



ROYAUME DU MAROC  
HAUT COMMISSARIAT AU PLAN  
INSTITUT NATIONAL DE STATISTIQUE  
ET D'ÉCONOMIE APPLIQUÉE

INSEA



*Sujet:*

# PROJET DE TARIFICATION EN SAS

Réalisé par : EL YOUSEFI Ahmed  
Encadré par : MARRI Fouad



# Table des matières

<b>I</b>	<b>TARIFICATION</b>	<b>5</b>
<b>1</b>	<b>CHAPITRE INTRODUCTIF</b>	<b>7</b>
1	Le marché de l'Assurance au Maroc . . . . .	7
2	Le marché de l'Assurance Automobile au Maroc . . . . .	7
2.1	Chiffres d'affaires . . . . .	7
2.2	Types de contrats . . . . .	8
2.3	Coefficient de réduction majoration . . . . .	9
3	Notions Clefs de marché automobile . . . . .	9
3.1	L'antisélection : . . . . .	9
3.2	Assurance multirisque habitation : . . . . .	9
3.3	L'assurance RC Décennale : . . . . .	10
3.4	L'assurance garantie décennale . . . . .	10
<b>2</b>	<b>IMPRTATION NETTOYAGE ET MANIPULATION DES DONNÉES</b>	<b>11</b>
1	Prétraitement des bases . . . . .	11
2	Importation de base : sinistres . . . . .	12
2.1	Importation . . . . .	12
2.2	Nettoyage . . . . .	13
2.3	Manipulation . . . . .	15
3	Importation de base : production . . . . .	18
3.1	Importation . . . . .	18
3.2	Manipulation . . . . .	19
3.3	Nettoyage . . . . .	20
4	Création de table récupilative . . . . .	22
4.1	Jointudre des tables . . . . .	22
4.2	Ajout de frequence et de severité . . . . .	22
<b>3</b>	<b>STATISTIQUES DESCRIPTIVES</b>	<b>25</b>
1	Statistiques univariées . . . . .	25
1.1	Pour chaque variables . . . . .	25
1.2	Étude de l'effet de chaque variable sur le nombre de sinistre et la severtité . . . . .	27

<b>4</b>	<b>MODELISATION ET CALCUL DES PRIMES</b>	<b>31</b>
1	Modélisation de nombre de sinistres . . . . .	31
1.1	Modèles sans "zéro inflation" . . . . .	31
1.2	Modèles avec zero inflation . . . . .	35
1.3	Comparaison des modèles : . . . . .	36
2	Modélisation de sévérité . . . . .	37
2.1	Modèle Gamma . . . . .	37
2.2	Modèle Log-normal . . . . .	40
2.3	Comparaison des modèles . . . . .	40
3	Calcul de la prime . . . . .	41
3.1	Calcul de la fréquence . . . . .	41
3.2	Calcul de la sévérité . . . . .	43
3.3	Calcul de la prime . . . . .	43
<b>5</b>	<b>ANNEXE</b>	<b>47</b>
<b>II</b>	<b>PROVISIONNEMENT</b>	<b>53</b>
<b>1</b>	<b>IMPORTATION DES DONNÉES ET DEFINITIONS DES MODÈLES</b>	<b>55</b>
1	Importation, nettoyage et manipulation . . . . .	55
2	Fonctions de modélisation . . . . .	56
2.1	ChaineLadder . . . . .	56
2.2	Mack . . . . .	57
2.3	GLM . . . . .	58
2.4	Regression Log-normal . . . . .	58
3	Autres fonctions de manipulation . . . . .	59
3.1	Fonction de calcul des pertes cumulées . . . . .	59
3.2	Fonction de calcul des réserves . . . . .	59
<b>2</b>	<b>MODÉLISATION ET CALCUL DES RÉSERVES</b>	<b>61</b>
1	Modélisation . . . . .	61
1.1	ChainLadder . . . . .	61
1.2	Mack . . . . .	62
1.3	GLM . . . . .	63
1.4	regression log-normal . . . . .	64
2	Calcul des réserves . . . . .	64
3	Comparaison . . . . .	65

Première partie

**TARIFICATION**



# Chapitre 1

## CHAPITRE INTRODUCTIF

### 1 Le marché de l'Assurance au Maroc

Le marché des assurances au Maroc est un secteur clé de l'économie marocaine qui a connu une croissance remarquable depuis l'an 2000. Il est constitué de divers intervenants, tels que les assureurs, les sociétés d'assurances, les mutuelles, les institutions de prévoyance, l'autorité de contrôle, les compagnies de réassurances, et d'intermédiaires réglementés tels que les courtiers, les agents d'assurance, les mandataires d'assurance et les mandataires d'intermédiaire d'assurance.

En 2022, les primes globales enregistrées par les compagnies d'assurances et de réassurance ont atteint 54,89 milliards de DH, en hausse de 9.3% sur un an. Cette croissance est principalement due à l'assurance vie, dont les primes ont augmenté de 11.5% à 25,56 milliards de DH, devant l'activité non-vie qui a progressé de 7.5% à 29,32 milliards. Les primes de l'assurance décès ont enregistré une petite hausse de 0.6% à 3,23 milliards de DH.

En termes de réglementation, le secteur d'assurance a connu plusieurs changements grâce aux lois qui lui sont imposées, telles que la loi n° 43.05 pour lutter contre le blanchiment des capitaux, la loi n°18.12 portant sur l'indemnisation des accidents du travail, la loi n°110.14 instituant un régime de couverture des conséquences d'événements catastrophiques, et la loi n°31.08 se concentrant sur la protection des assurés.

Malgré l'amélioration du marché des assurances, le taux de pénétration de l'assurance au Maroc reste faible, représentant à peine 4% du PIB selon l'ACAPS, ce qui positionne le Maroc en troisième position en Afrique, derrière l'Afrique du Sud et la Namibie.

### 2 Le marché de l'Assurance Automobile au Maroc

#### 2.1 Chiffres d'affaires

Le marché marocain de l'assurance automobile est un secteur clé du marché de l'assurance, avec des entreprises telles que Saham, Wafa Assurance et RMA générant des

chiffres d'affaires significatifs. Les données sont présentées dans le tableau ci-dessous :

Compagnies	Chiffre d'affaires 2022 (1)	Chiffre d'affaires 2021	Evolution 2021-2022	Parts 2022
Wafa Assurance	10 425 200	9 088 900	14.7%	19.12%
RMA	8 076 100	7 680 700	5.15%	14.82%
Mutuelle Taamine Chaabi	7 345 100	6 308 400	16.43%	13.47%
Axa Assurance Maroc	6 027 800	5 567 400	8.27%	11.06%
Sanlam Assurance	5 954 500	5 621 100	5.93%	10.92%
AtlantaSanad	5 403 600	5 400 800	0.05%	9.91%
MCMA	2 270 100	2 067 000	9.83%	4.16%
Marocaine Vie	2 268 700	2 339 100	-3.01%	4.16%
Allianz Assurance Maroc	1 486 100	1 426 800	4.16%	2.73%
MAMDA	1 278 000	1 172 100	9.04%	2.34%
CAT	1 263 700	1 177 200	7.35%	2.32%
MATU	862 700	714 100	20.81%	1.58%
Total général	54 510 500	50 206 900	8.57%	100%

L'assurance automobile représente une part significative de l'industrie de l'assurance, avec une part de marché de 69% pour les 5 principaux assureurs. Bien que le niveau de concentration soit moyen avec 15 compagnies d'assurance, la branche non-vie représente 53% du chiffre d'affaires global du secteur. En ce qui concerne l'assurance automobile, il s'agit principalement de l'assurance responsabilité civile obligatoire, qui couvre la responsabilité civile du souscripteur, du propriétaire et de toute personne autorisée à conduire le véhicule. Bien que cette assurance couvre les dommages matériels et corporels causés aux victimes lorsqu'un accident est de la responsabilité du souscripteur, elle ne couvre pas les blessures personnelles ou les dégâts matériels causés au véhicule. En conséquence, d'autres garanties facultatives sont souvent souscrites en plus de la RC automobile, telles que les garanties contre les dommages de collision, l'incendie, le vol, les bris de glace, les tiers et la défense. Il convient également de noter que la RC automobile est libérale en termes de réglementation des tarifs, car les compagnies d'assurance ont plus de liberté pour fixer les primes d'assurance responsabilité civile automobile sans être soumises à une réglementation gouvernementale stricte. (Source : ACAPS).

## 2.2 Types de contrats

Voici quelques types de contrats d'assurance automobile courants :

- **Assurance tous risques** : Ce type de contrat offre une couverture complète pour les dommages causés à votre véhicule, ainsi que pour les dommages causés à des tiers. Il peut également inclure des garanties supplémentaires telles que l'assistance routière, la protection juridique, etc.
- **Assurance responsabilité civile** : Ce type de contrat est obligatoire dans la plupart des pays. Il couvre les dommages que vous pourriez causer à des tiers, y compris les blessures corporelles et les dommages matériels.
- **Assurance tiers plus** : Ce type de contrat est une forme intermédiaire d'assurance automobile. Il offre une couverture pour les dommages causés à des tiers, ainsi qu'une couverture pour les dommages causés à votre propre véhicule dans certaines circonstances.



- **Assurance incendie, vol et bris de glace** : Ce type de contrat offre une couverture pour les dommages causés à votre véhicule en cas d'incendie, de vol ou de bris de glace.

Il existe également d'autres types de contrats d'assurance automobile qui peuvent offrir des garanties spécifiques en fonction de vos besoins, tels que l'assurance tous risques avec franchise, l'assurance kilométrage illimité, etc.

## 2.3 Coefficient de réduction majoration

Le coefficient de réduction majoration (CRM) est un indicateur utilisé en assurance automobile pour déterminer le niveau de risque d'un conducteur. Il est basé sur l'historique de conduite du conducteur et est calculé en fonction du nombre d'années sans accident responsable (coefficients de réduction) ou avec accident responsable (coefficients de majoration) qu'il a accumulés. Le CRM est utilisé pour calculer la prime d'assurance automobile, qui est ajustée en fonction du niveau de risque du conducteur.

Le coefficient de réduction est un pourcentage qui diminue la prime d'assurance automobile du conducteur. Plus le nombre d'années sans accident responsable est élevé, plus le coefficient de réduction est élevé, et plus la prime d'assurance est réduite. À l'inverse, le coefficient de majoration est un pourcentage qui augmente la prime d'assurance automobile du conducteur. Plus le nombre d'années avec accident responsable est élevé, plus le coefficient de majoration est élevé, et plus la prime d'assurance est majorée.

Le CRM est donc un système de bonus-malus qui encourage les conducteurs à conduire prudemment et à éviter les accidents. Les conducteurs qui ont un historique de conduite sans accident responsable bénéficieront d'une prime d'assurance automobile moins élevée que ceux qui ont eu des accidents responsables. Les conducteurs qui ont un historique de conduite avec des accidents responsables devront payer une prime d'assurance automobile plus élevée pour couvrir le risque plus élevé associé à leur historique de conduite.

# 3 Notions Clefs de marché automobile

## 3.1 L'antisélection :

L'antisélection, également connue sous le nom de sélection adverse, se produit lorsque les personnes les plus à risque sont plus susceptibles de souscrire une assurance que les personnes moins à risque. En conséquence, les primes d'assurance sont ajustées à la hausse pour tenir compte de ce risque accru, ce qui peut entraîner une spirale de l'antisélection où seules les personnes les plus à risque continuent à souscrire une assurance.

## 3.2 Assurance multirisque habitation :

L'assurance multirisque habitation est une assurance qui couvre les dommages causés à votre maison et à vos biens personnels. Elle couvre les dommages causés par les incendies, les inondations, les cambriolages, les actes de vandalisme et d'autres événements

indésirables. Cette assurance est généralement souscrite par les propriétaires de maisons, mais elle peut également être souscrite par les locataires.

### 3.3 L'assurance RC Décennale :

L'assurance responsabilité civile décennale (RC décennale) est une assurance obligatoire pour les constructeurs et les entrepreneurs qui effectuent des travaux de construction en France. Elle couvre les dommages causés à l'ouvrage construit pendant une période de dix ans à compter de la réception des travaux. Cette assurance est destinée à protéger les propriétaires contre les dommages causés par des défauts ou des vices de construction qui pourraient rendre l'ouvrage inhabitable ou dangereux.

### 3.4 L'assurance garantie décennale

La garantie décennale couvre les dommages qui affectent la solidité de l'ouvrage ou le rendent inhabitable ou impropre à sa destination pendant une période de 10 ans à compter de la réception des travaux. Elle s'applique aux travaux de construction, de rénovation, d'extension, de transformation ou de réhabilitation qui concernent des bâtiments existants ou neufs, ainsi que les travaux d'aménagement extérieur, comme les voiries, les réseaux et les équipements publics.

Les dommages couverts par la garantie décennale sont notamment les suivants :

- Les fissures, les affaissements et les déformations des murs, des planchers et des plafonds
- Les défauts d'étanchéité, les infiltrations d'eau et les fuites
- Les problèmes liés à l'isolation thermique ou phonique
- Les défauts de conception, de fabrication ou de mise en œuvre des éléments de construction, tels que les fondations, les charpentes, les couvertures, les revêtements de sol, les menuiseries extérieures, etc.
- Les désordres affectant les installations techniques, telles que les canalisations, les réseaux électriques, les installations de chauffage, de climatisation, de ventilation, de plomberie, etc.

Il est important de noter que la garantie décennale ne couvre pas les dommages qui résultent d'une mauvaise utilisation de l'ouvrage ou d'un défaut d'entretien de la part du propriétaire. Elle ne s'applique pas non plus aux travaux de décoration, d'aménagement intérieur, de nettoyage ou d'entretien courant.

