



Mémoire de fin d'Etudes

Thème

La tarification en assurance automobile

Vers l'hypothèse de la libéralisation des tarifs

Présenté et soutenu par :

Soumaya Ben Boubaker

Encadré par :

Mr. Farouk Kriaa

Etudiant(e) parrainé(e) par :

GAT ASSURANCES

REMERCIEMENTS

Au terme de ce travail, je voudrais exprimer mon affection, mon respect et ma profonde gratitude à l'Institut de Financement du Développement du Maghreb Arabe (IFID) ainsi que son corps professoral et administratif.

Mes premiers remerciements s'adressent à mon encadrant académique, Monsieur Farouk Kriaa pour ses précieux conseils, ses encouragements et sa disponibilité tout au long de la réalisation de ce mémoire.

Mes remerciements les plus distingués s'adressent à Monsieur Mohamed Dkhili, Président-Directeur Général du «GAT Assurances» ainsi que mon encadrant professionnel Monsieur Abdelaziz Abidi pour ses conseils judicieux qui ont assuré le bon déroulement de ce travail.

J'adresse mes sincères remerciements pour toutes les personnes qui m'ont reçu au sein de mon entreprise de parrainage «GAT Assurances» et plus particulièrement ceux de la direction Automobile.

Je remercie également toute l'équipe de la direction Risk Management et Actuariat qui m'ont aidé par leur savoir et savoir-faire, en particulier Monsieur Badi Draoui.

Enfin, sans oublier les membres du jury qui m'ont honoré d'avoir accepté d'évaluer ce travail.

SOMMAIRE

REMERCIEMENTS

SOMMAIRE

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABREVIATIONS

INTRODUCTION GENERALE 1

CHAPITRE PREMIER : LA TARIFICATION EN ASSURANCE AUTOMOBILE : UNE
ANALYSE THEORIQUE 4

INTRODUCTION 4

Section 1 : La tarification en assurance automobile 4

1.1. Le calcul de la prime d'assurance 4

1.2. Typologie de tarification en assurance automobile 7

1.3. Le principe de la mutualisation et la segmentation 11

Section 2: Les innovations et les opportunités possibles en assurance automobile 14

2.1. L'utilisation des variables innovantes 15

2.2. Une segmentation illimitée 16

2.3. Le concept d'Usage Based Insurance (UBI) 17

Section 3 : Les travaux empiriques sur les modèles de tarification 19

3.1. Les outils statistiques 19

3.2. Modélisation du risque en assurance automobile 25

3.3. Les travaux empiriques relatifs à la tarification en assurance automobile 26

CONCLUSION..... 28

CHAPITRE DEUX : MODELISATION DE LA FREQUENCE ET DU COÛT MOYEN DES
SINISTRES (APPLICATION EMPIRIQUE : CAS DU GAT ASSURANCES) 29

INTRODUCTION 29

Section1 : Situation actuelle du marché tunisien d'assurances automobile 29

| | |
|---|----|
| 1.1. Le marché d'assurance en Tunisie : focus sur la branche automobile | 29 |
| 1.2. La tarification de la RC automobile en Tunisie | 31 |
| 1.3. Les enjeux réglementaires et les perspectives futures | 34 |
| Section 2 : GAT ASSURANCES et les chiffres clés | 36 |
| 2.1 Présentation de la GAT Assurances | 36 |
| 2.2 Positionnement de la GAT Assurances | 37 |
| 2.3. Les principaux indicateurs de la GAT Assurances..... | 37 |
| 2.4. Focus sur la branche automobile | 38 |
| Section 3 : Traitement préliminaire des données de l'étude empirique | 39 |
| 3.1. La méthodologie appliquée | 39 |
| 3.2 Présentation de la base de données | 40 |
| 3.3 Traitement des variables | 42 |
| 3.4 Analyse univariée et bivariée des variables | 48 |
| Section 4 : Modélisation de la fréquence et du coût moyen..... | 53 |
| 4.1. Élimination des sinistres graves | 53 |
| 4.2. La modélisation de la fréquence..... | 55 |
| 4.3. La modélisation du coût moyen | 63 |
| 4.4. Synthèse des résultats | 68 |
| CONCLUSION..... | 69 |
| CONCLUSION GENERALE | 71 |
| BIBLIOGRAPHIE..... | 73 |
| LISTE DES ANNEXES..... | 77 |
| LES ANNEXES..... | 78 |
| LES CODES UTILISES SOUS R..... | 84 |

LISTE DES TABLEAUX

| | |
|--|----|
| Tableau 1 : Les autres variables de la tarification a priori | 9 |
| Tableau 2 : Les primes proposées par les deux assureurs | 13 |
| Tableau 3 : Les avantages du modèle PHYD | 19 |
| Tableau 4: Les lois de la famille exponentielle utilisée en GLM | 22 |
| Tableau 5 : Les fonctions de liens associés aux lois usuelles de la famille exponentielle..... | 24 |
| Tableau 6: Fiche de présentation de la GAT ASSURANCES | 36 |
| Tableau 7 : Classement des trois premières compagnies d'assurances selon le chiffre d'affaires en M.D | 37 |
| Tableau 8: Les indicateurs clés de GAT Assurances en M.D | 38 |
| Tableau 9 : Les indicateurs clés de la branche automobile de la GAT Assurances en M.D..... | 38 |
| Tableau 10 : Liste des variables du fichier production | 41 |
| Tableau 11 : Liste préliminaire des variables tarifaires | 42 |
| Tableau 12 : Classement d'âge de véhicule | 44 |
| Tableau 13 : Classement final d'âge du véhicule | 45 |
| Tableau 14 : Classement final des variables quantitatives..... | 45 |
| Tableau 15 : Les modalités finales de la variable marque de véhicule..... | 47 |
| Tableau 16 : Présentation des variables avant et après traitement..... | 48 |
| Tableau 17 : Analyse univariée des variables | 49 |
| Tableau 18 : Test du Khi-deux d'indépendance (p-value)..... | 51 |
| Tableau 19 : Test de Kruskal..... | 53 |
| Tableau 20 : Statistiques descriptives des coûts unitaires de sinistres..... | 54 |
| Tableau 21 : Nuage de points des coûts unitaires de sinistres avant et après élimination des sinistres graves..... | 55 |
| Tableau 22 : Espérance et variance de la variable nombre de sinistres | 57 |

LISTE DES FIGURES

| | |
|---|----|
| Figure 1 : Décomposition de la prime totale | 5 |
| Figure 2 : Les effets et objectifs de la segmentation..... | 12 |
| Figure 3 : Les flux d'assurés provoqués par l'utilisation ou non d'une segmentation | 14 |
| Figure 4 : Les variables tarifaires innovantes..... | 15 |
| Figure 5 : Fonctionnement du modèle Pay As You Drive | 18 |
| Figure 6 : La composition du modèle GLM..... | 24 |
| Figure 7 : Répartition des primes émises totales par branches d'assurance..... | 30 |
| Figure 8 : Répartition des indemnisations réglées par branches d'assurance | 31 |
| Figure 9 : Statistiques descriptives de la variable âge du véhicule..... | 43 |
| Figure 10 : Variance intra en fonction du nombre de cluster K | 43 |
| Figure 11 : Partitionnement d'âge des véhicules par la méthode de K-means..... | 44 |
| Figure 12 : Dendrogramme présentant le découpage des marques de véhicules | 46 |
| Figure 13 : Découpage en 4 classes de la marque de véhicule pour la fréquence | 47 |
| Figure 14 : Test du Khi-deux d'indépendance entre la variable M et BM | 51 |
| Figure 15 : Test de Kruskal pour la variable BM | 52 |
| Figure 16 : Ajustement des fréquences selon la loi de Poisson..... | 57 |
| Figure 17 : Test du Chi-2 d'adéquation pour la loi de Poisson | 58 |
| Figure 18 : Ajustement des fréquences selon la loi Binomiale Négative | 59 |
| Figure 19 : Test du Chi-2 d'adéquation pour la loi Binomiale Négative..... | 59 |
| Figure 20 : Estimation des coefficients de régression du modèle de fréquence..... | 60 |
| Figure 21 : Procédure de sélection des variables du modèle de fréquence | 61 |
| Figure 22 : Calcul de l'AIC et du BIC selon la loi du Gamma et Log-Normale..... | 64 |
| Figure 23 : Ajustement des coûts moyens par une loi Gamma | 64 |
| Figure 24 : Ajustement des coûts moyens par une loi Log-Normale | 65 |
| Figure 25 : Estimation des coefficients de régression du modèle de coût moyen..... | 66 |
| Figure 26 : Procédure de sélection des variables du modèle de coût moyen | 66 |

LISTE DES ABREVIATIONS

| | |
|-------|---|
| AIC | Critère d'Information d'AKAIKE |
| BIC | Bayesian Information Criterion |
| CAH | Classification Hiérarchique Ascendante |
| CGA | Comité Général des Assurances |
| CRBM | Centrale des Risques Bonus-Malus |
| FTUSA | Fédération Tunisienne des Sociétés d'Assurances |
| GAT | Groupe des assurances en Tunisie |
| GLM | Generalized Linear Model |
| GPS | Global Positionning System |
| MTPL | Motor Third Party Liability |
| OMS | Organisation Mondiale de Santé |
| PAYD | Pay As You Drive |
| PHYD | Pay How You Drive |
| PIB | Produit Intérieur Brut |
| RC | Responsabilité civile |
| SBM | Système Bonus-Malus |
| UBI | Usage Based Insurance |
| UE | Union Européenne |

INTRODUCTION GENERALE

Le monde dans lequel nous vivons est plein d'incertitudes et de risques. L'assurance est un produit financier qui réduit ou élimine le coût des pertes causées par différents types de risques.

L'activité d'assurance est une activité très particulière qui se caractérise par un cycle de production inversé dont lequel l'assureur s'engage à couvrir des risques pour lesquels il ne connaît pas avec certitude le coût et la date de réalisation en contrepartie d'une prime fixée à la souscription du contrat. (Paulin, 2007)

L'assurance automobile est la première catégorie d'assurance avec laquelle le grand public est familiarisé et l'assurance la plus dominante en termes de parts de marché. D'après le rapport 3/2021 de «Swiss Re Institute», les primes d'assurance automobile dans le monde représentent environ 30% de l'ensemble des primes non vies.

La croissance exponentielle du nombre des véhicules dans le monde a été suivie par un accroissement des accidents de la route. Selon l'Organisation Mondiale de la Santé (OMS), ces accidents entraînent chaque année environ 1,3 million de décès et coûtent à leur pays environ 3 % de leur produit intérieur brut (PIB).

De nos jours, l'importance de la branche automobile rend ce marché hyperconcurrentiel et renvoie la nécessité de posséder à une tarification précise et adaptée au portefeuille de chaque assureur. Cette tarification doit s'appuyer sur une segmentation poussée toute en étant pertinente et opérationnelle. Ainsi, une même segmentation ne doit pas rester invariante dans le temps et il est nécessaire d'effectuer un suivi régulier. (Gonnet, 2010)

Le contrat d'assurance automobile est basé principalement sur une garantie dite « Responsabilité Civile». Cette garantie est obligatoire par la loi pour tout véhicule en circulation. Elle permet de garantir des indemnisations aux victimes des accidents. La tarification de cette garantie représente depuis toujours une problématique récurrente dans toutes les compagnies d'assurance et elle est différente d'un pays à un autre.

Selon une étude réalisée par la Banque Mondiale en 2009, les primes de la RC automobile dans les pays développés sont libres et obéissent seulement à la forte concurrence. La libéralisation du marché de l'assurance automobile indique une liberté de prix, ce qui signifie l'élimination de l'intervention de l'État dans la fixation des tarifs et d'adapter pour chaque compagnie son propre

système de tarification en se basant sur un appui statistique et une information chiffrée précise relative à son portefeuille. Alors que, dans la plupart des pays émergents la tarification de la RC automobile est fixée par l'État. En général, cette garantie est réglementée dans le but d'offrir une assurance accessible avec des prix raisonnables.

Selon la même étude, cette réglementation peut être considérée comme un frein pour la durabilité et le développement des compagnies d'assurances surtout si ce tarif n'est pas mise à jour avec l'évolution de la sinistralité et qui peut causer un déficit de la branche.

La Tunisie n'a pas fait l'exception à cette réalité puisque son tarif est aussi réglementé. Actuellement, cette garantie est structurellement déficitaire en raison d'un tarif obligatoire inférieur au coût réel. Cette situation critique de la garantie RC automobile est évoquée par une étude réalisée le CGA en collaboration avec la Banque Mondiale en 2015 et qui nuit les assurés et les compagnies d'assurances. Le système de tarification de la RC automobile utilisé en Tunisie, se base sur l'usage du véhicule, la puissance fiscale et un système bonus-malus.

De nos jours, suite aux réclamations des assureurs et aux recommandations des institutions internationales comme la Banque Mondiale, le débat sur la libéralisation et la tarification de la garantie RC automobile attire aussi bien l'attention des chercheurs que celle des praticiens à cause de la tarification actuelle et obligatoire qui n'est pas à jour et qui ne prend pas en considération d'autres critères de tarification qui peuvent influencer la sinistralité.

Le présent projet de fin études, s'intéresse à cette problématique en faisant le lien entre la tarification de la RC automobile et les différents facteurs qui peuvent intervenir. De ce fait, notre problématique consiste à répondre essentiellement aux questions suivantes :

- ✓ **Comment présenter un modèle de tarification pertinent de la garantie RC automobile en adéquation avec la sinistralité constatée des assurés ?**
- ✓ **Quels sont les critères déterminants susceptibles d'influencer la prime pure de la garantie RC automobile en distinguant entre les deux modèles de fréquence et de coût moyen?**
- ✓ **Comment déterminer la prime pure optimale de la garantie RC automobile à partir du modèle ?**

Pour répondre à cette problématique, nous avons comme objectif central d'analyser le système de tarification de l'assurance RC automobile tunisien à l'aide des modèles de tarification appliqués

par des actuaires partout dans le monde. Ainsi, les conclusions tirées des résultats numériques obtenus ouvrent le débat sur la libéralisation et l'innovation de cette branche.

À partir des données relatives aux contrats d'assurances automobiles, nous allons construire un modèle de fréquence et de coût moyen des sinistres en fonction d'un certain nombre de variables en utilisant le modèle linéaire généralisé (GLM). Pour ce faire, nous allons proposer un tarif répondant aux diverses classes de risque existant dans le portefeuille automobile pour l'usage utilitaire de la compagnie de parrainage GAT Assurances durant l'année 2019.

Plus précisément, dans le cadre de ce mémoire, nous visons les objectifs suivants :

- ✓ **Présenter une synthèse théorique sur la tarification en assurance automobile**
- ✓ **Mener une analyse empirique sur des données tunisiennes permettant de tarifier la garantie RC automobile à travers des différents critères observables.**

Dans le but d'atteindre ces objectifs, il convient de répartir ce mémoire en deux chapitres. Un premier chapitre qui présente des généralités sur les concepts clés ainsi que le fondement théorique de la tarification en assurance automobile. Il expose également les innovations, les outils d'analyse et les résultats des travaux empiriques sur ce thème. Un deuxième chapitre est consacré à la méthodologie adaptée dans la partie empirique, la modélisation et la synthèse des résultats obtenus.

CHAPITRE PREMIER : LA TARIFICATION EN ASSURANCE AUTOMOBILE : UNE ANALYSE THEORIQUE

INTRODUCTION

L'activité d'assurance de manière générale se distingue par rapport aux autres activités économiques par des fondements et des mécanismes théoriques très spécifiques pour chaque branche.

Pour cela, il est nécessaire de mettre au point un certain nombre de concepts de base permettant de comprendre l'objet et la finalité de ce travail. Dans le souhait d'atteindre ce but et pour bien mener cette étude, le choix a été de commencer par un chapitre premier qui présente une analyse théorique des concepts clés sur l'assurance automobile dont lequel, nous nous intéressons à une notion très importante au sein d'une compagnie d'assurance à savoir : la tarification.

De ce fait, ce chapitre sera composé de trois sections. La première section traite le principe de tarification et la segmentation en assurance automobile. La deuxième section aborde les innovations et les opportunités possibles de la tarification en assurance automobile. Une troisième section traite les travaux empiriques et les outils statistiques permettant la tarification en assurance automobile.

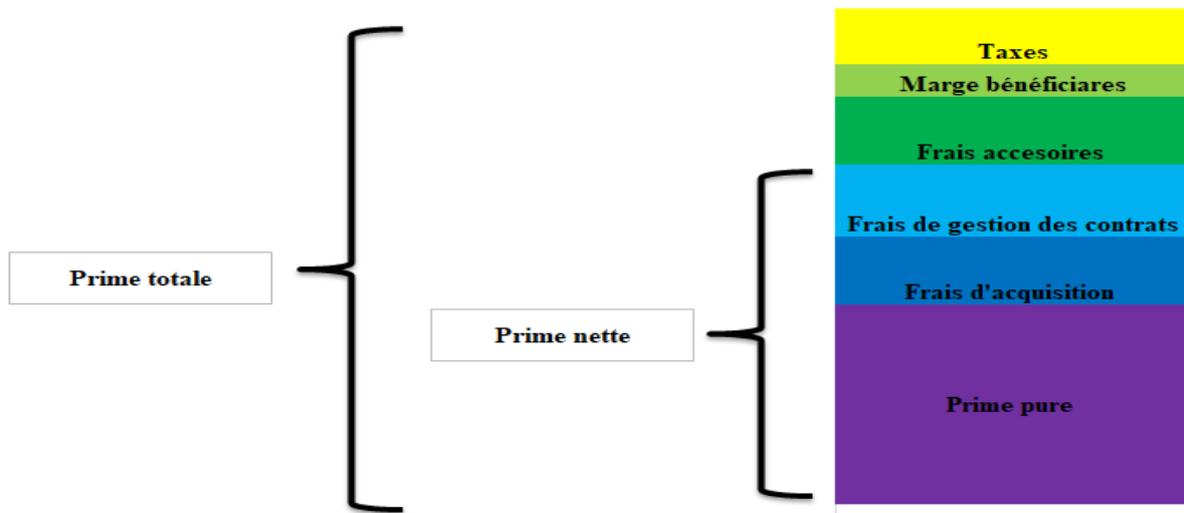
Section 1 : La tarification en assurance automobile

La première section de ce chapitre a pour objet de développer les concepts fondamentaux en assurance automobile, à savoir le calcul de prime, la tarification et la segmentation.

1.1. Le calcul de la prime d'assurance

D'une manière générale, la prime d'assurance est constituée de plusieurs éléments, la prime pure qui permet de couvrir l'engagement de l'assureur envers l'assuré, les frais de chargement et accessoires, la marge bénéficiaire et les taxes.

Figure 1 : Décomposition de la prime totale



Source : Paulin, 2007

1.1.1. La prime pure

Elle est appelée aussi la prime de risque, c'est-à-dire le prix du risque calculé par les actuaires. L'actuaire procède à une analyse la plus fine possible afin de la déterminer. La prime pure correspond à un montant que l'assureur doit avoir à sa disposition dont le but d'indemniser les assurés suite à la survenance des sinistres. (Paulin, 2007).

Techniquement, le montant de la prime pure ne permet pas à l'assureur de dégager ni un bénéfice ni un déficit, la totalité des primes pures doivent faire en sorte que l'assureur soit en mesure de faire face à ses engagements. (Charpentier et Denuit, 2009).

D'un point de vue mathématique, le calcul de la prime pure, correspondant à l'espérance mathématique du coût annuel des sinistres déclarés à l'assureur. Cette prime pure se décompose toutefois en deux facteurs de sinistralité : la fréquence et le coût moyen, en notant simplement (Noël, 2007).

Notations :

N : le nombre de risques

n : le nombre de sinistres pour un exercice donné

X_i: le coût du $i^{\text{ème}}$ sinistre

C : le coût moyen d'un sinistre

S : le coût total des sinistres à la charge de l'assurance

Nous avons alors :

$$E[S] = E[N] \times E[X]$$

Prime pure = fréquence x coût moyen

Sous l'hypothèse que les montants de sinistres X_1, X_2, \dots sont indépendants et suivent la même loi et indépendants de leur nombre N .

À ce niveau, il faut bien comprendre et savoir calculer ses différentes notions.

- **Le nombre de sinistres**

Cette variable représente le nombre de sinistres déclarés par l'assuré. Cette information doit être connue avec précision pour qu'elle soit extrêmement fiable pour les analyses statistiques.

- **La durée d'exposition au risque**

C'est le nombre de jours ou le contrat est exposé à un risque. Concrètement, elle représente la durée qui découle entre la date d'effet et la date de fin du contrat. Cette variable permet de mesurer l'exposition d'un portefeuille automobile à un risque donné sur une période donnée. En assurance automobile la durée d'exposition est généralement annuelle.

