

Projet Assurance Non-vie II

April 2023

Nom & Prénom : CHARET Mohamed

Encadré par : MARRI Fouad

Contents

1	Partie 1	2
1.1	Introduction	2
1.2	Définitions	2
2	Partie 2 : Tarification (RC automobile usage tourisme)	3
2.1	Importation de données(Production)	3
2.2	Nettoyage de base de donnée	4
2.3	Sinistre	6
2.3.1	Importation de donnée	6
2.3.2	Nettoyage de base de données	6
2.4	Exploration de données	9
2.4.1	statistique descriptives	9
2.4.2	Generer un ensembles de diagrammes en utilisant sgplot	15
2.5	Modelisation	30
2.5.1	Modelisation de "nombre de sinistre"	30
2.5.2	Modelisation des coût individuel de sinistre (CM)	35
3	Partie 3: Provisionnement	39
3.1	Importation des des données	39
3.2	Application de la méthode ChainLadder pour le calcul des réserves de sinistres.	39
3.3	Modèle de Mack	40
3.4	Modele de GLM	41

1 Partie 1

1.1 Introduction

– Le marché de l'Assurance au Maroc est un secteur en pleine expansion, avec de nombreux acteurs et intermédiaires. Les compagnies d'assurances, les courtiers, les agents généraux, les intermédiaires en assurances et les experts en assurance sont autant d'intervenants qui jouent un rôle crucial dans ce marché. La répartition du chiffre d'affaires du secteur des assurances est un indicateur important de la santé du marché, avec des évolutions propres à chaque branche. L'Assurance Vie et l'Assurance Non-Vie sont les deux principales branches du marché, avec des chiffres d'affaires en constante évolution. Le cadre réglementaire au Maroc encadre l'activité des compagnies d'assurances et des intermédiaires pour garantir la protection des consommateurs. Le taux de pénétration de l'Assurance au Maroc reste relativement faible, mais connaît une augmentation régulière. Le volume des primes émises est un élément clé du marché de l'Assurance au Maroc, reflétant les besoins et la demande de la population. Enfin, la réassurance joue un rôle important dans ce marché, offrant aux compagnies d'assurances une protection financière et une expertise pour gérer les risques.

– Le marché de l'Assurance Automobile au Maroc est un secteur clé de l'industrie de l'assurance. Avec une part importante du chiffre d'affaires total de l'assurance, cette branche est en constante évolution, reflétant les tendances et les besoins du marché automobile marocain. La responsabilité civile (RC) automobile est l'une des garanties les plus courantes dans cette branche, couvrant les dommages causés à des tiers par un véhicule assuré. La libéralisation de la RC automobile a permis une plus grande concurrence sur ce marché, offrant aux consommateurs une plus grande variété d'options. Les risques couverts par cette garantie peuvent varier en fonction des contrats d'assurance automobile, qui peuvent également inclure d'autres garanties telles que l'assurance tous risques ou l'assistance en cas de panne ou d'accident. Le coefficient de réduction majoration (CRM) est un élément clé de l'assurance automobile au Maroc, permettant de moduler le prix de l'assurance en fonction du profil de l'assuré.

Catégorie	Chiffre d'affaires (en milliards de dirhams)
Assurance Vie	26,2
Assurance Non-Vie	18,1
Total du secteur des assurances	44,3
Volume des primes émises	27,9
Taux de pénétration	3,3%
Réassurance	5,5

Table 1: Chiffres d'affaires du marché de l'Assurance au Maroc en 2020

Catégorie	Chiffre d'affaires (en milliards de dh)
Chiffre d'affaires de l'Assurance Automobile	10,8
Part de la branche automobile dans le chiffre d'affaires total de l'Assurance	24,4%
Nombre de contrats d'Assurance Automobile	2,1 millions
Coefficient de Réduction-Majoration (CRM) moyen	1,25

Table 2: Chiffres clés du marché de l'Assurance Automobile au Maroc

1.2 Définitions

- **L'antisélection** est un phénomène qui se produit lorsqu'une personne qui présente un risque plus élevé que la moyenne cherche à souscrire une assurance. Cette personne a une meilleure connaissance de son propre risque que l'assureur, ce qui peut conduire à une sélection adverse en faveur de l'assuré et au détriment de l'assureur. Cela peut entraîner une augmentation des primes d'assurance pour tous les assurés et rendre l'assurance moins accessible pour les personnes présentant un risque moyen ou faible.
- **L'Assurance Multirisque Habitation** est une assurance qui couvre les dommages causés à un logement et à son contenu. Cette assurance couvre généralement les risques liés aux incendies, aux dégâts des eaux, aux catastrophes naturelles, aux vols, aux bris de glace, aux responsabilités civiles et aux pertes d'exploitation. Elle peut être souscrite par les propriétaires, les locataires ou les copropriétaires d'un bien immobilier.
- **L'Assurance Responsabilité Civile Décennale** est une assurance obligatoire pour les professionnels du bâtiment, tels que les architectes, les entrepreneurs et les artisans. Elle couvre les dommages qui peuvent survenir dans les dix

années suivant la réception d'un ouvrage de construction, tels que les défauts de construction, les malfaçons ou les vices cachés. Cette assurance permet aux professionnels du bâtiment de garantir leur responsabilité et de protéger leurs clients en cas de sinistre.

- **La garantie décennale couvre** les dommages qui peuvent affecter la solidité et la stabilité d'un ouvrage de construction, ou qui le rendent inhabitable ou impropre à sa destination. Cela peut inclure des fissures importantes ou affaissements, des défauts de fondation, des infiltrations d'eau, des défauts de conception ou de réalisation de l'ouvrage, ou des vices cachés qui affectent la solidité ou la stabilité de l'ouvrage. La garantie décennale est en vigueur pendant 10 ans à compter de la réception des travaux.

2 Partie 2 : Tarification (RC automobile usage tourisme)

2.1 Importation de données(Production)

Au cour de l'importation de données soient pour la production ou la base de donnée "sinistre", j'ai re-contré un problème sur la cohérence des données dans SAS, donc j'ai converti les deux bases en fichiers excel.

```
*****importation de base de données*****;
/*1-creation de la bibliothèque*/
libname chrt 'C:\Users\hp\Desktop\projet assurance non vie\examenAss';
***** production*****;
/*2-importation de données*/
proc import
    datafile='C:\Users\hp\Desktop\projet assurance non vie\examenAss\production.xlsx' dbms=xlsx
    out=chrt.production replace;
    getnames=yes;

run;
```

Listing 1: Importation

```
/*Afficher les variables*/
PROC print DATA=chrt.production;
RUN;
```

Obs	VAR1	numepolice	exercice	DMC	DOB	Sexe	Zone	Comubsution	PF	exposition
1	1	880898	2011	2002-10-05	1972-04-05	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.6179260032
2	2	703479	2010	1994-12-08	1959-05-13	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.8011278689
3	3	186906	2012	1998-11-15	1981-01-07	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.3863229074
4	4	41168	2016	1992-04-28	1986-02-05	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.7923478724
5	5	69867	2017	2002-02-04	1989-11-13	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.2411030659
6	6	681674	2015	1996-06-19	1973-01-03	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.8221188765

Figure 1: Affichage des variables

2.2 Nettoyage de base de donnée

```
/*Edition les nomes de variables*/
data chrt.production;
    rename DOB=date_de_naissance
           DMC=date_de_mise_en_circulation
           PF=Puissance_Fiscale;
set chrt.production;
run;
proc print data=insea.production (obs=3);
run;
/*Verification de la presence de valeurs manquantes */
proc means data=chrt.production nmiss;
output out=chrt.tableau_nmiss nmiss=;
run;
/*supression quelques valeurs*/
data chrt.Production;
    set chrt.production;
    if date_mise_en_circulation=. then
        delete;
    if date_de_naissance=. then
        delete;
    if numepolice = . then
        delete;
    if exercice =. then
        delete;
    if exposition=. then
        delete;

run;
```

Dans ce code suivant permet à calculer l'âge du conducteur et l'âge de mise en circulation.

```
/*Ajouter des nouvelles variables */
data chrt.Production;
    set chrt.Production;
    age_du_conducteur = intck("year", date_de_naissance,mdy(1,1,exercice));
    age_mise_en_circulation = intck("year", date_de_mise_en_circulation,mdy(1,1,exercice));
run;
```

```
/*Test s'il existe un individu d'age <18(>100); */
PROC means DATA=chrt.production;
var age_du_conducteur ;
RUN;
*****;
* on constate qu'il n y a pas d'age < à 18 (>100) *
*****;
*detection des doublants;
proc sql;
create table chrt.dub as
select *, count(*) as dubs
from chrt.Production
group by numepolice, exercice
order by dubs DESC;
quit;
proc means data=chrt.dub;
var dubs;
run;
```

```

/*suppression des variables (date_de_naissance et date_de_mise_en_circulation) */
DATA chrt.production;
  SET chrt.production (DROP=date_de_naissance date_mise_en_circulation);
RUN;

```

The MEANS Procedure

Variable	N Miss
VAR1	0
numepolice	0
exercice	0
date_mise_en_circulation	0
date_de_naissance	0
exposition	0

The MEANS Procedure

Analysis Variable : age_du_conducteur				
N	Mean	Std Dev	Minimum	Maximum
100000	38.9612300	9.6771836	18.0000000	59.0000000

The MEANS Procedure

Analysis Variable : dubs				
N	Mean	Std Dev	Minimum	Maximum
100000	1.0000000	0	1.0000000	1.0000000

Obs	VAR1	numepolice	exercice	Sexe	Zone	Comubstution	Puissance_Fiscale	exposition	age_du_conducteur	age_mise_en_circulation
1	1	880898	2011	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.6179260032	39	9
2	2	703479	2010	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.8011278689	51	16
3	3	186906	2012	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.3863229074	51	14
4	4	41168	2016	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.7923478724	30	24
5	5	69867	2017	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.2411030659	28	15
6	6	681674	2015	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.8221188765	42	19



Interprétation

- Le 1^{ere} tableau affirme qu'il n'y a pas des valeurs manquantes;
- Le 2^{eme} tableau indique l'absence des ages qui dépassent 100 ou inferieur à 18;
- Le 3^{eme} tableau indique l'absncence des doublants.

2.3 Sinistre

2.3.1 Importation de donnée

```
***PART 2 : Importation de données sinistre***;  
/*1-importation de données */  
  
proc import datafile='C:\Users\hp\Desktop\projet assurance non vie\examenAss\sinistre.xlsx' dbms=xlsx  
  
    out=chrt.sinistre replace;  
    getnames=yes;  
  
run;  
/*2- affichage des données */  
proc print data=chrt.sinistre (obs=3);  
run;
```

2.3.2 Nettoyage de base de données

```
*****Nettoyage*****;  
/*1- supprimer la première colonne */  
    DATA chrt.sinistre;  
    SET chrt.sinistre (DROP=var1);  
RUN;  
  
/*2-verification de la presence des valeurs manquantes */  
proc means data=chrt.sinistre nmiss;  
output out=chrt.tab_nmiss nmiss=;  
run;  
*****;  
*Pas des valeurs manquantes *;  
*****;  
/*3- verification sur les variables */  
data chrt.sinistre;  
set chrt.sinistre;  
if numepolice=. or numepolice<=0 then  
delete;  
if annee_reference=. or annee_reference<=0 then  
delete;  
  
if n_sinistre=. or n_sinistre<=0 then  
delete;  
if montantsinistre=. or montantsinistre<=0 then  
delete;  
run;  
/*4-affichage des variables après nettoyage */  
proc print data= chrt.sinistre;  
run;
```

```
/* 5- detection des doublants */  
proc sql;  
create table chrt.dubb as  
select *, count(*) as dubbs  
from chrt.sinistre  
group by numepolice, annee_reference  
order by dubbs DESC;  
quit;  
proc means data=chrt.dubb;  
var dubbs;  
run;
```

le nombre de valeurs manquantes par chaque variable

The MEANS Procedure

Variable	N Miss
numepolice	0
annee_reference	0
n_sinistre	0
montantsinistre	0

Affichage des 6 premières observations apres nettoyage

Obs	numepolice	annee_reference	n_sinistre	montantsinistre
1	703479	2010	20122123	2269.038909
2	186906	2012	20122124	318.56930342
3	665352	2013	20122125	191.06868647
4	655353	2014	20122126	21898.294476
5	655353	2014	20122127	204.10113316
6	47415	2015	20122128	211.65011445

Affichage des 6 premières observations apres nettoyage

The MEANS Procedure

Analysis Variable : dubbs				
N	Mean	Std Dev	Minimum	Maximum
61235	1.8279252	0.9561870	1.0000000	6.0000000

Puisque '**Maximum = 6**', on en déduit qu'ils existent des lignes doublées, alors supprimant ces lignes par ce code suivant:

```
/*6-suppression des doublants */
* le nombre de lignes;
proc sql;
    select count(*) as nrows
    from chrt.sinistre;
quit;
*****;
*avant la suppression nrows=65536, *;
* Now : nrows=61235 *;
*****;
```

-Passant maintenant à calculer le nombre de sinistre(nbr_sinistre) par chaque contrat.

```
*/1-calculer le nombre de sinistres par contrat/;
proc sql;
    create table chrt.sinistres as
    select numepolice,annee_reference,count(*) as nbr_sinistre, sum(montantsinistre) as totalmontant
    from chrt.sinistre
    group by annee_reference, numepolice;
quit;
```

```

/*3- fusionner les deux base de données(production et sinistres) */

proc sql;
  create table chrt.Prod_Sin as
  select p.*, s.nbr_sinistre, s.totalmontant
  from chrt.production as p
  left join chrt.sinistres as s
  on p.numepolice= s.numepolice
  and s.annee_reference=p.exercice;
quit;
/* affichage après la fusion */
proc print data=chrt.prod_sin;
run

```

Obs	VAR1	numepolice	exercice	Sexe	Zone	Comubstution	Puissance_Fiscale	exposition	age_du_conducteur	age_mise_en_circulation	nbr_sinistre	totalmontant
1	23573	2	2009	M	Fès-Meknès	Essence	[00-08]	0.8505040312	28	7	1	1986.91
2	34126	5	2016	M	Rabat-Salé-Kénitra	Essence	[00-08]	0.4869068514	27	10	.	.
3	69527	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	28	6	2	3625.85
4	3923	9	2012	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.0609243414	47	19	.	.
5	40470	24	2008	M	Rabat-Salé-Kénitra	Diesel	[08-10]	0.5667409257	27	4	1	174.72
6	18469	36	2011	M	Fès-Meknès	Essence	[00-08]	0.1621531127	50	7	.	.