# Chapitre 2

## IMPORTATION NETTOYAGE ET MANIPULATION DES DONNÉES

### 1 Prétraitement des bases

On doit traiter en avance la base de données avant de l'importer en sas, en fait on procede comme suite :

1. On supprime tous les guillemets :

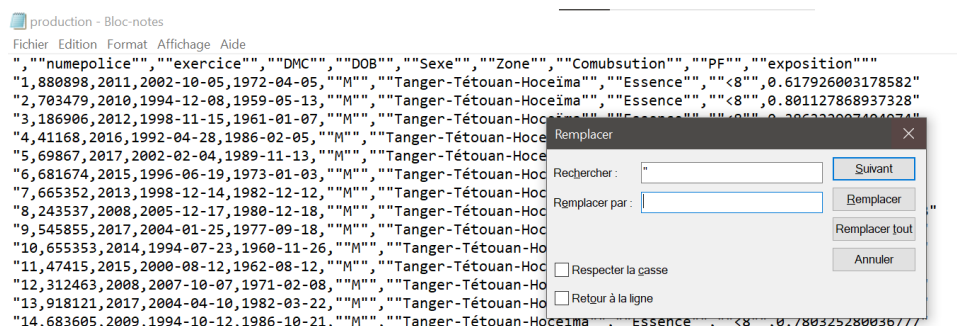


FIGURE 1.1 – Suppression des guillemets

2. On remplace les PF par des modalités plus simple :

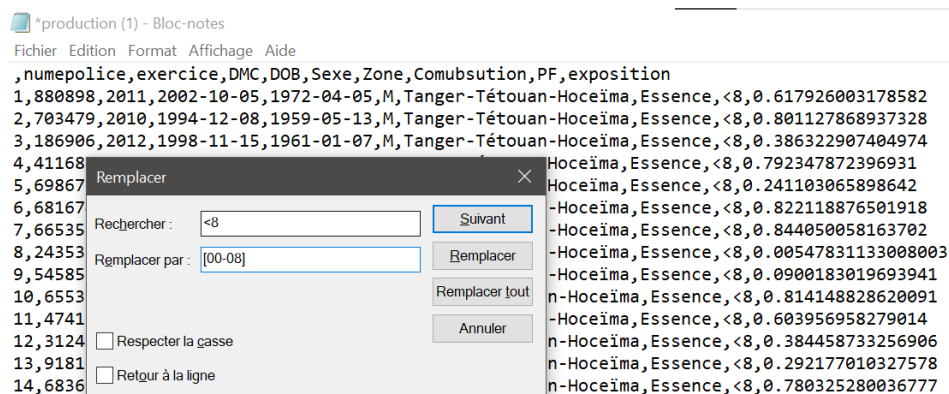


FIGURE 1.2 – Remplacement des PF

## 2 Importation de base : sinistres

### 2.1 Importation

On importe à travers le code :

```

1 FILENAME REFFILE '/home/u63328955/projet 2023/sinistre.csv'
  ↳ encoding="LATIN2";
2
3 PROC IMPORT DATAFILE=REFFILE
4     DBMS=CSV
5     OUT=tabs.sinistre replace;
6     GETNAMES=YES;
7 RUN;
```

Et on obtient :

Obs.	VAR1	numepolice	annee_reference	n_sinistre	montantsinistre
1	2	703479	2010	20122123	2269.038909
2	3	186906	2012	20122124	318.56930342
3	7	665352	2013	20122125	191.06868647
4	10	655353	2014	20122126	21898.294476
5	10.1	655353	2014	20122127	204.10113316
6	11	47415	2015	20122128	211.65011445
7	13	918121	2017	20122129	336.98594546
8	13.1	918121	2017	20122130	1136.0258708
9	16	670251	2015	20122131	38.024265433
10	19	809064	2011	20122132	1508.2637996

FIGURE 2.1 – Output du code

## 2.2 Nettoyage

On vérifie d'abord les valeurs manquantes :

```
1 proc means data=Tabs.sinistre NMISS;
2 run;
```

### La procédure MEANS

Variable	Nbre manquant
VAR1	0
numepolice	0
annee_reference	0
n_sinistre	0
montantsinistre	0

FIGURE 2.2 – Output du code

On dispose des finalement colonnes suivantes :

- **Numepolice** : c'est le numéro de police
- **Année de reference** : c'est l'année d'exercice
- **n\_sinistre** : numéro de sinistre qui doit être unique
- **Montant sinistre** : le montant de sinistre doit être positif

On éliminera tout les observations qui ne vérifie pas ces contraintes

#### 1. Vérification puis Élimination des doublons :

```
1 proc sql;
2     create table tab_test as
3     select n_sinistre, count(*) as repetition from
4         ↪ tabs.sinistre
5     group by n_sinistre
6     order by repetition desc;
7 quit;
```

Ainsi on ne trouvera aucune doublon pour N\_SINISTRE :

#### 2. Vérification puis élimination des montants négatifs :

Obs.	n_sinistre	repetition
1	20171274	1
2	20163082	1
3	20146698	1
4	20150794	1
5	20158986	1
6	20142602	1
7	20167178	1
8	20134410	1
9	20148746	1
10	20136458	1

FIGURE 2.3 – Output du code

```
1 proc sql;  
2     create table tab_test as  
3     select n_sinistre, montantsinistre from tabs.sinistre  
4     order by montantsinistre asc ;  
5 quit;
```

On trouver alors que toute les montants sont positifs :

Obs.	n_sinistre	montantsinistre
1	20168311	0.0042901856
2	20136766	0.0366163168
3	20155653	0.0496743261
4	20174641	0.0814288004
5	20122315	0.0940014301
6	20133633	0.108451845
7	20153203	0.1344633134
8	20180239	0.1782566053
9	20143235	0.1938000318
10	20169443	0.2293223484

FIGURE 2.4 – Output du code

## 2.3 Manipulation

On créera un nouveau table SINISTRES qui va avoir :

- **Le numepolice** : numéro de la police
- **L'exercice** : qui est l'année de reference
- **Nombre** : qui est le nombre des sinistre réalisées par numepolice dans un exercice donnée
- **Somme** : qui est la somme des montants des sinistres réalisées par numepolice dans un exercice donnée

Cela se fait par le code :

```

1 proc sql;
2     create table tabs.sinistresT as
3     select numepolice ,annee_reference as exercice, count(*) as
      ↪ nombre , sum(montantsinistre) as somme

```

```

4      from tabs.sinistre
5      group by annee_reference, numepolice;
6 quit;

```

Obs.	numepolice	exercice	nombre	somme
1	24	2008	1	174.72
2	370	2008	1	83.82
3	612	2008	2	9594.89
4	758	2008	1	19.91
5	974	2008	2	1232.61
6	989	2008	1	335.50
7	1259	2008	1	1147.13
8	1471	2008	1	10.45
9	1691	2008	3	5157.97
10	1836	2008	1	590.70

FIGURE 2.5 – Output du code

Cependant, on doit pas travailler sur la totalité de la base de donnée puisque cela fait intervenir même les grandes sinistres, donc on va créer une autre base dont on conditionne juste les montants  $< Quantile(99\%)$ .

```

1 proc univariate data=tabs.sinistre;
2 var montantsinistre;
3 output out=quantiles pctlpre=p_
4      pctlpts=5 25 50 75 95 97 99 pctlname=percentile;
5 run;

```

Ainsi, on obtient :



Quantiles (Définition 5)	
Niveau	Quantile
<b>100Max 100%</b>	5.50719E+05
<b>99%</b>	7.51402E+04
<b>95%</b>	2.92383E+04
<b>90%</b>	1.66434E+04
<b>75% Q3</b>	6.05297E+03
<b>50% Médiane</b>	1.88127E+03
<b>25% Q1</b>	5.53929E+02
<b>10%</b>	1.78686E+02
<b>5%</b>	8.38157E+01
<b>1%</b>	1.53991E+01
<b>0% Min</b>	4.29019E-03

FIGURE 2.6 – Les quantile de variable montant sinistre

Puis, on recalcule la tables des sinistres agrégées :

```

1 proc sql;
2     create table tabs.sinistres as
3     select numepolice ,annee_reference as exercice, count(*) as
      ↪ nombre , sum(montantsinistre) as somme
4     from tabs.sinistre
5     where montantsinistre< 80000
6     group by annee_reference, numepolice;
7 quit;
```

Obs.	numepolice	exercice	nombre	somme
1	24	2008	1	174.72
2	370	2008	1	83.82
3	612	2008	2	9594.89
4	758	2008	1	19.91
5	974	2008	2	1232.61
6	989	2008	1	335.50
7	1259	2008	1	1147.13
8	1471	2008	1	10.45
9	1691	2008	3	5157.97
10	1836	2008	1	590.70

FIGURE 2.7 – Montant sinistre standard agrégées

## 3 Importation de base : production

### 3.1 Importation

On importe à travers le code :

```

1 FILENAME REFFILE '/home/u63328955/projet 2023/production.csv'
  ↳ encoding="LATIN2";
2
3 PROC IMPORT DATAFILE=REFFILE
4     DBMS=CSV
5     OUT=tabs.production replace;
6     GETNAMES=YES;
7 RUN;
```

Et on obtient :

On dispose des finalement colonnes suivantes :

- **Numepolice** : c'est le numéro de police doit être unique pour chaque exercice
- **Exercice** : c'est l'année d'exercice
- **Exposition** : C'est l'exposition
- **Zone** : zone de circulation
- **DOB** : année de naissance
- **DMC** : date de mise en circulation

Obs.	VAR1	numepolice	exercice	DMC	DOB	Sexe	Zone	Comubsution	PF	exposition
1	1	880898	2011	2002-10-05	1972-04-05	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.6179260032
2	2	703479	2010	1994-12-08	1959-05-13	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.8011278689
3	3	186906	2012	1998-11-15	1961-01-07	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.3863229074
4	4	41168	2016	1992-04-28	1986-02-05	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.7923478724
5	5	69867	2017	2002-02-04	1989-11-13	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.2411030659
6	6	681674	2015	1996-06-19	1973-01-03	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.8221188765
7	7	665352	2013	1998-12-14	1982-12-12	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.8440500582
8	8	243537	2008	2005-12-17	1980-12-18	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.0054783113
9	9	545855	2017	2004-01-25	1977-09-18	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.090018302
10	10	655353	2014	1994-07-23	1960-11-26	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.8141488286

FIGURE 3.1 – Output du code

- **PF** : Puissance fiscale
- **Comubsution** : type combustible
- **Sexe** : sexe de conducteur

## 3.2 Manipulation

On va ajouter des nouvelle variables

- **ageV** : age de véhicule à l'année de l'exercice, qui doit être positif
- **ageC** : age de conducteur à l'année de l'exercice, qui doit être positif

Cela se fait par le code :

```

1 proc sql;
2     alter table tabs.production add ageV int;
3     alter table tabs.production add ageC int;
4
5     update tabs.production
6     set ageV = exercice - year(DMC), ageC = exercice - year(DOB);
7
8     alter table tabs.production drop DOB;
9     alter table tabs.production drop DMC;
10    alter table tabs.production drop var1;
11 quit;
```

Obs.	numepolice	exercice	Sexe	Zone	Comubsution	PF	exposition	ageV	ageC
1	880898	2011	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.6179260032	9	39
2	703479	2010	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.8011278689	16	51
3	186906	2012	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.3863229074	14	51
4	41168	2016	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.7923478724	24	30
5	69867	2017	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.2411030659	15	28
6	681674	2015	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.8221188765	19	42
7	665352	2013	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.8440500582	15	31
8	243537	2008	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.0054783113	3	28
9	545855	2017	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.090018302	13	40
10	655353	2014	M	Tanger-Tétouan-Hoced'ma	Essence	[00-08]	0.8141488286	20	54

FIGURE 3.2 – Output du code

### 3.3 Nettoyage

On vérifie d'abord les valeurs manquantes :

```
1 proc means data=tabs.production NMISS;
2 run;
```

#### La procédure MEANS

Variable	Nbre manquant
numepolice	0
exercice	0
exposition	0
ageV	0
ageC	0

FIGURE 3.3 – Output du code

Puis on vérifie les doublons :

```
1 proc sql;
2     create table tab_test as
3     select numepolice,exercice, count(*) as repetition from
4         ↪ tabs.production
5     group by numepolice,exercice
6     order by repetition desc;
quit;
```

Obs.	numepolice	exercice	repetition
1	792073	2010	1
2	956110	2011	1
3	667580	2012	1
4	833121	2015	1
5	915153	2017	1
6	750819	2009	1
7	996695	2012	1
8	606647	2012	1
9	812202	2017	1
10	688970	2011	1

FIGURE 3.4 – Output du code

Dans la table finale on doit vérifier les contraintes déjà vu :

```
1 proc sql;  
2     select count(*) as nombre_des_ages_neg  
3     from tabs.production  
4     where ageV<18 or ageC<0 ;  
5 quit;
```

nombre_des_ages_inferieur_à_18
0

FIGURE 3.5 – Output du code

## 4 Création de table récupilative

### 4.1 Jointudre des tables

On va joindre les deux tables SINISTRES et PRODUCTION, Cela se fait par le code suivante :

```

1 proc sql;
2 create table tabs.data as
3 SELECT p.* , s.nombre , s.somme
4 FROM tabs.production AS p
5 LEFT JOIN tabs.sinistres AS s
6 ON p.numepolice = s.numepolice and p.exercice = s.exercice;
7 alter table tabs.data drop var1;
8 update tabs.data
9     set somme=0,nombre=0
10    where somme =.;
11 quit;

```

Obs.	numepolice	exercice	Sexe	Zone	Comubsution	PF	exposition	ageV	ageC	nombre	somme
1	2	2009	M	Fcs-Mekncs	Essence	[00-08]	0.8505040312	7	28	1	1966.91
2	5	2016	M	Rabat-Salé-Kénitra	Essence	[00-08]	0.4869068514	10	27	0	0.00
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	6	28	2	3625.85
4	9	2012	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.0609243414	19	47	0	0.00
5	24	2008	M	Rabat-Salé-Kénitra	Diesel	[08-10]	0.5667409257	4	27	1	174.72
6	36	2011	M	Fcs-Mekncs	Essence	[00-08]	0.1621531127	7	50	0	0.00
7	65	2014	M	Casablanca-Settat	Diesel	[08-10]	0.920984172	16	25	1	50.47
8	68	2016	M	Oriental	Essence	[00-08]	0.6330778915	13	34	1	504.15
9	69	2016	F	Souss-Massa	Diesel	[10-14]	0.3843837266	9	32	1	371.25
10	74	2015	F	Marrakech-Safi	Diesel	[10-14]	0.6421667349	17	26	2	1441.47

FIGURE 4.1 – Output du code

### 4.2 Ajout de frequence et de severité

On ajoute la frequence et la severité à travers le code :

```

1 proc sql;
2
3 alter table tabs.data add frequence float(4);
4 alter table tabs.data add severity float(4);
5
6 update tabs.data
7     set frequence = nombre/exposition,severity = somme/nombre;
8

```

```

9 update tabs.data
10     set severity = 0
11     where severity = .;
12
13 quit;

```

Obs.	numepolice	exercice	Sexe	Zone	Comubstation	PF	exposition	ageV	ageC	nombre	somme	frequence	severity
1	2	2009	M	Fcs-Mekncs	Essence	[00-08]	0.8505040312	7	28	1	1966.91	1.17577	1966.91
2	5	2016	M	Rabat-Salé-Kénitra	Essence	[00-08]	0.4869068514	10	27	0	0.00	0.00000	0.00
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	6	28	2	3625.85	4.16185	1812.93
4	9	2012	M	Tanger-Tétouan-Hocedma	Essence	[00-08]	0.0609243414	19	47	0	0.00	0.00000	0.00
5	24	2008	M	Rabat-Salé-Kénitra	Diesel	[08-10]	0.5667409257	4	27	1	174.72	1.76447	174.72
6	36	2011	M	Fcs-Mekncs	Essence	[00-08]	0.1621531127	7	50	0	0.00	0.00000	0.00
7	65	2014	M	Casablanca-Settat	Diesel	[08-10]	0.920984172	16	25	1	50.47	1.08579	50.47
8	68	2016	M	Oriental	Essence	[00-08]	0.6330778915	13	34	1	504.15	1.57958	504.15
9	69	2016	F	Souss-Massa	Diesel	[10-14]	0.3843837266	9	32	1	371.25	2.60157	371.25
10	74	2015	F	Marrakech-Safi	Diesel	[10-14]	0.6421667349	17	26	2	1441.47	3.11446	720.74
11	75	2009	F	Laâyoune-Sakia El Hamra	Diesel	[14-oo]	0.4530977344	3	35	0	0.00	0.00000	0.00
12	87	2012	F	Casablanca-Settat	Diesel	[08-10]	0.1810813574	20	36	0	0.00	0.00000	0.00
13	90	2011	M	Fcs-Mekncs	Essence	[00-08]	0.1458816098	4	28	0	0.00	0.00000	0.00
14	95	2017	M	Rabat-Salé-Kénitra	Essence	[00-08]	0.3933973815	17	40	1	1228.62	2.54196	1228.62
15	110	2012	M	Casablanca-Settat	Diesel	[08-10]	0.6113797408	7	35	2	7124.20	3.27129	3562.10
16	112	2011	F	Marrakech-Safi	Diesel	[10-14]	0.5669078722	11	31	1	384.56	1.76396	384.56
17	131	2012	M	Oriental	Essence	[00-08]	0.6677749872	17	28	0	0.00	0.00000	0.00
18	132	2009	M	BéniMellal-Khénifra	Diesel	[08-10]	0.7394534245	17	33	2	22354.70	2.70470	11177.35
19	152	2016	F	Souss-Massa	Diesel	[10-14]	0.7752906904	18	56	2	18369.73	2.57968	9184.87