- **La fréquence**

La fréquence correspond au rapport sur une période donnée, entre le nombre de sinistres et la durée d'exposition. Elle permet de mesurer l'occurrence d'un sinistre. Ce ratio est produit par année de la période d'observation et par nature de sinistre. (Ray 2008)

- **Le coût total**

Cette variable correspond au coût total mis à la charge de la compagnie d'assurance. L'étude des coûts doit tenir compte du nombre de sinistres engendrés.

- **Le coût moyen**

C'est le rapport, pour une période donnée entre le coût total des sinistres nets de recours et de franchise, et le nombre des sinistres. Ce ratio doit être aussi produit par année d'assurance et par nature de sinistre. (Ray 2008)

1.1.2. La prime nette

Elle est appelée aussi la prime commerciale. La prime nette est égale à la prime pure en ajoutant les chargements qui sont constitués des frais d'acquisition et les frais des contrats. (Paulin, 2007).

1.1.3. La prime totale

Elle correspond à la prime nette en ajoutant les frais accessoires et les taxes.

La principale marge de manœuvre des compagnies d'assurances est d'assurer sa rentabilité sur un produit à travers une marge bénéficiaire. Elle doit être importante pour que l'assureur soit rentable sans qu'elle soit trop importante afin de ne pas pousser les clients à la concurrence.

1.2. Typologie de tarification en assurance automobile

Bressand (1993), a défini la tarification automobile comme un système qui combine l'affectation de l'assuré dans une grille de classification a priori selon les caractéristiques objectives observables, avec un mécanisme de modulation de la prime selon son antécédent au volant.

Selon Riad 2015, la tarification en assurance automobile est un processus instantané d'évaluation des risques où la prime doit être proportionnelle au risque encouru par les assurés.

L'effet de conduire une voiture ne dépend pas forcément de critères techniques observables lors de la conclusion du contrat, tels que la puissance du véhicule et l'usage, mais également d'autres facteurs difficilement observables comme la consommation d'alcool pour le conducteur, ou la connaissance du code de la route par exemple. Par conséquent, nous pouvons avoir des assurés identiques du point de vue de l'assureur et qui peuvent très bien s'avérer des risques complètement incomparables.

C'est pour cette raison que, d'une manière générale, la tarification de l'assurance automobile se fait par deux méthodes. Une tarification a priori permettant de fixer le prix dès la souscription du contrat en fonction des caractéristiques de l'assuré et de son véhicule. L'autre type de tarification est la tarification a posteriori qui est basée sur la connaissance ultérieure des antécédents en matière de sinistre permettant de corriger la prime a priori.

Ces deux méthodes de tarification utilisées sont complémentaires. Lorsque le tarif a priori est mis en place, nous pouvons l'affiner à l'aide de la tarification a posteriori et ce avec le biais du système Bonus-Malus par exemple.

1.2.1. La tarification a priori

Généralement, la tarification a priori est basée sur la segmentation du portefeuille en classes de risques homogènes. Cette méthode de tarification se repose sur les caractéristiques de l'assuré (le sexe, l'âge,...) et de son véhicule (l'ancienneté du véhicule, la puissance fiscale du véhicule...). Elle consiste à prévoir le nombre des sinistres en fonction de ses caractéristiques. Le but de cette tarification est de créer des classes de risque dans lesquelles les assurés paient la même prime au moment de la souscription du contrat.

La tarification a priori a été définie, par Charpentier et Denuit (2009), comme l'idée de séparer les contrats et les assurés en plusieurs catégories, de façon qu'à l'intérieur d'une catégorie les risques puissent être considérés comme équivalents.

Selon Bressand (1993), il existe des variables tarifaires qui sont utilisées dans certains pays, tandis qu'elles sont absentes dans d'autres et certains d'entre elles étant parfois interdites : il s'agit principalement de caractéristiques individuelles telles que le sexe, la situation patrimoniale et la prise en compte de l'état civil.

Cette tarification consiste à déterminer une prime de risque qui s'appuie sur l'observation de certaines variables relatives à l'assuré et son véhicule et qui influencent réellement sur le risque automobile. (Riad, 2015)

Les principales caractéristiques observables sont les suivantes (la liste n'est pas exhaustive):

❖ **Les caractéristiques liées à l'assuré**

- **Le sexe du conducteur** : d'un point de vue des assureurs les degrés de prudence entre les deux sexes ne sont plus les mêmes et donc un nombre différent des accidents, de gravité et des coûts.
- **L'âge du conducteur** : l'âge permet aux assureurs de connaître l'influence de l'âge sur la survenance des sinistres. Une personne moins âgée est potentiellement plus risquée par conséquent plus sinistrée.
- **La catégorie socioprofessionnelle** : chaque catégorie socioprofessionnelle représente un niveau de risque et par conséquent un niveau de prime correspond. La catégorie socioprofessionnelle n'est donc qu'un regroupement selon la profession (fonctionnaire, retraité, profession libérale...) et qui nous permet de savoir l'utilisation effective du véhicule. Ainsi, un fonctionnaire n'utilise pas son véhicule avec un usage intensif à l'inverse des transporteurs et des chauffeurs de qui ont un usage intensif, ce qui fait augmenter le risque d'avoir un accident.
- **Le lieu de résidence** : certaines zones géographiques notamment les zones urbaines sont considérées comme des zones plus dangereuses que d'autres. En effet, la prime pure d'un assuré hors de la ville doit être moins chère car il y aura moins de chance d'avoir un accident.

❖ **Les caractéristiques liées au véhicule:**

Ces variables dépendent de plusieurs paramètres liés au véhicule et à son utilisation, qui sont :

- **La puissance fiscale du véhicule**: elle est exprimée en nombre de chevaux ou le nombre de places en ce qui concerne les véhicules destinés au transport public de voyageurs. Un effet, un véhicule puissant il a plus de risques d'être acteur d'accident, il est logique donc que sa prime soit plus élevée.
- **L'âge du véhicule** : c'est un critère qui permet de voir l'impact de l'état de la voiture sur la survenance des accidents.

- **L'usage du véhicule** : représente l'activité dans laquelle le véhicule assuré est utilisé. Plus le véhicule circule plus il est exposé aux accidents, donc plus risqué d'être impliqué dans un accident.
- **D'autres variables** : elles sont représentées dans le tableau ci-suitant.

Tableau 1 : Les autres variables de la tarification a priori

| Caractéristiques du véhicule | Caractéristiques de l'assuré |
|------------------------------|------------------------------------|
| Poids du véhicule | Consommation de drogue, d'alcool |
| La marque | L'ancienneté du permis de conduire |
| Le modèle | La situation de famille |
| La valeur | Autres conducteurs... |
| Énergie utilisée | |
| Présence d'alarme... | |

Bressand C. (1993)

Néanmoins, ces critères ne sont pas universels et leurs adoptions varient d'un pays à un autre. Dans l'annexe n°1, nous présenterons les différents critères de tarification adoptés par un échantillon de pays.

Cependant, malgré la multitude de critères utilisés pour cette la tarification, il est possible d'observer une certaine variabilité au sein d'une même classe de risque. Le problème de cette tarification est qu'elle ne représente pas parfois le mérite de chaque assuré en fonction de ses sinistralités donc elle n'est pas équitable contrairement à la tarification a posteriori qui se fonde sur la base des actes effectifs de l'assuré. D'où l'utilité de la tarification a posteriori.

1.2.2. La tarification a posteriori

Malheureusement, dans la tarification a priori plusieurs caractéristiques des assurés ne peuvent pas être prises en considération soit parce qu'elles ne sont pas observables, ou parce qu'elles sont difficilement mesurables. Nous pouvons citer, l'agressivité au volant, la conduite sous l'effet de l'alcool et la qualité de réflexions et les réactions lors de la conduite...

Charpentier et Denuit (2009), ont bien expliqué que ces variables cachées peuvent avoir des impacts significatifs sur le nombre de réclamations des assurés. Par conséquent, le portefeuille après la classification dans la tarification a priori reste encore hétérogène malgré l'utilisation de plusieurs variables.

D'après Ghali (2001) la tarification a posteriori est une technique qui permet de réviser la tarification a priori, en ajustant les informations sur les critères de classification des risques. En

effet, l'expérience montre que l'utilisation des variables observables pour estimer le risque d'un assuré ne fournit pas toujours une segmentation assez précise de la population. Les classes de risque sont encore hétérogènes après tarification a priori.

C'est pour cette raison que les actuaires utilisent la méthode de tarification basée sur l'expérience sinistre des assurés. Cette tarification est connue sous le nom de tarification a posteriori. C'est une méthode utilisée par les actuaires qui permet de tenir compte des différences individuelles de chaque assuré dans le portefeuille.

L'analyse de tarification a posteriori est aléatoire en fonction du nombre de réclamations de l'assuré, elle permet de réévaluer la prime a priori dans le but de refléter le risque réel que représente l'assuré. Ce tarif est justifié par l'amélioration de la classe de risques et de son calcul. Donc, la tarification a posteriori est basée sur des systèmes ou des concepts mathématiques pour déterminer la prime des assurés en fonction de leur sinistralité.

Tout simplement, cette tarification permet de s'assurer que chaque assuré doit payer une prime qui couvre exactement le risque encouru selon le nombre de sinistres dont il est responsable avec une complémentarité entre la tarification a priori et la tarification a posteriori. (Riad, 2015)

1.2.3. Le système Bonus-Malus

C'est une forme de la tarification a posteriori à travers l'ajustement des informations sur les critères de tarification des risques. C'est un système de tarification basée exclusivement sur le nombre de sinistres déclarés par l'assuré. Ce système correspond à un système à classes dans lesquels les assurés grimpent ou descendent d'un certain niveau en fonction du nombre de sinistres déclarés ou du nombre d'années d'assurance sans réclamation.

Le système Bonus-Malus consiste à accorder des réductions ou des majorations en fonction de sinistralités causées par l'assuré dont il est responsable, il peut être défini comme étant une note personnelle qui reflète l'historique de conducteur. Ce système minore ou majore les primes de base. Dans ce cas, la prime payée par l'assuré est égale à la prime de base prédéterminée selon les règles de tarification a priori est multipliée par la relativité du niveau occupé par l'assuré dans l'échelle bonus-malus. (Riad, 2015)

Ce système est représenté dans le monde selon deux approches (Prévoit, 2005) :

- Un changement de classe de risque, dont lequel l'assuré peut passer d'une catégorie de risques à une autre et donc d'une prime de base à une autre soit supérieure ou inférieure en fonction de son taux de sinistralité;
- Une multiplication de la prime de base dont lequel le taux de sinistralité est converti en un coefficient de réduction ou de majoration multipliée, par la suite, à la prime de base.

Le bonus-malus est un système qui permet d'améliorer la tarification a posteriori en utilisant l'information révélée par les accidents passés de l'assuré, et donc de rendre les classes de risque plus homogènes. Il permet aussi de maintenir les incitations à la prudence et de réduire les inefficacités associées au risque.

Chaque nouvel assuré dispose d'un degré de l'échelle et un réajustement annuel de la position de l'assuré sur cette échelle est effectué en fonction de la sinistralité de cet assuré conformément aux règles de transition de ce système. De plus, le niveau qu'occupera l'assuré dans le système bonus-malus à l'année $n + 1$ ne dépend que du niveau qu'il occupait durant l'année n et du nombre de sinistres déclarés durant l'année n .

Selon Henriet et Rocher (1986), le système Bonus-Malus possède les trois objectifs fondamentaux suivants :

- La responsabilisation du conducteur, puisque la prime accroît en cas de sinistre;
- Un meilleur affinage des sélections de risques;
- Une réduction des coûts de gestion, à travers l'incitation des assurés à ne pas déclarer les petits sinistres.

De ce fait, la préoccupation majeure des assureurs est de disposer d'une base de données exploitable qui permet d'approcher une modélisation satisfaisante sur l'historique de survenance des accidents en estimant l'évolution du portefeuille automobile, de façon à déterminer une prime équitable permettant de couvrir le risque assumé. (Riad, 2015).

L'adoption du système Bonus-Malus varie d'un pays à un autre. Dans l'annexe n°2, nous présentons quelques modèles de système Bonus-Malus adoptés par un échantillon de pays.

1.3. Le principe de la mutualisation et la segmentation

Le but lors de l'établissement d'un tarif sur une population d'individus, est d'assurer un bon équilibre entre la mutualisation des individus d'un même groupe homogène de risques et la segmentation en classes de risques.

1.3.1. Le principe de la mutualisation

Selon le Petit Larousse, le verbe mutualiser signifie « faire passer un risque, une dépense à la charge d'une mutualité ou d'une collectivité ». Dans le domaine de l'assurance, elle peut être définie comme une mise en commun d'un partage des risques individuels. Elle peut s'interpréter comme étant une application de la loi des grands nombres au sein d'une population de risques homogènes.

Dans ce cas, les individus peuvent être considérés comme des variables aléatoires indépendantes qui suivent la même loi de probabilité, c'est-à-dire que ces individus se sont exposés aux mêmes risques et ont la même probabilité de subir un sinistre.

En effet, la loi des grands nombres permet d'estimer à la compagnie d'assurance une indemnité moyenne par assuré lorsque les dommages sont distribués de manière identique et indépendante. (Magali, 2016)

Le principe de la mutualisation est basé sur la répartition du coût de la réalisation du risque couvert entre les membres de groupe soumis au même risque et qui pourraient toucher certains d'entre eux. Il s'agit alors une compensation entre les risques sinistrés dont lesquels l'assureur reçoit une prime sans avoir à régler la prestation. Donc la masse des primes collectées permet de verser les indemnités aux sinistrés.

À ce niveau, il est important de souligner la distinction entre les deux notions suivantes, la solidarité et la mutualisation entre les assurés. En effet, la solidarité signifie que tous les assurés paieraient une prime identique à l'assureur quel que soit le profil de risque et cela est en contradiction avec le fondement de l'assurance.

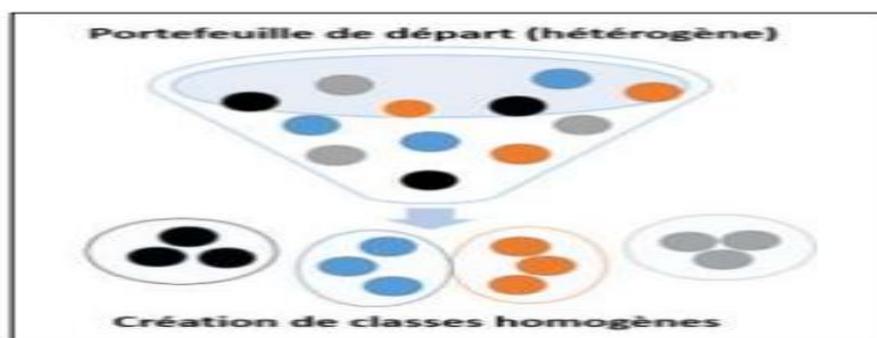
1.3.2. Le principe de la segmentation

L'hétérogénéité des risques au sein d'une population globale oblige les assureurs à segmenter les risques. C'est-à-dire découper le portefeuille en plusieurs groupes d'assurés en fonction de leurs profils et des risques semblables.

Dans le contexte actuel qui se caractérise par un marché très concurrentiel, la segmentation devient incontournable pour tous les assureurs, d'une part pour fidéliser les clients déjà existants et d'autre part pour attirer de nouveaux clients représentant des bons risques.

Le principe de la segmentation repose sur le fait que chaque individu doit payer une prime qui est fonction de son propre risque. C'est une pratique est utilisée depuis des années par plusieurs secteurs. La figure suivante explique les effets et les objectifs de la segmentation:

Figure 2 : Les effets et objectifs de la segmentation



Source : Rachel, 2016

Plusieurs définitions ont été accordées à la notion de la segmentation en assurance. D'après Charpentier et Denuit (2009) « On désigne par une segmentation les outils ou techniques que l'assureur exploite pour différencier les composantes de son portefeuille, et éventuellement aussi la couverture, en fonction d'un certain nombre de caractéristiques spécifiques du risque à assurer, et ce pour obtenir à une meilleure cohérence ».

Nous retenons une autre définition instaurée par Niny (2013), qui a défini la segmentation comme étant « la partition d'un portefeuille d'assurés en sous portefeuilles, appelés encore classes de risques, par l'étude de caractères distinctifs et qui permet à l'assureur de déterminer et d'appliquer une prime différenciée selon la classe ».

De nos jours, les compagnies d'assurance partout dans le monde utilisent la segmentation des clients dans leurs quotidiens afin de disposer pour chaque assuré plusieurs segments et sous-segments que ce soit pour les particuliers ou bien pour les professionnels à travers une base de données rationnelle.

Compte tenu de la structure de marché d'assurance, chaque assureur doit posséder à une tarification basée sur la segmentation de son portefeuille. Ainsi chaque catégorie de risques aura un niveau de tarification.

La nécessité de segmenter

Dans un but comprendre la nécessité de segmenter, prenons un exemple qui matérialise cette notion.

- Soit un assureur A qui décide de ne pas effectuer aucune segmentation
- Un assureur B, qui utilise la segmente en fonction des caractéristiques de l'assuré

Chaque assureur propose un contrat ayant les mêmes garanties. Le tableau ci-dessous présente les données relatives aux primes proposées par chacun des assureurs :

Tableau 2 : Les primes proposées par les deux assureurs

| | Mauvais conducteurs | Bon conducteur |
|------------|----------------------------|-----------------------|
| Assureur A | 700 € | 700 € |
| Assureur B | 1500 € | 500 € |

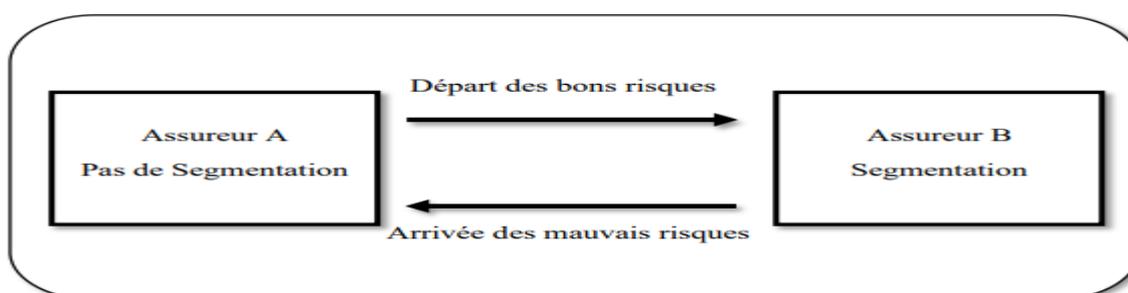
Suite à cette segmentation, les assurés de bons risques, c'est-à-dire ayant une faible probabilité d'avoir un sinistre, décideront de s'assurer auprès de l'assureur B, puisqu'il propose des primes plus faibles. En revanche, suite à la non segmentation de son portefeuille l'assureur A propose aux assurés un tarif unique bien moins intéressant que celui de l'assureur B.

Dans ce contexte, la prime perçue par l'assureur A pour les bons conducteurs sert, en partie à couvrir les sinistres des mauvais conducteurs. Les mauvais conducteurs de l'assureur B payent une prime supérieure du fait de leur exposition au risque et les bons conducteurs payent une prime plus faible, en accord avec leur profil de risque.

La mise en concurrence de ces deux assureurs entraîne des migrations des assurés d'un portefeuille vers l'autre. En effet, les assurés s'orienteront automatiquement vers l'assureur leur proposant le meilleur tarif. Ainsi, les mauvais conducteurs souscriront leurs contrats chez l'assureur A et les bons conducteurs chez l'assureur B. L'assureur A ne disposera donc plus de bons conducteurs pour couvrir les mauvais et son modèle économique ne sera plus viable.

La figure ci-dessous exprime bien la nécessité pour l'assureur de segmenter. Nous pouvons affirmer également que la segmentation, permet de diminuer les conditions tarifaires et en contrepartie, elle augmente les tarifs appliqués aux autres jugés risqués. Cela permettra à la compagnie d'assurance d'orienter sa politique vers les segments de portefeuilles les plus rentables et de leur proposer des offres de plus en plus adéquates ce qui va permettre de les fidéliser et aussi d'anticiper les affaires nouvelles qui ont un rendement plus intéressant.

Figure 3 : Les flux d'assurés provoqués par l'utilisation ou non d'une segmentation



Source : Gonnet (2010)

D'après tout ce qui précède, nous pouvons affirmer la nécessité de segmentation d'un portefeuille. Il faut savoir à ce niveau, que la technique de segmentation des risques n'est pas en contradiction avec le principe de mutualisation. En effet, nous pouvons diviser une population en plusieurs sous populations, tout en gardant la mutualisation valable au sein de chacune de ces classes et dans ce cas la loi des grands nombres reste toujours applicable.

Section 2: Les innovations et les opportunités possibles en assurance automobile

Dans cette section nous allons présenter une nouvelle piste utilisée par les assureurs pour l'évaluation du risque automobile qui est l'assurance comportementale. Cette nouvelle logique

constitue une innovation et permet d'appliquer une tarification individualisée en fonction du comportement du conducteur et grâce aux nouvelles technologies.