Figure 2: affichage des variables après la fusionne

- Nettoyage de cette base donnée (prod_sin) engendré par la fusionne des deux tables production et sinistre.

```

/* compter le nombre des doublants */
proc sql;
  create table chrt.dubb as
  select *, count(*) as dubbs
  from chrt.prod_sin
  group by numepolice, exercice
  order by dubbs DESC;
quit;

/* 4- Résumé sur les nombres de valeurs manquantes*/
proc means data=chrt.Prod_Sin nmiss;
output out=chrt.table_nmiss nmiss=;
run;

/* supprimer les var: var1, annee_reference, montantsinistre */
PROC SQL;
ALTER TABLE chrt.Prod_Sin
DROP var1, annee_reference, montantsinistre;
QUIT;
/* supprimer des valeurs manquantes des deux variables ajoutées */
proc sql;
update chrt.prod_sin
set nbr_sinistre=0, totalmontant=0
where nbr_sinistre = . and totalmontant = .;
quit;

```


2.4 Exploration de données

2.4.1 statistique descriptives

```
*statistique descriptives ;
proc means data= chrt.prod_sin;
var nbr_sinistre totalmontant;
run;
title "Analyse de la distribution des nombres de sinistres";
proc univariate data=chrt.prod_sin;
var nbr_sinistre;
run;

title "Analyse de la distribution de'totalmontant'";
proc univariate data=chrt.prod_sin;
var totalmontant;
run;

/* tables de contingences */
proc freq data=chrt.prod_sin noprint ;
tables age_du_conducteur* nbr_sinistre / out=freqout;
run;
/* affichage de la table freqout */
title"Calcul des fréquences croisées entre les variables age_du_conducteur et nbr_sinistre";
proc print data=freqout (obs=10);
run;

/* Calcul des fréquences croisées entre les variables sexe et nbr_sinistre */
proc freq data=chrt.prod_sin noprint;
tables sexe*nbr_sinistre / out=freqout1;
run;

/* Affichage des résultats dans une table */
title "Tableau croisé entre la variable sexe et le nombre de sinistres";
proc print data=freqout1;
run;

proc freq data=chrt.prod_sin noprint ;
tables zone* nbr_sinistre / out=freqout2;
run;

title" Tableau croisé entre la variable catégorielle Zone et le nombre de sinistre";
proc print data= freqout2 (obs=10);
run;

proc freq data=chrt.prod_sin noprint ;
tables age_mise_en_circulation* nbr_sinistre / out=freqout3;
run;

title" Tableau croisé entre la variable age_mise_en_circulation et le nombre de sinistre";
proc print data= freqout3 (obs=10);
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
nbr_sinistre	100000	0.6123500	0.8627659	0	6.0000000
totalmontant	100000	4172.12	20959.75	0	1056627.39

Analyse de la distribution des nombres de sinistres

The UNIVARIATE Procedure Variable: nbr_sinistre

Moments			
N	100000	Sum Weights	100000
Mean	0.61235	Sum Observations	61235
Std Deviation	0.86276586	Variance	0.74436492
Skewness	1.57098113	Kurtosis	2.69831955
Uncorrected SS	111933	Corrected SS	74435.7477
Coeff Variation	140.894236	Std Error Mean	0.00272831

Basic Statistical Measures			
Location		Variability	
Mean	0.612350	Std Deviation	0.86277
Median	0.000000	Variance	0.74436
Mode	0.000000	Range	6.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	224.4434	Pr > t 	<.0001
Sign	M	21087.5	Pr >= M 	<.0001
Signed Rank	S	4.4469E8	Pr >= S 	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	6
99%	3
95%	2
90%	2
75% Q3	1
50% Median	0

Quantiles (Definition 5)	
Level	Quantile
25% Q1	0
10%	0
5%	0
1%	0
0% Min	0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	100000	6	90692
0	99999	6	91466
0	99996	6	92804
0	99995	6	93402
0	99994	6	97371

Analyse de la distribution de'totalmontant'

The UNIVARIATE Procedure
Variable: totalmontant

Moments			
N	100000	Sum Weights	100000
Mean	4172.11671	Sum Observations	417211671
Std Deviation	20959.7494	Variance	439311095
Skewness	15.5354108	Kurtosis	398.307054
Uncorrected SS	4.56713E13	Corrected SS	4.39307E13
Coeff Variation	502.376872	Std Error Mean	66.2805473

Basic Statistical Measures			
Location		Variability	
Mean	4172.117	Std Deviation	20960
Median	0.000	Variance	439311095
Mode	0.000	Range	1056627
		Interquartile Range	1064

Tests for Location: Mu0=0		
Test	Statistic	p Value

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	62.94632	Pr > t	<.0001
Sign	M	21087.5	Pr >= M	<.0001
Signed Rank	S	4.4469E8	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	1056627.39
99%	78943.39
95%	18720.86
90%	6966.88
75% Q3	1063.99
50% Median	0.00
25% Q1	0.00
10%	0.00
5%	0.00
1%	0.00
0% Min	0.00

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	100000	760644	85471
0	99999	765634	89777
0	99996	920874	7968
0	99995	970448	91466
0	99994	1056627	21752

Calcul des fréquences croisées entre les variables age_du_conducteur et nbr_sinistre

Obs	age_du_conducteur	nbr_sinistre	COUNT	PERCENT
1	18	0	1	0.001
2	19	0	205	0.205
3	19	1	66	0.066
4	19	2	26	0.026
5	19	3	10	0.010
6	19	4	3	0.003

Obs	age_du_conducteur	nbr_sinistre	COUNT	PERCENT
7	20	0	361	0.361
8	20	1	192	0.192
9	20	2	58	0.058
10	20	3	24	0.024

Tableau croisé entre la variable sexe et le nombre de sinistres

Obs	Sexe	nbr_sinistre	COUNT	PERCENT
1	F	0	23148	23.148
2	F	1	11284	11.284
3	F	2	4089	4.089
4	F	3	1133	1.133
5	F	4	268	0.268
6	F	5	68	0.068
7	F	6	10	0.010
8	M	0	34677	34.677
9	M	1	16783	16.783
10	M	2	6175	6.175
11	M	3	1810	1.810
12	M	4	448	0.448
13	M	5	95	0.095
14	M	6	12	0.012

Tableau croisé entre la variable catégorielle Zone et le nombre de sinistre

Obs	Zone	nbr_sinistre	COUNT	PERCENT
1	BéniMellal-Khénifra	0	2858	2.858
2	BéniMellal-Khénifra	1	1462	1.462
3	BéniMellal-Khénifra	2	489	0.489
4	BéniMellal-Khénifra	3	152	0.152
5	BéniMellal-Khénifra	4	27	0.027
6	BéniMellal-Khénifra	5	10	0.010
7	BéniMellal-Khénifra	6	2	0.002
8	Casablanca-Settat	0	11513	11.513
9	Casablanca-Settat	1	5640	5.640

Obs	Zone	nbr_sinistre	COUNT	PERCENT
10	Casablanca-Settat	2	2062	2.062

Tableau croisé entre la variable age_mise_en_circulation et le nombre de sinistre

Obs	age_mise_en_circulation	nbr_sinistre	COUNT	PERCENT
1	0	0	2	0.002
2	1	0	392	0.392
3	1	1	153	0.153
4	1	2	58	0.058
5	1	3	21	0.021
6	1	4	1	0.001
7	2	0	798	0.798
8	2	1	334	0.334
9	2	2	106	0.106
10	2	3	24	0.024

2.4.2 Générer un ensemble de diagrammes en utilisant sgplot

