes et très" then
estimation_SEGMENTclasse = 0.1413;
    if SEGMENTclasse = "micro-citadines" then
estimation_SEGMENTclasse = 0;

    if GROUPE_SRA_p = "20 à 27" then estimation_GROUPE_SRA_p = -
0.2737;
    if GROUPE_SRA_p = "29" then estimation_GROUPE_SRA_p = -0.2468;
    if GROUPE_SRA_p = "30" then estimation_GROUPE_SRA_p = -0.1947;
    if GROUPE_SRA_p = "31" then estimation_GROUPE_SRA_p = -0.2078;
    if GROUPE_SRA_p = "32" then estimation_GROUPE_SRA_p = -0.1613;
    if GROUPE_SRA_p = "33" then estimation_GROUPE_SRA_p = -0.1586;
    if GROUPE_SRA_p = "34" then estimation_GROUPE_SRA_p = -0.0909;
    if GROUPE_SRA_p = "35" then estimation_GROUPE_SRA_p = -0.0841;
    if GROUPE_SRA_p = "36" then estimation_GROUPE_SRA_p = -0.0093;
    if GROUPE_SRA_p = "37 à 46" then estimation_GROUPE_SRA_p = 0;

    if REGION = "Auvergne-Rhône-Alpes" then estimation_REGION =
0.3008;
    if REGION = "Bourgogne Franche-Co" then estimation_REGION =
0.3784;
    if REGION = "Bretagne" then estimation_REGION = 0.2156;
    if REGION = "Centre Val-De-Loire" then estimation_REGION =
0.2722;
    if REGION = "Corse" then estimation_REGION = 0.7819;
    if REGION = "Grand Est" then estimation_REGION = 0.5113;
    if REGION = "Hauts-de-France" then estimation_REGION = 0.5045;
    if REGION = "Ile de France" then estimation_REGION = 0.5232;
    if REGION = "Normandie" then estimation_REGION = 0.7940;
    if REGION = "Nouvelle Aquitaine" then estimation_REGION =
0.2821;
    if REGION = "Occitanie" then estimation_REGION = 0.2382;
    if REGION = "Pays de la Loire" then estimation_REGION =
0.29618;
    if REGION = "Provence-Alpes-Côte" then estimation_REGION = 0;

```

```

0.3008;      if ALIMENTATION_p = "Autre" then estimation_ALIMENTATION_p =
0.0988;      if ALIMENTATION_p = "IDS" then estimation_ALIMENTATION_p =
0.0809;      if ALIMENTATION_p = "INJ" then estimation_ALIMENTATION_p =
if ALIMENTATION_p = "INS" then estimation_ALIMENTATION_p = 0;

if USAGE = "DOMICILE-TRAVAIL" then estimation_USAGE = -0.1406;
if USAGE = "PRIVE" then estimation_USAGE = -0.2786;
if USAGE = "PRIVE-PRO" then estimation_USAGE = -0.0406;
if USAGE = "TOURNEE" then estimation_USAGE = 0;

= 0.4867;    if AGE_VEHICULE_p = "12 à 16 an" then estimation_AGE_VEHICULE_p
= 0.3309;    if AGE_VEHICULE_p = "16 à 20 an" then estimation_AGE_VEHICULE_p
= 0.7332;    if AGE_VEHICULE_p = "4 à 8 ans" then estimation_AGE_VEHICULE_p
= 0.5939;    if AGE_VEHICULE_p = "8 à 12 ans" then estimation_AGE_VEHICULE_p
0.7559;    if AGE_VEHICULE_p = "<4 ans" then estimation_AGE_VEHICULE_p =
0;          if AGE_VEHICULE_p = ">20 ans" then estimation_AGE_VEHICULE_p =

if GARAGE = "BOX CLOS" then estimation_GARAGE = -0.0100;
if GARAGE = "BOX OU JARDIN CLOS" then estimation_GARAGE = -
0.1713;
if GARAGE = "JARDIN CLOS" then estimation_GARAGE = 0.0651;
if GARAGE = "PARKING" then estimation_GARAGE = -0.2783;
if GARAGE = "PARKING CLOS" then estimation_GARAGE = -0.0219;
if GARAGE = "PARKING OUVERT" then estimation_GARAGE = 0.0272;
if GARAGE = "VOIE PUBLIQUE" then estimation_GARAGE = 0;