La tarification du risque automobile en utilisant cette formule est basée sur de nouvelles informations non connues auparavant par l'assureur tel que le kilométrage, comportement de conduite, et le moment et les lieux de circulation

En effet, l'assurance comportementale est connue sous plusieurs noms, Assurance télématique, Assurance connectée, ou les termes anglo-saxons Usage-Based Insurance (UBI), Pay As You Drive (PAYD), et Pay How You Drive (PHYD). Nous allons dans la suite revenir sur chacun de ces points afin de les étudier en profondeur.

2.1. L'utilisation des variables innovantes

Dans un premier temps, il n'est pas question que les assureurs suppriment les anciennes variables tarifaires. Au contraire, ce nouveau système de tarification devrait contenir plus que le précédent système.

Actuellement, les assureurs ont tourné vers le Big Data pour l'élaboration de leurs tarifs avec la capacité de construire des modèles complexes et de plus en plus personnalisés. Certains assureurs ont fait le choix de totalement modifier l'esprit des variables utilisées pour la tarification.

Les assureurs ont commencé à surveiller le comportement du conducteur par le risque individuel. L'idée est donc d'utiliser des nouvelles variables, qui sont fortement corrélées non pas au véhicule ou à l'assuré mais aux habitudes de conduite. Ces nouvelles variables sont présentées dans la figure ci-dessous :

Figure 4 : Les variables tarifaires innovantes



Source : Gonnet (2010)

Le nombre de kilomètres parcourus annuellement est fréquent dans le questionnaire de souscription en assurance automobile et il est fortement corrélé au tarif. Plus nous utilisons le véhicule plus le risque d'accident est important et par conséquent plus la prime commerciale augmente. Dans ce cas, Il ne s'agit pas d'indiquer une fourchette comme précédemment mais d'avoir une mesure exacte au kilomètre. Les assureurs s'intéressent également sur les heures d'utilisation du véhicule, puisque statistiquement une conduite de nuit est plus dangereuse qu'une conduite en plein jour. (Atia, 2016)

2.2. Une segmentation illimitée

De nos jours, l'une des pistes explorées par les assureurs est l'application d'une tarification individualisée. La segmentation poussée à l'extrême, pour lequel l'assureur aura une analyse plus fine de son risque mais met en cause le principe fondamental de l'assurance à savoir la mutualisation des risques.

Il devient possible de faire payer l'assuré en fonction de son comportement en temps réel au volant. Il est possible aussi de contrôler le kilométrage parcouru du conducteur, de connaître l'heure d'utilisation du véhicule ainsi que le type de routes empruntées (Nationale, Autoroute, Ville ...). Les assureurs peuvent également s'intéresser à la vitesse, au temps de voyage, aux chocs, de coups de volant ou de freinages brusques.... (Gonnet, 2010)

Selon cette logique, un assuré conduisant principalement de nuit et sur des routes connues dangereuses s'acquittera d'une prime supérieure à un autre effectuant ses trajets essentiellement de jour et sur des autoroutes statistiquement moins dangereuses.

Nous remarquons, que la plupart de ces variables sont quantitatives continues et qualitatives avec un nombre important de modalités. Dans ce cas, les modèles traditionnellement utilisés dans la modélisation de la tarification en assurance, qui sont les Modèles Linéaires Généralisés (GLM) ne sont pas adaptés à l'intégration de telles variables.

Les assureurs doivent employer des modèles innovants pour pouvoir utiliser ces variables, il peut être par exemple l'utilisation d'Arbres de Régression. Si les assureurs parviennent à le faire, ça sera le début d'une tarification individualisée et la fin de la mutualisation puisque les classes tarifaires seront disparues. Dans ce cas, l'assuré peut payer uniquement pour ce qu'il consomme. (Atia, 2016)

2.3. Le concept d'Usage Based Insurance (UBI)

Le concept d'Usage Based Insurance (UBI), signifie une assurance basée sur l'utilisation. Ce concept a été introduit sur le marché de l'assurance automobile il y a quelques années. Au lieu de fixer les primes d'assurance en fonction de la marque, le modèle du véhicule, l'âge du conducteur, son expérience... L'UBI évalue les primes en fonction du temps d'utilisation, de la distance parcourue, du comportement de conduite et des endroits de conduite de l'assuré. (Ernst & Young, 2016)

Actuellement, il existe quatre catégories distinctes d'offres technologiques de l'UBI disponibles sur le marché :

- **Dongle USB** : L'assureur installe lui-même un appareil sous forme d'un port matériel dans le véhicule.
- **Boîte noire (black box)** : Une boîte noire installée par un professionnel est fixée de manière permanente sur le véhicule assuré. Cette solution est considérée comme l'une des plus sûres et des plus fiables.
- **Embarqué (Embedded)** : à la fin de l'année 2013, 11 constructeurs automobiles fournissent un équipement télématique embarqué pour les véhicules, fournissait des services tels que le diagnostic à distance.
- **Les smartphones** : La technologie de télécommunication mobile est le dernier outil de la télématique, les smartphones fonctionnant comme des appareils autonomes ou reliés aux systèmes des véhicules pour transmettre une variété d'informations.

Il existe actuellement deux modèles d'UBI :

- ✓ **Pay As You Drive (PAYD)**
- ✓ **Pay How You Drive (PHYD)**

2.3.1. La solution Pay As You Drive (PAYD)

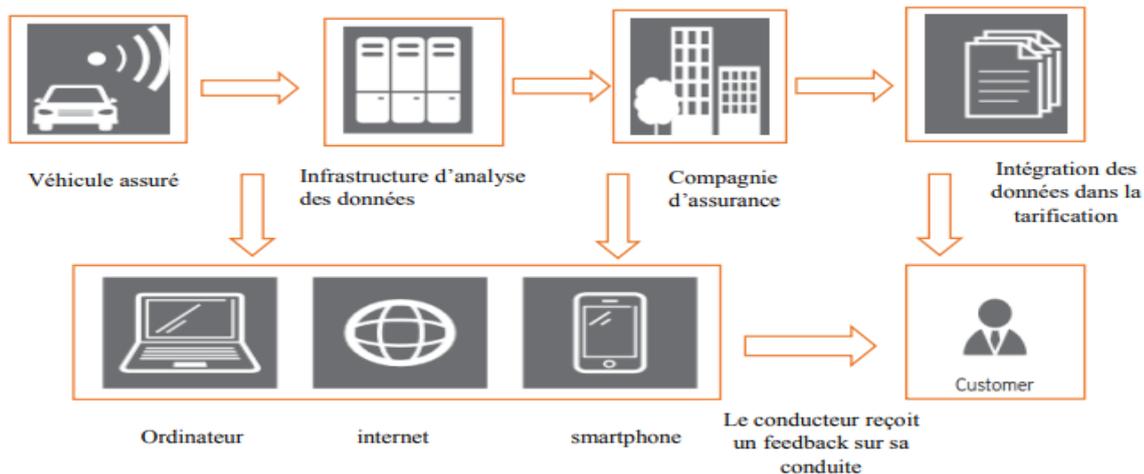
La solution « Pay as you drive » veut dire « payez pour ce que vous conduisez ». Appelée aussi assurance automobile au kilomètre. C'est une innovation dans le domaine d'assurance automobile dont laquelle elle permet aux assurés de payer selon l'utilisation réelle de leur voiture à travers la digitalisation. (Bordoff et Noel, 2008)

Le succès de ce type d'assurance n'a pas tardé à se développer dans le monde de l'assurance. Ce système consiste à installer dans le véhicule de l'assuré un instrument de mesure appelé odomètre, qui permet de connaître la distance parcourue. Cet élément est couplé d'un GPS (Global

Positioning System). Les données enregistrées sont ensuite transmises via le réseau de téléphonie mobile. La quantité des données est alors très importante. L'assureur peut alors à partir de ces données établir un tarif personnalisé pour chaque assuré. (Gonnet , 2010)

La figure qui suit décrit ce système et son fonctionnement :

Figure 5 : Fonctionnement du modèle Pay As You Drive



Source : Ernst & Young (2016)

Cette solution a été utilisée pour la première fois en France par Solly Azar, qui est le deuxième courtier du pays. Cette offre est désignée pour une population cible qui est les jeunes conducteurs car ils sont les plus familiarisés avec les nouvelles technologies. La tranche d'âge cible varie entre 18 à 25 ans. Cette offre est connue sous le nom d' « Easy Drive » en 2008.

2.3.2. La solution Pay How You Drive (PHYD)

La solution « Pay How You Drive » (PHYD), signifie littéralement «payez comme vous conduisez ». À ce niveau, il est très intéressant de distinguer entre les deux concepts PAYD et PHYD. Le concept PHYD succède au PAYD. Le principe du PAYD est similaire au principe du PHYD à la différence que seul le kilométrage est pris en compte. Le PHYD signifie que l'assuré paie une prime fonction du nombre de kilomètres parcourus et aussi de son comportement au volant (agressivité, freinage,..). Ce comportement peut être relevé à l'aide d'un boîtier installé dans le véhicule. Les avantages de ce modèle sont multiples pour l'assuré et pour l'assureur. Ces avantages sont représentés dans le tableau ci-dessous :

Tableau 3 : Les avantages du modèle PHYD

| Les avantages pour les assurés | Les avantages pour les assureurs |
|--|---|
| -Accès à des récompenses telles que des réductions sur les primes -Récupération de véhicules volés - Corriger la classification des risques... | - Lutter contre les réclamations frauduleuses - Réduire les coûts des sinistres - Réduire le temps de réponse aux accidents - Suivi et récupération des véhicules volés - Attirer les bons conducteurs... |

Source : Ernst & Young, 2016

En 2014, Direct Assurance, filiale d’Axa a fait un premier pas en France, dans l’assurance automobile tarifée en fonction du comportement du conducteur en lançant son contrat « You Drive» pour les assurés de 18 à 24 ans.

Section 3 : Les travaux empiriques sur les modèles de tarification

3.1. Les outils statistiques

Les travaux empiriques réalisés sur la tarification des primes en assurance peuvent se distinguer par leurs approches spécifiques et leurs modèles actuariels utilisés ainsi que leurs variables explicatives. Cependant, ils sont présentés dans un cadre commun qui sert à la modélisation générale et aux traitements statistiques et économétriques.

En effet, ces travaux s’intéressent à relier la variable à expliquer (Y) à des variables observées (X) qui se présentent sous la forme :

$$Y = f(X_1, \dots, X_n)$$

De manière générale en assurance, les variables X_i désignent souvent, le sexe, l’âge... Alors que la variable Y représente soit la fréquence ou le coût moyen.

3.1.1. Le modèle de régression linéaire

Les modèles de régression linéaires sont les modèles utilisés pour mesurer l’impact des variables explicatives sur le phénomène d’étude.

Soit n individus observés pour lesquels nous disposons :

- d’une variable réponse (Y_i) $i=1, \dots, n$, aussi appelée variable à expliquer
- de p variables explicatives (x_{i1}, \dots, x_{ip}) $i=1, \dots, n$

L’objectif est d’établir une relation entre la variable à expliquer Y_i et les p variables explicatives (x_{i1}, \dots, x_{ip}). Supposons l’existence d’une relation linéaire entre la variable à expliquer et les variables explicatives du type.

$$y_i = a_0 + a_1 x_{1i} + \dots + a_k x_{ki} + \varepsilon_i$$

Pour $i=1, \dots, n$

Nous pouvons adopter une écriture qui rend la lecture et la manipulation de l'ensemble plus facile sous forme des équations suivantes :

$$\begin{cases} y_1 = a_0 + a_1 x_{1,1} + \dots + a_p x_{1,p} + \varepsilon_1 \\ y_2 = a_0 + a_1 x_{2,1} + \dots + a_p x_{2,p} + \varepsilon_2 \\ \dots \\ y_n = a_0 + a_1 x_{n,1} + \dots + a_p x_{n,p} + \varepsilon_n \end{cases}$$

Ces équations peuvent être résumées avec la notation matricielle suivante :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

En notant :

- y : de dimension $(n, 1)$
- x : de dimension $(n, p+1)$
- a : de dimension $(p+1, 1)$
- ε : de dimension $(n, 1)$

Où nous pouvons écrire matriciellement : $Y = AX + \varepsilon$

$$\text{Avec } E(\varepsilon) = 0 \text{ et } V(\varepsilon) = \sigma^2 I$$

Notons : I la matrice d'identité

Cependant, les modèles linéaires s'avèrent faciles à mettre en œuvre, mais ils ne sont pas entièrement adaptés à la tarification en assurance. Ces modèles supposent que la moyenne est une fonction linéaire des variables puisque les variables que l'on cherche à modéliser en assurance ce sont des coûts qui sont des valeurs dans \mathbb{R}_+ et le nombre de sinistres qui sont des valeurs dans \mathbb{N} .

Pour pallier ces restrictions, les modèles linéaires généralisés (GLM) représentent une généralisation de la régression linéaire et ils permettent d'étudier la liaison entre une variable réponse Y et un ensemble de variables explicatives quelle que soit la nature de Y . (Charpentier et Denuit, 2009).

3.1.2. Le Modèle Linéaire Généralisé (GLM)

Le modèle linéaire généralisé (GLM) est une extension du modèle linéaire. Il a été introduit en 1972, par Nelder et Wedderburn. Il permet d'étudier le lien entre une variable à expliquer et un ensemble de variables explicatives, qualitatives ou quantitatives. (Gourieroux, et al., 1984).

Ce modèle permet de s'affranchir de l'hypothèse de normalité, en traitant de manière unifiée des réponses dont la loi fait partie de la famille exponentielle linéaire (loi Normale, Gamma, Poisson et la loi Binomiale). (Charpentier et Denuit, 2009).

Le modèle GLM est utilisé dans le cas où la variable réponse n'est pas continue ou bien tout simplement le cas où la relation entre la variable endogène et les variables exogènes n'est pas linéaire. Une fonction de lien adéquatement choisie permet de mieux modéliser l'effet des variables exogènes sur la variable endogène dans un modèle GLM.

Le modèle GLM se distingue du modèle linéaire par les trois composantes suivantes :

- Distribution de famille exponentielle,
- Les variables explicatives,
- La fonction de lien

a) Distribution de famille exponentielle

La loi de Y appartenait à la famille exponentielle, c'est-à-dire sa densité s'écrit sous la forme suivante:

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right), \quad y \in \mathcal{S},$$

Avec :

S : un sous-ensemble de N ou de R

θ : un paramètre

ϕ : un paramètre de dispersion.

Le tableau ci-dessous s'intéresse à présentation de la famille des modèles linéaires généralisés :

Tableau 4: Les lois de la famille exponentielle utilisée en GLM

| Lois | $\mathcal{Y}; \Omega$ | Densité/Loi | ϕ | $b(\theta)$ | $c(y, \phi)$ | $\mu = b'(\theta)$ | $V(\mu)=b''(\theta)$ | $\text{Var}(Y)=\phi V(\mu)$ | Lien canonique |
|--|---|--|------------------|----------------------|---|---------------------------------|------------------------------|-----------------------------|--|
| Loi Normale $\mathcal{N}(\mu, \sigma^2)$ | $\mathbb{R}; \mathbb{R} \times \mathbb{R}^{++}$ | $\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$ | σ^2 | $\frac{\theta^2}{2}$ | $-\frac{\left(\frac{y^2}{\phi} + \log(2\pi\phi)\right)}{2}$ | θ | 1 | σ^2 | Identité |
| Loi Gamma $\mathcal{G}(\mu, \tau)$ | $\mathbb{R}^{++}; \mathbb{R}^{++} \times \mathbb{R}^{++}$ | $\frac{\left(\frac{\tau}{\mu}\right)^\tau y^{\tau-1} \exp\left(-\frac{\tau}{\mu}y\right)}{\Gamma(\tau)}$ | $\frac{1}{\tau}$ | $-\ln(-\theta)$ | $\frac{\tau \ln(\tau y) - \ln(y)}{-\ln(\Gamma(\tau))}$ | $-\frac{1}{\theta}$ | $\frac{1}{\theta^2} = \mu^2$ | $\frac{\mu^2}{\tau}$ | $-\frac{1}{\mu}$ (Inverse) |
| Loi Poisson $\mathcal{P}(\mu)$ | $\mathbb{N}; \mathbb{R}^{++}$ | $P(Y = y) = \exp(-\mu) \frac{\mu^y}{y!}$ | 1 | $\exp(\theta)$ | $-\ln(y!)$ | $\exp(\theta)$ | $\exp(\theta) = \mu$ | μ | $\ln(\mu)$ (Log) |
| Loi Binomiale $\frac{\mathcal{B}(n, \mu)}{n}$ | $\left[\frac{0; n}{n}\right]; \mathbb{N} \times [0; 1]$ | $P(Y = y) = \binom{n}{ny} \mu^{ny} (1-\mu)^{n-ny}$ | $\frac{1}{n}$ | $\ln(1 + e^\theta)$ | $\ln\left(\binom{n}{ny}\right)$ | $\frac{e^\theta}{1 + e^\theta}$ | $\mu(1-\mu)$ | $\frac{\mu(1-\mu)}{n}$ | $\ln\left(\frac{\mu}{1-\mu}\right)$ (Logit) |

Source : Charpentier et Denuit, 2009

✓ **Critères de sélection de la loi**

Dans certains cas, plusieurs lois sont pertinentes pour la modélisation de la variable à expliquer. Dans ce cas, il existe de nombreux critères qui sont utilisés pour choisir le modèle qui s'ajuste le mieux aux données

Les deux critères les plus utilisés sont :

- L'AIC (Critère d'Information d'AKAIKE);
- le BIC (Bayesian Information Criterion).

Ces deux critères permettent de choisir le meilleur modèle en termes de qualité d'ajustement aux données et également en termes de variables à retenir. Ils font intervenir dans leur calcul le nombre de paramètres du modèle ainsi que la Log Vraisemblance LL maximisée, cette dernière est fournie par l'expression suivante:

$$\max_{\beta} LL(y, \theta(\beta), \phi) = \max_{\beta} \ln \left(\prod_{i=1}^n f(y_i | \theta, \phi) \right)$$

• **Le critère AIC :**

C'est un critère de comparaison des modèles qui s'applique aux modèles estimés par Maximum de Vraisemblance. Le choix du meilleur modèle est celui qui présente le critère AIC minimal. Le principe consiste à calculer pour chacun des modèles le critère selon la formule ci-dessous:

$$\mathbf{AIC} = -2 \mathbf{LL} + 2 (\mathbf{p} + 1)$$

Notons :

-LL : est la Log Vraisemblance maximisée du modèle ;

$-p + 1$: correspond au nombre de paramètres du modèle

- **Critère BIC**

De même, c'est un critère de comparaison de modèle calculée selon la formule ci-dessous :

$$\mathbf{BIC} = -2 \mathbf{LL} + (\mathbf{p} + 1) \mathbf{log(n)}$$

Notons :

-LL : est la Log Vraisemblance maximisée du modèle ;

$-p + 1$: correspond au nombre de paramètres du modèle ;

$-n$: est le nombre d'individus composant l'échantillon.

Le choix du meilleur modèle est celui qui présente le critère BIC minimal.

b) Les variables explicatives

Comme pour le modèle linéaire classique, ce modèle est exprimé par une combinaison des variables explicatives nommées prédicteur linéaire :

$$\mathbf{E}(Y_i) = \mathbf{g}(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$$

Avec g est la fonction de lien

c) La fonction de lien

La fonction de lien, comme son nom l'indique, contient le lien entre l'espérance de la variable réponse et la composante déterministe. Au lieu de modéliser la moyenne directement, nous allons introduire une fonction de lien strictement monotone et dérivable donc inversible. Contrairement aux modèles linéaires simples, il s'agit dans ce cas, de modéliser une transformation de l'espérance de la variable réponse.

L'espérance de Y notée u dépend de (X) à travers une fonction de liens notée $g(\cdot)$:

$$g(u) = \eta(X)$$

Donc, pour la construction d'un GLM, il est primordial d'effectuer deux choix, un choix concerne la loi de la variable à expliquer, un deuxième choix qui porte sur la fonction de lien. Le tableau ci-dessous présente les fonctions de lien classiquement utilisées:

Tableau 5 : Les fonctions de liens associés aux lois usuelles de la famille exponentielle

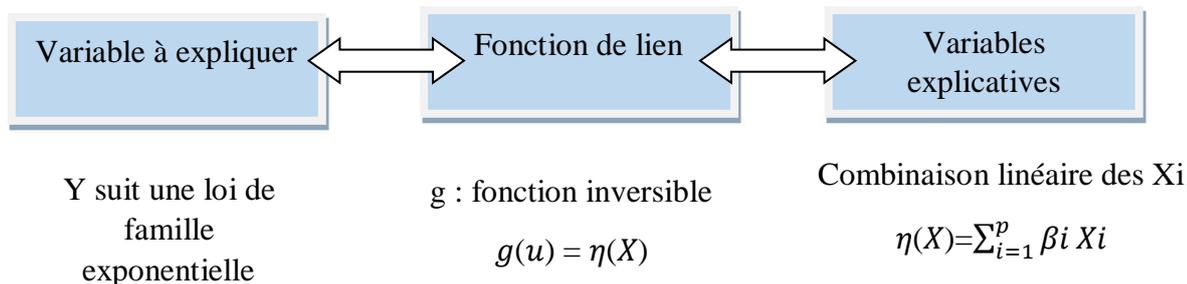
| Loi de probabilité | Fonction de lien canonique | |
|---------------------------|--|----------|
| Bernoulli | $\eta = \ln\left(\frac{\mu}{1-\mu}\right)$ | Logit |
| Binomiale négative | $\eta = \ln(p)$ | Log |
| Gamma | $\eta = \frac{1}{\mu}$ | Inverse |
| Poisson | $\eta = \ln(\mu)$ | Log |
| Normale | $\eta = \emptyset$ | Identité |

Source : Atia, 2016

Le choix du lien peut également être déterminé par l'existence d'études passées et par une connaissance préalable du problème.