```
***distribution de nbr_sinistre;
title"Analyse de la distribution du nombre de sinistres ";
proc univariate data=chrt.prod_sin;
var nbr_sinistre;
histogram ; /* Création d'un histogramme */
ods select histogram basicstats; /* Affichage des statistiques de base et de l'histogramme */
***distribution de totalmontant;
title "Graphique à barres verticales de la variable totalmontat ";
proc univariate data=chrt.prod_sin;
var totalmontant;
histogram ; /* Création d'un histogramme */
ods select histogram basicstats; /* Affichage des statistiques de base et de l'histogramme */
* 5-2-1 age_du_conducteur & nbr_sinistre;
proc sgplot data=chrt.prod_sin;
  vbar age_du_conducteur / response=nbr_sinistre stat=mean groupdisplay=stack;
run;
*5-2-2 age_mise_en_circulation & nbr_sinistre;
proc sgplot data=chrt.prod_sin;
  vbar age_mise_en_circulation / response=nbr_sinistre groupdisplay=stack;
run;
*5-2-3 sexe & nbr_sinistre;
proc sgplot data=chrt.prod_sin;
  vbar sexe / response=nbr_sinistre groupdisplay=stack;
run;
*5-2-4 zone & nbr_sinistre;
proc sgplot data=chrt.prod_sin;
  vbar zone / response=nbr_sinistre groupdisplay=stack;
run;
*5-2-5 age_du_conducteur & totalmontant;
proc sgplot data=chrt.prod_sin;
  vbar age_du_conducteur / response=totalmontant stat=mean groupdisplay=stack;
run;
*5-1-6 age_mise_en_circulation & totalmontant;
proc sgplot data=chrt.prod_sin;
  vbar age_mise_en_circulation / response=totalmontant stat=mean groupdisplay=stack;
run;
*5-1-7 sexe & totalmontant;
proc sgplot data=chrt.prod_sin;
  vbar sexe / response=totalmontant groupdisplay=stack;
run;
*5-2-8 zone & totalmontant;