FIGURE 4.2 – Output du code





# Chapitre 3

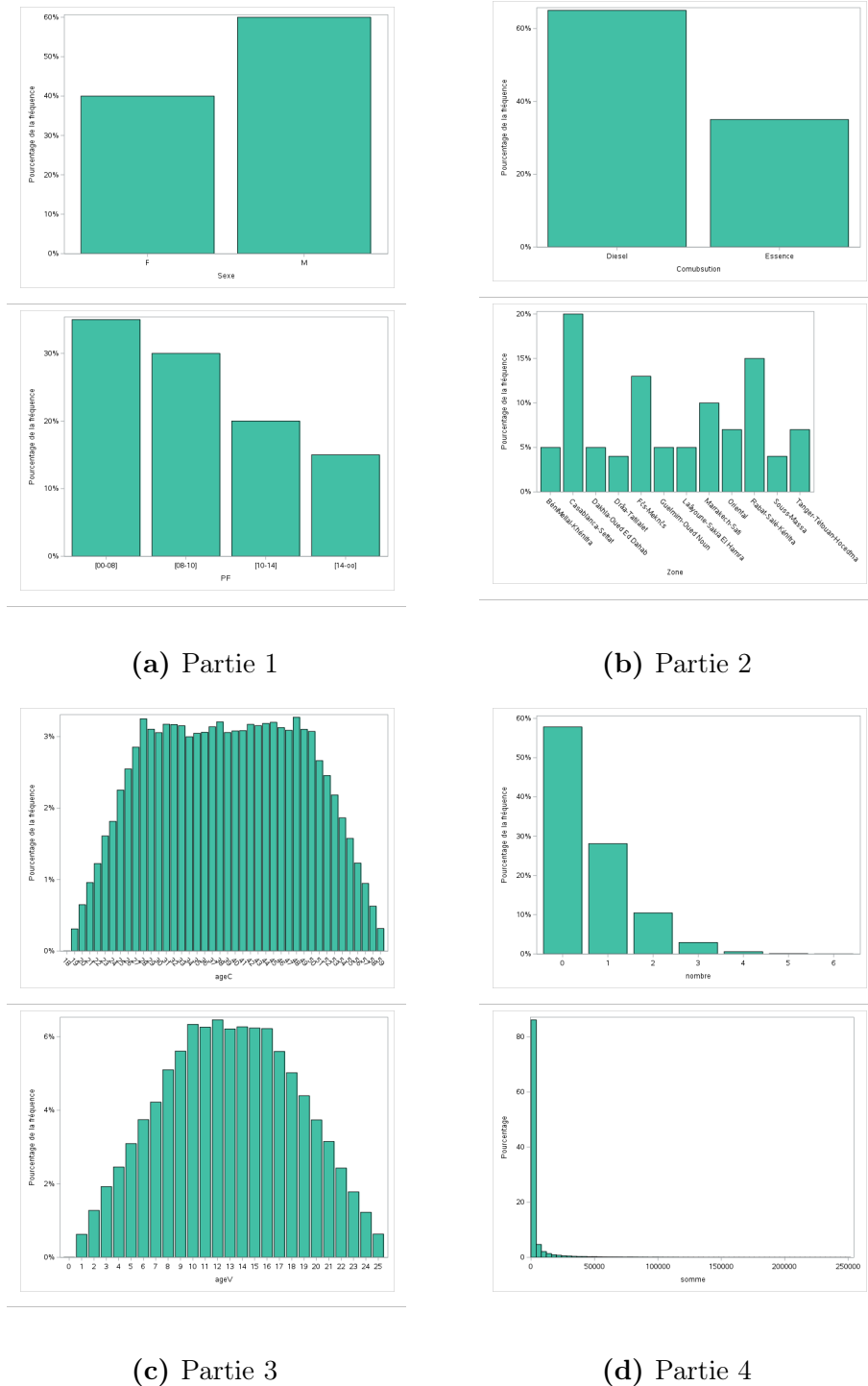
## STATISTIQUES DESCRIPTIVES

### 1 Statistiques univariées

#### 1.1 Pour chaque variables

On peut pour chaque variables faire un histogramme/bar plot visualisons les frequences (Voir page suivante) :

```
1 proc sgplot data=tabs.data;          vbar sexe / stat=percent
  ↳ fillattrs=(color=CX42bfa5);run;
2 proc sgplot data=tabs.data;          vbar pf / stat=percent
  ↳ fillattrs=(color=CX42bfa5);run;
3 proc sgplot data=tabs.data;          vbar comubsution / stat=percent
  ↳ fillattrs=(color=CX42bfa5);run;
4 proc sgplot data=tabs.data;          vbar zone/ stat=percent
  ↳ fillattrs=(color=CX42bfa5);run;
5 proc sgplot data=tabs.data;          vbar ageC / stat=percent
  ↳ fillattrs=(color=CX42bfa5);run;
6 proc sgplot data=tabs.data;          vbar ageV / stat=percent
  ↳ fillattrs=(color=CX42bfa5)run;
7 proc sgplot data=tabs.data;          vbar nombre / stat=percent
  ↳ fillattrs=(color=CX42bfa5);run;
8 proc sgplot data=tabs.data;          histogram somme
  ↳ /fillattrs=(color=CX42bfa5) ;run;
9
```



**FIGURE 1.1** – Distributions des variables de la base de données

On tire comme interprétation ce qui suit :

1. Les modalités sont bien enregistrées on ne trouve pas des modalités par exemple comme ('F', 'Féminin', ..., 'Autre')
2. On remarque le nombre de sinistre est similaire au distribution de *Poisson*, alors cette distribution serait la base de notre modélisation après.

## 1.2 Étude de l'effet de chaque variable sur le nombre de sinistre et la severité

On utilise le macro `plot_stats` suivant (le macro est défini en **Annexe 1**) :

### 1. Effet de la zone :

```
1 %plot_stats(cat_var=zone, quant_var=nombre);
2 %plot_stats(cat_var=zone, quant_var=severity);
```



(a) Effet sur le nombre



(b) Effet sur la severité

On remarque que il y a pas une grande difference entre les zones, donc on les répartie en deux zones (ZONE HIGH RISK, ZONE LOW RISK) pour améliorer la corrélaion avec la variable NOMBRE. Ainsi on va choisir :

— ZONE LOW RISK : ("Tanger-Tétouan-Hoced'ma", "Laâyoune-Sakia El Hamra", "Dakhla-Oued Ed Dahab", "Marrakech-Safi")

— ZONE HIGH RISK : les autres

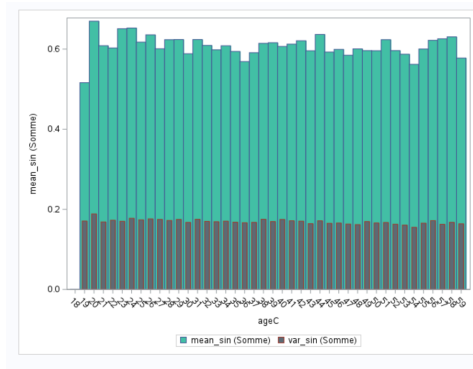
Même logique, on va choisir une autre découpage pour améliorer la corrélation avec la variable SEVERITE, ainsi on choisit (voir découpage en annexe) :

— ZONE LOW RISK : ("Tanger-Tétouan-Hoced'ma", "Laâyoune-Sakia El Hamra", "Dakhla-Oued Ed Dahab", "Marrakech-Safi")

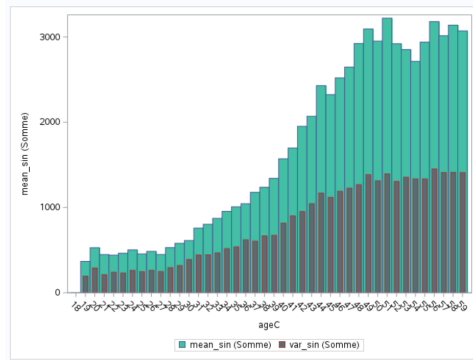
— ZONE HIGH RISK : les autres

### 2. Effet de la age conducteur :

```
1 %plot_stats(cat_var=ageC, quant_var=nombre);
2 %plot_stats(cat_var=ageC, quant_var=severity);
```



(a) Effet sur le nombre



(b) Effet sur la severité

On remarque qu'il y a pas un effet tendenciel de l'age conducteur sur le nombre de sinistre, donc on propose de découper les ages en des nouvelles tranche d'ages pour améliorer la corrélation :

- AGE CONDUCTEUR HIGH RISK :  $[30 - 55]$
- AGE CONDUCTEUR LOW RISK :  $[18 - 30] \cup [56 - +\infty]$

Cela se fait à travers ce code<sup>1</sup> :

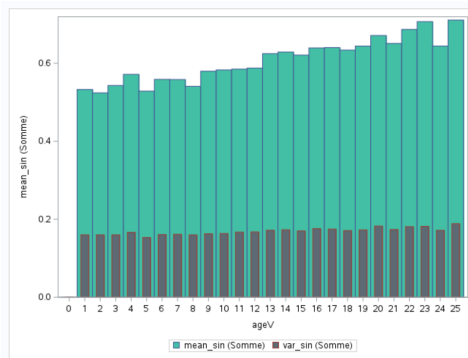
```
1 proc sql;
2     alter table tabs.data add ageCnew char(100) ;
3     update tabs.data set ageCnew = "[18-30] U [56-oo]" where
4         ↪ (ageC between 18 and 28) or (ageC >= 56);
5     update tabs.data set ageCnew = "[29-50]" where ageC between
6         ↪ 29 and 55;
7 quit;
```

### 3. Effet de la age Véhicule :

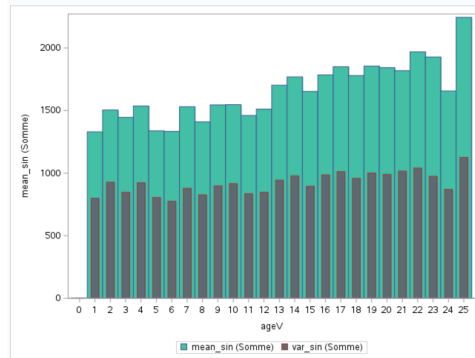
```
1 %plot_stats(cat_var=ageV, quant_var=nombre);
2 %plot_stats(cat_var=ageV, quant_var=severity);
```

On remarque alors qu'il y a une tendance entre l'age de vehicule et le nombre des sinistres/séverité. Donc ne fait pas de découpage de cette variable.

1. On remarque qu'un individus d'age moyenne est moins suspect à faire une sinistre par rapport au gens âgées (problèmes sanitaires) et les gens jeunes (manque d'expérience)



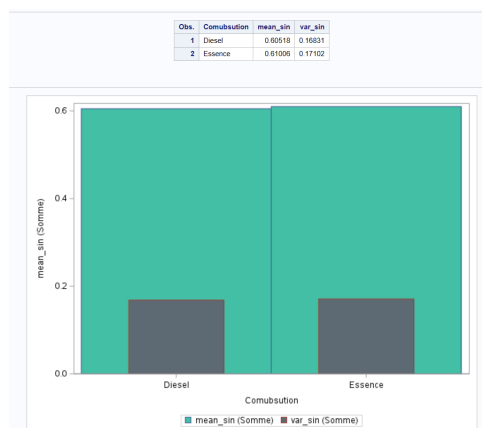
(a) Effet sur le nombre



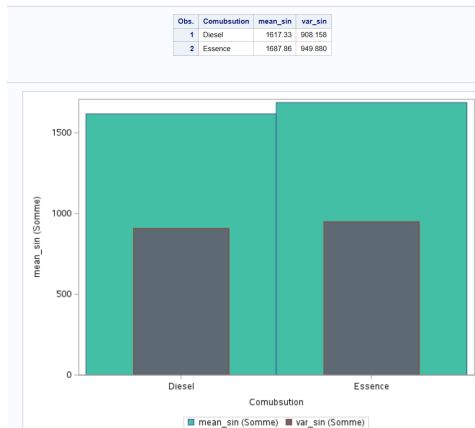
(b) Effet sur la severité

#### 4. Effet de la comubsution :

```
1 %plot_stats(cat_var=comubsution, quant_var=nombre);
2 %plot_stats(cat_var=comubsution, quant_var=severity);
```



(a) Effet sur le nombre



(b) Effet sur la severité

On remarque que la comubsution n'a pas de grande effet sur le nombre de sinistre, cependant on voit un effet intéressant sur la sévérité.

#### 5. Effet de Puissance fiscale :

```
1 %plot_stats(cat_var=PF, quant_var=nombre);
2 %plot_stats(cat_var=PF, quant_var=severity);
```

On remarque que il y a une seule puissance fiscal qui fait une différence significative, donc on regroupe les modalités deux (PF HIGH RISK, PF LOW RISK) pour améliorer la significativité.