La logique de composition de ce modèle peut être résumée dans la figure suivante :

Figure 6 : La composition du modèle GLM



Source : Gonnet, 2010

3.1.3. Estimations des paramètres par maximum de vraisemblance

L'estimation des paramètres du GLM se fait par la méthode de maximum de vraisemblance. Cette méthode est définie comme étant, un produit de fonctions de densité.

Pour savoir le maximum, il suffit de déterminer la valeur du paramètre de la fonction qui annule sa dérivée tout en s'assurant que la dérivée seconde négative. En pratique, on utilise la dérivée le logarithme de la vraisemblance. On dérive donc, une somme plutôt qu'un produit.

Ainsi, comme la fonction logarithme est strictement croissante, maximiser le logarithme de la fonction revient à maximiser la fonction. Concrètement, n variables indépendantes avec Y_i de loi appartenant à la famille exponentielle, de paramètre θ_i et de densité f. Soient (y_1, y_2, \dots, y_n) n observations de cette variable.

La vraisemblance s'écrit de la manière suivante:

$$l(y; \theta; \phi) = \prod_{i=1}^n f(y; \theta; \phi)$$

Et la log-vraisemblance s'écrit de la manière suivante:

$$L(y; \theta; \phi) = \ln l(y; \theta; \phi) = \sum_{i=1}^n \ln f(y; \theta; \phi)$$

3.2. Modélisation du risque en assurance automobile

Nous avons présenté précédemment la démarche à suivre dans le cadre des modèles GLM. À ce niveau, il est intéressant d'entrer plus en détails à la modélisation du risque en assurance automobile dont lequel nous évoquons la fréquence et du coût moyen des sinistres. La prime pure est ensuite calculée en multipliant la fréquence par le coût moyen. (Gonnet, 2010)

En règle générale, pour la modélisation des fréquences de sinistres nous observons un processus de comptage, pour lequel nous désirons de modéliser une proportion qui est la fréquence. Un modèle de comptage comme son nom l'indique, il s'agit d'un nombre d'évènements dans notre cas c'est le nombre de sinistre survenu dans une période. Les lois les plus fréquentes sont la loi de Poisson et la loi Binomiale Négative. L'utilisation d'une loi de Poisson suppose que la variance est égale à la moyenne et qui est rarement vérifiée. Le modèle de Poisson est généralement contraignant car il suppose l'équi-dispersion des données et qui est rarement à vérifier. (Denuit et Charpentier, 2009)

En ce qui concerne les montants de sinistres, ils sont généralement modélisés à partir d'une loi Log-Normale et la loi de Gamma. En effet, ces derniers correspondent à une distribution continue et définie sur des réels positifs. La modélisation du coût moyen, est plus compliquée. Toutefois, la fréquence annuelle des sinistres est faite sur toutes les polices par contre le coût moyen il est clair que seules les polices sinistrées peuvent être utilisées pour modéliser les montants de sinistres car ces deux lois n'acceptent pas des valeurs nulles. Dans ce cas, un actuair e dispose un nombre très limité d'observations. (Gonnet, 2010)

De plus, les coûts de sinistres sont beaucoup plus difficiles à modéliser car il existe des sinistres dont le remboursement nécessite des délais assez longs pour être clôturés. Par exemple, un sinistre RC automobile avec des dégâts corporels, il faut souvent attendre que l'état de victime soit stabilisé pour pouvoir déterminer le montant des indemnités. (Denuit et Charpentier, 2009)

De même, la modélisation des sinistres nécessite un traitement particulier des sinistres dits «graves ». Les sinistres graves sont des sinistres avec des montants exceptionnels et qui font l'objet d'une attention très particulière. Souvent, moins de 20% des sinistres causent plus de 80% des dépenses de la compagnie. Ces sinistres sont modélisés d'une manière séparée. (Denuit et Charpentier, 2009)

3.3. Les travaux empiriques relatifs à la tarification en assurance automobile

Les études sur la tarification en assurance automobile sont à la fois nombreuses et diversifiées. Nous allons présenter une vue d'ensemble des études antérieures qui portent sur ce sujet. Globalement les résultats des recherches antérieurs sont similaires avec quelques différences dans les détails. Cette divergence peut être expliquée par les contextes des études différents.

✓ Les travaux de Ghali (2001)

Cet article porte sur l'évaluation du système de tarification de l'assurance automobile en Tunisie à l'aide d'un modèle de tarification optimale. Cette étude présente un modèle optimal de tarification basé sur les caractéristiques des assurés, son véhicule et aussi sur le nombre d'accidents passés, c'est-à-dire sur la base d'un modèle a priori et a posteriori.

Pour ce faire, Ghali (2001), a utilisé un modèle de tarification basé sur une base de données qui s'étale sur cinq ans de 1990 jusqu'à 1995 et qui contient 46337 observations. Cette base de données de cinq ans est composée des variables suivantes: sexe, région de résidence par gouvernorat, puissance du véhicule, marque de véhicule, nombre d'accidents.

La méthodologie utilisée pour l'estimation de la fréquence d'accidents est celle proposée par Dionne et Vanasse en 1992, et qui consiste à utiliser les modèles de comptage avec la distribution de Poisson et Binomiale Négative comme composante de régression. Elle a aussi développé un système bonus-malus qui intègre en même temps les informations a priori et a posteriori sur une base individuelle.

Les résultats obtenus de cette étude sont très intéressants. Ils ont montré la non-optimalité du système de tarification tunisien pour l'usage privé. De plus, l'existence des variables significatives pour expliquer le nombre d'accidents comme la marque, puissance et âge. L'apport de cet article est l'utilisation des modèles de comptage dont l'usage est très récent dans les problèmes de tarification en assurance automobile.

✓ Les travaux de Riad (2015)

Dans le contexte Algérien, Riad a utilisé la même méthodologie appliquée par Ghali (2001). De même, l'objectif est d'analyser le système de tarification de l'assurance automobile en Algérie avec une proposition d'un nouveau système Bonus-Malus basé sur le nombre d'accidents dont l'assuré est le responsable.

Les données statistiques utilisées par cet article proviennent d'une compagnie d'assurance Algérienne étatique, et qui détient une part du marché importante. Pour chaque assuré, elle a pris en considération à partir de la police d'assurance les caractéristiques suivantes : l'âge, le sexe de l'individu, la puissance du véhicule l'âge du permis, la puissance du véhicule, l'usage et ainsi la zone de résidence que le genre du véhicule, les garanties souscrites, le nombre d'accidents, la marque de véhicule.

Les résultats obtenus de cette étude ont démontré le non optimalité du système de tarification algérien et que plusieurs variables sont statistiquement significatives pour expliquer le nombre d'accidents comme la région des assurés, les garanties auxquelles ils souscrivent et les caractéristiques du véhicule (marque, puissance et âge). De plus, l'auteur a constaté la non optimalité du système Bonus-Malus algérien, ce qui incite à proposer d'autres critères.

✓ **Les travaux de Bendarag et Aboukherraz (2011)**

La libéralisation de la garantie RC automobile au Maroc a incité les compagnies d'assurances à mettre en œuvre leur propre système de tarification a priori et a posteriori adapté à leur portefeuille. Dans ce cadre, cette étude propose un modèle de tarification adéquate dans une compagnie d'assurance marocaine pour l'usage touristique.

La première étape de cette étude consiste à élaborer un tarif technique pour le portefeuille RC automobile à travers l'adoption du modèle GLM (Generalized Linear Model) pour la modélisation de la fréquence des sinistres corporels et matériels ainsi que le montant des sinistres corporels et matériels.

La deuxième étape consiste à proposer un SBM personnalisé de la prime relative à la garantie RC en automobile. Le SBM est basé sur le modèle Binomial Négative.

✓ **Les travaux de Gonnet (2010)**

Ce mémoire a pour objectif d'étudier de la tarification et de la segmentation en assurance automobile. Pour ce faire, l'auteur a utilisé le modèle GLM pour la modélisation des fréquences et des coûts moyens des sinistres graves et non graves partir du logiciel R. La modélisation des coûts moyens de sinistres est faite à partir de la loi Gamma et pour les fréquences à partir de la loi Poisson. Afin de quantifier concrètement le modèle de tarification, Gonnet a comparé les données de la prime pure théorique à la prime pure réelle pour chaque catégorie de risque.

Les apports de cette étude sont multiples, principalement la détection de segments sur ou sous tarifés. Ainsi que, la proposition à court terme de mettre en place des évolutions autour des tarifs en vigueur. D'autre part à moyen ou long terme l'introduction des variables jugées pertinentes dans le modèle en intégrant des modifications importantes de la logique de calcul des tarifs.

✓ Les travaux d'Atia (2016)

La mise en application de la Gender Directive en 2012 a obligé les assureurs à utiliser d'autres critères que le genre de l'assuré pour l'élaboration des tarifs. Cette information est jugée discriminante.

Suite à cette suppression, Atia a étudié les impacts tarifaires d'une telle décision. Pour ce faire, il a présenté un premier modèle règlementaire et un autre un modèle différant du premier par l'intégration de la variable genre dans la modélisation. Par la suite, il a effectué une comparaison de ces deux modèles. Pour la modélisation des deux modèles il a utilisé la loi gamma pour la modélisation des coûts moyens et la loi Binomiale Négative pour la modélisation des fréquences. En réponse à ce nouveau challenge, l'auteur a proposé une solution à travers l'utilisation des modèles de tarification innovants, avec l'utilisation de nouvelles variables tarifaires, également de nouveaux principes pour combler la suppression de variable clé genre et de disposer d'une analyse plus fine du risque.

CONCLUSION

Afin de mieux maîtriser le risque, l'assureur doit procéder à une analyse fine de son risque. Cette analyse est basée sur une segmentation de son portefeuille en catégories homogènes de risque, c'est-à-dire un sous-ensemble d'individus qui partagent des comportements similaires en matière de sinistralité. Pour définir ces sous-ensembles, les assureurs doivent disposer de nombreuses informations, concernant l'assuré, son véhicule, et aussi le passé de sinistralité. De plus il ne faut pas oublier de connaître l'importance du comportement du conducteur dans l'évaluation du risque automobile, passant d'une assurance conventionnelle à l'assurance comportementale.

Tout au long de ce chapitre, nous avons présenté une analyse théorique de la tarification en assurance automobile dont laquelle nous avons présenté ces typologies et ces principes de base. Par la suite, nous avons présenté les innovations en matière de tarification et finalement pour avons exposé les outils statistiques et un petit aperçu sur les travaux empiriques réalisés par d'autres chercheurs. Le prochain chapitre, sera consacré à la modélisation des deux dimensions de la tarification à savoir le coût moyen et la fréquence à travers une base de données de la GAT Assurances.

CHAPITRE DEUX : MODELISATION DE LA FREQUENCE ET DU COÛT MOYEN DES SINISTRES (APPLICATION EMPIRIQUE : CAS DU GAT ASSURANCES)

INTRODUCTION

Après avoir présenté dans le chapitre précédent, le positionnement des concepts clés, les fondements théoriques, ainsi que les outils et la revue de la littérature, nous allons présenter dans ce chapitre la partie empirique par laquelle nous nous intéresserons à la modélisation de la fréquence et du coût moyen de sinistres RC automobile.

À ce titre, l'objectif principal de ce chapitre est de modéliser la prime pure de la RC automobile pour l'usage utilitaire de l'échantillon du portefeuille de la GAT Assurances. Cette modélisation est fondée vers l'hypothèse de la libéralisation des tarifs à travers un modèle collectif séparant l'étude de la fréquence et du coût moyen de sinistre.

De ce fait, ce chapitre est composé de quatre sections. La première section est consacrée à la présentation de la situation actuelle du marché tunisien d'assurance automobile. La deuxième section se focalise sur la présentation de l'entreprise de parrainage GAT Assurances. Une troisième section pour le traitement préliminaire des données dont laquelle nous allons exposer la méthodologie de recherche et une analyse univariée et bivariée des grandeurs endogènes et exogènes. Enfin, la quatrième section fait l'objet de la modélisation de la prime pure de la garantie RC automobile.

Les résultats empiriques obtenus à partir de la modélisation centrée sur l'identification des déterminants de la fréquence et du coût moyen offrent un cadre idéal de discuter sur l'hypothèse de la libéralisation des tarifs.

Section1 : Situation actuelle du marché tunisien d'assurances automobile

1.1. Le marché d'assurance en Tunisie : focus sur la branche automobile

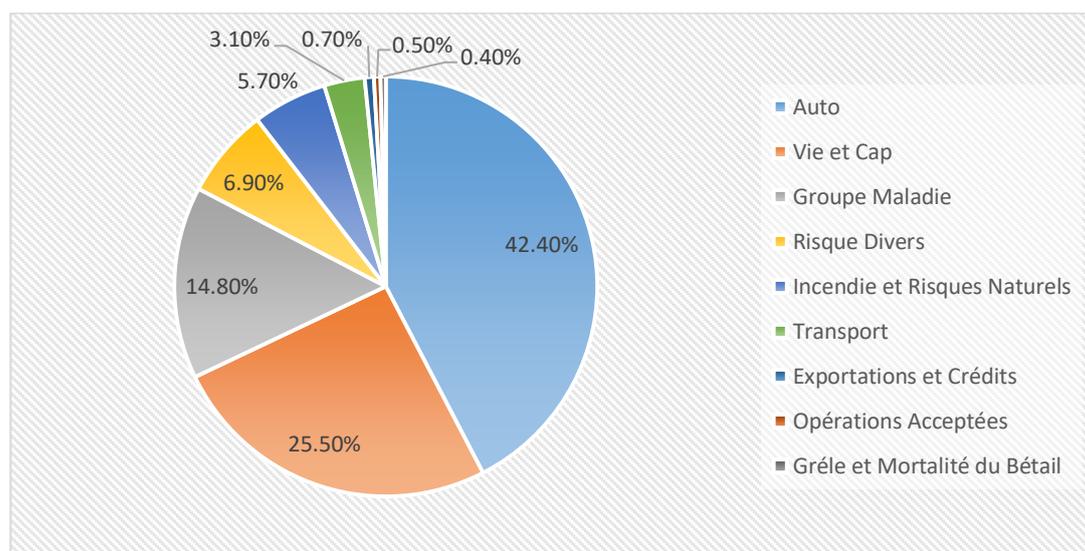
À la fin de l'année 2021, le marché tunisien des assurances se compose de 24 compagnies réparties entre des compagnies d'assurances et de réassurances résidentes, opérant sous le statut de société anonyme et mutuelle et celles non résidentes appelée « Off-shore ».

Le marché dispose d'un faible taux de pénétration de 2,5% en 2021, la Tunisie est toujours en dessous des niveaux enregistrés dans des pays émergents, assez loin de l'Afrique du Sud ainsi que de la moyenne mondiale qui est de 7% dans la même année.

Selon le dernier rapport annuel publié par le Comité Général des Assurances « CGA », le marché des assurances tunisien a enregistré un taux de croissance du chiffre d'affaires significatif de 10,2%.

Ce même rapport indique que le marché tunisien des assurances est toujours prédominé par la branche automobile qui représente la police la plus vendue. C'est ainsi que, le nombre de véhicules immatriculés évolue considérablement d'une année à une autre. Cette branche représente 42,4% des primes émises comme montrée dans la figure suivante.

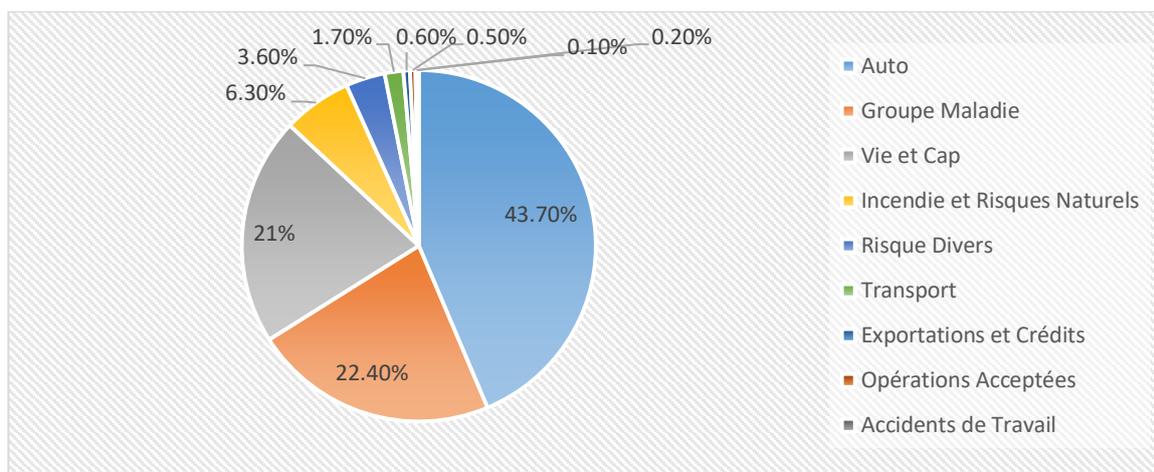
Figure 7 : Répartition des primes émises totales par branches d'assurance



Source : Rapport CGA 2021

Selon le même rapport, le taux de sinistres réglé a connu une augmentation remarquable de 21,3 % contre une baisse importante de 5,1% en 2020. La branche automobile domine avec un taux d'indemnisation que représente 43,7% du total des indemnisations de secteur.

Figure 8 : Répartition des indemnités réglées par branches d'assurance



Source : Rapport CGA 2021

Le marché de l'assurance automobile en Tunisie est l'un des plus importants marchés d'assurance, comme le témoignent les rapports annuels du Comité Général des Assurances (CGA) et de la Fédération Tunisienne des Sociétés d'Assurances (FTUSA). En effet, il est le marché le plus développé de l'assurance non vie. L'importance de cette branche d'assurance s'explique en grande partie par l'obligation d'assurance automobile minimale en responsabilité civile. De plus, ce marché n'a pas cessé de grandir d'une année à une autre.

1.2. La tarification de la RC automobile en Tunisie

La garantie Responsabilité Civile représente la composante principale de tout contrat d'assurance automobile, elle est notée RC et connue sous l'abréviation anglophone MTPL (Motor Third Party Liability) est une assurance obligatoire et réglementée dans plusieurs pays. Elle couvre les dommages causés aux tiers lorsque la responsabilité est engagée.

Dans le cas, d'un assuré qui provoque un accident avec son véhicule, son assureur indemniser la partie adverse. La garantie RC doit couvrir les dommages subis aux tiers tant pour leur personne c'est-à-dire des dommages corporels que pour leurs biens c'est-à-dire des dommages matériels lorsqu'un véhicule terrestre à moteur se trouve impliqué. Les sinistres peuvent être corporels et/ou des matériels. Depuis 1960, l'assurance RC automobile est obligatoire selon l'article 110 du Code des Assurances Tunisien. Cette obligation pèse sur les deux parties le propriétaire du véhicule et l'assureur.

Le système de tarification de la RC automobile utilisé en Tunisie, est basé sur l'usage, la puissance fiscale et un système bonus-malus. Dans ce cas, cette tarification est basée sur deux types de

tarification a priori et posteriori comme nous avons vu dans la première section de la partie théorique.

1.2.1 La tarification a priori de la RC automobile en Tunisie

L'usage du véhicule représente l'activité dans laquelle le véhicule assuré est utilisé et la puissance fiscale du véhicule est exprimée en nombre de chevaux ou le nombre de places pour les véhicules destinés au transport public. La segmentation actuelle en fonction de l'usage et la puissance sur laquelle reposent les tarifs obligatoires n'est pas à jour.

En Tunisie, trois types de tarifs automobiles sont appliqués:

- ❖ Tarif administré fixé par le ministère des Finances.
- ❖ Tarif fixé en vertu des conventions conclues entre la FTUSA et certaines associations professionnelles, il s'agit essentiellement des usages suivants, taxi, louage et transport rural.
- ❖ Tarif fixé par les compagnies d'assurance dans le cadre de la FTUSA et soumis au CGA, qui concerne le cyclomoteur grosse cylindrée, les voitures de location, les ambulances et les véhicules sanitaires.