proc sgplot data=chrt.prod_sin;
  vbar zone / response=totalmontant groupdisplay=stack;
run;
*5-2-9 puissance_fiscale & totalmontant;
proc sgplot data=chrt.prod_sin;
  vbar puissance_fiscale/ response=totalmontant groupdisplay=stack;
run;
*5-2-10 puissance_fiscale & nbr_sinistre;
proc sgplot data=chrt.prod_sin;
  vbar puissance_fiscale/ response=nbr_sinistre groupdisplay=stack;
run;
*****;
*Interpretation:
*****;
```

Graphique à barres verticales de la variable totalmontat

The UNIVARIATE Procedure

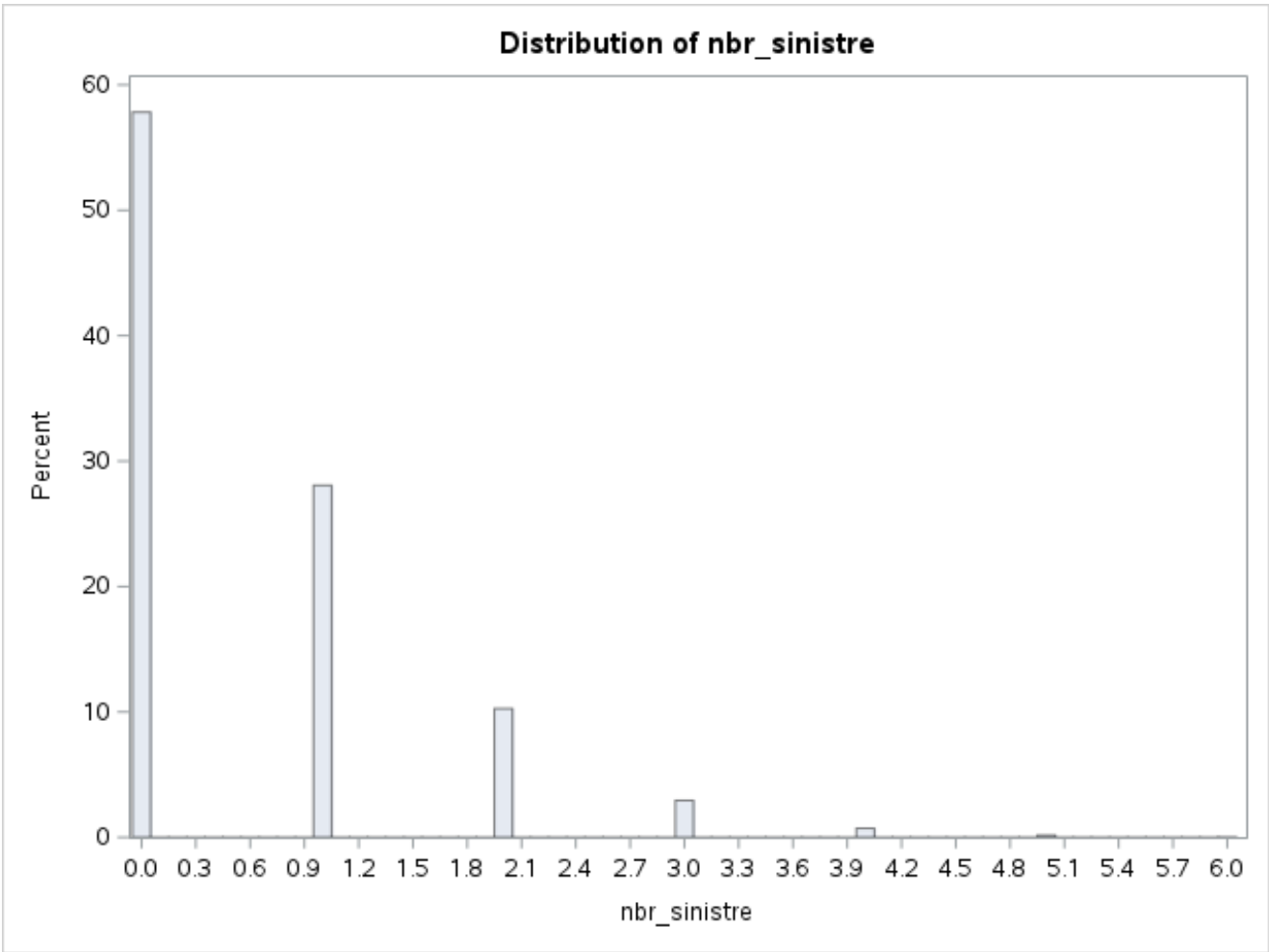
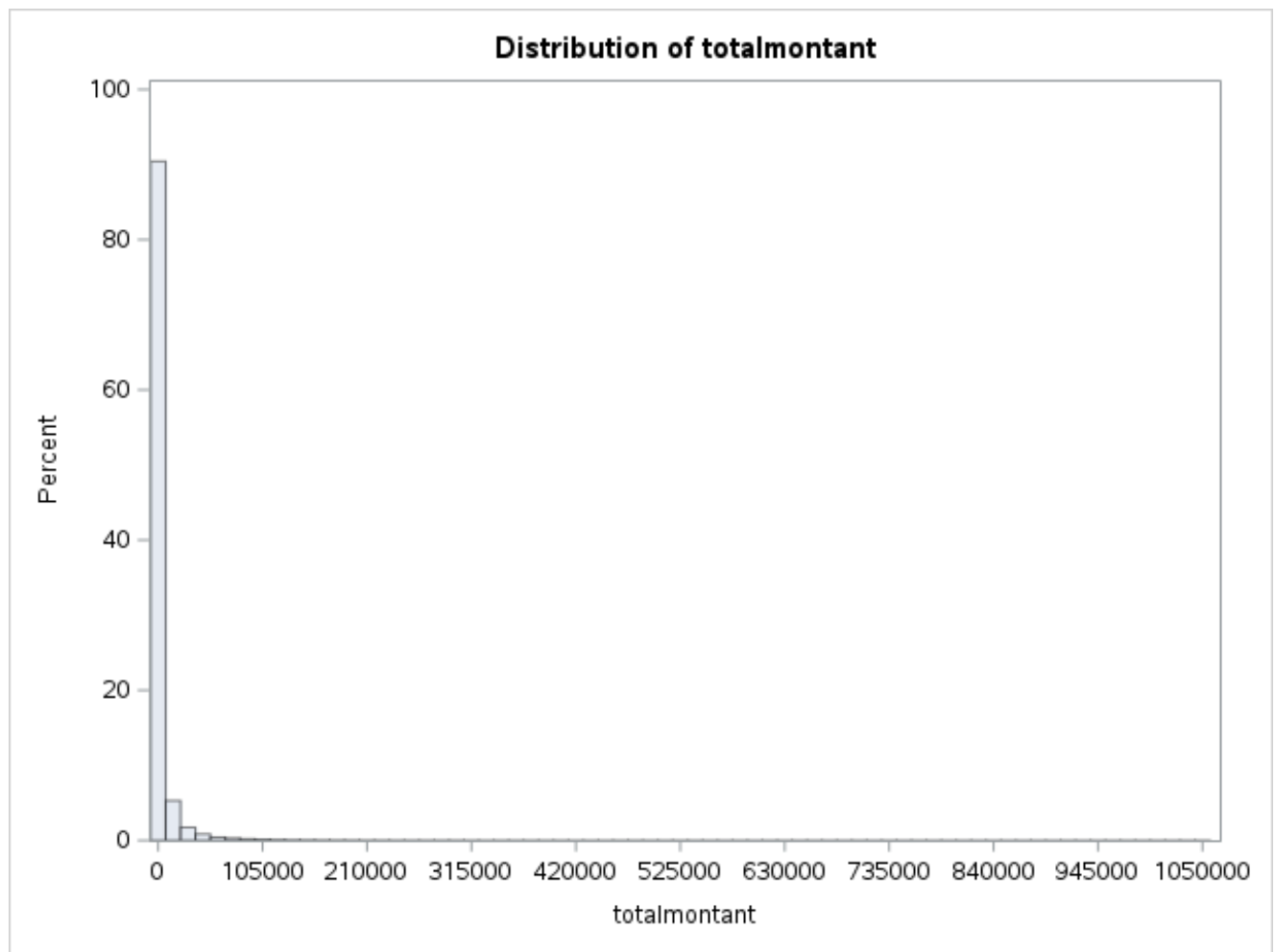


Diagramme à barres empilées du nombre de sinistres par age du conducteur

The UNIVARIATE Procedure



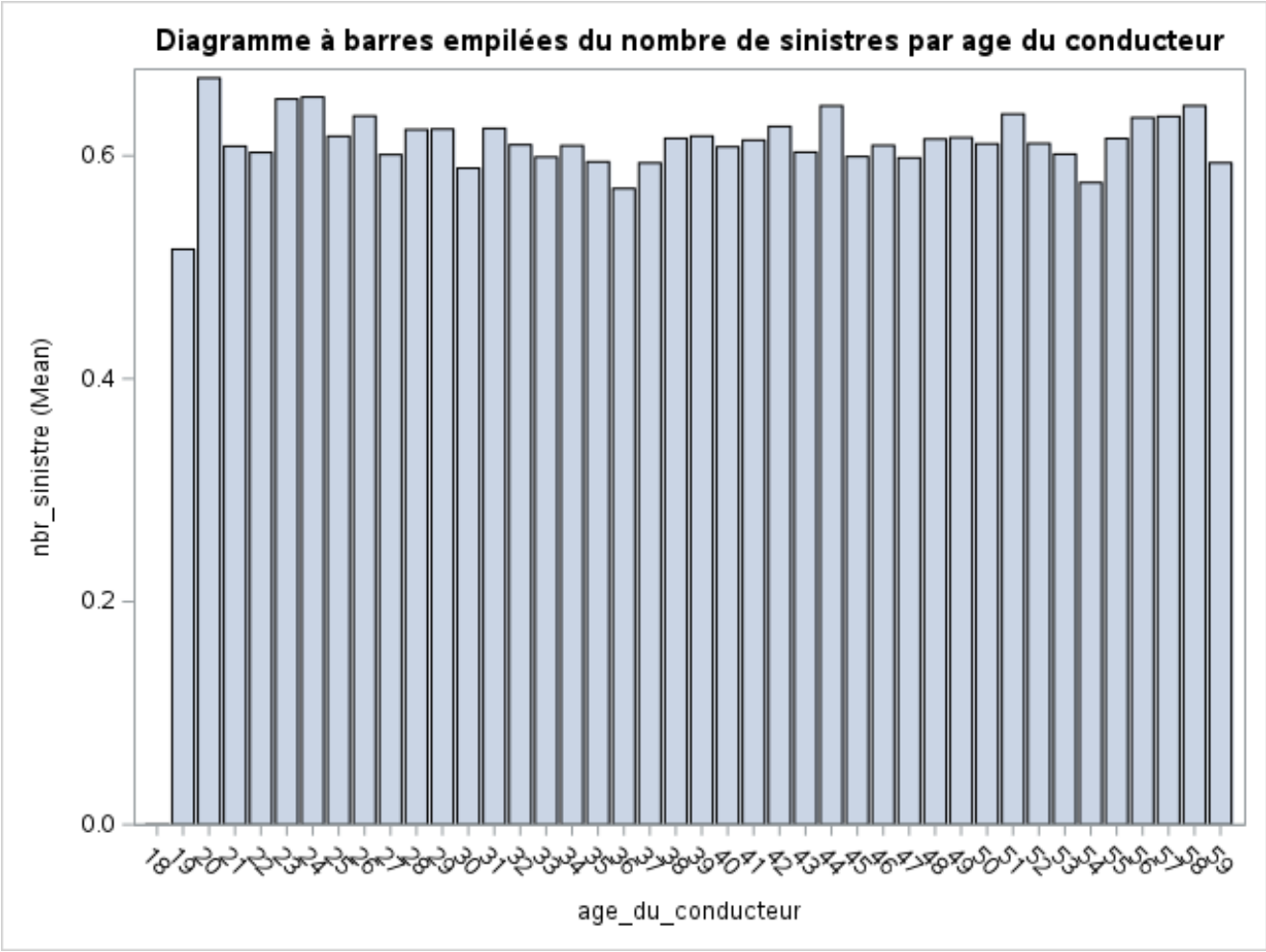


Diagramme à barres empilées du nombre de sinistres par age de mise en circulation

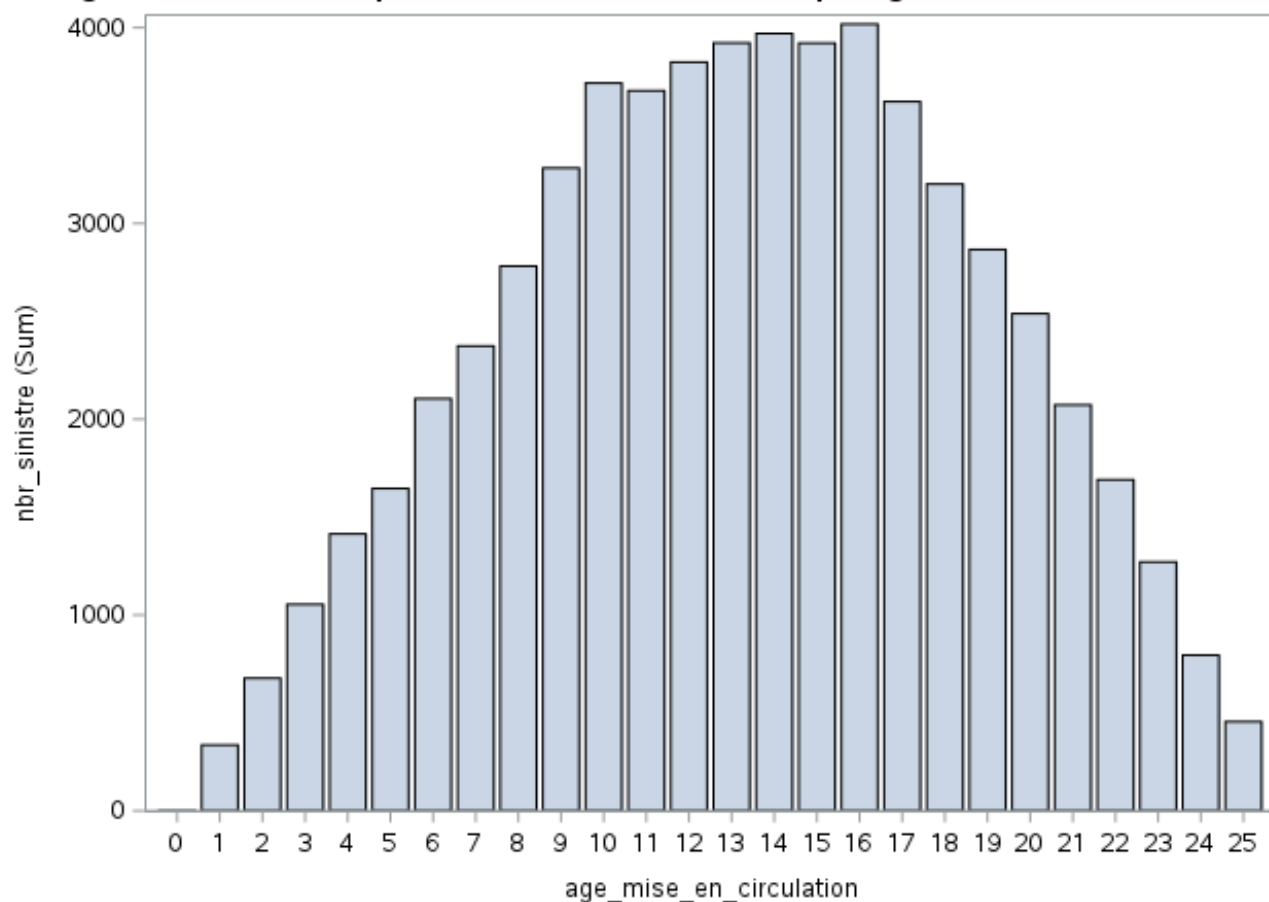


Diagramme à barres empilées du nombre de sinistres par sexe

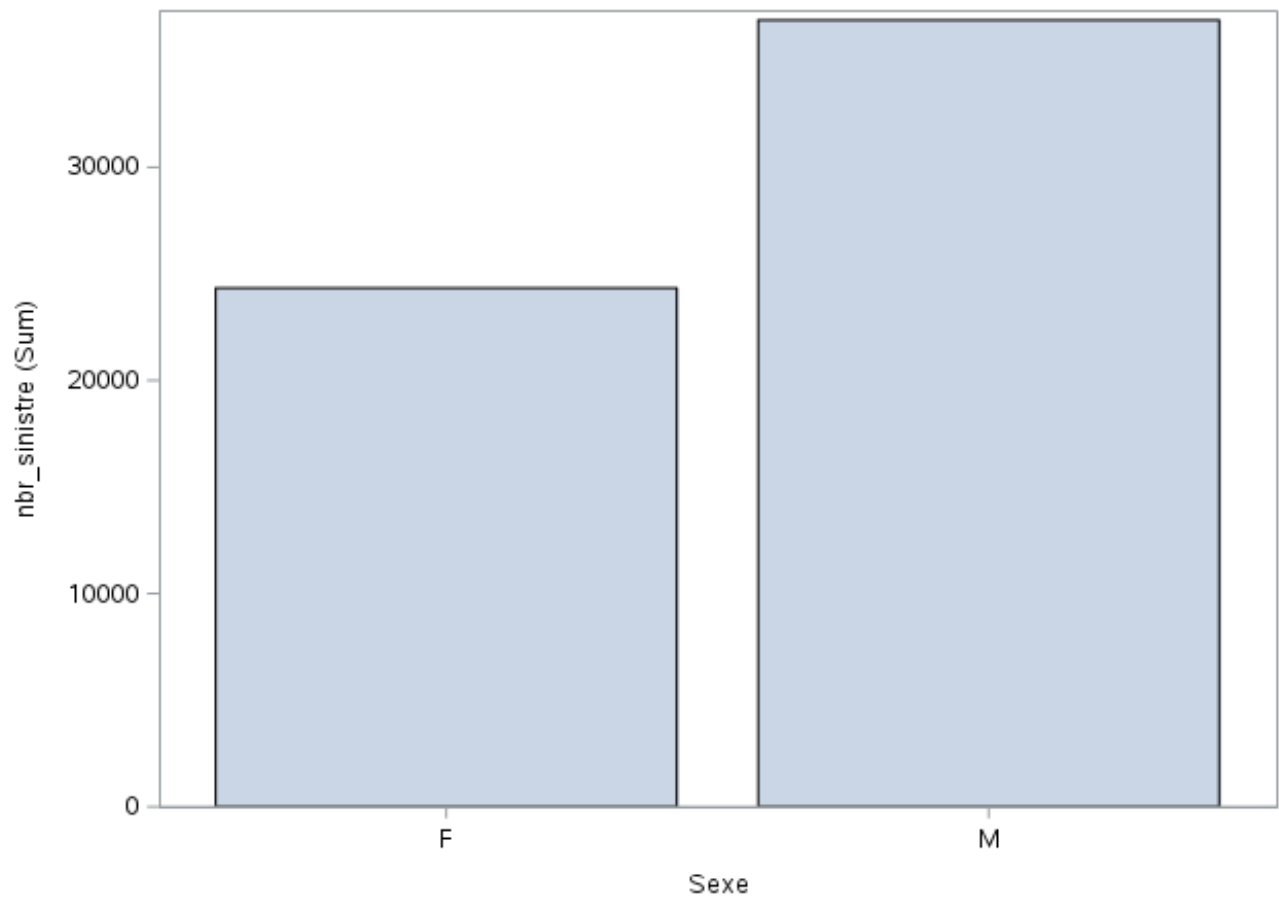


Diagramme à barres empilées du nombre de sinistres par zone

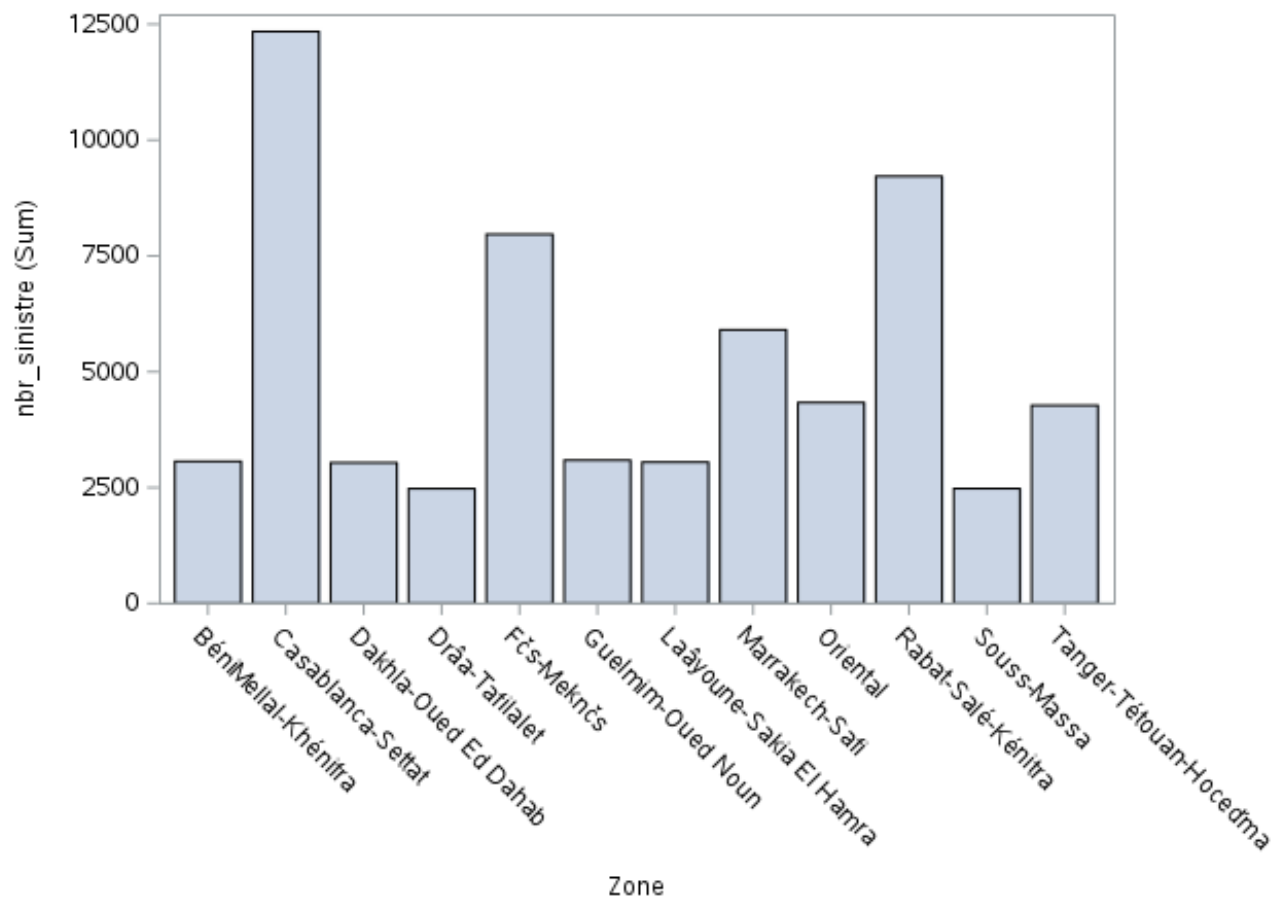


Diagramme à barres empilées du totalmontant par age du conducteur

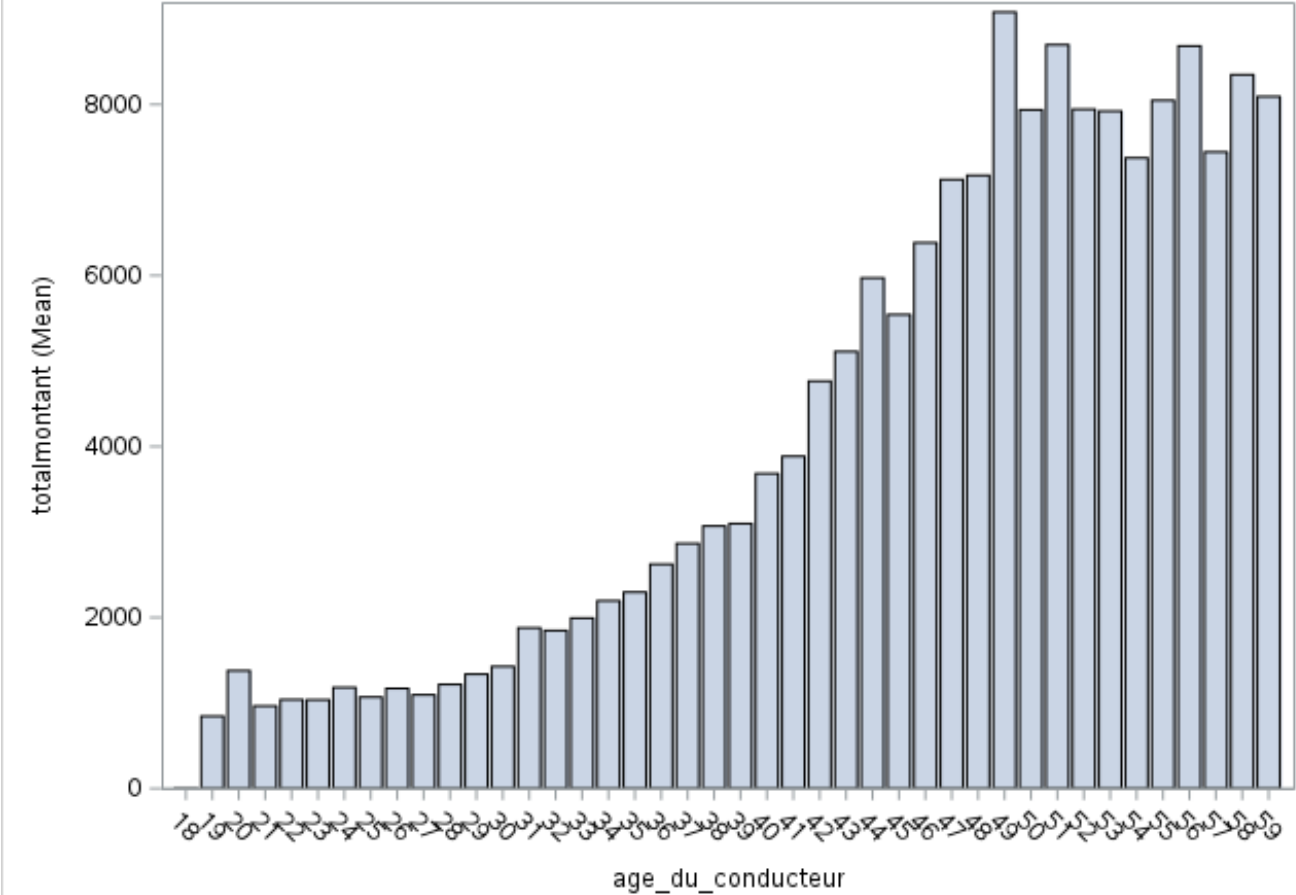


Diagramme à barres empilées du totalmontant par age de mise en circulation

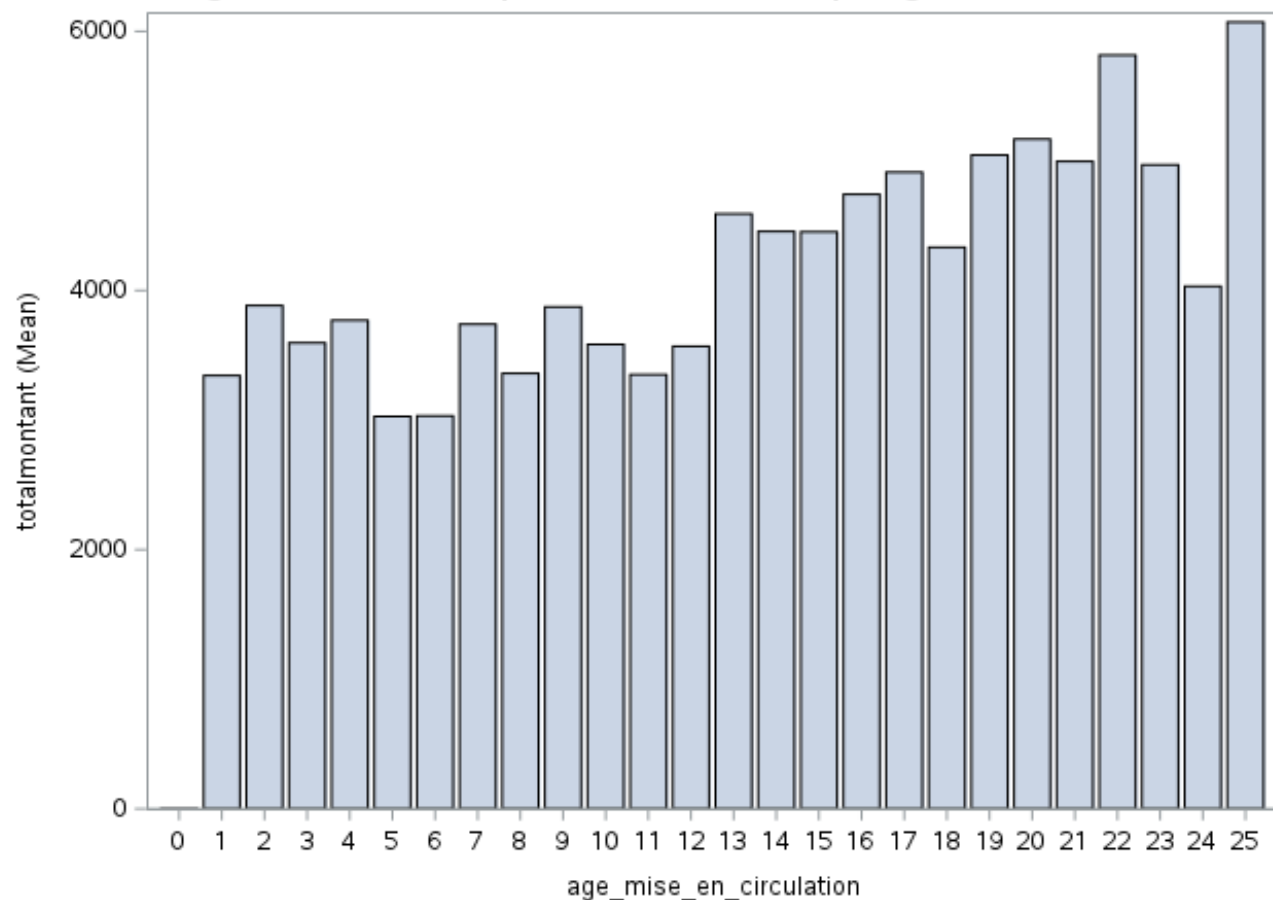




Diagramme à barres empilées du totalmontant par zone

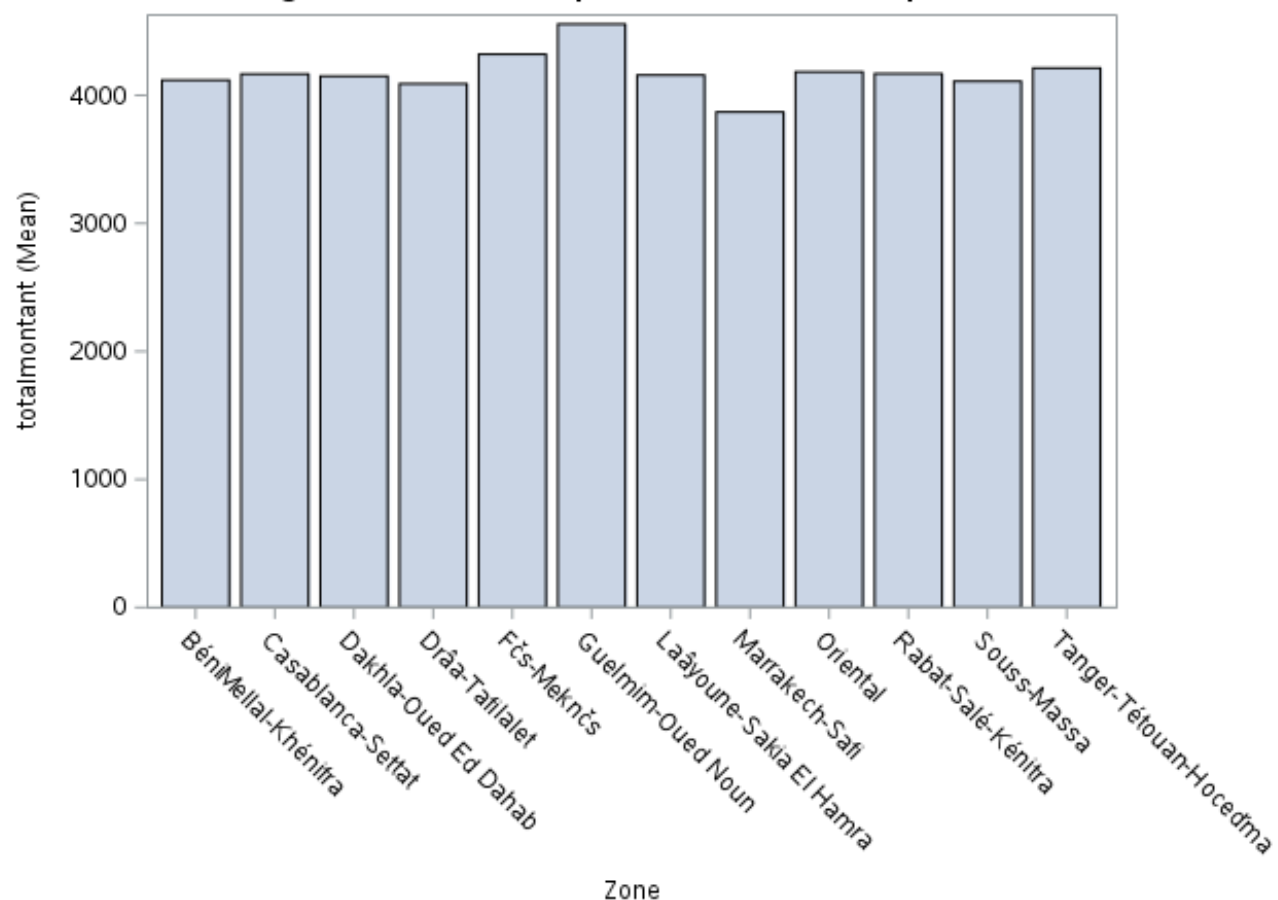


Diagramme à barres empilées du totalmontant par puissance fiscale

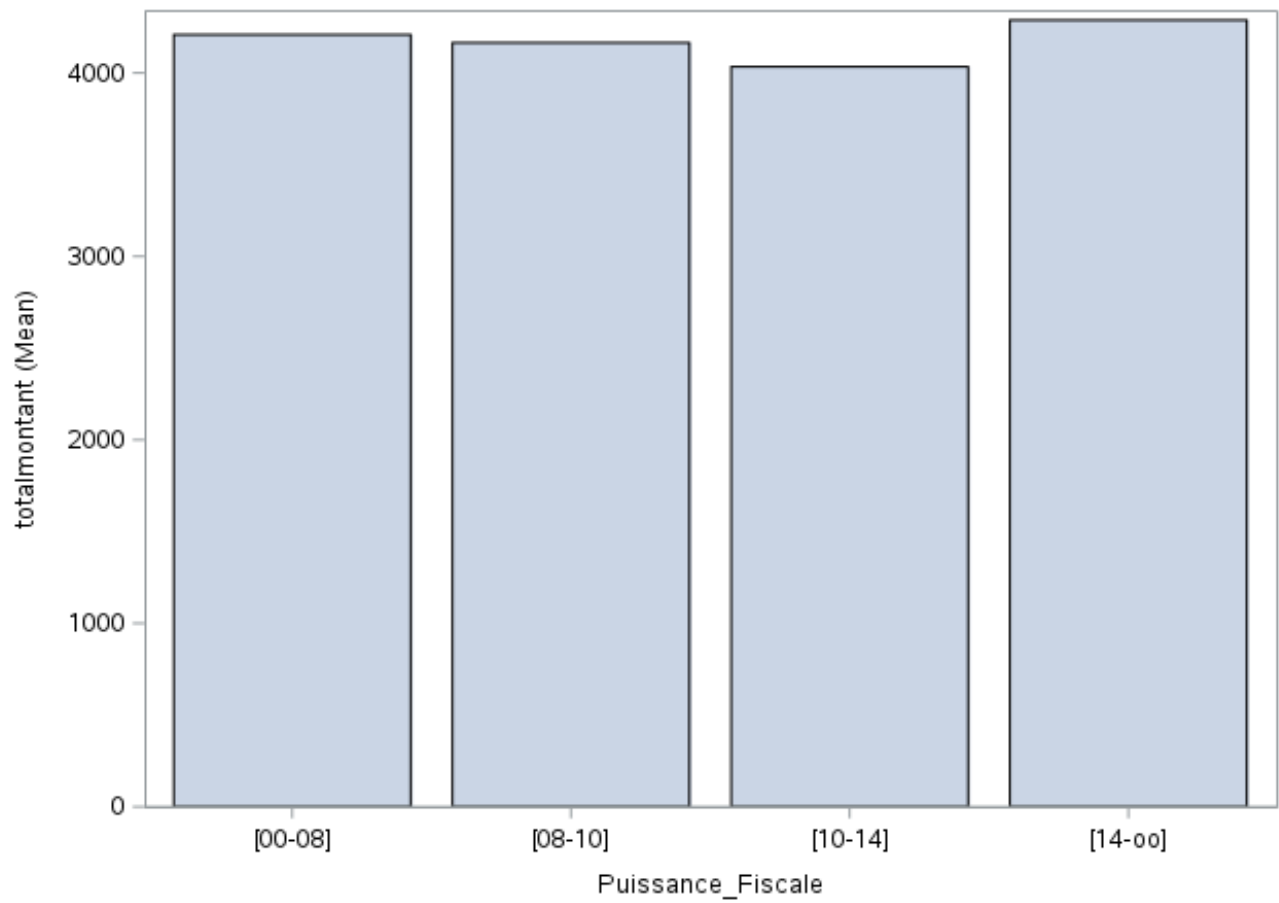
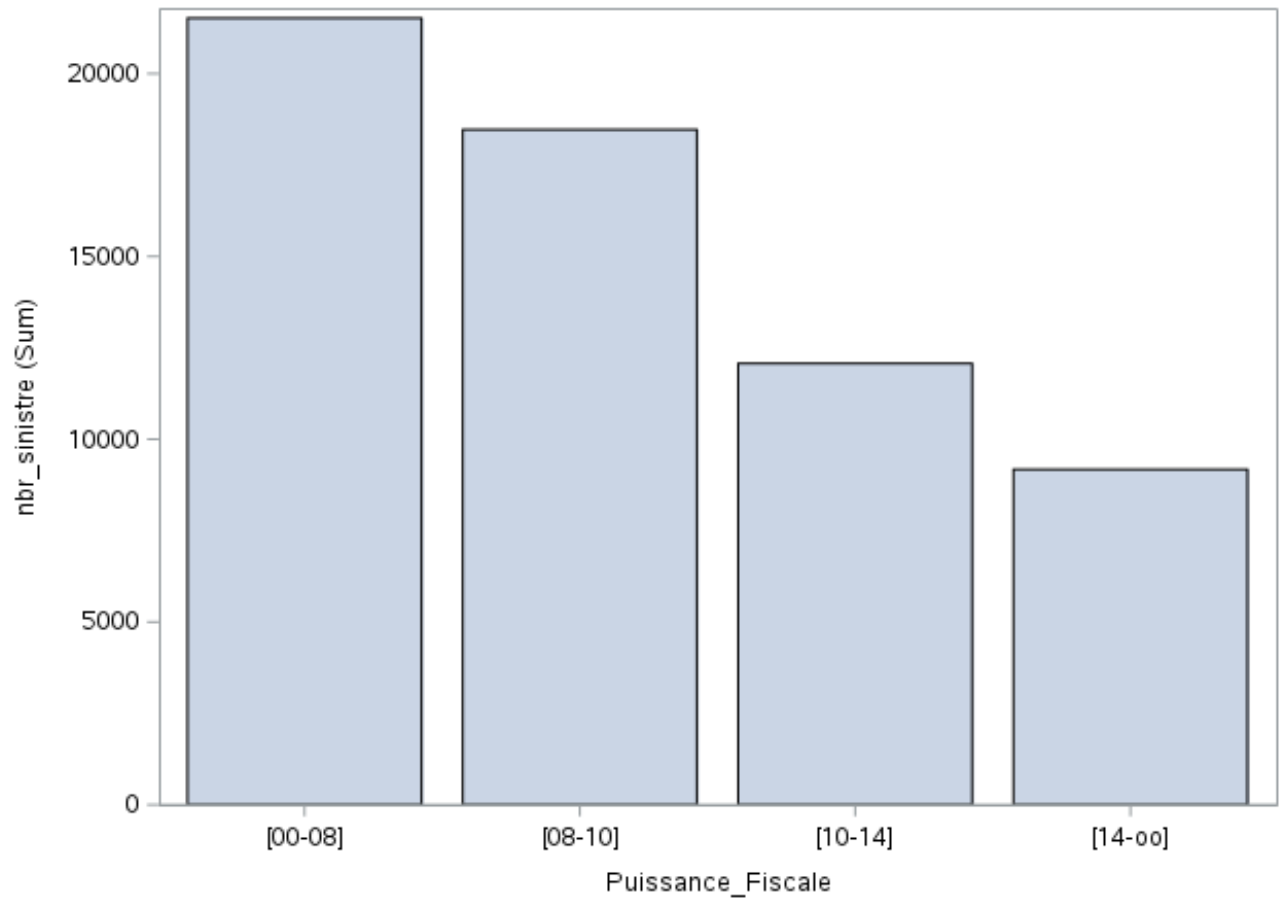


Diagramme à barres empilées du nombre de sinistre par puissance fiscale