Ainsi on va choisir pour la variable dépendante NOMBRE :



(a) Effet sur le nombre



(b) Effet sur la severité

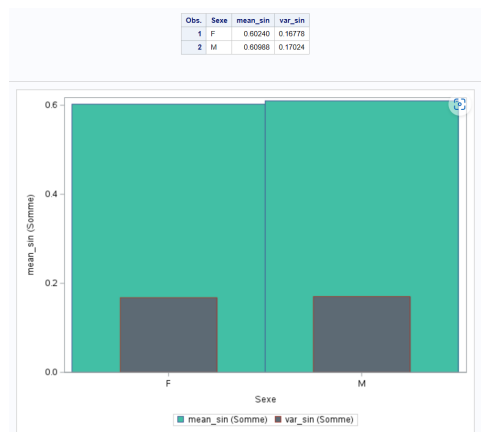
- PF HIGH RISK :  $[10 - 14]$
- PF LOW RISK :  $[0 - 10] \cup [14 - +\infty]$

Ainsi on va choisir pour la variable dépendante SEVERITÉ :

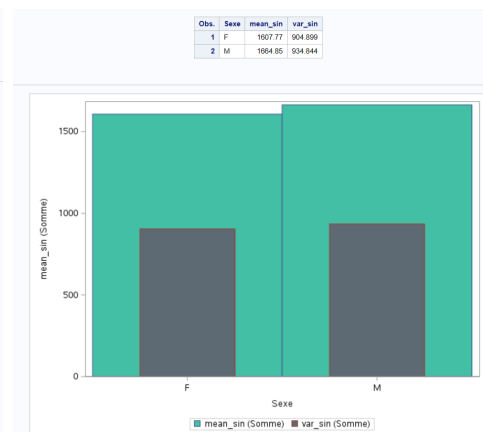
- PF HIGH RISK :  $[00 - 08]$
- PF LOW RISK :  $[08 - +\infty]$

## 6. Effet de la sexe :

```
1 %plot_stats(cat_var=sexe, quant_var=nombre);
2 %plot_stats(cat_var=sexe, quant_var=severity);
```



(a) Effet sur le nombre



(b) Effet sur la severité

# Chapitre 4

## MODELISATION ET CALCUL DES PRIMES

### 1 Modélisation de nombre de sinistres

On va modéliser notre base de données par un modèle de *Poisson* avec inflation et avec zéro inflation, notre choix de modèle poissonienne vient de la forme de fonction de distribution de variable NOMBRE. Cependant cela ne nous empêchera pas à comparer le modèle avec le modèle binomial négatif.

On doit d'abord calculer le logarithme de l'exposition qui va être l'*Offset* des modèles.

```
1 proc sql;
2     alter table tabs.data add logexp float(4) ;
3     update tabs.data set logexp = log(exposition);
4 quit;
```

#### 1.1 Modèles sans "zéro inflation"

On va d'abord essayer un modèle naïf qui inclut toutes les variables explicatives, puis on l'améliore au fur et à mesure selon les outputs qu'on obtient.

```
1 proc genmod data=tabs.data;
2 Class sexe zone    comubsution PF ;
3 Model nombre = sexe comubsution PF zone ageC ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 1";
6 ods output modelfit = pois1;
7 run;
```

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%	Khi-2 de Wald	Pr > khi-2
Intercept		1	0.1078	0.0236	0.0616	0.1540	20.94 <.0001
Sexe	F	1	-0.0160	0.0213	-0.0577	0.0258	0.56 0.4537
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	. .
Comubsution	Diesel	1	-0.0148	0.0320	-0.0775	0.0478	0.22 0.6428
Comubsution	Essence	0	0.0000	0.0000	0.0000	0.0000	. .
PF	[00-08]	0	0.0000	0.0000	0.0000	0.0000	. .
PF	[08-10]	1	0.0011	0.0382	-0.0738	0.0760	0.00 0.9764
PF	[10-14]	1	0.0211	0.0443	-0.0658	0.1080	0.23 0.6337
PF	[14-oo]	0	0.0000	0.0000	0.0000	0.0000	. .
Zone	BéniMellal-Khénifra	1	0.0085	0.0317	-0.0536	0.0707	0.07 0.7877
Zone	Casablanca-Settat	1	0.0136	0.0283	-0.0419	0.0691	0.23 0.6301
Zone	Dakhla-Oued Ed Dahab	1	0.0147	0.0258	-0.0358	0.0652	0.33 0.5674
Zone	Drâa-Taflalet	1	0.0106	0.0452	-0.0780	0.0991	0.05 0.8152
Zone	Fès-Meknès	1	0.0011	0.0190	-0.0363	0.0384	0.00 0.9558
Zone	Guelmim-Oued Noun	1	0.0413	0.0257	-0.0090	0.0916	2.58 0.1079
Zone	Laâyoune-Sakia El Hamra	0	0.0000	0.0000	0.0000	0.0000	. .
Zone	Marrakech-Safi	1	-0.0251	0.0425	-0.1084	0.0581	0.35 0.5539
Zone	Oriental	1	0.0071	0.0216	-0.0353	0.0495	0.11 0.7428
Zone	Rabat-Salé-Kénitra	1	0.0048	0.0210	-0.0363	0.0459	0.05 0.8173
Zone	Souss-Massa	1	0.0234	0.0452	-0.0651	0.1120	0.27 0.6041
Zone	Tanger-Tétouan-Hocedma	0	0.0000	0.0000	0.0000	0.0000	. .
ageC		1	-0.0020	0.0004	-0.0028	-0.0012	22.06 <.0001
ageV		1	0.0128	0.0008	0.0113	0.0143	284.25 <.0001
Echelle		0	1.0000	0.0000	1.0000	1.0000	

FIGURE 1.1 – Modèle poisson 1

On remarque que les puissances fiscales sont pas significatifs, et il ont des pvalues qui dépassent 90%, donc on va essayer en premiere tempe à remedier ce problème. On va alors travailler avec NEWPF

```

1 proc genmod data=tabs.data;
2 Class sexe zone comubsution newPF ;
3 Model nombre = sexe comubsution newPF zone ageC ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 2";
6 ods output modelfit = pois2;
7 run;

```

On voit bien que ce modèle améliore le pvalue de puissance fiscale mais sans qui le rendre significative, on le laisse dans le modèle jusqu'à remedier le problème de zone qui parait plus sérieux, on travaille alors sur la variable NEWZONE.



Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Pr > khi-2
Intercept		1	0.1278	0.0415	0.0466	0.2091	0.0020
Sexe	F	1	-0.0160	0.0213	-0.0577	0.0258	0.4537
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.
Comubsution	Diesel	1	-0.0137	0.0210	-0.0548	0.0274	0.43
Comubsution	Essence	0	0.0000	0.0000	0.0000	0.0000	.
newPF	PF_high_risk	1	-0.0200	0.0341	-0.0868	0.0468	0.34
newPF	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
Zone	BéniMellal-Khénifra	1	0.0085	0.0317	-0.0536	0.0707	0.07
Zone	Casablanca-Settat	1	0.0136	0.0283	-0.0419	0.0691	0.23
Zone	Dakhla-Oued Ed Dahab	1	0.0136	0.0382	-0.0613	0.0885	0.13
Zone	Drâa-Tafilalet	1	0.0106	0.0452	-0.0780	0.0991	0.05
Zone	Fcs-Meknès	1	0.0011	0.0190	-0.0363	0.0384	0.00
Zone	Guelmim-Oued Noun	1	0.0401	0.0382	-0.0347	0.1149	1.11
Zone	Laâyoune-Sakia El Hamra	1	-0.0011	0.0382	-0.0760	0.0738	0.00
Zone	Marrakech-Safi	1	-0.0251	0.0425	-0.1084	0.0581	0.35
Zone	Oriental	1	0.0071	0.0216	-0.0353	0.0495	0.11
Zone	Rabat-Salé-Kénitra	1	0.0048	0.0210	-0.0363	0.0459	0.05
Zone	Souss-Massa	1	0.0234	0.0452	-0.0651	0.1120	0.27
Zone	Tanger-Tétouan-Hocedma	0	0.0000	0.0000	0.0000	0.0000	.
ageC		1	-0.0020	0.0004	-0.0028	-0.0012	22.06
ageV		1	0.0128	0.0008	0.0113	0.0143	284.25
Echelle		0	1.0000	0.0000	1.0000	1.0000	.

FIGURE 1.2 – Poisson 2

```

1 proc genmod data=tabs.data;
2 Class sexe newzone comubsution newPF ;
3 Model nombre = sexe comubsution newPF newzone ageC ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 3";
6 ods output modelfit = pois3;
7 run;

```

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Pr > khi-2
Intercept		1	0.0931	0.0245	0.0451	0.1412	14.45
Sexe	F	1	0.0010	0.0132	-0.0249	0.0268	0.01
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.
Comubsution	Diesel	1	-0.0106	0.0108	-0.0318	0.0105	0.97
Comubsution	Essence	0	0.0000	0.0000	0.0000	0.0000	.
newPF	PF_high_risk	1	0.0008	0.0129	-0.0245	0.0260	0.00
newPF	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
newzone	zone_high_risk	1	0.0212	0.0103	0.0009	0.0414	4.20
newzone	zone_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
ageC		1	-0.0020	0.0004	-0.0028	-0.0012	22.11
ageV		1	0.0128	0.0008	0.0113	0.0143	284.14
Echelle		0	1.0000	0.0000	1.0000	1.0000	.

Note: Le paramètre d'échelle a été fixé.

FIGURE 1.3 – Poisson 3

Lorsque On a ajouté une deuxième variable explicative (newzone) au modèle, la p-value de newPF augmente considérablement, passant de 55% à 95% Cela peut signifier que

l'ajout de newzone en tant que variable explicative a en quelque sorte "expliqué" une partie de la variation dans la variable dépendante qui était auparavant attribuée à newPF. Cela peut également indiquer une corrélation entre newPF et newzone, ce qui peut rendre difficile la distinction de l'effet de chacune de ces variables sur la variable dépendante. On peut essayer de faire un modèle qui fait intervenir l'interaction entre les deux variables :

```

1 proc genmod data=tabs.data;
2 Class sexe newzone comubsution newPF ;
3 Model nombre = sexe comubsution newPF*newzone ageC ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 4";
6 ods output modelfit = pois4;
7 run;

```

Analyse des paramètres estimés du maximum de vraisemblance									
Paramètre			DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Khi-2 de Wald	Pr > khi-2
Intercept			1	0.0858	0.0251	0.0365	0.1350	11.65	0.0006
Sexe	F		1	-0.0038	0.0137	-0.0306	0.0230	0.08	0.7833
Sexe	M		0	0.0000	0.0000	0.0000	0.0000	.	.
Comubsution	Diesel		1	-0.0088	0.0109	-0.0301	0.0126	0.65	0.4205
Comubsution	Essence		0	0.0000	0.0000	0.0000	0.0000	.	.
newzone*newPF	zone_high_risk	PF_high_risk	1	0.0275	0.0171	-0.0060	0.0610	2.59	0.1074
newzone*newPF	zone_high_risk	PF_low_risk	1	0.0415	0.0183	0.0056	0.0773	5.14	0.0234
newzone*newPF	zone_low_risk	PF_high_risk	1	0.0158	0.0171	-0.0177	0.0494	0.86	0.3541
newzone*newPF	zone_low_risk	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.	.
ageC			1	-0.0020	0.0004	-0.0028	-0.0012	22.16	<.0001
ageV			1	0.0128	0.0008	0.0113	0.0143	284.11	<.0001
Echelle			0	1.0000	0.0000	1.0000	1.0000		

Note: Le paramètre d'échelle a été fixé.

FIGURE 1.4 – Poisson 4

On remarque qu'on a réduit de toutes les variables, donc ça serait intéressant de modéliser avec ce modèle (On va se baser sur l'AIC et la deviance pour décider plus tard). Maintenant on va procéder sur l'ageC, on a déjà remarqué que ça peut être intéressant de modéliser par des tranches d'âges que d'un seul variable quantitative (voir statistiques descriptives 2)

```

1 proc genmod data=tabs.data;
2 Class sexe newzone comubsution agecnew newPF ;
3 Model nombre = sexe comubsution newPF*newzone ageCnew ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 5";
6 ods output modelfit = pois5;
7 run;

```

On constate que cette décomposition n'a rien amélioré, on fait elle a décortiquer le modèle, donc on élimine cette idée.

Analyse des paramètres estimés du maximum de vraisemblance									
Paramètre			DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Khi-2 de Wald	Pr > khi-2
Intercept			1	0.0046	0.0205	-0.0356	0.0448	0.05	0.8209
Sexe	F		1	-0.0036	0.0137	-0.0304	0.0232	0.07	0.7942
Sexe	M		0	0.0000	0.0000	0.0000	0.0000	.	.
Comubsution	Diesel		1	-0.0088	0.0109	-0.0302	0.0125	0.66	0.4171
Comubsution	Essence		0	0.0000	0.0000	0.0000	0.0000	.	.
newzone*newPF	zone_high_risk	PF_high_risk	1	0.0278	0.0171	-0.0057	0.0613	2.64	0.1042
newzone*newPF	zone_high_risk	PF_low_risk	1	0.0412	0.0183	0.0053	0.0770	5.07	0.0244
newzone*newPF	zone_low_risk	PF_high_risk	1	0.0158	0.0171	-0.0177	0.0494	0.86	0.3542
newzone*newPF	zone_low_risk	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.	.
ageCnew	[18-30] U [56-oo]		1	0.0385	0.0100	0.0190	0.0581	14.95	0.0001
ageCnew	[29-50]		0	0.0000	0.0000	0.0000	0.0000	.	.
ageV			1	0.0124	0.0008	0.0110	0.0139	273.79	<.0001
Echelle			0	1.0000	0.0000	1.0000	1.0000		

Note: Le paramètre d'échelle a été fixé.

FIGURE 1.5 – Poisson 5

Cependant le sexe reste moins significative donc on l'élimine du modèle :

```

1 proc genmod data=tabs.data;
2 Class sexe newzone comubsution newPF ;
3 Model nombre = comubsution newzone*newpf ageC ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 6";
6 ods output modelfit = pois6;
7 run;

```

Analyse des paramètres estimés du maximum de vraisemblance									
Paramètre			DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Khi-2 de Wald	Pr > khi-2
Intercept			1	0.0837	0.0239	0.0368	0.1305	12.24	0.0005
Comubsution	Diesel		1	-0.0104	0.0091	-0.0283	0.0075	1.30	0.2549
Comubsution	Essence		0	0.0000	0.0000	0.0000	0.0000	.	.
newzone*newPF	zone_high_risk	PF_high_risk	1	0.0300	0.0146	0.0013	0.0586	4.20	0.0403
newzone*newPF	zone_high_risk	PF_low_risk	1	0.0415	0.0183	0.0056	0.0773	5.14	0.0234
newzone*newPF	zone_low_risk	PF_high_risk	1	0.0167	0.0168	-0.0162	0.0497	0.99	0.3197
newzone*newPF	zone_low_risk	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.	.
ageC			1	-0.0020	0.0004	-0.0028	-0.0012	22.17	<.0001
ageV			1	0.0128	0.0008	0.0113	0.0143	284.12	<.0001
Echelle			0	1.0000	0.0000	1.0000	1.0000		

FIGURE 1.6 – Poisson 6

Étant donnée qu'on a éliminer juste le sexe pour obtenir des pvalues assez bons comme cela (juste la comubsution qui est supérieur à 5%) on conververa ce modèle.