La circulaire n°1/2017 du 28 février portant la fixation du tarif de la RC automobile du fait de l'usage des véhicules terrestres à moteur et qui annule et remplace la circulaire n°1/2014 du 20 octobre 2014 portant aussi sur la fixation du tarif de l'assurance de la RC automobile du fait de l'usage des véhicules terrestres à moteur.

Selon la circulaire n°1/2017 du 28 février, le barème des tarifs de l'assurance RC pour les utilisateurs de véhicules terrestres à moteur est fixé selon l'usage et la puissance. Dans ce cas, nous disposons pour chaque usage et puissance d'une prime correspondante, dans l'annexe 3, nous présenterons le barème des tarifs de l'assurance responsabilité civile comme fixé dans la circulaire.

Ce tarif administré est fixé par le ministère des Finances selon la circulaire n°1/2017 et qui touche 90% du parc automobile et qui concerne les usages suivants:

- ✓ Usage privé: (promenade et affaires) le véhicule est utilisé dans les déplacements privés ou professionnels de l'assuré,
- ✓ Usages utilitaires: le véhicule est utilisé dans des activités commerciales.
- ✓ Usage agricole: camionnette destinée aux activités agricoles (quitus fiscal obligatoire)
- ✓ Usage engins de chantier et engins agricoles
- ✓ Usage 2 roues : inférieur ou supérieur à 125 cm

1.2.2 La tarification a posteriori de la RC automobile en Tunisie

Le système bonus-malus consiste à appliquer un coefficient de réduction et de majoration de la prime RC automobile dont le but de favoriser la prévention routière et sanctionner les mauvais conducteurs pour une période de 12 mois consécutifs.

La dernière version de ce système a été mise en vigueur depuis mai 2019. Ce système est applicable pour tous types d'usage à l'exception des deux roues. La prime d'assurance est déterminée en multipliant la prime de base pour la responsabilité civile hors taxes par un coefficient de réduction ou de majoration tel que présenté dans l'annexe 4.

En 2019, le CGA a créé une centrale des risques bonus-malus (CRBM). Il s'agit d'une plateforme chargée de gérer le système bonus-malus et de suivre son application par toutes les compagnies d'assurance tunisiennes pour les contrats individuels. Elle réunit les données relatives aux automobiles assurées et aux sinistres et permet aux assureurs d'accéder aux données des conducteurs comme l'historique des accidents, la classe bonus-malus à appliquer.

Les classes relatives aux nouvelles souscriptions se feront en fonction d'une consultation de la centrale des risques auprès du CGA sont les suivantes:

- ✓ Classe 8 (200%) pour l'usage privé et classe 5 (150%) pour les autres usages et pour les assurés dont le permis de conduire date de moins de 2 ans ou ceux dont le de conduire date de plus de 2 ans mais qui ne peuvent justifier leur souscription d'assurance antérieure,
- ✓ Classe 4 pour l'usage affaire (100 %) pour un 2ème véhicule,
- ✓ Classe 3 pour les autres usages (100 %) pour un 2ème véhicule,
- ✓ Classe 4 (100%) si l'assuré prouve qu'il possédait une voiture de fonction.

Ce système fonctionne de la manière suivante :

- ✓ Descente d'une classe si aucun sinistre dont la responsabilité totale ou partielle est engagée n'a été déclaré durant une période de 2 années consécutives
- ✓ Montée d'une classe pour tout sinistre ayant entraîné uniquement des dommages matériels,
- ✓ Montée de 2 classes pour tout sinistre ayant entraîné des dommages corporels accompagnés ou non des dégâts matériels, et de 3 classes pour chaque autre sinistre ayant entraîné des dommages corporels accompagnés ou non des dégâts matériels survenus durant la même année.

À ce niveau, il ne faut pas oublier les taxes et les fonds de l'assurance automobile qui sont les suivants :

- ✓ TUA : La Taxe Unique des Assurances (12%)
- ✓ FGAA: Fonds de Garantie des Victimes des Accidents de la Circulation (2%)
- ✓ FSSR: Fonds Spécial de la Sécurité Routière (0,300 d)
- ✓ FPAC: Fonds de Prévention des Accidents de la Circulation (0,500 d)
- ✓ FGA: Fonds de Garantie des Assurés (3 d)

1.3. Les enjeux réglementaires et les perspectives futures

L'étude réalisée par le CGA en 2015 et en collaboration avec la Banque Mondiale, a présenté le problème principal de l'assurance automobile en Tunisie qui est la sous tarification obligatoire de la garantie RC automobile. Cette situation a engendré un déséquilibre structurel de la RC qui se traduit, par des ratios des sinistres sur primes largement supérieures à 100%.

Un faible montant des primes RC réglementée, pour les différentes catégories de véhicules contre une hausse croissante des coûts. En général, ce secteur est connu par les longs délais de règlements des sinistres ce qui a engendré des problèmes de manque de clarté et de confiance entre les assureurs et les assurés. (Ghali, 2001)

Afin de lutter contre ce déséquilibre structurel les assureurs ont trouvé une compensation à travers une sur-tarification des garanties annexes, non obligatoires qui entravent leur développement et entraînent des comportements commerciaux de vente conditionnée de ces garanties. (CGA, 2015)

La situation alarmante de la RC automobile en Tunisie a poussé l'autorité de tutelle (le CGA) en 2015 à penser à un plan pour une perspective d'une réforme de l'assurance automobile. Premièrement, l'une des solutions proposées pour combler le déficit est l'adoption des augmentations progressives des tarifs. Actuellement, ce tarif est resté inchangé depuis 2017 contre des augmentations des coûts, par conséquent une inadéquation entre l'évolution des primes et celles des coûts.

Deuxièmement, suite aux réclamations des assureurs et aux recommandations des institutions internationales comme la Banque Mondiale, le CGA cherche à étudier la possibilité de mettre en place une libéralisation des primes de l'assurance RC auto. Selon le CGA, cette libéralisation doit se faire progressivement et avec un contrôle régulier. La progressivité de cette décision est primordiale pour mettre en place un cadre bien déterminé permettant la couverture des assurés.

Les expériences internationales dans la libéralisation du secteur d'assurance ont bien montré les effets d'une telle décision dans le secteur. Selon une étude effectuée par la Banque Mondiale en 2009 sur la libéralisation du secteur d'assurance dans le monde, les meilleures pratiques dans le processus de libéralisation sont d'établir un plan d'action composé de trois étapes:

❖ **Étape 1 : Élaboration d'un calendrier**

L'élaboration d'un calendrier permet de définir le délai nécessaire pour parvenir à un marché totalement libéralisé. Tout processus de libéralisation nécessite une période de temps relativement longue pour parvenir à un accord entre les parties notamment le gouvernement, les compagnies d'assurance, les autorités de contrôle, les associations professionnelles et les associations de protection des consommateurs, si elles existent dans le pays.

Par exemple depuis 1988, la directive européenne a annoncé le principe de la liberté de fixer les tarifs en assurance automobile dans tous les pays de l'UE. Pourtant, les différents pays ont poursuivi la libéralisation à des moments différents. L'Italie n'a commencé à discuter de la libéralisation des tarifs qu'en 1990 et qu'en 1994, une loi ministérielle a annulé les tarifs administrés.

❖ **Étape 2 : Étudier les réactions et les actions des différents acteurs**

Cette étape a pour but d'éviter les conflits entre les acteurs. Généralement, les différentes parties prenantes ont des points de vue différents, souvent contradictoires. Ces conflits doivent être résolus afin de limiter les problèmes à long terme par exemple entre les compagnies d'assurance et les associations de protection des consommateurs.

❖ **Étape 3 : Prévoir l'effet possible d'une telle décision sur les prix**

L'histoire de la tarification de l'assurance dans les pays de l'UE montre que les primes augmentent généralement après la libéralisation. À long terme, cependant, la libéralisation devrait entraîner une baisse progressive des prix.

D'après le même rapport, la libéralisation de la RC automobile n'a pas eu d'effet négatif sur le secteur. Au contraire, elle a fourni aux assureurs des incitations en matière d'efficacité et d'innovation. Cette libéralisation a entraîné une forte concurrence entre les compagnies, voire même des opérations de fusions et des acquisitions entre eux et a permis au consommateur de mieux comprendre le lien entre son comportement et son prix d'assurance.

D'après ce qui précède nous pouvons dire que la libéralisation des tarifs peut être considérée comme une solution de la situation actuelle de la RC automobile en Tunisie.

Section 2 : GAT ASSURANCES et les chiffres clés

2.1 Présentation de la GAT Assurances

GAT Assurances est une entreprise d'assurance et de réassurance opérant sur le marché tunisien depuis 1975, avec un capital social de 45 millions de dinars tunisien. Elle était classée toujours parmi les premières compagnies, en s'appuyant sur ses valeurs humaines, ses engagements respectés et la richesse de ses produits et de son expertise. C'est une société privée qui est représentée dans la plupart des régions de la Tunisie. Nous présentons ci-dessous quelques indications concernant la compagnie de parrainage.

Tableau 6: Fiche de présentation de la GAT ASSURANCES

| Dénomination | GAT ASSURANCES |
|------------------|---|
| Logo |  |
| Activités | Toutes branches d'assurances et de réassurances |
| Siège social |  92-94 avenue Hedi Cheker-1002 Tunis-Bélvédère |
| Date de création | Le 18/07/1975 |
| Forme juridique | Société anonyme de droit tunisien |
| Capital social | 45 000 000 DT |
| Site web | www.gat.com.tn |

Source : Site Web de la GAT Assurances

2.2 Positionnement de la GAT Assurances

Selon le dernier classement des compagnies en fonction de leurs chiffres d'affaires en 2021, la société GAT Assurances occupe la troisième place.

Tableau 7 : Classement des trois premières compagnies d'assurances selon le chiffre d'affaires en M.D

| Société | Chiffre d'affaires 2020 | Chiffre d'affaires 2021 | Taux d'évolution |
|---------|-------------------------|-------------------------|------------------|
| STAR | 361,4 | 368,2 | 1,9% |
| COMAR | 223,0 | 233,3 | 4,6% |
| GAT | 188,6 | 218,4 | 15,8% |

Source : Rapport CGA 2021

GAT Assurances a réalisé avec un chiffre d'affaires en 2021 de 218,4 MD, soit une croissance annuelle de 15,8% par rapport à l'année 2020 (188,6 MD).

2.3. Les principaux indicateurs de la GAT Assurances

Les primes totales émises en 2021 ont augmenté avec un taux de croissance égal à 15,80% contre une évolution de 10,29% entre 2019 et 2020.

Au niveau des indemnisations, l'année 2021 est connue par une évolution très importante par rapport à l'année précédente, de 22,47% contre une baisse enregistrée l'année précédente de 15,65%.

Les provisions techniques constituées pour faire face aux engagements envers leurs assurés ont enregistré une croissance au taux de 6,06% contre 5,34% en 2020.

GAT Assurances a dégagé en 2021 un résultat technique bénéficiaire d'ordre de 24,991 MD contre 16,93 MD l'année précédente.

Parallèlement à ce résultat technique bénéficiaire, l'exercice comptable 2021 a été clôturé avec un résultat net de 23,636 MD contre 15,579 MD en 2020.

Les placements cumulés de la GAT Assurances ont atteint 364,856 MD à la fin de l'année 2021 contre 325,455 MD en 2019, progressant ainsi de 12,11%.

Tableau 8: Les indicateurs clés de GAT Assurances en M.D

| Indicateur | 2019 | 2020 | Taux d'évolution 2019-2020 | 2021 | Taux d'évolution 2020-2021 |
|------------------------------|-------------|-------------|-----------------------------------|-------------|-----------------------------------|
| Chiffre d'affaires | 171 | 188,603 | 10,29% | 218,4 | 15,80% |
| Indemnisations payées | 121,031 | 102,095 | -15,65% | 125,039 | 22,47% |
| Provisions techniques | 270,41 | 284,856 | 5,34% | 302,122 | 6,06% |
| Résultats techniques | 11,185 | 16,93 | 51,36% | 24,991 | 47,61% |
| Résultats comptables | 12,85 | 15,579 | 21,24% | 23,636 | 51,72% |
| Placements | 296,025 | 325,455 | 9,94% | 364,856 | 12,11% |

Source : Rapport CGA, 2019, 2020 et 2021

D'après le tableau 8, nous remarquons l'évolution des principaux indicateurs du GAT Assurances durant les trois dernières années. Ainsi, l'année 2021 a dégagé une évolution très sensible des indemnisations payées.

2.4. Focus sur la branche automobile

La branche automobile monopolise plus de 50% du chiffre d'affaires de la GAT Assurances. Le tableau ci-dessous représente les principaux indicateurs de cette branche :

Tableau 9 : Les indicateurs clés de la branche automobile de la GAT Assurances en M.D

| Indicateur | 2018 | 2019 | Taux d'évolution 2018-2019 | 2020 | Taux d'évolution 2019-2020 |
|------------------------------|-------------|-------------|-----------------------------------|-------------|-----------------------------------|
| Chiffre d'affaires | 78,129 | 82,859 | 6,05% | 104,994 | 26,71% |
| Indemnisations payées | 46,282 | 53,166 | 14,87% | 48,583 | -8,62% |
| Provisions techniques | 158,451 | 164,004 | 3,50% | 182,188 | 11,09% |
| Résultats techniques | 9,524 | 11,101 | 16,56% | 16,717 | 50,59% |

Source : Rapport CGA, 2018, 2019 et 2020

D'après ce tableau, nous remarquons une évolution remarquable du chiffre d'affaires de 26,71%. Cette branche est marquée par un résultat technique positif.

Section 3 : Traitement préliminaire des données de l'étude empirique

3.1. La méthodologie appliquée

L'objectif de la méthodologie consiste à étudier la tarification de la responsabilité civile de la branche automobile au sein l'entreprise de parrainage (GAT Assurances) d'une part, et de tester l'existence d'autres critères tarifaires qui doivent être pris en considération pour la détermination des tarifs autres que l'usage, la puissance fiscale et le système bonus-malus.

Pour ce faire, la méthodologie à appliquer est la suivante, nous allons proposer un cadre pour la mise en place d'un modèle de tarification de la RC automobile en intégrant des variables et en se basant sur le principe du tarif égale à «Fréquence \times Coût» où la fréquence et le coût sont estimés par le Modèle Linéaire Généralisé (GLM).

La tarification exige une transformation des variables quantitatives contenues en classes exploitables et des variables qualitatives qui ayant un nombre important de modalités en nombre réduit afin de disposer d'un tarif homogène pour chaque classe de risque.

Dans ce cas, avant de commencer la modélisation il nécessaire de modifier et de regrouper ces variables. Il faudrait également, effectuer une analyse univariée et bivariée entre les variables. L'analyse bivariée est réalisée à travers des tests entre les variables explicatives d'une part et entre chaque variable explicative potentielle avec la variable à expliquer pour disposer d'une meilleure qualité d'ajustement.

Le modèle GLM consiste à étudier la liaison entre une variable dite à expliquer Y et un ensemble de variables explicatives x_1, \dots, x_p . Il s'agit, dans ce cas, de faire une hypothèse sur la loi suivie par les observations de la variable à expliquer. Les lois choisies dans l'assurance automobile, sont des lois appartiennent à la famille exponentielle, qui regroupe de nombreuses lois classiques (Binomiale, Poisson, Gamma, Normale, etc.). Nous avons déjà exposé dans le premier chapitre les étapes à respecter pour appliquer la modélisation GLM.

Le modèle GLM réside sur la séparation entre l'étude de la fréquence et du coût moyen. La première variable que nous cherchons à expliquer est la fréquence de sinistres. Nous rappelons que la fréquence observée est égale au nombre de sinistres déclarés divisés par l'exposition, car nous cherchons une fréquence exprimée par année. Pour la modélisation du coût, la même démarche sera adoptée pour le deuxième modèle.

Après la validation de la loi suivie, nous allons passer à l'estimation des coefficients attribués à chaque modalité de chaque variable. Suite aux résultats des deux modèles, fréquence et coût nous allons présenter le modèle de tarification qui nous permet de calculer la prime pure.

En bref, la méthodologie est appliquée à travers la réalisation des étapes suivantes:

- 1) Présentation de la base de données
- 2) Traitement des variables (regroupement de variables qualitatives à plusieurs modalités et transformations des variables quantitatives) ;
- 3) Analyse univariée et bivariée des variables ;
- 4) Le choix de la loi de probabilité (Poisson ou Binomiale Négative pour les fréquences et Log-Normale ou Gamma pour les coûts moyens) ;
- 5) L'estimation des coefficients du modèle GLM;
- 6) La sélection finale des variables;
- 7) La validation du modèle estimé
- 8) La dernière étape s'intéresse à faire une synthèse des résultats

3.2 Présentation de la base de données

La phase de constitution de la base de données est cruciale pour toute démarche de modélisation. La plupart des assureurs sont conscients de la nécessité de disposer des données en bonne qualité. La constitution des bases de données figure parmi les plus importantes préoccupations des actuaires. (Denuit et Charpentier, 2009)

La base de données de l'étude contient des informations provenant du portefeuille d'assurance automobile de la compagnie GAT assurances. Ces données ont été fournies sur deux fichiers différents. Un fichier relatif à un extrait de la production de l'année 2019 (afin d'éliminer l'effet du COVID-19) et qui contient des informations sur le contrat, l'assuré et son véhicule dont lequel chaque ligne correspond à un véhicule ayant souscrit la garantie RC automobile.

Le deuxième fichier représente un extrait des sinistres de l'année 2019. Ces fichiers concernent le transport de marchandises pour son propre compte et dont le poids est inférieur à 3,5 tonnes (usage utilitaire).

L'élaboration d'un tarif exige un appui statistique qui permet de relier le montant des sinistres aux caractéristiques du véhicule et de l'assuré. Ce tarif provient souvent d'un extrait de la production et la sinistralité.

Avant de commencer la démarche de modélisation, il est nécessaire d'assurer à la qualité des données à travers l'épure de la base brute. Le fichier initial contient 2682 polices. Après l'épuration notre échantillon final contient n= 2351 polices.

❖ **Le fichier production**

L'extrait de la base de données production est composé de 16 variables relatives à la police l'assuré et à son véhicule à savoir :

Tableau 10 : Liste des variables du fichier production

| Liste des variables relatives à la police | Liste des variables relatives à l'assuré | Liste des variables relatives au véhicule |
|--|--|--|
| -Numéro police -Numero Quittance -Date Début Situation -Date Fin Situation -Libellé Garantie - Code Garantie -Usage Véhicule -Prime | -Région | -Numéro Matricule -Marque Véhicule -Classe Bonus Malus -Poids Total Véhicule -Charge Utile Véhicule -Puissance Véhicule -Age du véhicule |

❖ **Le fichier sinistre**

Nous disposons également des informations sur l'historique de la sinistralité de l'assuré durant la période d'observation qui contient 11 variables à savoir :

- ✓ Numéro police
- ✓ Numéro de quittance
- ✓ Numéro Matricule
- ✓ Num Sinistre
- ✓ Date Survenance
- ✓ Etat Sinistre
- ✓ Nature Sinistre
- ✓ Resp Sinistre
- ✓ Usage Véhicule
- ✓ Code Garantie
- ✓ Coût Sinistre

Après avoir présenté les deux fichiers, nous avons effectué un rapprochement entre les deux pour construire la base de données à traiter, nous avons laissé seulement les variables tarifaires pour l'étude qui sont présentées dans le tableau ci-dessous :

Tableau 11 : Liste préliminaire des variables tarifaires

| Variable | Type et nature | Nature | Nb de valeurs ou de modalités |
|---------------------------------------|-----------------------|---------------|--------------------------------------|
| M : marque de véhicule | Qualitative | Nominale | 10 |
| BM : système bonus-malus | Quantitative | Continue | 7 |
| PV : puissance du véhicule | Quantitative | Continue | 9 |
| PTV : poids totale du véhicule | Quantitative | Continue | - |
| CU : charge utile | Quantitative | Continue | - |
| A : Age du véhicule | Quantitative | Continue | 23 |
| R : Région | Qualitative | Nominale | 5 |

D'après ce tableau, nous constatons la nécessité de transformer les variables quantitatives (BM, PV, CU, A) en variables qualitatives et d'effectuer des regroupements pour la variable M afin de réduire le nombre de modalités. Nous présentons dans l'annexe 5 les acronymes des variables retenues et leurs modalités.

3.3 Traitement des variables

La tarification d'un point de vue actuarielle, consiste à présenter un montant de prime pour chaque classe de risque homogène. Dans cette logique, il est nécessaire de regrouper un ensemble de modalités dans un même groupe ayant des caractéristiques homogènes. De même, le modèle GLM exige la transformation des variables quantitatives contenues en des classes exploitables.

Il existe plusieurs méthodes de classification dont nous allons utiliser deux variantes afin de répondre au problème posé. Le choix de la méthode de classification à utiliser est en fonction du type de variable.

- ✓ La méthode des K-means faisant partie des méthodes de classification par partitionnement pour les variables quantitatives,
- ✓ La méthode de Classification Hiérarchique Ascendante (CAH), qui s'inscrit parmi les méthodes de classification hiérarchique pour les variables quantitatives.