```

/* Creation du format d'age */
/* creation d'une table summary de l'age du conducteur pour en déduire les classes adéquates
   en basant sur les ages ayant meme pourcentage de presentation dans la base de donnée */
proc freq data=chrt.prod_sin;
tables age_du_conducteur / nocol norow nocum out =tabb;
run;
proc format;
    value ageFormat
        18-19=" [18,19] "
        20-24=" [20,24] "
        25-30=" [25,30] "
        30-49=" [30,49] "
        50-52=" [50,52] "
        53-55=" [53,55] "
        56-57=" [56,57] " ;
run;
data chrt.prod_sin_age;
set chrt.prod_sin;
if age_du_conducteur>=18 and age_du_conducteur<=19 then
ageFormat=" [18,19] ";
if age_du_conducteur>=20 and age_du_conducteur<=24 then
ageFormat=" [20,24] ";
if age_du_conducteur>=25 and age_du_conducteur<=30 then
ageFormat=" [25,30] ";
if age_du_conducteur>=31 and age_du_conducteur<=49 then
ageFormat=" [31,49] ";
if age_du_conducteur>=50 and age_du_conducteur<=52 then
ageFormat=" [50,52] ";
if age_du_conducteur>=53 and age_du_conducteur<=57 then
ageFormat=" [53,57] ";
run;

```