## 1.2 Modèles avec zero inflation

On essaie d'améliorer les bons modèles trouvées par *Poisson* (poisson 4 et poisson 6) avec l'ajout de phénomen de zéro inflation :

```

1 proc genmod data=tabs.data;
2 Class newzone    comubsution  newPF  ;
3 Model nombre =   comubsution newzone*newpf  ageC ageV/
4 dist = zip link = log offset=logexp ;
5 zeromodel  / link = logit  ;
6 title "ZIP 4";
7 ods output modelfit = zip4;
8 run;
9
11 proc genmod data=tabs.data;
12 Class newzone    comubsution  newPF  ;
13 Model nombre =   comubsution newzone*newpf  ageC ageV/
14 dist = zip link = log offset=logexp ;
15 zeromodel  / link = logit  ;
16 title "ZIP 6";
17 ods output modelfit = zip6;
18 run;

```

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald 95%	Khi-2 de Wald	Pr > khi-2
Intercept		1	0.0837	0.0239	0.0368	0.1305	12.24
Combustion	Diesel	1	-0.0104	0.0091	-0.0283	0.0075	1.30
Combustion	Essence	0	0.0000	0.0000	0.0000	0.0000	0.2549
newzone*newPF	zone_high_risk PF_high_risk	1	0.0300	0.0146	0.0013	0.0586	4.20
newzone*newPF	zone_high_risk PF_low_risk	1	0.0415	0.0183	0.0056	0.0773	5.14
newzone*newPF	zone_low_risk PF_high_risk	1	0.0187	0.0168	-0.0162	0.0487	0.99
newzone*newPF	zone_low_risk PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	0.3197
ageC		1	-0.0020	0.0004	-0.0028	-0.0012	22.17
ageV		1	0.0128	0.0008	0.0113	0.0143	284.12
Echelle		0	1.0000	0.0000	1.0000	1.0000	<.0001

Note: Le paramètre d'échelle a été fixé.

Paramètres estimés par l'analyse du maximum de vraisemblance							
Paramètre	DDL	Estimation	Erreur type	Intervalle de confiance de Wald 95%	Khi-2 de Wald	Pr > khi-2	
Intercept	1	-21.6304	2677.747	-5209.92	5226.658	0.00	

(a) ZIP 4

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald 95%	Khi-2 de Wald	Pr > khi-2
Intercept		1	0.0837	0.0239	0.0368	0.1305	12.24
Combustion	Diesel	1	-0.0104	0.0091	-0.0283	0.0075	1.30
Combustion	Essence	0	0.0000	0.0000	0.0000	0.0000	-
newzone*newPF	zone_high_risk PF_high_risk	1	0.0300	0.0146	0.0013	0.0586	4.20
newzone*newPF	zone_high_risk PF_low_risk	1	0.0415	0.0183	0.0056	0.0773	5.14
newzone*newPF	zone_low_risk PF_high_risk	1	0.0187	0.0168	-0.0162	0.0487	0.99
newzone*newPF	zone_low_risk PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	-
ageC		1	-0.0020	0.0004	-0.0028	-0.0012	22.17
ageV		1	0.0128	0.0008	0.0113	0.0143	284.12
Echelle		0	1.0000	0.0000	1.0000	1.0000	<.0001

Note: Le paramètre d'échelle a été fixé.

Paramètres estimés par l'analyse du maximum de vraisemblance							
Paramètre	DDL	Estimation	Erreur type	Intervalle de confiance de Wald 95%	Khi-2 de Wald	Pr > khi-2	
Intercept	1	-21.6304	2677.747	-5209.92	5226.658	0.00	

(b) ZIP 6

On remarque que *l'intercepte* est significativement null et que ce modèle ne diffère rien de modèle poisson, donc on ne travaille pas avec ZIP.

### 1.3 Comparaison des modèles :

On fait une comparaison statistiques à l'aide de la deviance et l'AIC :

Obs.	Criterion	DF	Value	ValueDF
1	Ecart	1E5	93538.3287	0.9355
2	Déviance normalisée	1E5	93538.3287	0.9355
3	Khi2 de Pearson	1E5	97873.1617	0.9788
4	Pearson normalisé X2	1E5	97873.1617	0.9788
5	Log-vraisemblance	—	-79165.1146	—
6	Log-vraisemblance complète	—	-94056.9202	—
7	AIC (préférer les petites valeurs)	—	188129.8405	—
8	AICC (préférer les petites valeurs)	—	188129.8419	—
9	BIC (préférer les petites valeurs)	—	188205.9439	—

(a) Poisson 4

Obs.	Criterion	DF	Value	ValueDF
1	Ecart	1E5	93538.4044	0.9354
2	Déviance normalisée	1E5	93538.4044	0.9354
3	Khi2 de Pearson	1E5	97873.2434	0.9788
4	Pearson normalisé X2	1E5	97873.2434	0.9788
5	Log-vraisemblance	—	-79165.1525	—
6	Log-vraisemblance complète	—	-94056.9581	—
7	AIC (préférer les petites valeurs)	—	188127.9161	—
8	AICC (préférer les petites valeurs)	—	188127.9173	—
9	BIC (préférer les petites valeurs)	—	188194.5066	—

(b) Poisson 6

Les modèles en tant presque similaires au niveau statistique, on va faire une autre comparaison *Macro* qu'on verrait après.

## 2 Modélisation de sévérité

Dans la suite on va modéliser la sévérité en utilisant les deux modèles gamma et lognormal :

### 2.1 Modèle Gamma

On commence toujours avec un modèle très naïf, qui inclut toutes les variables :

```
1 proc genmod data=tabs.data;
2 Class sexe zone Comubsution PF;
3 Model severity = sexe zone Comubsution PF ageC ageV/
4 dist = gamma link = log offset=logexp ;
5 ods output modelfit = Gam1;
6 run;
```

Analyse des paramètres estimés du maximum de vraisemblance								
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à95%		Khi-2 de Wald	Pr > khi-2
Intercept		1	5.9338	0.0355	5.8643	6.0034	27961.0	<.0001
Sexe	F	1	-0.0806	0.0314	-0.1422	-0.0191	6.59	0.0102
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.	.
Zone	BéniMellal-Khénifra	1	-0.0589	0.0467	-0.1505	0.0327	1.59	0.2078
Zone	Casablanca-Settat	1	-0.0552	0.0418	-0.1372	0.0268	1.74	0.1872
Zone	Dakhla-Oued Ed Dahab	1	-0.1855	0.0656	-0.3140	-0.0570	8.01	0.0047
Zone	Drâa-Tafilalet	1	-0.1702	0.0667	-0.3010	-0.0395	6.51	0.0107
Zone	Fès-Meknès	1	0.0248	0.0282	-0.0305	0.0801	0.77	0.3796
Zone	Guelmim-Oued Noun	1	-0.1424	0.0656	-0.2709	-0.0139	4.72	0.0299
Zone	Laâyoune-Sakia El Hamra	1	-0.1617	0.0655	-0.2899	-0.0334	6.10	0.0135
Zone	Marrakech-Safi	1	-0.1553	0.0627	-0.2782	-0.0324	6.14	0.0132
Zone	Oriental	1	0.1020	0.0321	0.0391	0.1649	10.10	0.0015
Zone	Rabat-Salé-Kénitra	1	-0.0162	0.0310	-0.0769	0.0445	0.27	0.6014
Zone	Souss-Massa	1	-0.1913	0.0666	-0.3219	-0.0608	8.25	0.0041
Zone	Tanger-Tétouan-Hocedma	0	0.0000	0.0000	0.0000	0.0000	.	.
Comubsution	Diesel	1	0.2371	0.0591	0.1213	0.3529	16.11	<.0001
Comubsution	Essence	0	0.0000	0.0000	0.0000	0.0000	.	.
PF	[00-08]	0	0.0000	0.0000	0.0000	0.0000	.	.
PF	[08-10]	1	-0.2024	0.0503	-0.3009	-0.1038	16.19	<.0001
PF	[10-14]	0	0.0000	0.0000	0.0000	0.0000	.	.
PF	[14-oo]	0	0.0000	0.0000	0.0000	0.0000	.	.
ageC		1	0.0732	0.0006	0.0720	0.0745	12700.6	<.0001
ageV		1	-0.0109	0.0011	-0.0131	-0.0086	91.50	<.0001
Echelle		1	0.6602	0.0039	0.6527	0.6678		

FIGURE 2.1 – Gamma1

On remarque que la majorité des variables sont significatives, sauf quelques-unes. On commence à régler le problème de puissance fiscale, on utilise alors la décomposition trouvée précédemment NEWSPF :

```

1 proc genmod data=tabs.data;
2 Class sexe zone Comubsution newPF;
3 Model severity = sexe zone Comubsution newPF ageC ageV/
4 dist = gamma link = log offset=logexp ;
5 ods output modelfit = Gam2;
6 run;

```

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		
Intercept		1	6.1362	0.0614	6.0159	6.2565	9989.53
Sexe	F	1	-0.0806	0.0314	-0.1422	-0.0191	6.59
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.
Zone	BéniMellal-Khénifra	1	-0.0589	0.0467	-0.1505	0.0327	1.59
Zone	Casablanca-Settat	1	-0.0552	0.0418	-0.1372	0.0268	1.74
Zone	Dakhla-Oued Ed Dahab	1	0.0168	0.0565	-0.0939	0.1276	0.09
Zone	Drâa-Tafilet	1	-0.1702	0.0667	-0.3010	-0.0395	6.51
Zone	Fès-Meknès	1	0.0248	0.0282	-0.0305	0.0801	0.77
Zone	Guelmim-Oued Noun	1	0.0599	0.0565	-0.0508	0.1707	1.12
Zone	Laâyoune-Sakia El Hamra	1	0.0407	0.0564	-0.0698	0.1512	0.52
Zone	Marrakech-Safi	1	-0.1553	0.0627	-0.2782	-0.0324	6.14
Zone	Oriental	1	0.1020	0.0321	0.0391	0.1649	10.10
Zone	Rabat-Salé-Kénitra	1	-0.0162	0.0310	-0.0769	0.0445	0.27
Zone	Souss-Massa	1	-0.1913	0.0666	-0.3219	-0.0608	8.25
Zone	Tanger-Tétouan-Hocedma	0	0.0000	0.0000	0.0000	0.0000	.
Comubsution	Diesel	1	0.0347	0.0310	-0.0260	0.0955	1.26
Comubsution	Essence	0	0.0000	0.0000	0.0000	0.0000	.
newPF	PF_high_risk	1	-0.2024	0.0503	-0.3009	-0.1038	16.19
newPF	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
ageC		1	0.0732	0.0006	0.0720	0.0745	12700.6
ageV		1	-0.0109	0.0011	-0.0131	-0.0086	91.50
Echelle		1	0.6602	0.0039	0.6527	0.6678	.

FIGURE 2.2 – Gamma2

Cela règle le problème de la puissance fiscale, passons maintenant aux zones : on utilise également la décomposition NEWSZONE :

```

1 proc genmod data=tabs.data;
2 Class sexe newszone Comubsution newPF;
3 Model severity = sexe newszone Comubsution newPF ageC ageV/
4 dist = gamma link = log offset=logexp ;
5 ods output modelfit = Gam3;
6 run;

```

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Pr > khi-2
Intercept		1	5.9702	0.0358	5.9000	6.0403	27825.8 <.0001
Sexe	F	1	-0.0184	0.0180	-0.0536	0.0169	1.04 0.3077
Sexe	M	0	0.0000	0.0000	0.0000	0.0000	.
newszone	zone_high_risk	1	0.0422	0.0158	0.0112	0.0732	7.10 0.0077
newszone	zone_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
Comubsution	Diesel	1	-0.0149	0.0177	-0.0497	0.0198	0.71 0.4003
Comubsution	Essence	0	0.0000	0.0000	0.0000	0.0000	.
newPF	PF_high_risk	1	-0.0417	0.0194	-0.0796	-0.0037	4.62 0.0316
newPF	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
ageC		1	0.0732	0.0006	0.0719	0.0745	12687.1 <.0001
ageV		1	-0.0109	0.0011	-0.0131	-0.0087	91.71 <.0001
Echelle		1	0.6599	0.0039	0.6524	0.6675	

Note: Le paramètre d'échelle a été estimé par le maximum de vraisemblance.

FIGURE 2.3 – Gamma 3

On remarque que le regroupement de zone, a augmenté les pvalues des autres variables, pour régler ce problème on propose de faire ajouter un effet d'interaction entre le sexe et la comubsution :

```

1 proc genmod data=tabs.data;
2 Class sexe newszone Comubsution newPF;
3 Model severity = newszone sexe* Comubsution newPF ageC ageV/
4 dist = gamma link = log offset=logexp ;
5 ods output modelfit = Gam4;
6 run;

```

On trouve :

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		Pr > khi-2
Intercept		1	5.9702	0.0358	5.9000	6.0403	27825.8 <.0001
newszone	zone_high_risk	1	0.0422	0.0158	0.0112	0.0732	7.10 0.0077
newszone	zone_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
Sexe*Comubsution	F Diesel	1	-0.0333	0.0190	-0.0704	0.0039	3.08 0.0792
Sexe*Comubsution	M Diesel	1	-0.0149	0.0177	-0.0497	0.0198	0.71 0.4003
Sexe*Comubsution	M Essence	0	0.0000	0.0000	0.0000	0.0000	.
newPF	PF_high_risk	1	-0.0417	0.0194	-0.0796	-0.0037	4.62 0.0316
newPF	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
ageC		1	0.0732	0.0006	0.0719	0.0745	12687.1 <.0001
ageV		1	-0.0109	0.0011	-0.0131	-0.0087	91.71 <.0001
Echelle		1	0.6599	0.0039	0.6524	0.6675	

FIGURE 2.4 – Gamma 4

On peut également supprimer la variable sexe de modèle :

```

1 proc genmod data=tabs.data;
2 Class newszone Comubsution newPF;
3 Model severity = newszone Comubsution newPF ageC ageV/

```

```

4 dist = gamma link = log offset=logexp ;
5 ods output modelfit = Gam5;
6 run;

```

Analyse des paramètres estimés du maximum de vraisemblance							
Paramètre		DDL	Estimation	Erreur type	Intervalle de confiance de Wald à 95%		
Intercept		1	5.9599	0.0344	5.8926	6.0273	30069.1
newszone	zone_high_risk	1	0.0425	0.0158	0.0115	0.0736	7.23
newszone	zone_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
Comubsubtion	Diesel	1	-0.0228	0.0159	-0.0540	0.0085	2.04
Comubsubtion	Essence	0	0.0000	0.0000	0.0000	0.0000	.
newPF	PF_high_risk	1	-0.0315	0.0166	-0.0641	0.0011	3.58
newPF	PF_low_risk	0	0.0000	0.0000	0.0000	0.0000	.
ageC		1	0.0732	0.0006	0.0719	0.0745	12685.6
ageV		1	-0.0109	0.0011	-0.0131	-0.0087	91.82
Echelle		1	0.6599	0.0039	0.6524	0.6675	

FIGURE 2.5 – Gamma 5

On dispose alors de deux modèles qui semblent assez bons, on les compare ultérieurement à l'aide de l'AIC.

## 2.2 Modèle Log-normal

On ajoute d'abord le logarithme de la sévérité :

```

1 proc sql;
2     alter table tabs.data add logSev float(4);
3     update tabs.data set logsev = log(severity);
4 quit;

```

À l'aide d'une démarche analogique on trouve toutes les modèles possibles non significatives, voir l'annexe 5 :

## 2.3 Comparaison des modèles

On ne compare que les modèles : Gamma4 et Gamma5,

Autre fois, on ne remarque aucune différence significative entre les deux modèles, donc on va faire une comparaison macro en dernière temps.