3.3.1 La méthode de regroupement par K-means

Le regroupement par K-means est une méthode statistique qui permet de diviser un ensemble de données en groupes homogènes. Cette méthode est généralement utilisée pour classer les

variables quantitatives continues en groupes, à travers de nombreux algorithmes permettant le partitionnement des données. Dans le cadre de ce mémoire, nous allons partitionner les variables quantitatives avec l’algorithme des K-means.

✓ **Application : Partitionnement de l’âge du véhicule**

La variable A (âge du véhicule) est une variable quantitative continue. Dans cette application, nous allons la transformer en variable qualitative à travers la création des classes d’âge de véhicules par la méthode de K-means.

Nous rappelons des statistiques descriptives de cette variable :

Figure 9 : Statistiques descriptives de la variable âge du véhicule

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|--------|
| 0.000 | 1.000 | 3.000 | 3.396 | 5.000 | 22.000 |

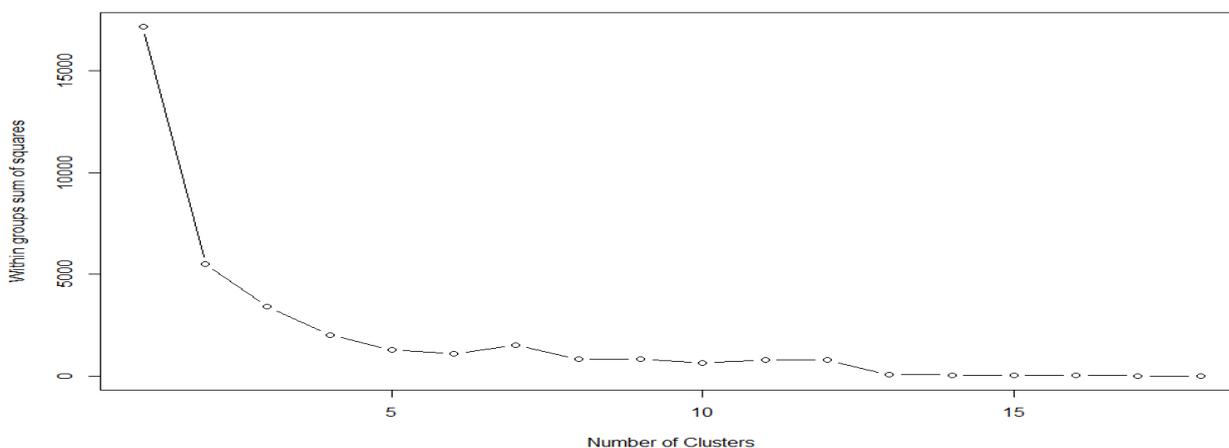
Source : Sortie du logiciel R

Nous remarquons que la moyenne et la médiane sont proches ce qui est témoin d’une distribution d’âge de véhicule symétrique et une population équilibrée.

Le nombre de classes par la méthode de K-Means se fait par le choix des différentes valeurs de K et de calculer la variance des différentes classes. Ainsi, on cherche à trouver un nombre de clusters K qui minimise la distance entre leurs centres appelés la minimisation de la distance intra-classe.

En pratique, nous allons fixer le nombre classe à l’aide du logiciel R par une présentation graphique à travers la fonction wssplot() en reliant les différents nombres de clusters K avec la variance. Le graphique ci-dessous présente ce rapport :

Figure 10 : Variance intra en fonction du nombre de cluster K



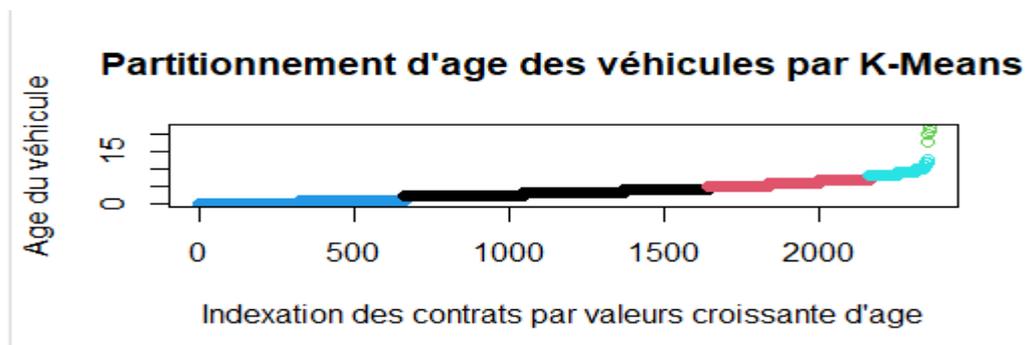
Source : Sortie du logiciel R

Interprétation :

Nous remarquons, que le graphique possède une forme de bras. Le nombre optimal de clusters est le point qui représente le coude. Dans ce cas, le coude représente par K valant 5 ($k=5$). Donc 5 peut être considéré comme le nombre optimal de clusters.

L'étape suivante consiste à chercher la délimitation des classes d'âge du véhicule à l'aide de la fonction `kmeans()` sous R. dans ce cas, l'algorithme `kmeans` est lancé 25 fois pour assurer la robustesse à l'aide de l'argument `nstart`.

Figure 11 : Partitionnement d'âge des véhicules par la méthode de K-means



Source : Sortie du logiciel R

Tableau 12 : Classement d'âge de véhicule

| Numéro de cluster | Borne inf | Borne Sup | % de l'échantillon |
|-------------------|-----------|-----------|--------------------|
| 1 | 0 | 1 | 28,15 |
| 2 | 2 | 3 | 30,14 |
| 3 | 4 | 6 | 26,74 |
| 4 | 7 | 11 | 14,71 |
| 5 | 12 | 22 | 0,43 |
| Total | | | 100 |

Les classes constituées ont une importante représentativité dans l'échantillon sauf la classe 5 qui représente un pourcentage inférieur à 5%. Dans ce cas, la classe 5 sera regroupée à la classe précédente afin de garantir un niveau suffisant par classe. Finalement, nous retenons seulement les quatre classes d'âge suivantes :

Tableau 13 : Classement final d'âge du véhicule

| Classe d'âge | Borne inf | Borne Sup | % de l'échantillon |
|--------------|-----------|-----------|--------------------|
| A1 | 0 | 1 | 28,15 |
| A2 | 2 | 3 | 30,14 |
| A3 | 4 | 6 | 26,74 |
| A4 | 7 | 22 | 15,14 |
| Total | | | 100 |

Cette démarche sera répétée pour les autres variables quantitatives, les détails de calcul sont présentés dans les annexes de 6 à 9 et les résultats sont présentés dans le tableau ci-dessous.

Tableau 14 : Classement final des variables quantitatives

| Num de classe | A | % | BM | % | PV | % | PTV | % | CU | % |
|---------------|----|-------|-----|-------|-----|-------|------|-------|-----|-------|
| 1 | A1 | 28,15 | BM1 | 33,42 | PV1 | 36,44 | PTV1 | 33,50 | CU1 | 33,50 |
| 2 | A2 | 30,14 | BM2 | 25,43 | PV2 | 29,59 | PTV2 | 27,42 | CU2 | 27,42 |
| 3 | A3 | 26,74 | BM3 | 26,19 | PV3 | 34,10 | PTV3 | 25,00 | CU3 | 25,00 |
| 4 | A4 | 15,14 | BM4 | 15,14 | - | - | PTV4 | 15,39 | CU4 | 15,39 |
| | | | | 100 | | 100 | | 100 | | 100 |

D'après le tableau 14, l'ensemble de la variable dispose de quatre modalités à l'exception de la variable PV qui désigne la puissance fiscale.

3.3.2 La méthode de Classification Hiérarchique Ascendante (CAH)

C'est une méthode qui fait partie des méthodes de classification hiérarchique. Dans la tarification, nous ne pouvons pas utiliser un nombre trop important de modalités pour une même variable car cela fait augmenter le nombre de paramètres à estimer.

Parmi les variables disponibles qui ont un nombre important de modalités dans notre cas nous avons la variable M qui désigne la marque de véhicule. Afin de répondre à cette situation, nous allons utiliser la méthode de CAH dont le but consiste à regrouper l'ensemble des modalités présentant des similarités au sein d'une même classe.

Cette variable qualitative sera attribuée aux variables à expliquer, à savoir :

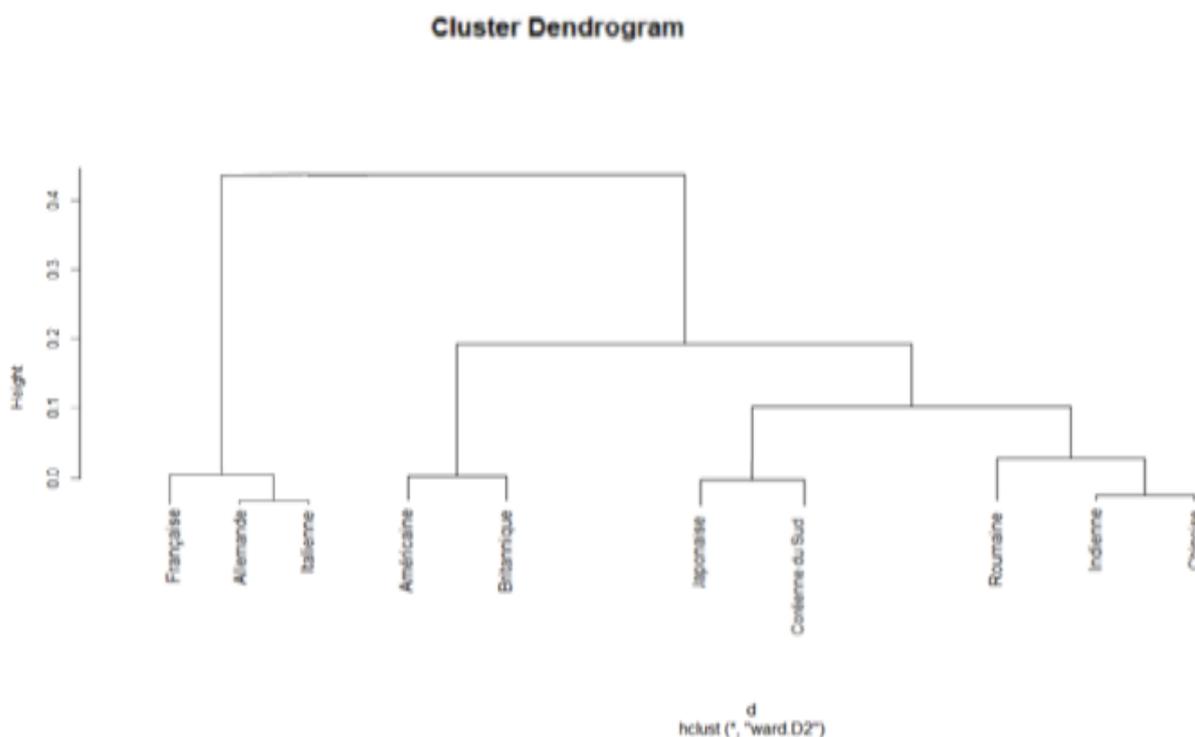
- ✓ les fréquences dans le cas où la variable à expliquer est la fréquence
- ✓ les coûts moyens dans le cas où la variable à expliquer est le coût moyen.

Application de la méthode de CAH pour la variable M :

Nous avons dans la variable M 10 modalités. Afin d'éviter de retrouver ce nombre important de classes, nous allons appliquer la méthode de CAH. En pratique, cette méthode sera utilisée à travers la méthode de Ward qui est une distance euclidienne pour l'obtention du dendrogramme. Donc, il s'agit d'une représentation visuelle appelée arbre de classification ou dendrogramme. Cette représentation est faite sous le logiciel R à l'aide de la fonction `hclust()`.

- ✓ **Pour l'étude de la fréquence :**

Figure 12 : Dendrogramme présentant le découpage des marques de véhicules



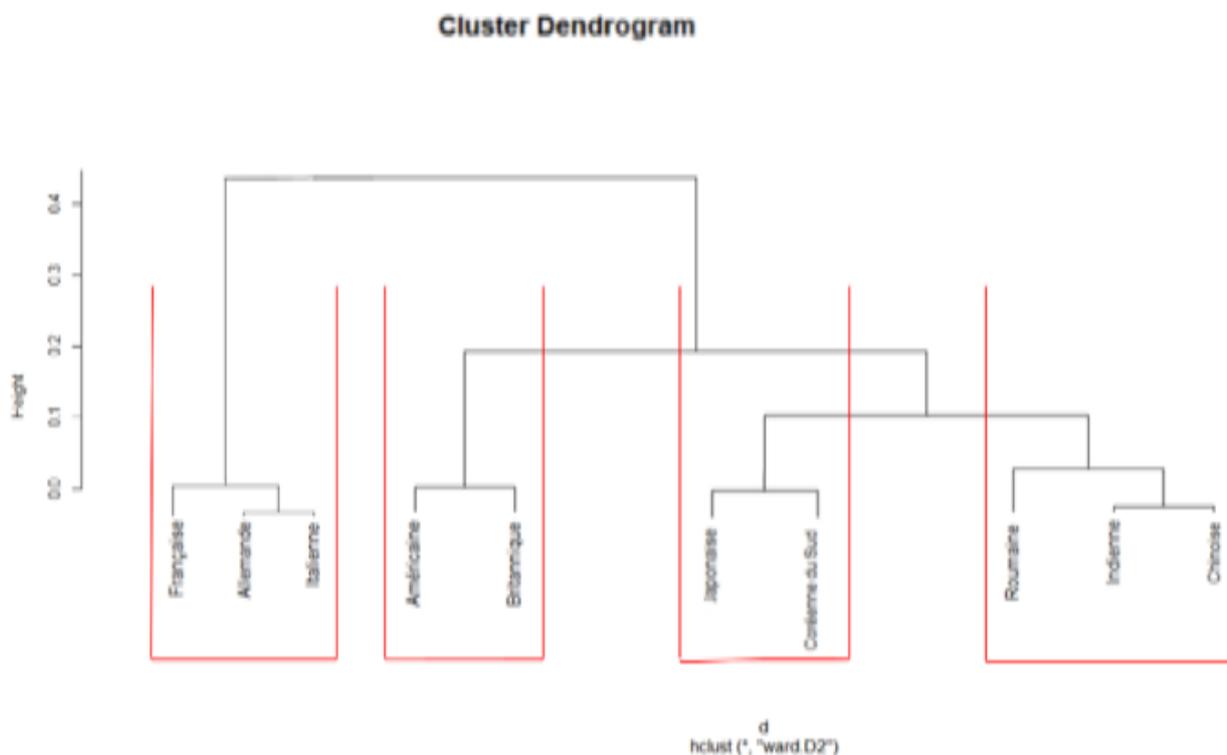
Source : Sortie du logiciel R

Interprétation :

D'après ce graphique, nous remarquons une forte perte d'inertie inter-classes est observée en passant quatre classes. Les modalités sont donc fusionnées de manière à obtenir quatre classes pour

la variable marque de véhicule. Ces différentes classes sont délimitées en rouge sur le dendrogramme ci-dessous à l'aide de la fonction `rect.hclust()` sous R

Figure 13 : Découpage en 4 classes de la marque de véhicule pour la fréquence



Source : Sortie du logiciel R

D'après cette méthode cette variable admet finalement quatre modalités : M1_f, M2_f, M3_f et M4_f pour le modèle de fréquence comme montré dans le dendrogramme. À ce niveau, nous nous sommes intéressés à savoir si l'ensemble des polices du portefeuille était réparti de manière homogène parmi les quatre nouvelles classes ou non.

Tableau 15 : Les modalités finales de la variable marque de véhicule

| Classe de la marque | % de l'échantillon |
|---------------------|--------------------|
| M1_f | 59,21 |
| M2_f | 1,74 |
| M3_f | 25,65 |
| M4_f | 13,53 |
| Total | 100 |

D'après ce tableau, nous remarquons certes l'existence d'une hétérogénéité entre les modalités et parfois même des effectifs insuffisants pour la modalité M2 car inférieur à 5 %. Cependant, en se basant sur cette méthode, nous n'allons pas regrouper les classes obtenues car celles-ci traduisent l'écart réel de fréquence observée entre les différentes marques de véhicules.

✓ **Pour l'étude du coût moyen :**

La même méthode sera appliquée d'une manière similaire sur le coût moyen. Le résultat de cette méthode indique le même nombre de classes quatre et le même regroupement par rapport à la fréquence. Donc, la variable marque est la même modalité pour les deux modèles.

Le tableau ci-dessous représente les variables avant et après traitement :

Tableau 16 : Présentation des variables avant et après traitement

| Variable | Avant traitement | | | Après traitement | | |
|------------|------------------|----------|-------------------------------|------------------|----------|----------------|
| | Type et nature | Nature | Nb de valeurs ou de modalités | Type et nature | Nature | Nb de modalité |
| M | Qualitative | Nominale | 10 | Qualitative | Nominale | 4 |
| BM | Quantitative | Continue | 7 | | | 4 |
| PV | Quantitative | Continue | 9 | | | 3 |
| PTV | Quantitative | Continue | - | | | 3 |
| CU | Quantitative | Continue | - | | | 3 |
| A | Quantitative | Continue | 23 | | | 4 |
| R | Qualitative | Nominale | 5 | | | 5 |

Ce traitement nous a permis de rendre les variables exploitables pour la modélisation. Nous avons réduit le nombre des modalités pour la variable Marque et nous avons effectué des regroupements pour les variables quantitatives.

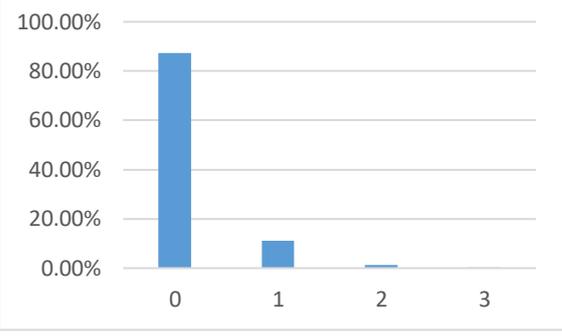
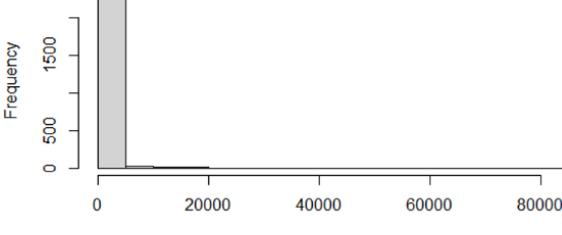
3.4 Analyse univariée et bivariée des variables

3.4.1 Analyse univariée

Suite aux traitements effectués ci-dessus, nous disposons à présent exclusivement des variables qualitatives que nous essayons d'étudier des différentes modalités possibles par variable.

Tableau 17 : Analyse univariée des variables

| Variable | Interprétation | Représentation graphique |
|-------------------------------|---|--|
| La marque | La modalité M1 représente la modalité la plus importante avec presque 60% et qui regroupe les voitures Françaises, Allemandes et Italiennes. | <p>Détails du graphique circulaire :</p> <ul style="list-style-type: none"> M1 : 59% M2 : 2% M3 : 26% M4 : 13% |
| Le système bonus-malus | La modalité BM1 contient les véhicules qui ont une classe de système bonus-malus une et qui représente la modalité la plus représentative dans l'échantillon. | <p>Détails du graphique à barres :</p> <ul style="list-style-type: none"> BM1 : 33% BM2 : 25% BM3 : 26% BM4 : 15% |
| Puissance fiscale | De même, la modalité PV1 est la modalité la plus représentative dans l'échantillon. | <p>Détails du graphique à barres :</p> <ul style="list-style-type: none"> PV1 : 36% PV2 : 29% PV3 : 34% |
| La région | Les 24 gouvernorats de la Tunisie ont été répartis sur cinq régions, Cap Bon, Grand Tunis, Centre, Nord-Ouest et le Sud. La région R2 qui est la région de Grand Tunis représente presque la moitié de l'échantillon. | <p>Détails du graphique circulaire :</p> <ul style="list-style-type: none"> R1 : 4.13% R2 : 48.62% R3 : 4.21% R4 : 29.39% R5 : 13.65% |
| Age du véhicule | Nous remarquons d'après le graphique que la modalité A4 est la moins représentative et qui concerne les véhicules dont l'âge varie entre 7 ans et plus. | <p>Détails du graphique à barres :</p> <ul style="list-style-type: none"> A1 : 28% A2 : 30% A3 : 26% A4 : 15% |

| | | |
|-------------------------------|--|--|
| Nombre de sinistre | Dans l'échantillon, le nombre de sinistres annuel observé est variable qui varie entre 0 et 3 sinistres. De plus nous remarquons que presque 90% des contrats ne sont pas sinistrés. |  |
| Montants des sinistres | Les montants de sinistres varient entre 0 et 82.236 DT. La plupart des contrats ne sont pas sinistrés ce qui explique que plus de la majorité des montants égale à 0. |  |

Ce tableau décrit de façon synthétique les variables de l'échantillon du portefeuille de la GAT Assurances. Nous faisons cette analyse descriptive des variables en vue d'identifier leur lien potentiel avec le tarif à proposer.