```

/* creation de la table prod_sin_ts qui va conetenir le taux de sinistralité Notée ts */
proc sql;
create table chrt.prod_sin_ts as
    select sexe , ageFormat, sum(nbr_sinistre*exposition)/sum(exposition) as ts
    from chrt.prod_sin_age
    group by ageFormat, sexe;
run;
/* affichage de la table */
title"Création d'une table pour le taux de sinistralité par sexe et par tranche d'âge";
proc print data=chrt.prod_sin_ts;
run;

```

Le taux de sinistralité selon le sexe et l'âge est une mesure qui permet d'évaluer la fréquence des sinistres (accidents) pour des groupes de conducteurs de différents sexes et tranches d'âge. Cette mesure est importante car les caractéristiques démographiques et les comportements de conduite diffèrent souvent entre les groupes d'âge et de sexe, ce qui peut influencer le risque d'accident. Le calcul du taux de sinistralité pondéré qui prend en compte les différences dans l'exposition entre les groupes d'âge et de sexe. Cela permet de comparer les taux de sinistralité entre les groupes de manière plus précise et de mieux évaluer les risques associés à la conduite pour différents groupes de population.

Création d'une table pour le taux de sinistralité par sexe et par tranche d'âge

Obs	Sexe	ageFormat	ts
1	F		0.78462
2	M		0.81999
3	F	[18,19]	0.62958
4	M	[18,19]	0.82481
5	F	[20,24]	0.83628
6	M	[20,24]	0.85093
7	F	[25,30]	0.81548
8	M	[25,30]	0.81963
9	F	[31,49]	0.80792
10	M	[31,49]	0.81999
11	F	[56,57]	0.82260
12	M	[56,57]	0.82818

Figure 3: table(sexe, ageFormat, ts)

Un graphique à nuage de points avec une ligne de tendance pour représenter la relation entre l'âge du conducteur et le taux de sinistralité

```
proc sgplot data=chrt.prod_sin_ts;
title "Taux de sinistralité vs Age du conducteur";
scatter x=ageFormat y=ts / filledoutlinedmarkers
    markerfillattrs=(color='darkred')
    markeroutlineattrs=(color='black' thickness=2)
    markerattrs=(symbol=circlefilled size =15);
xaxis grid label="age";
yaxis grid label="taux de sinistralité";
series x=ageFormat y =ts / lineattrs=(color='darkblue');
run;
```

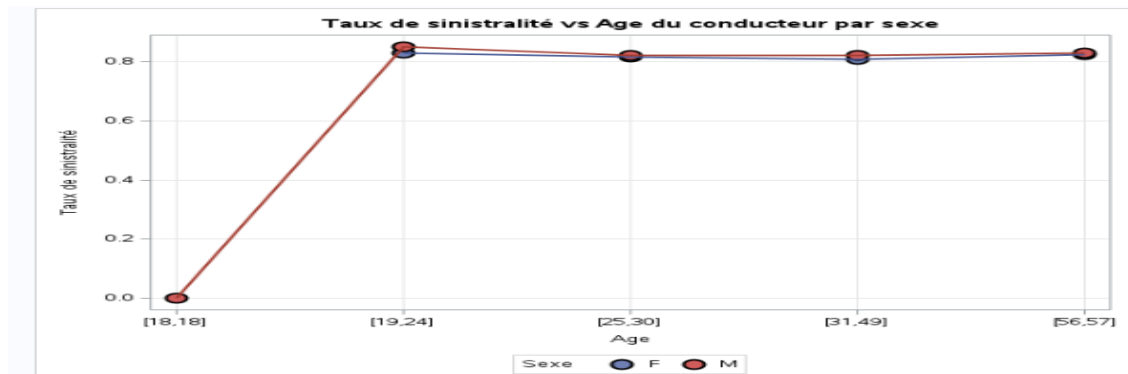


Figure 4: Evolution de taux de sinistralité selon les tranches d'âge et le sexe

Interprétation du graphe

Le graphique montre comment le taux de sinistralité varie en fonction de l'âge des conducteurs pour chaque groupe de sexe. Les points de données représentent le taux de sinistralité pour chaque groupe d'âge et de sexe, et la ligne de tendance représente la relation entre l'âge et le taux de sinistralité pour chaque groupe de sexe. -on peut dire que le taux de sinistralité semble diminuer avec l'âge jusqu'à les deux tranches d'âges [25,30] et [31,49] et puis une petite augmentation jusqu'à la tranche d'âge [56,57] pour les conducteurs de chaque sexe. De plus, le taux de sinistralité des conductrices semble être inférieur à celui des conducteurs pour chaque groupe d'âge, ce qui suggère une corrélation négative entre le sexe et le taux de sinistralité. et ceci dû le fait que les hommes sont plus agressif que les femmes.

2.5 Modelisation

2.5.1 Modelisation de "nombre de sinistre"

– Modèle de Poisson:

```
data chrt.prod_sin;
set chrt.prod_sin;
offset=log(exposition);
run;
proc genmod data = chrt.prod_sin;
    class sexe ;
    model nbr_sinistre= sexe age_du_conducteur age_mise_en_circulation /
    dist=poisson link=log offset=offset;
    ods output modelfit=fit;
run;
data fit;
    set fit;
    pval=1-cdf("chisq", value,df);
run;
```

The GENMOD Procedure

Model Information	
Data Set	CHRT.PROD_SIN
Distribution	Poisson
Link Function	Log
Dependent Variable	nbr_sinistre
Offset Variable	offset

Number of Observations Read	100000
Number of Observations Used	100000

Class Level Information		
Class	Levels	Values
Sexe	2	F M

Algorithm converged.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1E5	94837.7931	0.9484
Scaled Deviance	1E5	94837.7931	0.9484
Pearson Chi-Square	1E5	99721.4982	0.9973
Scaled Pearson X2	1E5	99721.4982	0.9973
Log Likelihood		-79206.7721	
Full Log Likelihood		-94794.9741	
AIC (smaller is better)		189597.9482	
AICC (smaller is better)		189597.9486	
BIC (smaller is better)		189635.9999	

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.0810	0.0186	0.0445	0.1176	18.90	<.0001
Sexe	F	1	-0.0124	0.0083	-0.0285	0.0038	2.24	0.1345
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.	.
age_du_conducteur		1	-0.0011	0.0004	-0.0019	-0.0002	6.44	0.0111
age_mise_en_circulat		1	0.0128	0.0008	0.0113	0.0143	285.56	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.



Interpretation

- On constate que AIC=189597.9482, les p-value sont tous < 0.05 sauf la variable binaire sexe qui n'est pas significative
- la déviance et la déviance ajustée ont des valeurs de 94837,7931 et un nombre de degrés de liberté de 100 000, ce qui donne une valeur de 0,9484 pour la déviance/df. Cette valeur est relativement faible, ce qui suggère que le modèle peut être bien ajusté aux données.
- la statistique de Pearson a une valeur de 99721,4982 et un nombre de degrés de liberté de 100 000, ce qui donne une valeur de 0,9973 pour la statistique de Pearson/df. Cette valeur est proche de 1, ce qui suggère que le modèle peut être bien ajusté aux données.

– Modele de Binomial-Negative:

Rappel:

La **surdispersion** se produit lorsque la variance de la variable dépendante(nbr_sinistre) est plus grande que ce qui serait attendu selon la distribution théorique choisie. Dans le cas de la régression de Poisson, la variance est égale à la moyenne, ce qui implique que la variance doit être égale à la moyenne pour que les données soient bien décrites par une distribution de Poisson. Cependant, dans certaines situations, la variance peut être plus grande que la moyenne, ce qui indique une **surdispersion**. -Test de surdispersion

$$\begin{cases} H_0 : \text{equidispersion}, \\ H_1 : \text{surdispersion} \end{cases}$$

On rejette H_0 si la p-value <0.05 , par suite on peut déduire qu'il y a une surdispersion ce qui signifie la variance est loin de moyenne.

```
/* modele de binomial negative */
proc genmod data=chrt.prod_sin;
  class sexe ;
  model nbr_sinistre = sexe age_du_conducteur age_mise_en_circulation /
    dist= negbin link=log offset=offset scale=0 noscale;
  ods output info;
run;
```

The GENMOD Procedure

Model Information	
Data Set	CHRT.PROD_SIN
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	nbr_sinistre
Offset Variable	offset

Number of Observations Read	100000
Number of Observations Used	100000

Class Level Information		
Class	Levels	Values
Sexe	2	F M

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	1E5	94837.7931	0.9484
Scaled Deviance	1E5	94837.7931	0.9484
Pearson Chi-Square	1E5	99721.4982	0.9973
Scaled Pearson X2	1E5	99721.4982	0.9973
Log Likelihood		-79206.7721	
Full Log Likelihood		-94794.9741	
AIC (smaller is better)		189597.9482	
AICC (smaller is better)		189597.9486	
BIC (smaller is better)		189635.9999	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.0810	0.0186	0.0445	0.1176	18.90	<.0001
Sexe	F	1	-0.0124	0.0083	-0.0285	0.0038	2.24	0.1345
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.	.
age_du_conducteur		1	-0.0011	0.0004	-0.0019	-0.0002	6.44	0.0111
age_mise_en_circulat		1	0.0128	0.0008	0.0113	0.0143	285.56	<.0001
Dispersion		0	0.0000	0.0000				

Note: The negative binomial dispersion parameter was held fixed.

Lagrange Multiplier Statistics

Parameter	Chi-Square	Pr > ChiSq	
Dispersion	1.5649	0.1055	*
* One-sided p-value			



Interpretation

- La déviance et la déviance ajustée ont des valeurs de 94837,7931 et un nombre de degrés de liberté de 100 000, ce qui donne une valeur de 0,9484 pour la **déviance/df**. Cette valeur est relativement faible, ce qui suggère que le modèle peut être bien ajusté aux données selon cette mesure d'ajustement.
- La statistique de test du test de Lagrange pour la surdispersion a une valeur de 1,5649 et une p-value de 0,1055. Cette p-value est supérieure au seuil de signification de 5%, ce qui implique l'acceptation de l'hypothèse H_0 .

–Modèle ZIP (Zero-Inflated Poisson)

```
/* Model ZIP */
proc genmod data = chrt.prod_sin;
  class sexe ;
  model nbr_sinistre= sexe age_du_conducteur age_mise_en_circulation /
  dist=zip link=log offset=offset;
  zeromodel /link=logit;
  ods output modelfit=fit;
run;
data fit;
  set fit;
  pval=1-cdf("chisq", value,df);
run;
```


The GENMOD Procedure

Model Information		
Data Set	CHRT.PROD_SIN	Predicted Values and Diagnostic Statistics
Distribution	Zero Inflated Poisson	
Link Function	Log	
Dependent Variable	nbr_sinistre	
Offset Variable	offset	
Zero Model Link Function	Logit	

Number of Observations Read	100000
Number of Observations Used	100000

Class Level Information		
Class	Levels	Values
Sexe	2	F M

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance		189588.8097	
Scaled Deviance		189588.8097	
Pearson Chi-Square	1E5	99372.9073	0.9938
Scaled Pearson X2	1E5	99372.9073	0.9938
Log Likelihood		-79206.2029	
Full Log Likelihood		-94794.4049	
AIC (smaller is better)		189598.8097	
AICC (smaller is better)		189598.8103	
BIC (smaller is better)		189646.3744	

Algorithm converged.

Analysis Of Maximum Likelihood
Zero Inflation Parameter Estimates

Parameter	DF	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square
Intercept	1	-5.1693	-7.0092 -3.3293	30.32

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.0867	0.0194	0.0486	0.1247	19.94	<.0001
Sexe	F	1	-0.0124	0.0083	-0.0286	0.0038	2.24	0.1347
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.	.
age_du_conducteur		1	-0.0011	0.0004	-0.0019	-0.0002	6.40	0.0114
age_mise_en_circulat		1	0.0128	0.0008	0.0113	0.0143	284.17	<.0001
Scale		0	1.0000	0.0000	1.0000	1.0000		

Note: The scale parameter was held fixed.



Interpretation

- la déviance et la déviance ajustée ont des valeurs de 189588,8097. Cette valeur est plus élevée que celle du modèle précédent, ce qui suggère que le modèle de régression avec une distribution de loi ZIP peut ne pas être bien ajusté aux données.
- Les résultats de l'analyse des paramètres d'inflation zéro montrent que l'intercept a une valeur estimée de -5,1693 avec une erreur standard de 0,9388. Cela suggère que la probabilité d'avoir un sinistre est significativement plus faible que la probabilité d'avoir zéro sinistre.

– Modele Zero-Inflated Negative Binomial(ZINB):