0.3361;    if CARROSSERIE_p = "Autre" then estimation_CARROSSERIE_p =
if CARROSSERIE_p = "B" then estimation_CARROSSERIE_p = 0.3137;
if CARROSSERIE_p = "C" then estimation_CARROSSERIE_p = 0;

0.3553;    if ENERGIE_p = "AU, EL, ES" then estimation_ENERGIE_p = -
if ENERGIE_p = "GO, GP" then estimation_ENERGIE_p = 0;

if GENRE = "VP" then estimation_GENRE = -0.1837;
if GENRE = "VU" then estimation_GENRE = 0;

= -0.0569;  if STATUT_LOGEMENT = "INCONNU" then estimation_STATUT_LOGEMENT
estimation_STATUT_LOGEMENT = 0.0581;
if STATUT_LOGEMENT = "PROPRIETAIRE" then
estimation_STATUT_LOGEMENT = 0;

Intercept_csin = 6.5590;

0.2102;    if CLASSEM = "ALLEMAGNE" then estimation_csin_CLASSEM = -

```

```

        if CLASSEM = "AUTRE" then estimation_csin_CLASSEM = -0.3929;
        if CLASSEM = "COREE DU SUD" then estimation_csin_CLASSEM = -
0.0272;
        if CLASSEM = "ESPAGNE" then estimation_csin_CLASSEM = -0.2087;
        if CLASSEM = "ETATS-UNIS" then estimation_csin_CLASSEM =
0.0290;
        if CLASSEM = "ITALIE" then estimation_csin_CLASSEM = -0.1164;
        if CLASSEM = "FRANCE" then estimation_csin_CLASSEM = -0.1391;
        if CLASSEM = "GRANDE-BRETAGNE" then estimation_csin_CLASSEM = -
0.1104;
        if CLASSEM = "JAPON" then estimation_csin_CLASSEM = 0;

        if CSPclasse = "Agriculteur/Pêcheur" then
estimation_csin_CSPclasse = -0.2446;
        if CSPclasse = "Artisan" then estimation_csin_CSPclasse =
0.0247;
        if CSPclasse = "Chef d'entreprise" then
estimation_csin_CSPclasse = -0.0147;
        if CSPclasse = "Commerçant" then estimation_csin_CSPclasse =
0.0349;
        if CSPclasse = "Etudiant" then estimation_csin_CSPclasse = -
0.1169;
        if CSPclasse = "Fonctionnaire de l'" then
estimation_csin_CSPclasse = 0.0090;
        if CSPclasse = "Fonctionnaire et as" then
estimation_csin_CSPclasse = -0.0366;
        if CSPclasse = "Policier/Militaire" then
estimation_csin_CSPclasse = -0.0420;
        if CSPclasse = "Profession libérale" then
estimation_csin_CSPclasse = -0.0199;
        if CSPclasse = "Professionnel du sp" then
estimation_csin_CSPclasse = -0.0980;
        if CSPclasse = "Retraité" then estimation_csin_CSPclasse = -
0.0504;
        if CSPclasse = "Salarié du privé: c" then
estimation_csin_CSPclasse = -0.0109;
        if CSPclasse = "Salarié du privé: e" then
estimation_csin_CSPclasse = -0.0016;
        if CSPclasse = "Salarié du privé: t" then
estimation_csin_CSPclasse = 0.0706;
        if CSPclasse = "Sans activité et en" then
estimation_csin_CSPclasse = -0.1516;
        if CSPclasse = "Sans activité profe" then
estimation_csin_CSPclasse = 0;

        if SEGMENTclasse = "camionnette four" then
estimation_csin_SEGMENTclasse = -0.0908;
        if SEGMENTclasse = "citadines-polyv" then
estimation_csin_SEGMENTclasse = -0.0684;
        if SEGMENTclasse = "compactes" then
estimation_csin_SEGMENTclasse = -0.0004;
        if SEGMENTclasse = "familiales" then
estimation_csin_SEGMENTclasse = 0.0162;
        if SEGMENTclasse = "grandes et très" then
estimation_csin_SEGMENTclasse = 0.0754;
        if SEGMENTclasse = "micro-citadines" then
estimation_csin_SEGMENTclasse = 0;