3.4.2 Analyse bivariée des variables

Avant de commencer toute étape relative à la modélisation, il est indispensable tout d'abord d'étudier les liens existant entre les variables et plus spécialement entre les variables endogènes d'un côté et les variables exogènes de l'autre côté. Pour ce faire, nous allons procéder à une analyse en vue de tester ces liens. Cette analyse se réalise en deux étapes.

La première consiste à tester si les variables explicatives sont indépendantes entre elles ou non. Ces variables seront testées deux à deux. Concernant, la deuxième étape, nous testons si les variables explicatives ont une influence ou pas sur la variable à expliquer.

3.4.2.1 Analyse bivariée entre les variables explicatives (test d'indépendance)

L'analyse bivariée est réalisée à travers une multitude de tests qui nous permettent d'étudier l'indépendance entre les variables explicatives. Le test d'indépendance Khi-deux est établi pour étudier le lien de dépendance potentielle entre les variables. Donc, nous allons recourir au test d'indépendance de Khi-deux.

Pour ce faire, nous allons exécuter la fonction `chisq.test()` sur R. Ce test est basé deux hypothèses:

Hypothèse nulle (H_0): les deux variables sont indépendantes

Hypothèse alternative (H_1) : les deux variables ne sont pas indépendantes

La décision est prise de la manière suivante si la p-value de ce test indique une valeur inférieure à 0,05 nous rejetons l'hypothèse H_0 . Nous allons effectuer ce test sur les deux variables M et BM comme la montre la figure ci-dessous :

Figure 14 : Test du Khi-deux d'indépendance entre la variable M et BM

```
> chisq.test(auto_$M ,auto_$BM)

Pearson's Chi-squared test

data: auto_$M, and auto_$BM
X-squared = 24.501, df = 9, p-value = 0.003575
```

Source : Sortie du logiciel R

À partir de ce test, nous allons construire un tableau récapitulatif qui contient les résultats de ce test pour le reste des variables.

Tableau 18 : Test du Khi-deux d'indépendance (p-value)

| | M | BM | PTV | CU | PV | A | R |
|-----|---|--------|--------|---------|---------|-----------|-----------|
| M | | 0.0003 | 0.1416 | 0.1372 | 2.2e-16 | 2.663e-07 | 0.04266 |
| MB | | | 0.0055 | 0.0080 | 0.0006 | 2.2e-16 | 0.0006737 |
| PTV | | | | 0,20125 | 2.2e-16 | 2.2e-16 | 1.921e-06 |
| CU | | | | | 2.2e-16 | 2.2e-16 | 0.0001466 |
| PV | | | | | | 2.2e-16 | 0.0003287 |
| A | | | | | | | 4.954e-06 |
| R | | | | | | | |

Source : Sortie du logiciel R

Ce test indique l'existence des p-values inférieures à 0.05, ce qui nous amène à rejeter pour ces variables l'hypothèse H_0 qui suppose l'indépendance de ces variables. Ainsi, nous constatons l'existence des p-values supérieures à 0.05 qui sont colorés en rouge ce qui signifie l'acceptation de l'hypothèse H_1 pour ces variables. Ce test permet de relever l'existence des dépendances entre quelques variables explicatives ce qui nous ramène à éliminer l'une de variable du modèle en gardant la deuxième, il s'agit dans notre cas le poids total du véhicule et la charge utile.

3.4.2.1 Analyse bivariée entre les variables explicatives et à expliquer

À présent, nous allons étudier de près l'influence des variables explicatives sur les variables à expliquer, la fréquence et le coût de sinistres. Cette étude est réalisée à travers le test Kruskal-Wallis.

Nous savons que toutes les variables explicatives sont des variables qualitatives et qui disposent des modalités varient entre trois et cinq. Nous réalisons le test de Kruskal-Wallis pour toutes les variables.

✓ Test de Kruskal-Wallis

Ce test est réalisé à l'aide de la fonction `kruskal.test()` sous R. Les hypothèses de ce test sont les suivants :

- ✓ Hypothèse nulle (H_0) : la variable explicative n'a aucune influence sur la fréquence (ou le coût) de sinistres.
- ✓ Hypothèse alternative (H_1) : la variable explicative influe sur la fréquence(ou le coût) de sinistres.

Nous allons présenter le test de la variable M et la même démarche sera faite pour toutes les autres variables.

Figure 15 : Test de Kruskal pour la variable BM

```
> kruskal.test(auto_$Freq~auto_$BM)
kruskal-wallis rank sum test
data: auto_$Freq by auto_$BM
kruskal-wallis chi-squared = 291.59, df = 3, p-value < 2.2e-16
```

Source : Sortie du logiciel R

La p-value présentée par ce test est inférieure au seuil de 5 % : nous pouvons donc affirmer que le système bonus-malus influe de manière significative sur la fréquence de sinistres. Ce résultat justifie la modélisation entre la fréquence et le système bonus-malus avec d'autres variables explicatives.

Cette étape va être effectuée sur la totalité des variables explicatives avec les deux variables à expliquer, la fréquence et le coût moyen. Le tableau ci-dessous représente les résultats de tests effectués sous le logiciel R.

Tableau 19 : Test de Kruskal

| | Fréquence | CM |
|-----------|-----------|-----------|
| BM | 2.2e-16 | 2.2e-16 |
| M | 0.0008544 | 0.0008893 |
| PV | 2.652e-14 | 1.515e-13 |
| A | 2.663e-07 | 2.543e-07 |
| R | 2.2e-16 | 2.2e-16 |

Source : Sortie du logiciel R

D'après ce tableau, nous remarquons que toutes les variables ont une influence significative sur la fréquence et le coût de sinistres avec des p-value < 0.05 donc, nous retenons toutes les variables évoquées ci-dessus.

Section 4 : Modélisation de la fréquence et du coût moyen

4.1. Élimination des sinistres graves

Avant de commencer la modélisation, il faut savoir que lors de l'occurrence d'un sinistre couvert par le contrat d'assurance automobile, l'assureur devrait indemniser son assuré en lui versant le montant du sinistre correspondant.

Cependant, certains sinistres peuvent avoir des montants très élevés et ayant un poids très important dans la charge totale de sinistre. Ces sinistres, n'ont ni une fréquence ni un coût semblable aux sinistres classiques, et donc doivent être traités de façon différente lors de la tarification.

C'est pour cette raison, que les études sur la tarification consistent à distinguer entre les sinistres classiques à ceux des sinistres graves (comme nous avons vu dans la partie théorique). Dans ce cas, une analyse des coûts des sinistres est alors primordiale. Cette analyse permet de préciser le seuil de gravité afin d'appliquer cette division sur les contrats étudiés.

Pour ce faire, nous allons utiliser les statistiques descriptives à l'aide du logiciel R et la technique de « nuage de points » pour détecter les valeurs aberrantes. Cette technique sera appliquée sur les charges des sinistres de l'échantillon pour l'année 2019 en vue de détecter les sinistres graves.

Les paramètres de position de la variable des coûts de sinistre définissent la tendance générale de la distribution. Parmi ces quatre paramètres présentés ci-dessous, notre attention se porte sur la médiane (0 dt) et la moyenne (436.2 dt) des coûts. L'écart entre ces deux valeurs témoigne d'une distribution du coût des sinistres asymétriques.

Tableau 20 : Statistiques descriptives des coûts unitaires de sinistres

| | Min | Median | Mean | Max |
|--------------------------|------------|---------------|-------------|------------|
| Avant élimination | 0 | 0 | 436,2 | 82236,6 |
| Après élimination | 0 | 0 | 345,3 | 18547,6 |

Source : Sortie du logiciel R

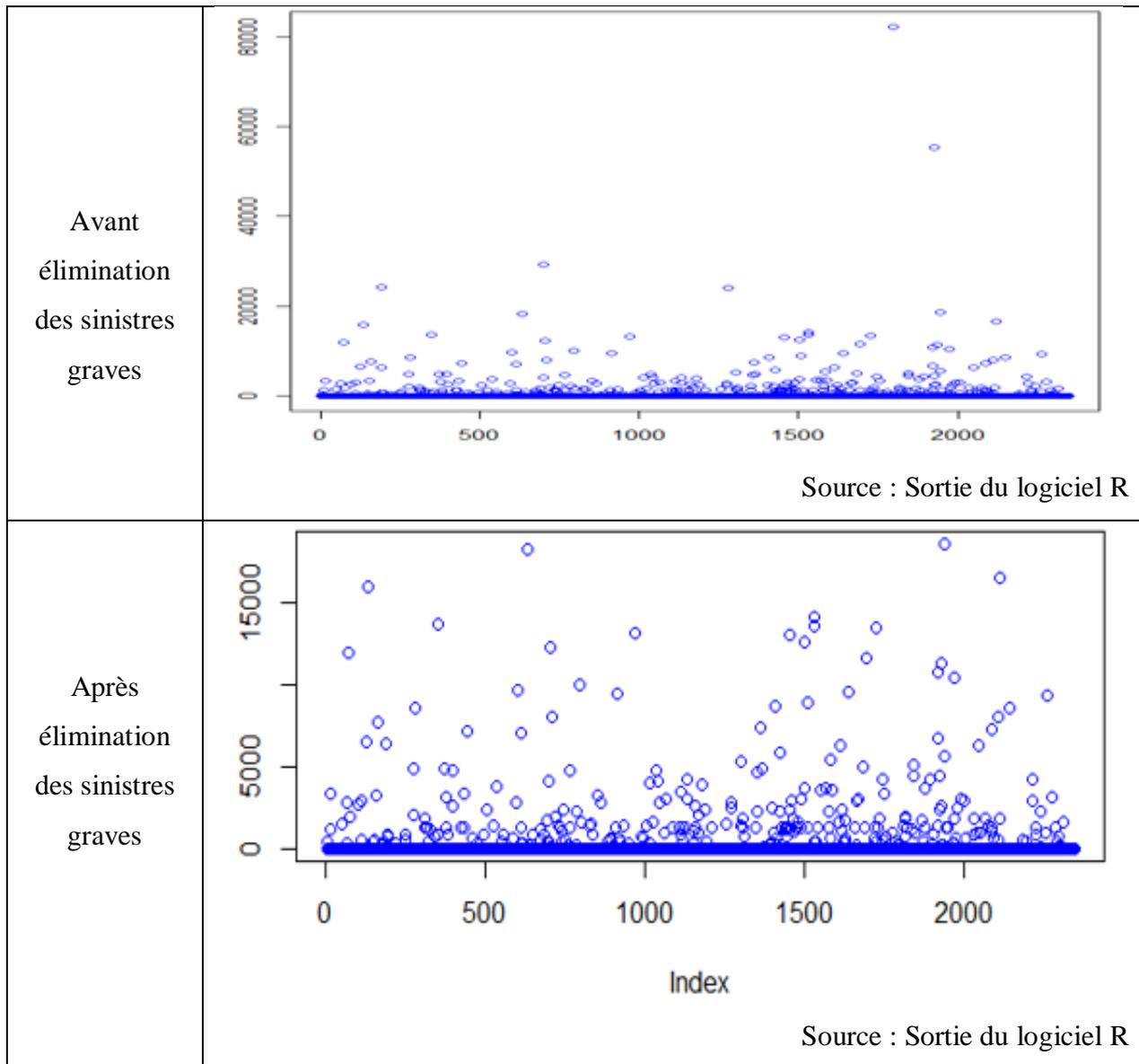
La différence entre ces deux paramètres peut être l'origine de deux possibilités soit :

- ✓ La présence de valeurs extrêmes qui pèsent à la hausse sur la moyenne
- ✓ L'importance de coûts faibles qui pèsent à la baisse sur la médiane

À ce niveau, nous devons analyser en détail la nature de la distribution du coût des sinistres pour pouvoir déterminer l'origine de cet écart. Nous observons dans la figure ci-après la présence de montants faibles et qui pèsent sur la valeur de la médiane et aussi l'existence de quelques valeurs très élevées et qui sont supérieures à 20 000 dt et leur poids est important dans la charge totale puisqu'ils influent de façon significative sur la moyenne.

Afin de résoudre ce problème, nous allons fixer à un niveau maximum, appelé seuil de gravité. Le choix de ce plafond est très délicat et peut conduire soit à une sous-estimation soit à une surestimation des sinistres classiques.

Tableau 21 : Nuage de points des coûts unitaires de sinistres avant et après élimination des sinistres graves



Nous remarquons d’après le nuage de points qu’au-dessus de 20.000 DT la dispersion de la charge des sinistres augmentent fortement. Dans ce cas, nous considérons alors 20.000 DT est le seuil de gravité et par la suite les sinistres dépassant ce seuil sont classés comme des sinistres graves et seront éliminés de l’étude pour ne pas biaiser les résultats.

4.2. La modélisation de la fréquence

La première variable que nous cherchons à expliquer est la fréquence de sinistres. Nous rappelons que la fréquence est égale au nombre de sinistres déclarés divisés par l’exposition, car nous cherchons une fréquence exprimée par année.

Comme déjà présenté dans la méthodologie, la modélisation de la fréquence suit quatre étapes :

- ✓ Choix de la loi probabilité;
- ✓ Estimation des coefficients de la régression ;
- ✓ Sélection des variables;
- ✓ Validation du modèle.

4.2.1. Étape 1 : Choix de la loi de probabilité

D'après la littérature en assurance, nous disposons deux lois de comptages qui sont généralement utilisées pour la modélisation des fréquences de sinistre comme nous avons expliqué dans la partie théorique.

- ✓ Un cas classique qui consiste à utiliser la loi de Poisson
- ✓ Un Cas de sur-dispersion des données qui consiste à utiliser la loi Binomiale Négative

Pour choisir la loi adéquate, nous allons vérifier trois critères qui vont être effectués sur les deux lois :

- ✓ Critère graphique
- ✓ Selon l'Espérance-Variance
- ✓ Selon le test du Chi-2 d'adéquation

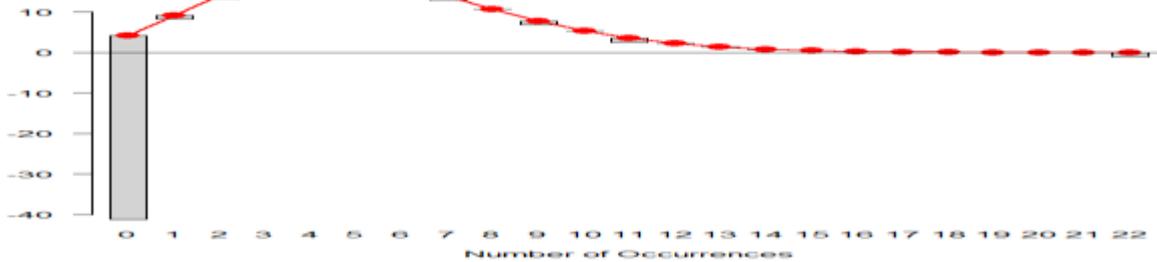
a) La loi de Poisson

- ✓ **Critère graphique**

La fonction "plot ()" du logiciel R, nous permet d'obtenir une vision graphique, qui affiche sur le même graphique les valeurs observées et les valeurs prévues par la distribution de la loi théorique avec laquelle nous testons son adéquation.

L'interprétation des graphiques se fait de la manière suivante, les points en rouge représentent la loi théorique à tester et les histogrammes représentent les fréquences observées, qui sont collées par le sommet avec la loi théorique. De plus, tout écart des histogrammes avec l'axe des abscisses indique un mauvais ajustement des observations par la loi théorique.

Figure 16 : Ajustement des fréquences selon la loi de Poisson



Source : Sortie du logiciel R

Interprétation :

À partir de ce graphique, nous remarquons l'existence :

- Des sous ajustements en 0 et 22
- Des sur ajustements entre 1 et 13
- Le reste des ajustements entre 14 et 21 sont correctement estimés.

✓ **Critère Espérance-Variance**

La loi de Poisson suppose une équi-dispersion c'est-à-dire une égalité entre l'espérance mathématique et la variance des données. Cependant, cette caractéristique n'est pas forcément compatible avec les données de fréquences en assurance automobile.

Le choix de loi est basé sur les valeurs de critères d'espérance-variance. Si N est la variable de comptage dont on cherche la loi, nous choisissons :

- ✓ La loi de Poisson si $E(N) = \text{Var}(N)$
- ✓ la loi Binomiale Négative si $E(N) \neq \text{Var}(N)$

Tableau 22 : Espérance et variance de la variable nombre de sinistres

| Espérance | Variance |
|------------------|------------------|
| 0.1402387 | 0.1581499 |

Source : Sortie du logiciel R

D'après le tableau ci-dessus, nous remarquons le voisinage de la valeur de l'espérance et à la valeur de la variance, mais nous ne savons pas si c'est statistiquement significatif ou pas. C'est pour cette raison que nous allons effectuer un test pour choisir la loi adéquate.

✓ Test du Chi-2 d'adéquation

Le test du Chi-2 d'adéquation est un test non paramétrique qui permet de déterminer si l'échantillon est distribué selon une loi choisie par nos soins ou non.

Ce test est basé sur deux hypothèses :

- ✓ H_0 : Hypothèse nulle, les données sont distribuées selon la loi indiquée
- ✓ H_1 : Hypothèse alternative les données ne sont pas distribuées selon la loi indiquée

En pratique, l'interprétation de ce test se fait de la manière suivante :

- ✓ si $p\text{-value} > \alpha$: la distribution des données est bien celle de la loi indiquée
- ✓ si $p\text{-value} < \alpha$: la distribution des données n'est pas celle de la loi indiquée

Avec $\alpha = 5\%$ pour l'ensemble des tests pour un seuil de confiance fixé de 95%

Ce test a été réalisé sous R à l'aide de la fonction `goodfit`. Cette fonction renvoie à la valeur de la statistique de test du Chi-2 ainsi que la p-value associée à ce test.

Figure 17 : Test du Chi-2 d'adéquation pour la loi de Poisson

```
Goodness-of-fit test for poisson distribution

      X^2 df P(> X^2)
Pearson 299199.5 21      0
```

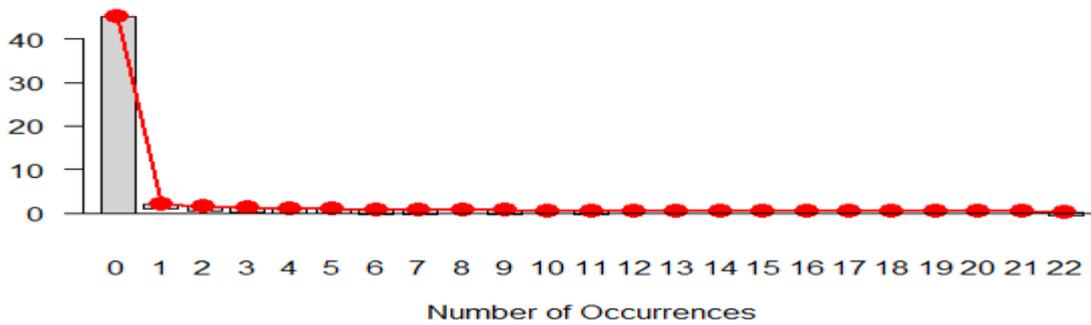
D'après ce test la p-value du test de la loi de Poisson est de 0.000 qui est inférieur à 0,05. Sur la base de cette valeur la loi de Poisson ne convient pas pour modéliser la fréquence.

En conclusion, d'après les trois critères, nous ne pouvons pas choisir la loi de Poisson. Dans ce cas, nous allons voir l'adéquation de la fréquence par la loi de Binomiale Négative selon ces critères.

b) La loi Binomiale Négative

- ✓ Critère graphique

Figure 18 : Ajustement des fréquences selon la loi Binomiale Négative



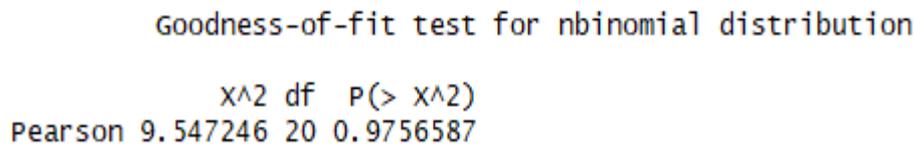
Source : Sortie du logiciel R

Interprétation :

À partir de ce graphique, nous remarquons que la majorité des ajustements de fréquences sont correctement estimés. Visuellement nous choisissons la loi Binomiale Négative car elle permet d'estimer mieux la fréquence.

✓ **Test du Chi-2 d'adéquation**

Figure 19 : Test du Chi-2 d'adéquation pour la loi Binomiale Négative



Source : Sortie du logiciel R

D'après ce test la p-value est de 0,976 qui est largement supérieure à 0,05. Sur la base de cette valeur nous choisissons la loi Binomiale Négative. En conclusion, d'après ces critères, nous retenons la loi Binomiale négative pour modéliser la fréquence de sinistre.