```
/* Model ZINB */
proc genmod data = chrt.prod_sin;
  class sexe ;
  model nbr_sinistre= sexe age_du_conducteur age_mise_en_circulation /
  dist=zinb link=log offset=offset;
  zeromodel / link=logit;
  ods output modelfit=fit;
run;
data fit;
  set fit;
  pval=1-cdf("chisq", value,df);
run;
```

The GENMOD Procedure

Model Information		
Data Set	CHRT.PROD_SIN	Predicted Values and Diagnostic Statistics
Distribution	Zero Inflated Negative Binomial	
Link Function	Log	
Dependent Variable	nbr_sinistre	
Offset Variable	offset	
Zero Model Link Function	Logit	

Number of Observations Read	100000
Number of Observations Used	100000

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance		189588.3599	
Scaled Deviance		189588.3599	
Pearson Chi-Square	1E5	99231.6616	0.9924
Scaled Pearson X2	1E5	99231.6616	0.9924
Log Likelihood		-94794.1799	
Full Log Likelihood		-94794.1799	
AIC (smaller is better)		189600.3599	
AICC (smaller is better)		189600.3607	
BIC (smaller is better)		189657.4374	

Class Level Information

Class	Levels	Values
Sexe	2	F M

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates						
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Pr > ChiSq
Intercept	1	0.0810	0.0206	0.0406	0.1213	15.47
Sexe	F 1	-0.0123	0.0083	-0.0286	0.0039	2.22
Sexe	M 0	0.0000	0.0000	0.0000	0.0000	.
age_du_conducteur	1	-0.0011	0.0004	-0.0019	-0.0002	6.38
age_mise_en_circulat	1	0.0128	0.0008	0.0113	0.0143	283.76
Dispersion	1	0.0080	0.0108	0.0006	0.1127	

Note: The negative binomial dispersion parameter was estimated by maximum likelihood.

Analysis Of Maximum Likelihood Zero Inflation Parameter Estimates						
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Pr > ChiSq
Intercept	1	-10.8695	448.2025	-889.330	867.5913	0.00

Interpretation

- La déviance et la déviance ajustée ont des valeurs de 189588,3599. Cette valeur est similaire à celle du modèle précédent, ce qui suggère que le modèle de régression avec une distribution de loi ZINB peut ne pas être bien ajusté aux données.

La statistique de Pearson a une valeur de 99231,6616 et un nombre de degrés de liberté de 100 000, ce qui donne une valeur de 0,9924 pour la statistique de Pearson/df. Cette valeur est proche de 1, ce qui suggère que le modèle peut être bien ajusté aux données selon cette mesure d'ajustement.

- L'intercept de ce modèle a une valeur estimée de -10,8695 avec une erreur standard de 448,2025. Cependant, la plage de confiance des intervalles de confiance à 95% est très large, allant de -889.330 à 867.5913. Cela suggère que les estimations pourraient ne pas être précises.

Conclusion :

D'après le tableau qui est en dessous on en déduit que les deux modèles acceptés sont le modèle de poisson et le modèle binomiale-négatif. Cependant, le test de dispersion pour le modèle de binôme négatif a une valeur de 1,5649 avec une valeur p de 0,1055, ce qui suggère que la surdispersion dans les données n'est pas statistiquement significative.

Lagrange Multiplier Statistics			
Parameter	Chi-Square	Pr > ChiSq	
Dispersion	1.5649	0.1055	*
* One-sided p-value			

Listing 2: Comparaison des modèles selon l'AIC

Modèle	AIC
Modèle de Poisson	189597.9482
Modèle de binôme négatif	189597.9482
Modèle de loi ZIP	189598.8097
Modèle de loi ZINB	189600.3599

2.5.2 Modélisation des coût individuel de sinistre (CM)

–Modèle de Log-Normal

La formule de calcul du coût moyen est donnée Par:

$$CM = \frac{\text{totalmontant}}{\text{nbr_sinistre}}$$

```
/* Calcul de coût moyen (CM) */
proc sql;
create table chrt.prod_sin as
select *, totalmontant/nbr_sinistre as CM
from chrt.prod_sin;
quit;
```

Après calcul de **Coût moyen**, on va faire la Mise à jour des valeurs manquantes de la variable CM à 0 dans la table prod.sin.

```
/* delete var1: ce variable n'a aucun utilité */
data chrt.prod_sin;
set chrt.prod_sin(drop=var1);
run;
/* Mise à jour des valeurs manquantes */
proc sql;
update chrt.prod_sin
set CM=0
where CM =. ;
quit;
```

Générant maintenant le modèle de **Log-normal**

```
data chrt.prod_sin_ln;
set chrt.prod_sin;
loga=log(CM);
run;
proc genmod data=chrt.prod_sin_ln;
class sexe ;
model loga=sexe age_du_conducteur age_mise_en_circulation /
dist=normal link=id;
title'Ajustement par lognormal';
ods output modelfit=logn;
output out=chrt.prod_sin p=pred_loga;
run;
proc sql;
alter table chrt.prod_sin add prime float(3);
update chrt.prod_sin set prime = exp(pred_loga)*pred_freq;
quit;
```

Ajutement par lognormal

CCC
The GENMOD Procedure

Model Information	
Data Set	CHRT.PROD_SIN_LN
Distribution	Normal
Link Function	Identity
Dependent Variable	loga

Number of Observations Read	100000
Number of Observations Used	42175
Missing Values	57825

Class Level Information		
Class	Levels	Values
Sexe	2	F M

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	42E3	100845.6240	2.3914
Scaled Deviance	42E3	42175.0000	1.0001
Pearson Chi-Square	42E3	100845.6240	2.3914
Scaled Pearson X2	42E3	42175.0000	1.0001
Log Likelihood		-78227.0400	
Full Log Likelihood		-78227.0400	
AIC (smaller is better)		156464.0801	
AICC (smaller is better)		156464.0815	
BIC (smaller is better)		156507.3280	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	4.3910	0.0346	4.3232	4.4589	16088.2	<.0001
Sexe	F	1	-0.0017	0.0154	-0.0318	0.0285	0.01	0.9139
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.	.
age_du_conducteur		1	0.0756	0.0008	0.0741	0.0772	9260.74	<.0001
age_mise_en_circulat		1	-0.0127	0.0014	-0.0155	-0.0100	81.67	<.0001
Scale		1	1.5463	0.0053	1.5359	1.5568		

Note: The scale parameter was estimated by maximum likelihood.

Ajutement par lognormal

Obs	numepolice	exercice	Sexe	Zone	Comubsution	Puissance_Fiscale	exposition	age_du_conducteur	age_mise_en_circulation	nbr_sinistre	totalmontant	offset
1	2	2009	M	Fès-Meknès	Essence	[00-08]	0.8505040312	28	7	1	1966.91	-0.16193
2	5	2016	M	Rabat-Salé-Kénitra	Essence	[00-08]	0.4869068514	27	10	0	0.00	-0.71968
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	28	6	2	3625.85	-0.73281
4	9	2012	M	Tanger-Tétouan-Hoceima	Essence	[00-08]	0.0609243414	47	19	0	0.00	-2.79812
5	24	2008	M	Rabat-Salé-Kénitra	Diesel	[08-10]	0.5667409257	27	4	1	174.72	-0.56785
6	36	2011	M	Fès-Meknès	Essence	[00-08]	0.1621531127	50	7	0	0.00	-1.81921



Interpretation

- Les résultats indiquent que le modèle ajusté est statistiquement significatif pour toutes les variables incluses, car la valeur p-value < 0,05 pour chaque variable. Cependant, la variable "Sexe" a des résultats peu significatifs, car la valeur p-value > 0.05 pour le test de Wald pour les deux niveaux de cette variable.
- Le test de déviance montre que le modèle ajusté explique une grande partie de la variation dans les données, car la déviance observée est beaucoup plus petite que la déviance attendue sous le modèle nul. Cela est confirmé par le test de chi-carré de Pearson, qui montre que le modèle ajusté est significativement meilleur que le modèle nul.

-Modèle de Gamma

Generant le Modele de **Gamma**

```
proc genmod data=chrt.prod_sin ;
class sexe comubsution;
model CM= sexe comubsution age_mise_en_circulation age_du_conducteur/
dist =gamma link=log offset = offset ;
title'Ajustement par loi de GAMMA';
ods output modelfit=Gam;
run;
```

Calculant la frequence(pred_freq) qui est donnée par la formule suivante:

$$pred_freq = \frac{nbr_sinistre}{exposition}$$

```
/* calcul frequence */
proc sql;
alter table chrt.prod_sin add pred_freq float(5);
update chrt.prod_sin set pred_freq = nbr_sinistr / exposition;
quit;
```

Modelisation du coût moyen (CM):

```
proc genmod data=chrt.prod_sin;
Class zone sexe comubsution puissance_fiscale;
Model CM = age_du_conducteur age_mise_en_circulation /
dist = gamma link = log offset=offset ;
ods output modelfit = Gamma;
output out=chrt.prod_sin p=pred_cm;
run;
```

Ainsi on calcul la **prime pure**

```
/* calcul de prime */
proc sql;
alter table chrt.prod_sin add prime float(3);
update chrt.prod_sin set prime = pred_cm*pred_freq;
quit;
/*affichage des données*/
proc print data=chrt.prod_sin (obs=10);
run;
```

The GENMOD Procedure

Model Information		
Data Set	CHRT.PROD_SIN	Predicted Values and Diagnostic Statistics
Distribution	Gamma	
Link Function	Log	
Dependent Variable	CM	
Offset Variable	offset	

Number of Observations Read	100000
Number of Observations Used	42175
Number of Invalid Responses	57825

Class Level Information		
Class	Levels	Values
Sexe	2	F M

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	42E3	80596.3050	1.9112
Scaled Deviance	42E3	51623.1863	1.2241
Pearson Chi-Square	42E3	139965.7493	3.3190
Scaled Pearson X2	42E3	89650.2382	2.1259
Log Likelihood		-381436.5493	
Full Log Likelihood		-381436.5493	
AIC (smaller is better)		762883.0987	
AICC (smaller is better)		762883.1001	
BIC (smaller is better)		762926.3466	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	5.8366	0.0287	5.7804	5.8927	41474.8	<.0001
Sexe	F	1	-0.0168	0.0124	-0.0412	0.0075	1.83	0.1756
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.	.
age_du_conducteur		1	0.0771	0.0007	0.0758	0.0784	13782.6	<.0001
age_mise_en_circulat		1	-0.0112	0.0012	-0.0135	-0.0090	94.67	<.0001
Scale		1	0.6405	0.0037	0.6333	0.6479		

Note: The scale parameter was estimated by maximum likelihood.

Obs	numepolice	exercice	Sexe	Zone	Comubstution	Puissance_Fiscale	exposition	age_du_conducteur	age_mise_en_circulation	nbr_sinistre	totalmontant	offset
1	2	2009	M	Fcs-Meknes	Essence	[00-08]	0.8505040312	28	7	1	1966.91	-0.16193
2	5	2016	M	Rabat-Sale-Kenitra	Essence	[00-08]	0.4869068514	27	10	0	0.00	-0.71968
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	28	6	2	3625.85	-0.73281
4	9	2012	M	Tanger-Tetouan-Hocedma	Essence	[00-08]	0.0609243414	47	19	0	0.00	-2.79812
5	24	2008	M	Rabat-Sale-Kenitra	Diesel	[08-10]	0.5667409257	27	4	1	174.72	-0.56785
6	36	2011	M	Fcs-Meknes	Essence	[00-08]	0.1621531127	50	7	0	0.00	-1.81921

Inter

- Le modèle ajusté est statistiquement significatif, car la p-value < 0,05 pour toutes les variables incluses sauf le sexe. Cela indique que toutes les variables ont un effet significatif sur la fréquence de sinistres.
- Le test de déviance suggère également que le modèle ajusté explique une grande partie de la variation dans les données, car la déviance observée est beaucoup plus petite que la déviance attendue sous le modèle nul. Cela est confirmé par le test de chi-carré de Pearson, qui montre que le modèle ajusté est significativement meilleur que le modèle nul.

-Comparaison:

Le modèle lognormal avait un **AIC de 156464.0801**, alors que le modèle gamma a un **AIC de 762883.0987**. Comparativement, le modèle lognormal a un AIC beaucoup plus petit que le modèle gamma, ce qui indique qu'il fournit un meilleur ajustement pour les données par rapport au modèle gamma. Cela peut impliquer que le modèle lognormal est préférable pour la prime par rapport au modèle gamma.

3 Partie 3: Provisionnement

3.1 Importation des des données

```
```{r}
library(readxl)
library(pander)
provisions=read_excel("provisions.xlsx")
```
```

Table 3: Les pertes incrémentées

| Année | Colonne1 | Colonne2 | Colonne3 | Colonne4 | Colonne5 | Colonne6 | Colonne7 | Colonne8 |
|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| 2011 | 1376384 | 1211168 | 535883 | 313790 | 168142 | 79972 | 39235 | 15030 |
| 2012 | 1576278 | 1437150 | 652445 | 342694 | 188799 | 76956 | 35042 | 17089 |
| 2013 | 1763277 | 1540231 | 678959 | 364199 | 177108 | 78169 | 47391 | 25288 |
| 2014 | 1779698 | 1498531 | 661401 | 321434 | 162578 | 84581 | 53449 | NA |
| 2015 | 1843224 | 1573604 | 613401 | 299473 | 176842 | 106296 | NA | NA |
| 2016 | 1962385 | 1520298 | 581932 | 347434 | 238375 | NA | NA | NA |
| 2017 | 2033371 | 1430541 | 633500 | 432257 | NA | NA | NA | NA |
| 2018 | 2072061 | 1458541 | 727098 | NA | NA | NA | NA | NA |
| 2019 | 2210754 | 1517501 | NA | NA | NA | NA | NA | NA |
| 2020 | 2206886 | NA | NA | NA | NA | NA | NA | NA |

3.2 Application de la méthode ChainLadder pour le calcul des réserves de sinistres.

La boucle "for" est utilisée pour calculer les pertes cumulatives pour chaque année. Les pertes cumulatives sont calculées en ajoutant la valeur de la colonne précédente à chaque élément de la colonne courante.