Obs.	Criterion	DF	Value	ValueDF
1	Ecart	42E3	77894.5167	1.8472
2	Déviante normalisée	42E3	51402.7123	1.2190
3	Khi2 de Pearson	42E3	131178.0063	3.1108
4	Pearson normalisé X2	42E3	86564.5697	2.0528
5	Log-vraisemblance	—	-380136.4036	—
6	Log-vraisemblance complète	—	-380136.4036	—
7	AIC (préférer les petites valeurs)	—	760288.8072	—
8	AICC (préférer les petites valeurs)	—	760288.8106	—
9	BIC (préférer les petites valeurs)	—	760358.0038	—

(a) Gamma 4

Obs.	Criterion	DF	Value	ValueDF
1	Ecart	42E3	77896.0922	1.8472
2	Déviante normalisée	42E3	51402.6376	1.2190
3	Khi2 de Pearson	42E3	131026.0154	3.1072
4	Pearson normalisé X2	42E3	86462.3963	2.0504
5	Log-vraisemblance	—	-380136.9234	—
6	Log-vraisemblance complète	—	-380136.9234	—
7	AIC (préférer les petites valeurs)	—	760287.8468	—
8	AICC (préférer les petites valeurs)	—	760287.8495	—
9	BIC (préférer les petites valeurs)	—	760348.3939	—

(b) Gamma 5

### 3 Calcul de la prime

La primes serait calculée comme suit :

$$Prime_{pure} = Prime_{stand} + Prime_{grand}$$

$$Prime_{pure} = Severité \times Frequence + \mathbb{P}(Grand \sin) \times \mathbb{E}(Mont_{grand})$$

- $Prime_{stand}$  correspond au primes qui vont couvrir les sinistres standards ;
- $Prime_{grand}$  correspond à une prime constante (appliquée à tous independamment de son profile assurée) qui va couvrir les sinistres de grandes montants.

On calcul chaque terme toute seule :

#### 3.1 Calcul de la frequence

On utilise ce code :

```

1 proc genmod data=tabs.data;
2 Class sexe newzone comubsution newPF ;
3 Model nombre = sexe comubsution newPF*newzone ageC ageV/
4 dist = pois link = log offset=logexp ;
5 title "Poisson 4";
6 ods output modelfit = pois4;
7 output out=tabs.tarif4 p=fitted_frequence;
8 run;
9
10 proc sql;
11     update tabs.tarif4 set fitted_frequence =
12         ↪ fitted_frequence/exposition;
13 quit;
14
15 proc genmod data=tabs.data;
16 Class newzone comubsution newPF ;
17 Model nombre = comubsution newzone*newpf ageC ageV/
18 dist = pois link = log offset=logexp ;

```

```

19 title "Poisson 6";
20 ods output modelfit = pois6;
21 output out=tabs.tarif6 p=fitted_frequence;
22 run;
23
24 proc sql;
25     update tabs.tarif6 set fitted_frequence =
        ↪ fitted_frequence/exposition;
26 quit;

```

On disposera à la sortie du code de deux tables TARIF4 et TARIF6 chaqu'une possède les frequences calculée selon un modèle donnée.

Comparons d'abord s'il la totalité des nombres de sinistres observé est proches des prédictions des modèles, Cela peut être fait à travers ce code :

```

1  proc sql;
2  CREATE TABLE amd (
3      nombre_total INT ,
4      nombre_tarif4 INT,
5      nombre_tarif6 INT
6  );
7
8  INSERT INTO amd (nombre_total, nombre_tarif4, nombre_tarif6) VALUES
9      (0, 0, 0);
10
11 UPDATE amd set
12 nombre_total = (select sum(nombre) from tabs.sinistres),
13 nombre_tarif4 = (select sum(fitted_frequence*exposition) from
    ↪ tabs.tarif4),
14 nombre_tarif6 = (select sum(fitted_frequence*exposition) from
    ↪ tabs.tarif6);
15 quit;

```

Qui nous donne cela :

Obs.	nombre_total	nombre_tarif4	nombre_tarif6
1	60689	60689.00	60689.00

FIGURE 3.1 – Comparaison des totaux de nombre sinistres

On remarque ainsi que les deux modèles ne diffèrent pas dans leur estimation de la totalité des nombre de sinistres, et ils sont adéquates au valeur de notre base.

## 3.2 Calcul de la severité

On utilise ce code :

```
1 proc genmod data=tabs.data;
2 Class sexe newszone Comubsution newPF;
3 Model severity = newszone sexe* Comubsution newPF ageC ageV/
4 dist = gamma link = log offset=logexp ;
5 ods output modelfit = Gam4;
6 output out=sev4 p=fitted_severity;
7 run;
8
9 proc genmod data=tabs.data;
10 Class newszone Comubsution newPF;
11 Model severity = newszone Comubsution newPF ageC ageV/
12 dist = gamma link = log offset=logexp ;
13 ods output modelfit = Gam5;
14 output out=sev5 p=fitted_severity;
15 run;
```

## 3.3 Calcul de la prime

### 3.i Prime standard

On dispose de deux modèles pour chaque de nombre et de severité, donc on aurait quatre tables de tarif, on donnera les code de création dans l'annexe : 6 On obtient alors les tables suivantes, affichée par ce code (même ordre) :

```
1 proc print data=tarif44(obs=5);run;
2 proc print data=tarif45(obs=5);run;
3 proc print data=tarif64(obs=5);run;
4 proc print data=tarif65(obs=5);run;
```

Obs.	numepolice	exercice	Sexe	Zone	Combustion	PF	exposition	ageV	ageC	nombre	somme	frequence	severity	newzone	newszone	ageCnew	newPF	newsPF	logexp	lsnull	logSev	fitted_frequence	fitted_severity	prime_stand
1	2	2009	M	Fcs-Mekrds	Essence	[00-08]	0.8505040312	7	28	1	1966.91	1.17577	1966.91	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15838	2396.73	2776.32
2	5	2016	M	Rabat-Sala-Kenitra	Essence	[00-08]	0.4869068514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.71968	0	.	1.20616	1234.34	1489.81
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	6	28	2	3625.85	4.16185	1812.93	zone_low_risk	zone_low_risk	[18-30] U [56-oo]	PF_low_risk	PF_high_risk	-0.73281	1	7.50270	1.09874	1323.53	1454.22
4	9	2012	M	Tanger-Tetouan-Hoceima	Essence	[00-08]	0.0609243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[29-50]	PF_high_risk	PF_low_risk	-2.79812	0	.	1.28543	580.29	745.93
5	24	2008	M	Rabat-Sala-Kenitra	Diesel	[08-10]	0.5667409257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10719	1510.89	1672.84

Obs.	numepolice	exercice	Sexe	Zone	Combustion	PF	exposition	ageV	ageC	nombre	somme	frequence	severity	newzone	newszone	ageCnew	newPF	newsPF	logexp	lsnull	logSev	fitted_frequence	fitted_severity	prime_stand
1	2	2009	M	Fcs-Mekrds	Essence	[00-08]	0.8505040312	7	28	1	1966.91	1.17577	1966.91	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15838	2397.12	2776.77
2	5	2016	M	Rabat-Sala-Kenitra	Essence	[00-08]	0.4869068514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.71968	0	.	1.20616	1234.52	1489.03
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	6	28	2	3625.85	4.16185	1812.93	zone_low_risk	zone_low_risk	[18-30] U [56-oo]	PF_low_risk	PF_high_risk	-0.73281	1	7.50270	1.09874	1323.66	1454.36
4	9	2012	M	Tanger-Tetouan-Hoceima	Essence	[00-08]	0.0609243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[29-50]	PF_high_risk	PF_low_risk	-2.79812	0	.	1.28543	580.08	745.65
5	24	2008	M	Rabat-Sala-Kenitra	Diesel	[08-10]	0.5667409257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10719	1499.33	1660.04

Obs.	numepolice	exercice	Sexe	Zone	Combustion	PF	exposition	ageV	ageC	nombre	somme	frequence	severity	newzone	newszone	ageCnew	newPF	newsPF	logexp	lsnull	logSev	fitted_frequence	fitted_severity	prime_stand
1	2	2009	M	Fcs-Mekrds	Essence	[00-08]	0.8505040312	7	28	1	1966.91	1.17577	1966.91	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15873	2396.73	2777.17
2	5	2016	M	Rabat-Sala-Kenitra	Essence	[00-08]	0.4869068514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.71968	0	.	1.20653	1234.34	1489.27
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	6	28	2	3625.85	4.16185	1812.93	zone_low_risk	zone_low_risk	[18-30] U [56-oo]	PF_low_risk	PF_high_risk	-0.73281	1	7.50270	1.09874	1323.53	1454.22
4	9	2012	M	Tanger-Tetouan-Hoceima	Essence	[00-08]	0.0609243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[29-50]	PF_high_risk	PF_low_risk	-2.79812	0	.	1.28382	580.29	744.99
5	24	2008	M	Rabat-Sala-Kenitra	Diesel	[08-10]	0.5667409257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10572	1510.89	1670.63

Obs.	numepolice	exercice	Sexe	Zone	Combustion	PF	exposition	ageV	ageC	nombre	somme	frequence	severity	newzone	newszone	ageCnew	newPF	newsPF	logexp	lsnull	logSev	fitted_frequence	fitted_severity	prime_stand
1	2	2009	M	Fcs-Mekrds	Essence	[00-08]	0.8505040312	7	28	1	1966.91	1.17577	1966.91	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15873	2397.12	2777.63
2	5	2016	M	Rabat-Sala-Kenitra	Essence	[00-08]	0.4869068514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_low_risk	-0.71968	0	.	1.20653	1234.52	1489.49
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4805559502	6	28	2	3625.85	4.16185	1812.93	zone_low_risk	zone_low_risk	[18-30] U [56-oo]	PF_low_risk	PF_high_risk	-0.73281	1	7.50270	1.09874	1323.66	1454.36
4	9	2012	M	Tanger-Tetouan-Hoceima	Essence	[00-08]	0.0609243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[29-50]	PF_high_risk	PF_low_risk	-2.79812	0	.	1.28382	580.08	744.71
5	24	2008	M	Rabat-Sala-Kenitra	Diesel	[08-10]	0.5667409257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-30] U [56-oo]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10572	1499.33	1657.85

FIGURE 3.2 – Les tables des primes

### 3.ii Prime pure

On doit d'abord calculer la cotisation qui va s'ajouter au primes standards :

```

1 proc sql;
2 CREATE TABLE cotisation (cot float(4));
3 INSERT INTO cotisation (cot) VALUES (0);
4 UPDATE cotisation set
5 cot = (select mean(montantsinistre)*count(*) from tabs.sinistre where
6      ↪ montantsinistre > 80000)/(select count(*) from tabs.sinistre);
7 quit;
```

Finalement, on l'ajoute à travers ce code :

Obs.	cot
1	1160.00

FIGURE 3.3 – Cotisation à ajouter au primes standard

```

1 proc sql;
2 alter table tarif44 add prime_pure float(4);
3 update tarif44 set prime_pure = prime_stand + (select cot from
  ↳ cotisation);
4 alter table tarif45 add prime_pure float(4);
5 update tarif45 set prime_pure = prime_stand + (select cot from
  ↳ cotisation) ;
6 alter table tarif64 add prime_pure float(4);
7 update tarif64 set prime_pure = prime_stand + (select cot from
  ↳ cotisation) ;
8 alter table tarif65 add prime_pure float(4);
9 update tarif65 set prime_pure = prime_stand + (select cot from
  ↳ cotisation) ;
10 quit;

```

Ce qui donne :

Ota	numspolice	exercice	Sexe	Zone	Combustion	PF	exposition	agrV	agrC	nombre	somme	frequence	severite	newzone	newzone	agrCnew	newPF	newPF	logsew	lenul	logSev	ftotd_frequence	ftotd_severite	prime_stand	prime_pure
1	2	2009	M	Fda-Ménets	Essence	[30-35]	0.8505040312	7	28	1	1986.91	1.17577	1986.91	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15838	2397.73	2776.32	3005.32
2	5	2010	M	Rabat-Sale-Kénitra	Essence	[30-35]	0.4889580514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.71968	0	-	1.20616	1234.34	1468.81	2948.81
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4889580502	6	28	2	3625.85	4.16185	1812.83	zone_low_risk	zone_low_risk	[18-35] U [35-60]	PF_low_risk	PF_high_risk	-0.73281	1	7.58270	1.08674	1323.53	1454.22	2814.22
4	9	2012	M	Tanger-Tétouan-Houassine	Essence	[30-35]	0.500243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[25-50]	PF_high_risk	PF_low_risk	-2.79812	0	-	1.28543	580.29	745.85	1955.85
5	24	2008	M	Rabat-Sale-Kénitra	Diesel	[35-10]	0.5687430257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10719	1510.89	1672.84	2832.84

Ota	numspolice	exercice	Sexe	Zone	Combustion	PF	exposition	agrV	agrC	nombre	somme	frequence	severite	newzone	newzone	agrCnew	newPF	newPF	logsew	lenul	logSev	ftotd_frequence	ftotd_severite	prime_stand	prime_pure
1	2	2009	M	Fda-Ménets	Essence	[30-35]	0.8505040312	7	28	1	1986.91	1.17577	1986.91	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15838	2397.72	2776.77	
2	5	2010	M	Rabat-Sale-Kénitra	Essence	[30-35]	0.4889580514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.71968	0	-	1.20616	1234.32	1468.83	
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4889580502	6	28	2	3625.85	4.16185	1812.83	zone_low_risk	zone_low_risk	[18-35] U [35-60]	PF_low_risk	PF_high_risk	-0.73281	1	7.58270	1.08674	1323.65	1454.36	
4	9	2012	M	Tanger-Tétouan-Houassine	Essence	[30-35]	0.500243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[25-50]	PF_high_risk	PF_low_risk	-2.79812	0	-	1.28543	580.08	745.85	
5	24	2008	M	Rabat-Sale-Kénitra	Diesel	[35-10]	0.5687430257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10719	1499.33	1660.04	

Ota	numspolice	exercice	Sexe	Zone	Combustion	PF	exposition	agrV	agrC	nombre	somme	frequence	severite	newzone	newzone	agrCnew	newPF	newPF	logsew	lenul	logSev	ftotd_frequence	ftotd_severite	prime_stand	prime_pure
1	2	2009	M	Fda-Ménets	Essence	[30-35]	0.8505040312	7	28	1	1986.91	1.17577	1986.91	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15873	2397.73	2777.17	
2	5	2010	M	Rabat-Sale-Kénitra	Essence	[30-35]	0.4889580514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.71968	0	-	1.20653	1234.34	1469.27	
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4889580502	6	28	2	3625.85	4.16185	1812.83	zone_low_risk	zone_low_risk	[18-35] U [35-60]	PF_low_risk	PF_high_risk	-0.73281	1	7.58270	1.08674	1323.53	1454.22	
4	9	2012	M	Tanger-Tétouan-Houassine	Essence	[30-35]	0.500243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[25-50]	PF_high_risk	PF_low_risk	-2.79812	0	-	1.28362	580.29	744.99	
5	24	2008	M	Rabat-Sale-Kénitra	Diesel	[35-10]	0.5687430257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10572	1510.89	1670.83	