        if GROUPE_SRA_p = "20 à 27" then estimation_csin_GROUPE_SRA_p =
-0.2445;

```

```

0.2583;      if GROUPE_SRA_p = "29" then estimation_csin_GROUPE_SRA_p = -
0.2289;      if GROUPE_SRA_p = "30" then estimation_csin_GROUPE_SRA_p = -
0.1645;      if GROUPE_SRA_p = "31" then estimation_csin_GROUPE_SRA_p = -
0.1570;      if GROUPE_SRA_p = "32" then estimation_csin_GROUPE_SRA_p = -
0.1536;      if GROUPE_SRA_p = "33" then estimation_csin_GROUPE_SRA_p = -
0.1325;      if GROUPE_SRA_p = "34" then estimation_csin_GROUPE_SRA_p = -
0.1770;      if GROUPE_SRA_p = "35" then estimation_csin_GROUPE_SRA_p = -
0.0592;      if GROUPE_SRA_p = "36" then estimation_csin_GROUPE_SRA_p = -
0;           if GROUPE_SRA_p = "37 à 46" then estimation_csin_GROUPE_SRA_p =

= 0.0290;    if REGION = "Auvergne-Rhône-Alpes" then estimation_csin_REGION
= -0.0185;    if REGION = "Bourgogne Franche-Co" then estimation_csin_REGION
-0.0949;      if REGION = "Bretagne" then estimation_csin_REGION = -0.0075;
              if REGION = "Centre Val-De-Loire" then estimation_csin_REGION =
0.0561;      if REGION = "Corse" then estimation_csin_REGION = 0.1196;
0.0513;      if REGION = "Grand Est" then estimation_csin_REGION = 0.0444;
              if REGION = "Hauts-de-France" then estimation_csin_REGION =
0.0001;      if REGION = "Ile de France" then estimation_csin_REGION =
0.0334;      if REGION = "Normandie" then estimation_csin_REGION = 0.0822;
              if REGION = "Nouvelle Aquitaine" then estimation_csin_REGION =
0.0001;      if REGION = "Occitanie" then estimation_csin_REGION = -0.0019;
              if REGION = "Pays de la Loire" then estimation_csin_REGION = -
0.0334;      if REGION = "Provence-Alpes-Côte" then estimation_csin_REGION =
0;           if ALIMENTATION_p = "Autre" then estimation_csin_ALIMENTATION_p
= 0.0749;    if ALIMENTATION_p = "IDS" then estimation_csin_ALIMENTATION_p =
0.0936;      if ALIMENTATION_p = "INJ" then estimation_csin_ALIMENTATION_p =
0.0557;      if ALIMENTATION_p = "INS" then estimation_csin_ALIMENTATION_p =
0;           if AGE_VEHICULE_p = "12 à 16 an" then
estimation_csin_AGE_VEHICULE_p = 0.0942;
              if AGE_VEHICULE_p = "16 à 20 an" then
estimation_csin_AGE_VEHICULE_p = 0.1027;
              if AGE_VEHICULE_p = "4 à 8 ans" then
estimation_csin_AGE_VEHICULE_p = 0.0162;
              if AGE_VEHICULE_p = "8 à 12 ans" then
estimation_csin_AGE_VEHICULE_p = 0.0712;
              if AGE_VEHICULE_p = "<4 ans" then
estimation_csin_AGE_VEHICULE_p = -0.0364;

```

```

        if AGE_VEHICULE_p = ">20 ans" then
estimation_csin_AGE_VEHICULE_p = 0;

        if GARAGE = "BOX CLOS" then estimation_csin_GARAGE = 0.0402;
        if GARAGE = "BOX OU JARDIN CLOS" then estimation_csin_GARAGE =
0.0847;
        if GARAGE = "JARDIN CLOS" then estimation_csin_GARAGE = 0.0492;
        if GARAGE = "PARKING" then estimation_csin_GARAGE = 0.0321;
        if GARAGE = "PARKING CLOS" then estimation_csin_GARAGE = -
0.0531;
        if GARAGE = "PARKING OUVERT" then estimation_csin_GARAGE =
0.0577;
        if GARAGE = "VOIE PUBLIQUE" then estimation_csin_GARAGE = 0;

        if classe_prix_p = "A à D" then estimation_csin_classe_prix_p =
-0.1880;
        if classe_prix_p = "E à H" then estimation_csin_classe_prix_p =
-0.2743;
        if classe_prix_p = "I à L" then estimation_csin_classe_prix_p =
-0.0905;
        if classe_prix_p = "M à P" then estimation_csin_classe_prix_p =
-0.0306;
        if classe_prix_p = "Q à T" then estimation_csin_classe_prix_p =
0.0325;
        if classe_prix_p = "U à V" then estimation_csin_classe_prix_p =
0;