```
```{r}
#Cumulative Loss
prov=provisions[, -1]
for (i in 2:10) {
 prov[, i] = prov[, i] + prov[, i - 1]
}
```
```

le code suivant donne la méthode "ChainLadder" qui est appliquée pour calculer les facteurs de développement "LAMBDA". Ces facteurs sont utilisés pour projeter les pertes futures à partir des pertes passées. Les facteurs "LAMBDA" sont calculés en prenant la somme des pertes des années suivantes divisée par la somme des pertes des années courantes.

```
```{r}
ChainLadder
LAMBDA = rep(NA, 10 - 1)
for (k in 1:(10 - 1)) {
 LAMBDA[k] = (sum(prov[1:(10 - k), k + 1]) / sum(prov[1:(10 - k), k]))
}
```
```

Ensuite, le triangle de paiement est calculé en utilisant les facteurs de développement "LAMBDA". Le triangle de paiement représente les paiements attendus pour chaque année et pour chaque niveau de développement.

```
```{r}
TRIANGLE = prov
for (i in 1:(10 - 1)) {
 TRIANGLE[(10 - i + 1):10, i + 1] = LAMBDA[i] * TRIANGLE[(10 - i + 1):(10), i]
}
TRIANGLE
```
```

Ainsi, les réserves sont calculées en soustrayant les paiements attendus pour chaque année des charges ultimes prévues. Les réserves sont stockées dans la variable appelé "reserve"

```

```{r}
triangle = as.data.frame(TRIANGLE)
tt = as.matrix(TRIANGLE)
chargeultime = TRIANGLE[, 10]
paiements = diag(tt[,10:1])
reserve = chargeultime - paiements
```

```

Table 4: Triangle d'incrément de pertes, réserves et paiements

| Année | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|-----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 2011 | 1376384 | 2587552 | 3123435 | 3437225 | 3605367 | 3685339 | 3724574 | 3739604 | 3750469 |
| 2012 | 1576278 | 3013428 | 3665873 | 4008567 | 4197366 | 4274322 | 4309364 | 4326453 | 4338960 |
| 2013 | 1763277 | 3303508 | 3982467 | 4346666 | 4523774 | 4601943 | 4649334 | 4674622 | 4688167 |
| 2014 | 1779698 | 3278229 | 3939630 | 4261064 | 4423642 | 4508223 | 4561672 | 4582319 | 4595597 |
| 2015 | 1843224 | 3416828 | 4030229 | 4329702 | 4506544 | 4612840 | 4660162 | 4681255 | 4694820 |
| 2016 | 1962385 | 3482683 | 4064615 | 4412049 | 4650424 | 4743616 | 4792280 | 4813971 | 4827920 |
| 2017 | 2033371 | 3463912 | 4097412 | 4529669 | 4732784 | 4827626 | 4877152 | 4899227 | 4913423 |
| 2018 | 2072061 | 3530602 | 4257700 | 4640885 | 4848987 | 4946158 | 4996900 | 5019517 | 5034062 |
| 2019 | 2210754 | 3728255 | 4455215 | 4856177 | 5073932 | 5175611 | 5228707 | 5252374 | 5267593 |
| 2020 | 2206886 | 3958267 | 4730076 | 5155775 | 5386965 | 5494917 | 5551289 | 5576415 | 5592573 |
| Réserves | 0 | 1166703 | 1668564 | 2028476 | 2158461 | 2172188 | 2113942 | 1930266 | 1527971 |
| Paiements | 1376384 | 1431174 | 537283 | 261790 | 170693 | 89103 | 42432 | 14963 | 10865 |

3.3 Modèle de Mack

```

```{r}
library('chainladder')
M=MackChainladder(prov, est.sigma="Mack")
pander(summary(M)$By) rigin, style="grid")
pander(summary(M)$Totals, style="grid")
```

```

Table 5: Résultats du modèle de Mack pour chaque triangle de pertes

| Latest | Dev.To.Date | Ultimate | IBNR | Mack.S.E | CV(IBNR) |
|---------|-------------|----------|---------|----------|-----------|
| 3754537 | 1 | 3754537 | 0 | 0 | NA |
| 4338960 | 0.9989 | 4343666 | 4706 | 0.744 | 0.0001581 |
| 4674622 | 0.996 | 4693252 | 18630 | 56.13 | 0.003013 |
| 4561672 | 0.9915 | 4600581 | 38909 | 4376 | 0.1125 |
| 4612840 | 0.9815 | 4699912 | 87072 | 8927 | 0.1025 |
| 4650424 | 0.9622 | 4833157 | 182733 | 16259 | 0.08898 |
| 4529669 | 0.9209 | 4918753 | 389084 | 34287 | 0.08812 |
| 4257700 | 0.8449 | 5039522 | 781822 | 62113 | 0.07945 |
| 3728255 | 0.707 | 5273306 | 1545051 | 99106 | 0.06414 |
| 2206886 | 0.3942 | 5598639 | 3391753 | 285564 | 0.08419 |

- Les résultats du modèle indiquent une estimation de l'ultime, c'est-à-dire la perte totale attendue pour chaque triangle de pertes. Les résultats montrent également une estimation de l'IBNR (incurred but not reported), qui est la perte attendue pour les sinistres qui ont été signalés mais qui n'ont pas encore été réglés, ainsi que l'erreur-type du modèle.

Table 6: Résultats totaux du modèle de Mack

| | Totals |
|------------|----------|
| Latest: | 41315565 |
| Dev: | 0.8652 |
| Ultimate: | 47755325 |
| IBNR: | 6439760 |
| Mack S.E.: | 322503 |
| CV(IBNR): | 0.05008 |

- Ensuite le modèle montre que l'ultime total pour l'ensemble des triangles de pertes est de 47 755 325 dollars. L'IBNR total est estimé à 6 439 760 dollars, ce qui signifie qu'il y a des pertes qui ont été signalées mais qui n'ont pas encore été réglées. L'erreur-type du modèle est estimée à 322 503 dollars, ce qui indique l'incertitude autour des estimations.
- triangle montrent une augmentation des pertes ultimes et de l'IBNR à mesure que le triangle avance dans le temps. Cela est cohérent avec l'idée que les pertes peuvent prendre du temps à être signalées et réglées, en particulier pour les sinistres graves.

3.4 Modèle de GLM

```

```{r}
#modele de GLM
pp=as.data.frame(prov)
ligne=rep(1:10,10)
colone=rep(1:10,each=10)
Y=c(pp[,1],pp[,2],pp[,3],pp[,4],pp[,5],pp[,6],pp[,7],pp[,8],pp[,9],pp[,10])

lig=as.factor(ligne)
col=as.factor(colone)
cbind(Y,lig,col)

log normal
reg=lm(log(Y)~lig+col)
summary(reg)

z=predict(reg, newdata=data.frame(lig,col),type="response")
sigma=summary(reg)$sigma
pred=exp(z+sigma^2/2)

#Gamma
regamma=glm(Y~lig+col,family=Gamma(link="log"))
summary(regamma)
x <- predict(regamma, newdata = data.frame(lig, col), type = "response")
x
sum(x[is.na(Y)==TRUE])
#
```

```

- Le modèle linéaire est utilisé avec des données logarithmiques et montre que chaque variable "lig" (indice de ligne) et "col" (indice de colonne) a une influence significative sur les valeurs de "Y" (les données de la matrice "prov"). Les coefficients indiquent l'effet de chaque niveau de "lig" et de "col" sur les valeurs de "Y". Par exemple, le coefficient pour "lig2" est de 0,14884, ce qui signifie que lorsque "lig" est égal à 2, les valeurs de "Y" ont tendance à être multipliées par $\exp(0,14884) = 1,16$ fois par rapport à "lig1". De même, le coefficient pour "col2" est de 0,58856, ce qui signifie que lorsque "col" est égal à 2, les valeurs de "Y" ont tendance à être multipliées par $\exp(0,58856) = 1,80$ fois par rapport à "col1".
- Le modèle gamma est utilisé avec les données d'origine et montre également que chaque variable "lig" et "col" a une influence significative sur les valeurs de "Y". Les coefficients ont la même interprétation que dans le modèle

linéaire, mais ils sont maintenant interprétés comme des multiplicateurs de la moyenne, car la famille de distribution Gamma a une moyenne qui dépend des paramètres de forme et d'échelle. Par exemple, le coefficient pour "lig2" est de 0,14893, ce qui signifie que lorsque "lig" est égal à 2, les valeurs de "Y" ont tendance à être multipliées par $\exp(0,14893) = 1,16$ fois la moyenne de "Y" pour ce niveau de "lig". Le modèle gamma a également une valeur de déviance résiduelle plus faible que le modèle linéaire, ce qui indique qu'il ajuste mieux les données.

Resultats

Residuals:

Min 1Q Median 3Q Max

-0.058680 -0.009634 0.000775 0.011086 0.046424

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 14.18042 0.01108 1280.31 < 2e-16 ***

lig2 0.14884 0.01081 13.77 6.48e-16 ***

lig3 0.23277 0.01131 20.59 < 2e-16 ***

lig4 0.22052 0.01185 18.61 < 2e-16 ***

lig5 0.24828 0.01249 19.88 < 2e-16 ***

lig6 0.27898 0.01331 20.97 < 2e-16 ***

lig7 0.29836 0.01441 20.70 < 2e-16 ***

lig8 0.32802 0.01606 20.43 < 2e-16 ***

lig9 0.39545 0.01888 20.95 < 2e-16 ***

lig10 0.42667 0.02547 16.75 < 2e-16 ***

col2 0.58856 0.01081 54.44 < 2e-16 ***

col3 0.77140 0.01131 68.23 < 2e-16 ***

col4 0.86006 0.01185 72.60 < 2e-16 ***

col5 0.90614 0.01249 72.55 < 2e-16 ***

col6 0.92865 0.01331 69.79 < 2e-16 ***

col7 0.94210 0.01441 65.37 < 2e-16 ***

col8 0.94985 0.01606 59.16 < 2e-16 ***

col9 0.95543 0.01888 50.61 < 2e-16 ***

col10 0.95806 0.02547 37.62 < 2e-16 ***

Signif. codes: 0 ' ' 0.001 ' ' 0.01 ' ' 0.05 ' ' 0.1 ' ' 1

Residual standard error: 0.02293 on 36 degrees of freedom

(45 observations deleted due to missingness)

Multiple R-squared: 0.9967, Adjusted R-squared: 0.995

F-statistic: 595.4 on 18 and 36 DF, p-value: < 2.2e-16

Call:

glm(formula = Y ~ lig + col, family = Gamma(link = "log"))

Deviance Residuals:

Min 1Q Median 3Q Max

-0.058856 -0.009711 0.000520 0.011007 0.045999

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 14.18109 0.01103 1285.78 < 2e-16 ***

lig2 0.14893 0.01077 13.83 5.59e-16 ***

lig3 0.23269 0.01126 20.67 < 2e-16 ***

lig4 0.22038 0.01180 18.68 < 2e-16 ***

lig5 0.24817 0.01244 19.95 < 2e-16 ***

lig6 0.27893 0.01325 21.05 < 2e-16 ***

lig7 0.29847 0.01435 20.80 < 2e-16 ***

lig8 0.32802 0.01599 20.52 < 2e-16 ***

lig9 0.39557 0.01880 21.04 < 2e-16 ***

lig10 0.42601 0.02536 16.80 < 2e-16 ***

col2 0.58805 0.01077 54.62 < 2e-16 ***

col3 0.77086 0.01126 68.47 < 2e-16 ***

```
col4 0.85949 0.01180 72.86 < 2e-16 ***
col5 0.90558 0.01244 72.81 < 2e-16 ***
col6 0.92808 0.01325 70.05 < 2e-16 ***
col7 0.94149 0.01435 65.60 < 2e-16 ***
col8 0.94918 0.01599 59.37 < 2e-16 ***
col9 0.95472 0.01880 50.78 < 2e-16 ***
col10 0.95739 0.02536 37.75 < 2e-16 ***
Signif. codes: 0 ' ' 0.001 ' ' 0.01 ' ' 0.05 ' ' 0.1 ' ' 1
```

(Dispersion parameter for Gamma family taken to be 0.0005215047)

```
apache
Copy
Null deviance: 4.89852 on 54 degrees of freedom
Residual deviance: 0.01888 on 36 degrees of freedom
(45 observations deleted due to missingness)
AIC: 1413.4
```

Number of Fisher Scoring iterations: 3

```
tap
Copy
  1      2      3      4      5
1441344 1672811 1818964 1796709 1847341
```