Ota	numspolice	exercice	Sexe	Zone	Combustion	PF	exposition	agrV	agrC	nombre	somme	frequence	severite	newzone	newzone	agrCnew	newPF	newPF	logsew	lenul	logSev	ftotd_frequence	ftotd_severite	prime_stand	prime_pure
1	2	2009	M	Fda-Ménets	Essence	[30-35]	0.8505040312	7	28	1	1986.91	1.17577	1986.91	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.16193	1	7.58422	1.15873	2397.72	2777.83	
2	5	2010	M	Rabat-Sale-Kénitra	Essence	[30-35]	0.4889580514	10	27	0	0.00	0.00000	0.00	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_low_risk	-0.71968	0	-	1.20653	1234.52	1469.49	
3	6	2010	F	Marrakech-Safi	Diesel	[10-14]	0.4889580502	6	28	2	3625.85	4.16185	1812.83	zone_low_risk	zone_low_risk	[18-35] U [35-60]	PF_low_risk	PF_high_risk	-0.73281	1	7.58270	1.08674	1323.65	1454.36	
4	9	2012	M	Tanger-Tétouan-Houassine	Essence	[30-35]	0.500243414	19	47	0	0.00	0.00000	0.00	zone_low_risk	zone_low_risk	[25-50]	PF_high_risk	PF_low_risk	-2.79812	0	-	1.28362	580.08	744.71	
5	24	2008	M	Rabat-Sale-Kénitra	Diesel	[35-10]	0.5687430257	4	27	1	174.72	1.76447	174.72	zone_high_risk	zone_high_risk	[18-35] U [35-60]	PF_high_risk	PF_high_risk	-0.56785	1	5.16319	1.10572	1499.33	1657.85	

FIGURE 3.4 – Tables des primes pures (même ordre que précédemment)

### 3.iii Comparaison macro des modèles :

On utilise ce code pour calculer les différents charges totaux et les primes totaux pour chaque tarif :

```

1 proc sql;
2 CREATE TABLE amd (
3 charges_totaux float(5) , tarifs44 float(5), tarifs45 float(5) ,
4 ↪ tarifs64 float(5), tarifs65 float(5)
5 );
6 INSERT INTO amd (charges_totaux, tarifs44, tarifs45, tarifs64, tarifs65)
7 ↪ VALUES
8 (0, 0, 0, 0, 0);
9 UPDATE amd set
10 charges_totaux = ( select sum(montantsinistre) from tabs.sinistre),
11 tarifs44 = (select sum(prime_pure) from tarif44),
12 tarifs45 = (select sum(prime_pure) from tarif45),
13 tarifs64 = (select sum(prime_pure) from tarif64),
14 tarifs65 = (select sum(prime_pure) from tarif65);
15 quit;

```

Obs.	charges_totaux	tarifs44	tarifs45	tarifs64	tarifs65
1	417211670.61	549312989.96	549306224.74	549312140.03	549307977.33

**FIGURE 3.5** – Comparaisons des Tarifs

On remarque que toutes les tarifs, assure la solvabilité de l'assureur, cependant il vaut mieux choisir celle le plus proche des charges réelles, donc on va choisir le tarif : TARIF45.

# Chapitre 5

## ANNEXE

1. Code pour le macro qui fait les statistique descriptives :

```
1  %macro plot_stats( cat_var=, quant_var=);
2
3  proc sql;
4  create table amd as
5  select &cat_var, mean(&quant_var) as mean_sin,
6         ↪ sqrt(var(&quant_var))/5 as var_sin
7  from tabs.data
8  group by &cat_var
9  order by mean_sin;
10 quit;
11
12 /* Print table */
13 proc print data=amd;
14 run;
15
16 /* Create bar chart */
17 proc sgplot data=amd;
18 vbar &cat_var / response=mean_sin barwidth=1
19     ↪ fillattrs=(color=CX42bfa5);
20 vbar &cat_var / response=var_sin barwidth=0.5
21     ↪ fillattrs=(color=CX5d6a74);
22 run;
23 %mend;
```

2. Découpage de la zone :

Voici le code de découpage pour le nombre sinistre<sup>1</sup> :

---

1. Dans ce découpage on ai basé sur le tableau des statistique descriptives, on ait choisit les 4 régions le moins générateurs des sinistres par individus

```

1 proc sql;
2     alter table tabs.data add newzone char(1000) ;
3     update tabs.data set newzone = zone;
4     update tabs.data
5     set newzone = "zone_low_risk"
6     where zone in ("Tanger-Tétouan-Hocedma", "Laâyoune-Sakia El
   ↪ Hamra", "Dakhla-Oued Ed Dahab", "Marrakech-Safi");
7
8     update tabs.data
9     set newzone = "zone_high_risk"
10    where newzone <> "zone_low_risk" ;
11 quit;

```

Voici le code de découpage pour la severité<sup>2</sup> :

```

1 proc sql;
2     alter table tabs.data add newszone char(1000) ;
3     update tabs.data set newszone = zone;
4
5     update tabs.data
6     set newszone = "zone_low_risk"
7     where zone in ("Laâyoune-Sakia El
   ↪ Hamra", "Marrakech-Safi", "Dakhla-Oued Ed
   ↪ Dahab", "Souss-Massa", "Tanger-Tétouan-Hocedma",
   ↪ "Drâa-Tafilalet", "BéniMellal-Khénifra",
   ↪ "Casablanca-Settat");
8
9     update tabs.data
10    set newszone = "zone_high_risk"
11    where newszone <> "zone_low_risk" ;
12 quit;

```

### 3. Découpage de Puissance fiscale

Voici le code de découpage pour le nombre sinistre<sup>3</sup> :

```

1 proc sql;
2     alter table tabs.data add newPF char(1000) ;
3     update tabs.data set newPF = PF;

```

2. Dans ce découpage on ai basé sur le tableau des statistique descriptives, on ait choisit les 8 régions générateurs des moindres montants par individus

3. Dans ce découpage on se base sur les graphes ou on voit clairement quelle modalité est plus risquée



```

4      update tabs.data
5      set newPF = "PF_low_risk"
6      where PF in ("[10-14]");
7      update tabs.data
8      set newPF = "PF_high_risk"
9      where newPF <> "PF_low_risk" ;
10 quit;

```

Voici le code de découpage pour la severité<sup>4</sup> :

```

1 proc sql;
2     alter table tabs.data add newsPF char(1000) ;
3     update tabs.data set newPF = PF;
4     update tabs.data
5     set newsPF = "PF_low_risk"
6     where PF in ("[00-08]");
7     update tabs.data
8     set newsPF = "PF_high_risk"
9     where newsPF <> "PF_low_risk" ;
10 quit;

```

#### 4. Decoupage des ages conducteurs Voici le code :

```

1 proc sql;
2     alter table tabs.data add ageCnew char(100) ;
3     update tabs.data set ageCnew ="[18-30] U [56-oo]" where
4     ↪ ageC between 18 and 28;
5     update tabs.data set ageCnew ="[29-50]" where ageC between
6     ↪ 29 and 55;
7     update tabs.data set ageCnew ="[18-30] U [56-oo]" where
8     ↪ ageC >= 56;
9 quit;

```

#### 5. Code démarche pour distribution lognormal

```

1 /* Model tout variable */
2 proc genmod data=tabs.data;
3 Class sexe zone Comubsution PF;
4 Model logsev = sexe zone Comubsution PF ageC ageV/

```

---

4. Dans ce découpage on se base sur les graphes ou on voit clairement quelle modalité est plus risquée

```

5  dist = normal link = identity offset=logexp ;
6  ods output modelfit = LGN1;
7  run;
8
9  /* Model avec decomposition pf */
10 proc genmod data=tabs.data;
11 Class sexe zone Comubsution newsPF;
12 Model logsev = sexe zone Comubsution newsPF ageC ageV/
13 dist = normal link = identity offset=logexp ;
14 ods output modelfit = LGN2;
15 run;
16
17 /* ne se regle pas , on elimine newspf */
18 proc genmod data=tabs.data;
19 Class sexe newzone Comubsution ;
20 Model logsev = sexe newzone Comubsution ageC ageV/
21 dist = normal link = identity offset=logexp ;
22 ods output modelfit = LGN3;
23 run;
24
25 /* On elimine newszone */
26 proc genmod data=tabs.data;
27 Class sexe Comubsution ;
28 Model logsev = sexe Comubsution ageC ageV/
29 dist = normal link = identity offset=logexp ;
30 ods output modelfit = LGN4;
31 run;
32
33 /* On elimine sexe */
34 proc genmod data=tabs.data;
35 Class newzone Comubsution ;
36 Model logsev = newzone Comubsution ageC ageV/
37 dist = normal link = identity offset=logexp ;
38 ods output modelfit = LGN3;
39 run;

```

## 6. Calcule des primes :

```

1  proc sql;
2      create table tarif44 as
3      select p4.*,s4.fitted_severity,
           ↪ s4.fitted_severity*p4.fitted_frequence as prime_stand

```

```
4      from tabs.tarif4 as p4
5      inner join sev4 as s4
6      on p4.numepolice = s4.numepolice and p4.exercice =
       ↪ s4.exercice;
7  quit;
8
9
10 proc sql;
11     create table tarif45 as
12     select p4.*,s5.fitted_severity,
       ↪ s5.fitted_severity*p4.fitted_frequence as prime_stand
13     from tabs.tarif4 as p4
14     inner join sev5 as s5
15     on p4.numepolice = s5.numepolice and p4.exercice =
       ↪ s5.exercice;
16 quit;
17
18
19 proc sql;
20     create table tarif64 as
21     select p6.*,s4.fitted_severity,
       ↪ s4.fitted_severity*p6.fitted_frequence as prime_stand
22     from tabs.tarif6 as p6
23     inner join sev4 as s4
24     on p6.numepolice = s4.numepolice and p6.exercice =
       ↪ s4.exercice;
25 quit;
26
27
28 proc sql;
29     create table tarif65 as
30     select p6.*,s5.fitted_severity,
       ↪ s5.fitted_severity*p6.fitted_frequence as prime_stand
31     from tabs.tarif6 as p6
32     inner join sev5 as s5
33     on p6.numepolice = s5.numepolice and p6.exercice =
       ↪ s5.exercice;
34 quit;
```



Deuxième partie

**PROVISIONNEMENT**



# Chapitre 1

## IMPORTATION DES DONNÉES ET DEFINITIONS DES MODÈLES

### 1 Importation, nettoyage et manipulation

D'abord, on importe la base de données à travers le code :

```
1 # Importation des bases de données :  
2 pIncrém = read.csv("C:/Users/DELL/Documents/R STUDIO/no  
  ↳ life/provision.csv", sep = ";")
```

Annee	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9
2011	1376384	1211168	535883	313790	168142	79972	39235	15030	10865	4068
2012	1576278	1437150	652445	342694	188799	76956	35042	17089	12507	NA
2013	1763277	1540231	678959	364199	177108	78169	47391	25288	NA	NA
2014	1779698	1498531	661401	321434	162578	84581	53449	NA	NA	NA
2015	1843224	1573604	613401	299473	176842	106296	NA	NA	NA	NA
2016	1962385	1520298	581932	347434	238375	NA	NA	NA	NA	NA
2017	2033371	1430541	633500	432257	NA	NA	NA	NA	NA	NA
2018	2072061	1458541	727098	NA	NA	NA	NA	NA	NA	NA
2019	2210754	1517501	NA	NA	NA	NA	NA	NA	NA	NA
2020	2206886	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE 1.1 – Les pertes incrémenté

On fait quelque nettoyage pour qu'il soit adapté aux fonctions qu'on dispose :

```
1 ## Nettoyage  
2 rownames(pIncrém) = pIncrém[,1]  
3 pIncrém = pIncrém[2:length(pIncrém[,1])]
```

De plus on calcule les pertes cumulées à travers ce code :

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9
2011	1376384	1211168	535883	313790	168142	79972	39235	15030	10865	4068
2012	1576278	1437150	652445	342694	188799	76956	35042	17089	12507	NA
2013	1763277	1540231	678959	364199	177108	78169	47391	25288	NA	NA
2014	1779698	1498531	661401	321434	162578	84581	53449	NA	NA	NA
2015	1843224	1573604	613401	299473	176842	106296	NA	NA	NA	NA
2016	1962385	1520298	581932	347434	238375	NA	NA	NA	NA	NA
2017	2033371	1430541	633500	432257	NA	NA	NA	NA	NA	NA
2018	2072061	1458541	727098	NA	NA	NA	NA	NA	NA	NA
2019	2210754	1517501	NA	NA	NA	NA	NA	NA	NA	NA
2020	2206886	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE 1.2 – Les pertes incrementé après nettoyage

```

1 ## Calcul de version cumulé
2 pCumul = cumul(pIncrem)
3 colnames(pCumul) = 1:length(colnames(pCumul))

```

	1	2	3	4	5	6	7	8	9	10
2011	1376384	2587552	3123435	3437225	3605367	3685339	3724574	3739604	3750469	3754537
2012	1576278	3013428	3665873	4008567	4197366	4274322	4309364	4326453	4338960	NA
2013	1763277	3303508	3982467	4346666	4523774	4601943	4649334	4674622	NA	NA
2014	1779698	3278229	3939630	4261064	4423642	4508223	4561672	NA	NA	NA
2015	1843224	3416828	4030229	4329702	4506544	4612840	NA	NA	NA	NA
2016	1962385	3482683	4064615	4412049	4650424	NA	NA	NA	NA	NA
2017	2033371	3463912	4097412	4529669	NA	NA	NA	NA	NA	NA
2018	2072061	3530602	4257700	NA	NA	NA	NA	NA	NA	NA
2019	2210754	3728255	NA	NA	NA	NA	NA	NA	NA	NA
2020	2206886	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE 1.3 – Les pertes cumulées

La modélisation serait fait dans le chapitre suivante, Discutons dans la section suivante les fonctions qu'on va utiliser.