        if COUT_PARE_BRISE_p = "200 à 400" then csin_COUT_PARE_BRISE_p
= -0.1108;
        if COUT_PARE_BRISE_p = "600 à 800" then csin_COUT_PARE_BRISE_p
= 0.1544;
        if COUT_PARE_BRISE_p = "<200" then csin_COUT_PARE_BRISE_p =
0.0108;
        if COUT_PARE_BRISE_p = ">800" then csin_COUT_PARE_BRISE_p =
0.3140;
        if COUT_PARE_BRISE_p = "400 à 600" then csin_COUT_PARE_BRISE_p
= 0;

        if TYPE_LOGEMENT = "APPARTEMENT" then csin_TYPE_LOGEMENT = -
0.0680;
        if TYPE_LOGEMENT = "INCONNU" then csin_TYPE_LOGEMENT = -0.0544;
        if TYPE_LOGEMENT = "MAISON" then csin_TYPE_LOGEMENT = 0;

        if densite_p = "0 à 100" then csin_densite_p = 0.0397;
        if densite_p = "100 à 200" then csin_densite_p = 0.0130;
        if densite_p = "200 à 300" then csin_densite_p = 0.0347;
        if densite_p = ">300" then csin_densite_p = 0;

        /* estimation de la fréquence de sinistre */

        mod_freqsin =
exp(Intercept)*exp(estimation_BONUS)*exp(estimation_CLASSEM)*exp(estimation
_CSPclasse)*exp(estimation_SEGMENTclasse)*exp(estimation_GROUPE_SRA_p)*exp(
estimation_REGION)*exp(estimation_ALIMENTATION_p)*exp(estimation_USAGE)*exp
(estimation_AGE_VEHICULE_p)*exp(estimation_GARAGE)*exp(estimation_CARROSSER
IE_p)*exp(estimation_ENERGIE_p)*exp(estimation_GENRE);

        /* estimation de la charge de sinistres de bris de glace */

        mod_csin =
exp(Intercept_csin)*exp(estimation_csin_CLASSEM)*exp(estimation_csin_CSPcla

```

```

sse)*exp(estimation_csin_SEGMENTclasse)*exp(estimation_csin_GROUPE_SRA_p)*e
xp(estimation_csin_REGION)*exp(estimation_csin_ALIMENTATION_p)*exp(estimati
on_csin_AGE_VEHICULE_p)*exp(estimation_csin_GARAGE)*exp(estimation_csin_cla
sse_prix_p)*exp(csin_COUT_PARE_BRISE_p)*exp(csin_TYPE_LOGEMENT)*exp(csin_de
nsite_p);

        /* estimation du montant de la prise en charge de sinistres */

        cout = mod_freqsin*mod_csin;

run;

/* calcul charge de la prime totale du modèle */

proc sql;
select sum(cout*expo) as prime, sum(ch_bdg) as charge from MODEL_FINAL;
run;

/* analyse exploratoire de la prime pure */

/* création de la variable prime (cout * expo) */

data model_prime;
set MODEL_FINAL;
prime = cout*expo;
run;

/* calcul charge de la prime totale du modèle */

proc sql;
select sum(cout*expo) as prime, sum(ch_bdg) as charge from MODEL_FINAL;
run;

/* distribution de la prime en fonction de tous les critères de
tarification */

%let var =
CLASSEM
ALIMENTATION_p
AGE_VEHICULE_p
GARAGE
classe_prix_p
COUT_PARE_BRISE_p
TYPE_LOGEMENT
REGION
COEFF_BONUS_MALUS
CSPclasse
SEGMENTclasse
GROUPE_SRA_p
REGION
USAGE
AGE_VEHICULE_p
GARAGE
CARROSSERIE_p
ENERGIE_p
GENRE
STATUT_LOGEMENT

```

```

densite_p
;

%macro boxplots2;
    %do i = 1 %to 21;
        %let variable = %scan( &var,&i);
        proc sgplot data = MODEL_FINAL;
            title "Variable : &variable";
            vbox prime/category = &variable;
            run;
        %end;
        title;
    %mend boxplots;

%boxplots2;

```