## 2 Fonctions de modélisation

### 2.1 Chaineladder

Voici le code pour faire cette modélisation, il prend en parametre une matrice des pertes cumulées, et il fait sortir un output qui contient, la matrice complét ee par le modèle et d'autre mesure qui fait evaluer la qualité de modèle :

```

1 ChainLadderr<- function(pCumul){
2   n=length(pCumul[,1])
3   L = rep(0,times= n)

```



```

4  Rsq = rep(0,times=n)
5  pertes = data.frame(pCumul)
6
7  for(k in 2:n){
8    n <- length(pCumul[,1])
9    y <- as.numeric(pCumul[1:(n-k+1),k])
10   x <- as.numeric(pCumul[1:(n-k+1),k-1])
11   model <- lm(y ~ 0 + x)
12   newdata <- data.frame(x = pertes[(n-k+2):n,k-1])
13   pertes[(n-k+2):n,k] = predict(model, newdata)
14   L[k]<-coef(model)[1]
15   Rsq[k]<-summary(model)$r.squared
16 }
17 return(list(L=L,Rsq=Rsq,pertes = pertes))
18 }

```

## 2.2 Mack

Voici le code pour faire cette modélisation, il prend en parametre une matrice des pertes cumulées, et il fait sortir un output qui contient, la matrice complétée par le modèle et d'autre mesure qui fait evaluer la qualité de modèle :

```

1  library(ChainLadder)
2  # La fonction est déjà défini dans cette library, on l'adapte à notre
3  ↪ besoin :
4  MackCumul <- function(pCumul){
5    M <- MackChainLadder(pCumul, est.sigma = "Mack")
6    df = as.data.frame(M$FullTriangle)
7    mat <- matrix(df$value, nrow = length(unique(df$origin)), ncol =
8      ↪ length(unique(df$dev)), byrow = TRUE)
9    mat=t(mat)
10   rownames(mat) = unique(df$origin)
11   colnames(mat) = unique(df$dev)
12
13   return(list(pertes = mat,sum_up = summary(M)))
14 }

```

## 2.3 GLM

Voici le code pour faire cette modélisation, il prend en parametre une matrice des pertes cumulées, et il fait sortir un output qui contient, la matrice complétée par le modèle et d'autre mesure qui fait evaluer la qualité de modèle :

```

1 glmCumul <- function(pCumul){
2   df <- expand.grid(x = rownames(pCumul),y=colnames(pCumul))
3   colnames(df) = c("rows","cols")
4   df$cumuls <- pCumul[cbind(df$rows, df$cols)]
5   model = glm(cumuls ~ rows+cols,data = df,family = Gamma(link="log"))
6   summary = summary(model)
7
8   df$predicted <- predict(model, newdata = df[, c("rows", "cols")], type
  ↪ = "response")
9   out <- data.frame(pCumul)
10  out[cbind(df$rows, df$cols)] <- df$predicted
11
12  return(list(pertes = out,sum_up=summary))
13 }
```

## 2.4 Regression Log-normal

Voici le code pour faire cette modélisation, il prend en parametre une matrice des pertes cumulées, et il fait sortir un output qui contient, la matrice complétée par le modèle et d'autre mesure qui fait evaluer la qualité de modèle :

```

1 regCumul <- function(pCumul){
2   df <- expand.grid(x = rownames(pCumul),y=colnames(pCumul))
3   colnames(df) = c("rows","cols")
4   df$cumuls <- pCumul[cbind(df$rows, df$cols)]
5   model = lm(log(cumuls)~rows+cols,data = df)
6   summary = summary(model)
7   std = summary$sigma
8
9   df$predicted <- exp(predict(model, newdata = df[, c("rows", "cols")],
  ↪ type = "response")+std^2/2)
10  out <- data.frame(pCumul)
11  out[cbind(df$rows, df$cols)] <- df$predicted
12
13  return(list(pertes = out,sum_up=summary))
```

```
14 }
```

## 3 Autres fonctions de manipulation

On utilise de plus ces deux fonctions :

### 3.1 Fonction de calcul des pertes cumulées

À partir d'une matrice des pertes incrémentées on peut calculer la matrice des pertes cumulées à travers cette fonction :

```
1 cumul<-function(Increm){
2   n=length(pCumul[,1])
3   Cumul = data.frame(pIncrem)
4   for (i in 2:n) {Cumul[,i] = Cumul[,i]+Cumul[,i-1]}
5   colnames(pCumul) = colnames(Increm)
6   return(Cumul)
7 }
```

### 3.2 Fonction de calcul des reserves

Après avoir modélisé et remplir la matrice des pertes cumulées, on peut calculer les reserves à travers cette fonction :

```
1 reserve <- function(pCumul){
2   tt = as.matrix(pCumul)
3   chargeultime = pCumul[,length(pCumul[,1])]
4   paiements= diag(tt[,10:1])
5   r = chargeultime-paiements
6   return(r)
7 }
```



# Chapitre 2

## MODÉLISATION ET CALCUL DES RÉSERVES

### 1 Modélisation

#### 1.1 ChainLadder

1 `ChainLadderr(pCumul)`

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
2011	1376384	2587552	3123435	3437225	3605367	3685339	3724574	3739604	3750469	3754537
2012	1576278	3013428	3665873	4008567	4197366	4274322	4309364	4326453	4338960	4343666
2013	1763277	3303508	3982467	4346666	4523774	4601943	4649334	4674622	4688165	4693250
2014	1779698	3278229	3939630	4261064	4423642	4508223	4561672	4582531	4595807	4600792
2015	1843224	3416828	4030229	4329702	4506544	4612840	4660219	4681529	4695091	4700184
2016	1962385	3482683	4064615	4412049	4650424	4743380	4792100	4814013	4827959	4833196
2017	2033371	3463912	4097412	4529669	4732326	4826919	4876497	4898796	4912988	4918317
2018	2072061	3530602	4257700	4639743	4847325	4944217	4994999	5017840	5032377	5037835
2019	2210754	3728255	4452468	4851988	5069065	5170389	5223495	5247380	5262582	5268290
2020	2206886	3937711	4702611	5124576	5353849	5460865	5516954	5542182	5558238	5564266

**TABLE 1.1** – Les pertes cumulé prévision ChainLadder

Les  $(\lambda_k)$  du modèle sont exprimées comme suit :

$\lambda$	0.000000	1.784284	1.194250	1.089730	1.044740	1.019989	1.010271	1.004573	1.002897	1.001085
-----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

## 1.2 Mack

```
1 M= MackChainLadder(pCumul, est.sigma = "Mack" )
```

	1	2	3	4	5	6	7	8	9	10
2011	1376384	2587552	3123435	3437225	3605367	3685339	3724574	3739604	3750469	3754537
2012	1576278	3013428	3665873	4008567	4197366	4274322	4309364	4326453	4338960	4343666
2013	1763277	3303508	3982467	4346666	4523774	4601943	4649334	4674622	4688167	4693252
2014	1779698	3278229	3939630	4261064	4423642	4508223	4561672	4582319	4595597	4600581
2015	1843224	3416828	4030229	4329702	4506544	4612840	4660162	4681255	4694820	4699912
2016	1962385	3482683	4064615	4412049	4650424	4743616	4792280	4813971	4827920	4833157
2017	2033371	3463912	4097412	4529669	4732784	4827626	4877152	4899227	4913423	4918753
2018	2072061	3530602	4257700	4640885	4848987	4946158	4996900	5019517	5034062	5039522
2019	2210754	3728255	4455215	4856177	5073932	5175611	5228707	5252374	5267593	5273306
2020	2206886	3958267	4730076	5155775	5386965	5494917	5551289	5576415	5592573	5598639

**TABLE 1.2** – Les pertes cumulé prévision Mack

Ainsi les erreurs d'estimation des réserves sont :

	Latest	Dev.To.Date	Ultimate	IBNR	Mack.S.E	CV(IBNR)
2011	3754537	1.0000000	3754537	0.000	0.000000e+00	NaN
2012	4338960	0.9989165	4343666	4706.315	7.439965e-01	0.0001581
2013	4674622	0.9960304	4693252	18630.153	5.612735e+01	0.0030127
2014	4561672	0.9915425	4600581	38909.321	4.376441e+03	0.1124780
2015	4612840	0.9814737	4699912	87071.896	8.926831e+03	0.1025225
2016	4650424	0.9621919	4833157	182732.708	1.625929e+04	0.0889785
2017	4529669	0.9208979	4918753	389083.668	3.428735e+04	0.0881233
2018	4257700	0.8448619	5039522	781822.111	6.211308e+04	0.0794466
2019	3728255	0.7070052	5273306	1545051.249	9.910553e+04	0.0641438
2020	2206886	0.3941826	5598639	3391752.891	2.855638e+05	0.0841936

**TABLE 1.3** – Erreurs d'estimation des réserves

## 1.3 GLM

1 `glmCumul(pCumul)`

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
2011	1441344	2595108	3115633	3404397	3564981	3646082	3695302	3723856	3744523	3754537
2012	1672811	3011860	3615977	3951113	4137486	4231611	4288736	4321875	4345861	4357483
2013	1818964	3275005	3931903	4296320	4498976	4601325	4663440	4699475	4725556	4738194
2014	1796709	3234936	3883797	4243755	4443932	4545028	4606384	4641977	4667740	4680223
2015	1847341	3326097	3993243	4363346	4569163	4673108	4736193	4772790	4799278	4812113
2016	1905042	3429987	4117971	4499633	4711879	4819071	4884126	4921866	4949182	4962418
2017	1942636	3497674	4199235	4588429	4804864	4914171	4980510	5018994	5046850	5060347
2018	2000889	3602558	4325157	4726021	4948946	5061531	5129859	5169498	5198188	5212090
2019	2140728	3854335	4627434	5056315	5294819	5415273	5488376	5530785	5561481	5576354
2020	2206886	3973450	4770442	5212577	5458452	5582628	5657991	5701710	5733354	5748687

**TABLE 1.4** – Les pertes cumulé prévision GLM

Ainsi les coeficient sont exprimé comme suite :

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	14.1810865	0.0110292	1285.77759	0
rows2012	0.1489297	0.0107652	13.83434	0
rows2013	0.2326911	0.0112585	20.66797	0
rows2014	0.2203808	0.0117971	18.68094	0
rows2015	0.2481714	0.0124375	19.95341	0
rows2016	0.2789281	0.0132496	21.05177	0
rows2017	0.2984700	0.0143518	20.79672	0
rows2018	0.3280159	0.0159888	20.51539	0
rows2019	0.3955701	0.0187995	21.04147	0
rows2020	0.4260066	0.0253604	16.79813	0
cols2	0.5880523	0.0107652	54.62521	0
cols3	0.7708565	0.0112585	68.46863	0
cols4	0.8594918	0.0117971	72.85624	0
cols5	0.9055828	0.0124375	72.81045	0
cols6	0.9280772	0.0132496	70.04554	0
cols7	0.9414864	0.0143518	65.60066	0
cols8	0.9491837	0.0159888	59.36564	0
cols9	0.9547183	0.0187995	50.78411	0
cols10	0.9573891	0.0253604	37.75140	0

**TABLE 1.5** – Coeficients de modèle GLM

## 1.4 regression log-normal

```
1 regCumul(pCumul)
```

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
2011	1440763	2595370	3116063	3404952	3565533	3646711	3696074	3724827	3745677	3755524
2012	1671982	3011885	3616140	3951391	4137743	4231948	4289233	4322600	4346797	4358225
2013	1818369	3275584	3932744	4297348	4500015	4602468	4664769	4701057	4727372	4739801
2014	1796234	3235711	3884870	4245036	4445236	4546442	4607984	4643831	4669826	4682103
2015	1846791	3326784	3994215	4364518	4570353	4674407	4737682	4774538	4801264	4813886
2016	1904364	3430496	4118734	4500581	4712833	4820132	4885379	4923384	4950943	4963959
2017	1941643	3497649	4199359	4588681	4805088	4914487	4981011	5019760	5047858	5061130
2018	2000090	3602934	4325768	4726809	4949730	5062422	5130948	5170864	5199808	5213479
2019	2139606	3854257	4627512	5056527	5294998	5415551	5488858	5531557	5562521	5577145
2020	2207466	3976500	4774279	5216902	5462936	5587313	5662944	5706998	5738944	5754032

TABLE 1.6 – Les pertes cumulé prévision regression log normal

Ainsi les coeficient sont exprimé comme suite :

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	14.1804202	0.0110758	1280.30914	0
rows2012	0.1488369	0.0108107	13.76756	0
rows2013	0.2327672	0.0113061	20.58777	0
rows2014	0.2205194	0.0118469	18.61407	0
rows2015	0.2482768	0.0124901	19.87793	0
rows2016	0.2789757	0.0133056	20.96680	0
rows2017	0.2983618	0.0144124	20.70173	0
rows2018	0.3280194	0.0160563	20.42932	0
rows2019	0.3954491	0.0188790	20.94655	0
rows2020	0.4266728	0.0254675	16.75363	0
cols2	0.5885565	0.0108107	54.44208	0
cols3	0.7713976	0.0113061	68.22849	0
cols4	0.8600583	0.0118469	72.59760	0
cols5	0.9061410	0.0124901	72.54888	0
cols6	0.9286530	0.0133056	69.79419	0
cols7	0.9420985	0.0144124	65.36719	0
cols8	0.9498477	0.0160563	59.15729	0
cols9	0.9554297	0.0188790	50.60819	0
cols10	0.9580553	0.0254675	37.61877	0

TABLE 1.7 – Coeficient de modèle log normal

## 2 Calcul des reserves

On calcul dans une table recupilative :

```
1 reserves = data.frame(
2   CHLadder = reserve(CH$pertes),
3   Mack = reserve(CH$pertes),
```



```

4   GLM = reserve(GLM$pertes),
5   LogNorm = reserve(LOGN$pertes))
6 rownames(reserves) = rownames(pCumul)

```

On obtient :

	CHLadder	Mack	GLM	LogNorm
2011	0.000	0.000	0.00	0.00
2012	4706.315	4706.315	11622.30	11427.99
2013	18627.699	18630.153	38719.64	38743.31
2014	39119.728	38909.321	73839.37	74118.63
2015	87344.072	87071.896	139004.81	139479.02
2016	182771.705	182732.708	250538.48	251125.88
2017	388647.694	389083.668	471917.39	472448.56
2018	780135.213	781822.111	886933.23	887711.18
2019	1540035.495	1545051.249	1722019.17	1722888.25
2020	3357380.473	3391752.891	3541801.35	3546565.24

TABLE 2.1 – Les reserves selon les differentes modèles

### 3 Comparaison

On compare graphiquement à travers ce code :

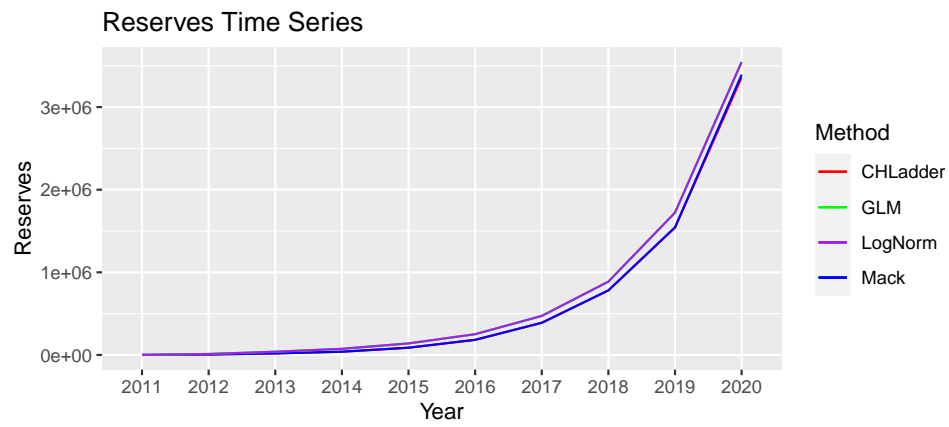
```

1 library(ggplot2)
2 # Create a time series plot for each methode :
3 ggplot(data = reserves, aes(x = rownames(reserves))) +
4   geom_line(aes(y = CHLadder, group = 1, color = "CHLadder")) +
5   geom_line(aes(y = Mack, group = 2, color = "Mack")) +
6   geom_line(aes(y = GLM, group = 3, color = "GLM")) +
7   geom_line(aes(y = LogNorm, group = 4, color = "LogNorm")) +
8   labs(title = "Reserves Time Series",
9         x = "Year",
10        y = "Reserves",
11        color = "Method") +
12   scale_color_manual(values = c("CHLadder" = "red",
13                                "Mack" = "blue",
14                                "GLM" = "green",

```

15

```
"LogNorm" = "purple"))
```



**FIGURE 3.1** – Comparaison des provision selon les differents modèles