Une variable aléatoire Y est dite de loi binomiale négative, notée $Y \sim N \text{ Bin}(\alpha, q)$, la loi de probabilité s'écrit de la manière suivante :

$$P(K = k) = \begin{cases} C_{r+k}^{r-1} p^r (1-p)^k \\ 0, \text{ si non} \end{cases}$$

Avec une espérance et variable se calcule successivement comme suit :

$$E(k) = \frac{r(1-p)}{p} \text{ et } V(k) = \frac{r(1-p)}{p^2}$$

4.2.2. Étape 2 : Estimation des coefficients de régression

Après le choix de la loi suivie par la fréquence de sinistre, c'est au tour des coefficients attribués que chaque modalité de variable sera estimée. L'estimation des coefficients du modèle GLM est réalisée à l'aide de la fonction `glm.nb()` disponible sous R.

Il ne faut pas oublier d'intégrer l'exposition comme une variable offset. Les résultats de cette estimation sont les suivants :

Figure 20 : Estimation des coefficients de régression du modèle de fréquence

```
Call:
glm.nb(formula = auto_$Nb ~ auto_$M + auto_$BM + auto_$PV + auto_$A +
  auto_$R + offset(log(exp)), data = auto_, init.theta = 1.145138804,
  link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4704  -0.4076  -0.2588  -0.1518   3.7297

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.8660     0.2795  -10.254 < 2e-16 ***
auto_$MM2    -0.0439     0.1624   -2.487 8.83e-06 ***
auto_$MM3    -0.1273     0.2797   -0.555 2.22e-08 ***
auto_$MM4     0.8001     0.1452    5.510 < 2e-16 ***
auto_$BMBM2  0.3385     0.2308    1.467 0.142408
auto_$BMBM3  0.4370     0.1817    2.405 4.93e-12 ***
auto_$BMBM4  0.6654     0.1797    3.703 < 2e-16 ***
auto_$PVPV2  0.2638     0.1742    1.514 0.000615 ***
auto_$PVPV3  0.2938     0.1661    1.769 0.000854 ***
auto_$AA2    0.0283     0.1701    0.166 0.500747
auto_$AA3    0.1998     0.1564    1.277 0.000937 ***
auto_$AA4    0.3254     0.1694    1.921 2.88e-07 ***
auto_$RR2    1.2155     0.1617    7.517 < 2e-16 ***
auto_$RR3    0.1768     0.2293    0.770 8.52e-08 ***
auto_$RR4   -0.1552     0.2843   -0.546 0.000546 ***
auto_$RR5     0.1112     0.2502    0.444 0.031174 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.1451) family taken to be 1)

Null deviance: 1455.4 on 2350 degrees of freedom
Residual deviance: 1148.5 on 2333 degrees of freedom
AIC: 1819.6
```

Source : Sortie du logiciel R

Interprétation du résultat du modèle :

- ✓ Estimate : représente les coefficients à estimer associés à chaque modalité.
- ✓ Std. Error : représente l'estimation de l'écart-type associé à chaque modalité.
- ✓ z value : représente la variable du test la statistique elle est obtenue en divisant le coefficient de régression par son écart-type (estimate/Std.error).
- ✓ Pr(>|z|) : représente la probabilité critique du z value dans un intervalle de confiance généralement 5%. L'interprétation de cette probabilité ne nous permet pas de juger de la significativité d'une variable mais d'une modalité.

- ✓ Nous remarquons l'absence d'une modalité pour chaque variable. Cette dernière est appelée la modalité de référence, car son coefficient vaut 0 par construction et sa valeur est fixe.
- ✓ Intercept : c'est le terme constant dans la modélisation.

4.2.3 Étape 3 : Procédure de sélection des variables

Pour réaliser cette étape, nous allons utiliser la fonction Backward sous R comme une méthode de sélection finale des variables. Elle nous permettra d'éliminer les variables les moins significatives.

Figure 21 : Procédure de sélection des variables du modèle de fréquence

```
> step=stepAIC(regBnlog,dir="backward")
Start: AIC=1819.6
auto_$Nb ~ auto_$M + auto_$BM + auto_$PV + auto_$A + auto_$R +
offset(log(exp))
```

| | Df | Deviance | AIC |
|-------------|----|----------|------|
| <none> | | 1185 | 1820 |
| - auto_\$PV | 2 | 1992 | 1825 |
| - auto_\$A | 3 | 1199 | 1831 |
| - auto_\$R | 4 | 1233 | 1836 |
| - auto_\$M | 3 | 1135 | 1952 |
| - auto_\$BM | 3 | 1146 | 1955 |

Source : Sortie du logiciel R

D'après cette procédure, nous remarquons que toutes les variables explicatives sont significatives pour la modélisation de la fréquence.

✓ Application numérique

Ces coefficients permettant de retrouver une fréquence moyenne de tous les individus par classe tarifaire déjà créée. Cinq variables tarifaires ont été retenues et la fonction de lien utilisée est la fonction log, nous avons alors :

$$\log[E(Nb|C > d)] = \beta_0 + \beta_1 M + \beta_2 BM + \beta_3 PV + \beta_4 A + \beta_5 R + 1 \times \log(\text{exposition})$$

⇔

$$(Nb |C > d) = \text{exposition} e^{(\beta_0 + \beta_1 M + \beta_2 BM + \beta_3 PV + \beta_4 A + \beta_5 R)}$$

Nous cherchons une fréquence annuelle moyenne, donc le niveau d'exposition égale à l'unité.

De ce fait, l'équation finale s'écrit de la manière suivante:

$$(fréquence|C > d) = e^{(\beta_0 + \beta_1 M + \beta_2 BM + \beta_3 PV + \beta_4 A + \beta_5 R)}$$

Par exemple, si un individu détient les caractéristiques suivantes, une marque de véhicule M2, appartient à un SBM, BM2, puissance du véhicule PV2, ayant un véhicule de classe d'âge A2, et sa région est R3. La fréquence moyenne annuelle de sinistre égale à :

$$(fréquence|C > d) = e^{(-2,8660-0,0439+0,3385+0,2638+0,0283+0,1768)} = 12,21\%$$

Pour un autre individu ayant les mêmes caractéristiques mais son véhicule appartient à la classe d'âge A3, sa fréquence moyenne égale:

$$(fréquence|C > d) = e^{(-2,8660-0,0439+0,3385+0,2638+0,1998+0,1768)} = 14,5\%$$

4.2.4. Étape 4 : Validation du modèle

La validation du modèle est une étape qui permet de comparer le modèle théorique avec la réalité économique. Cette étape se matérialise par la vérification des signes de coefficients estimés et de contrôler la légitimité de ce modèle par sa déviance.

a) Signe des coefficients

L'étude des signes des coefficients, se fait par la vérification de la position des coefficients des modalités des variables par rapport au coefficient de sa modalité de référence. En effet, par construction le coefficient de la modalité de référence est fixé à 0. À partir du lien logarithmique nous pouvons établir une relation entre chaque coefficient de la modalité de la variable x et le coefficient de sa modalité de référence. Cette relation est expliquée par la fonction exponentielle de la manière suivante :

Soit c le coefficient associé à la modalité :

$$\left\{ \begin{array}{l} \text{Si } c > 0 \Rightarrow \exp(c) > \exp(0) \\ \text{Sinon} \Rightarrow \exp(c) < \exp(0) \end{array} \right.$$

D'après cette relation, nous obtenons la règle suivante pour contrôler les signes des coefficients de du modèle de fréquence :

- ✓ Si le coefficient c_i associé à la modalité i est positif dans ce cas, la fréquence moyenne de cette modalité doit être supérieure à la fréquence moyenne de la modalité de référence correspondante.
- ✓ À l'inverse si le coefficient est négatif cela témoigne que la fréquence moyenne de la modalité afférente est inférieure à la fréquence moyenne de la modalité de référence correspondante.

Pour chacune des modalités retenues, la règle des signes des coefficients a été vérifiée :

- ✓ L'ensemble des coefficients obtenus pour la variable BM sont positifs, c'est-à-dire que la fréquence moyenne est supérieure à la fréquence moyenne de la modalité de référence, toute autre modalité égale par ailleurs.
- ✓ À l'inverse, pour les variables M, les deux premiers coefficients sont négatifs, preuve que les individus dont la marque de véhicule appartient à M2 et M3 ont une fréquence moyenne inférieure à celle des M1 qui est la modalité de référence, toute modalité égale par ailleurs.

b) La déviance du modèle

La déviance Standardisée permet de contrôler la légitimité du modèle et de savoir si notre modèle est en adéquation avec les données. La déviance du modèle vaut 1148,5 et le nombre de degré de liberté est de 2333. Pour standardiser la déviance en la divisant par le paramètre de dispersion, $\phi = 1,14$. Nous obtenons alors une déviance Standardisée de 1007,456 inférieure au nombre de degrés de liberté donc nous pouvons admettre que notre modèle est pertinent.

4.3. La modélisation du coût moyen

Dans cette section, nous allons nous intéresser au coût moyen suivant les mêmes étapes du paragraphe précédent en utilisant comme variable à expliquer pour ce modèle le coût moyen des sinistres.

4.3.1. Étape 1 : Choix de la loi de probabilité

Les modèles les plus classiques utilisés pour modéliser les coûts moyens sont réalisés à partir de deux lois : (Denuit et Charpentier, 2009) :

- ✓ La loi de Gamma
- ✓ La loi modèle Log-Normale

Le choix entre les deux lois est fait selon :

- ✓ Les critères l'AIC et BIC
- ✓ Le critère graphique

a) Critères de l'AIC et BIC

La fonction `fitdist()` sous R nous permet de déterminer l'AIC et le BIC de deux lois. Ces deux critères sont les plus adaptés pour le choix dans la tarification comme démontrer dans la partie théorique. Le choix porte sur la loi qui possède le plus petit AIC ou BIC afin de choisir la meilleure distribution théorique en matière d'ajustement aux données.

Pour ces deux critères nous avons déjà expliqué leurs utilités et leurs importances dans la partie théorique. À ce niveau, nous avons calculé les paramètres de loglikelihood comme le montre la figure ci-après.

Figure 22 : Calcul de l'AIC et du BIC selon la loi du Gamma et Log-Normale

```
> summary(fit1)
Fitting of the distribution ' gamma '
Loglikelihood: -2567.832   AIC:  5139.664   BIC:  5147.01
> summary(fit2)
Fitting of the distribution ' lnorm '
Loglikelihood: -2542.308   AIC:  5088.616   BIC:  5095.963
```

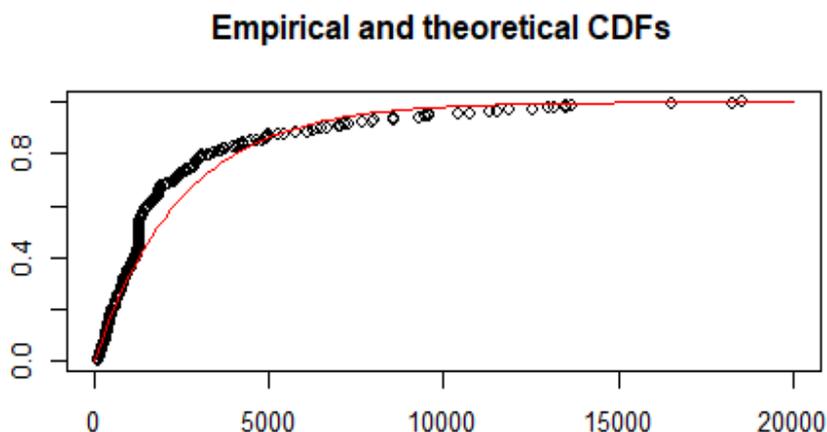
Source : Sortie du logiciel R

Selon les critères de l'AIC et BIC nous choisissons la loi Log-Normale car elle dispose les deux AIC et BIC les plus faibles.

b) Critère graphique

Les deux graphiques ci-dessous représentent l'ajustement de la densité du coût par la loi Gamma et la loi Log-Normale.

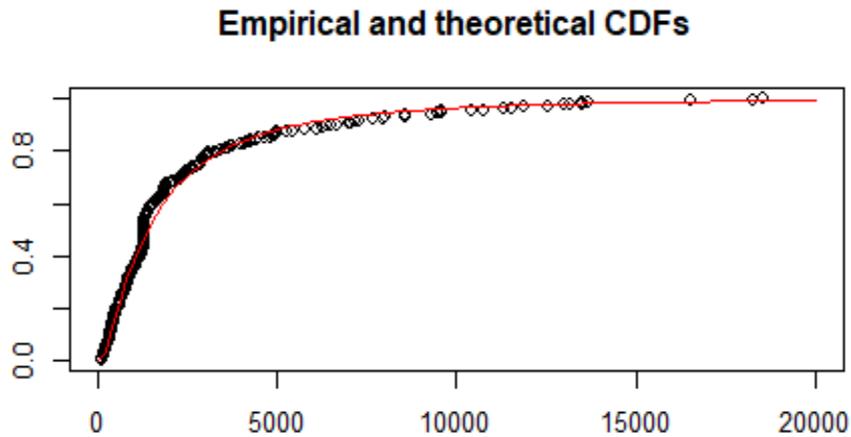
Figure 23 : Ajustement des coûts moyens par une loi Gamma



Source : Sortie du logiciel R

D'après ce graphique, nous remarquons que cette loi présente des écarts d'estimation entre les montants de 2500 et de 10000 DT.

Figure 24 : Ajustement des coûts moyens par une loi Log-Normale



Source : Sortie du logiciel R

D'après le graphique, nous remarquons que la modélisation par une loi Log-Normale présente que quelque peu d'écart observés. Selon les deux graphiques nous choisissons la loi Log-Normale. La décision finale est le choix de la loi Log-Normale selon les critères de BIC et AIC et aussi la représentation graphique.

Nous pouvons dire que $X \sim (m, \sigma^2)$, si $Y = \ln X \sim \mathcal{N}(m, \sigma^2)$. La fonction de densité de probabilité de X est la suivante :

$$f_X(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2\right), & \text{si } x > 0, \\ 0, & \text{sinon.} \end{cases}$$

Avec une espérance mathématique : $E(X) = E(e^Y) = e^{(m + \frac{\sigma^2}{2})}$

Et une Variance : $\text{Var}(X) = \text{Var}(e^Y) = (1 - e^{-\sigma^2}) (e^{2(m + \sigma^2)})$

4.3.2. Étape 2 : Estimation des coefficients de régression

Une fois que nous avons choisi la loi, il convient d'estimer les paramètres du modèle. Cette estimation est réalisée à l'aide de la fonction glm() sous R. De même, nous allons intégrer la durée d'exposition comme une variable offset. Les résultats de ce modèle sont les suivants :

Figure 25 : Estimation des coefficients de régression du modèle de coût moyen

```
> summary(logcout)

Call:
glm(formula = auto_C$logCM ~ auto_C$M + auto_C$BM + auto_C$PV +
     auto_C$A + auto_C$R + offset(log(Exposition)), family = gaussian(link = "identity"),
     data = auto_C)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.63560  -0.51464  -0.02352   0.41255   2.14729

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.52111    0.45627  18.676 < 2e-16 ***
auto_C$MM2  -0.22738    0.19815  -1.148 1.89e-12 ***
auto_C$MM3  -0.24765    0.27139  -0.913 5.49e-16 ***
auto_C$MM4   0.88014    0.19131   4.601 < 2e-16 ***
auto_C$BMBM2 0.02586    0.38133   0.068 0.000687 ***
auto_C$BMBM3 0.26387    0.38545   0.686 4.07e-14 ***
auto_C$BMBM4 0.77485    0.38447   3.198 < 2e-16 ***
auto_C$PVPV2 0.86464    0.24228   3.679 1.62e-05 ***
auto_C$PVPV3 0.94497    0.23504   3.199 0.000340 ***
auto_C$AA2   0.13403    0.29537   0.452 0.674373
auto_C$AA3   0.14058    0.29664   0.429 0.025865 *
auto_C$AA4   0.24519    0.32738   0.920 0.000259 ***
auto_C$RR2   0.99994    0.26654   3.481 < 2e-16 ***
auto_C$RR3   0.80924    0.28724   2.817 1.34e-06 ***
auto_C$RR4  -0.77485    0.31236   2.481 1.53e-05 ***
auto_C$RR5   0.72850    0.34601   0.070 2.85e-05 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.4429417)

Null deviance: 137.48  on 290  degrees of freedom
Residual deviance: 118.81  on 275  degrees of freedom
```

Source : Sortie du logiciel R

4.3.3. Étape 3 : Procédure de sélection des variables

À ce niveau, nous essaierons d'éliminer les variables les moins significatives du modèle. Pour ce faire, nous utiliserons la même méthode adoptée pour le premier modèle à l'aide de la fonction Backward sous R.

Figure 26 : Procédure de sélection des variables du modèle de coût moyen

```
> step1=step(logcout,dir="backward")
Start: AIC=5088.6
auto_C$logCM ~ auto_C$M + auto_C$BM + auto_C$PV + auto_C$A +
  auto_C$R + offset(log(Exposition))

      Df Deviance  AIC
<none>      119  5089
- auto_C$M    3   121  5095
- auto_C$R    2   126  5110
- auto_C$PV   4   134  5112
- auto_C$A    3   135  5119
- auto_C$BM   3   137  5124
```

Source : Sortie du logiciel R

D'après cette procédure, nous remarquons que toutes les variables explicatives sont significatives pour la modélisation du coût moyen.

✓ **Application numérique**

Les coefficients permettant de retrouver le coût moyen d'un sinistre pour chaque individu par classe tarifaire. Cinq variables tarifaires ont été retenues, nous avons alors :

$$E(CM|C > d) = e^{(\beta_0 + \beta_1 M + \beta_2 BM + \beta_3 PV + \beta_4 A + \beta_5 R)}$$

Par exemple un individu d'une marque de véhicule M3, appartient à un SBM BM2, puissance du véhicule PV1, ayant un véhicule de classe d'âge A2, et sa région est R4, son coût moyen est le suivant :

$$E(CM|C > d) = e^{8,52111 - 0,24765 + 0,02586 + 0 + 0,13403 - 0,77485} = 2118,577 \text{ DT}$$

Pour un autre individu ayant les mêmes caractéristiques mais l'âge du véhicule est A4, son coût moyen est égal :

$$E(CM|C > d) = e^{8,52111 - 0,24765 + 0,02586 + 0 + 0,24519 - 0,77485} = 2367,666 \text{ DT}$$

4.3.4 Étape 4 : Validation du modèle

a) Signe de coefficient

Comme précédemment, nous allons vérifier les signes des coefficients estimés. Si le coefficient associé à une modalité est positif, le coût moyen de cette modalité doit être supérieur au coût moyen de la modalité de référence correspondante et inversement, si ce coefficient est négatif. Le modèle de coût respecte bien la règle imposée par les signes : toutes les modalités pour lesquelles le coefficient associé est positif ont un coût moyen de sinistre inférieur à la modalité de référence et inversement.

b) Déviance du modèle

À ce niveau, nous allons vérifier l'adéquation du modèle à travers la déviance standardisée. Dans notre cas, la déviance standardisée est égale 268 qui est inférieure au nombre de degrés de liberté 275, donc notre modèle est pertinent.

4.4. Synthèse des résultats

Après avoir présenté les deux modèles de fréquence et du coût moyen pour l'usage utilitaire, nous allons présenter dans cette partie une synthèse des résultats. Nous savons que la prime pure se calcule selon la formule suivante :

$$\text{Prime pure} = \text{Fréquence} \times \text{Coût moyen}$$

Par la suite, nous utilisons l'hypothèse, que les chargements moyens utilisés dans le secteur sont de 30%. Notre prime commerciale se calcule de la manière suivante :

$$\text{Prime commerciale} = 1,30 \times \text{Prime pure}$$

Par la suite, nous allons effectuer un exemple explicatif pour calculer la prime pure et prime commerciale. Soit un individu dispose d'un véhicule d'usage utilitaire ayant les caractéristiques suivantes :

Une marque de véhicule M2, appartient à un SBM BM1, puissance du véhicule PV1, ayant un véhicule de classe d'âge A4, et sa région est R1.

Selon ces caractéristiques, la prime pure et commerciale selon le modèle se calculent de la manière suivante :

$$\text{Fréquence} = e^{(-2,866 - 0,1273 + 0,3385 + 0 + 0,3254 - 0,1552)} = 8,34\%$$

$$\text{Coût moyen} = e^{(8,52111 - 0,24765 + 0,02586 + 0 + 0,24519 - 0,77485)} = 2367,666 \text{ DT}$$

$$\text{Prime pure} = 8,34\% \times 2367,666 = 197,463 \text{ DT}$$

$$\text{Prime Commerciale} = 197,463 \times 1,3 = 256,702 \text{ DT}$$

Pour le même exemple, nous allons calculer la prime commerciale telle qu'elle est fixée par la circulaire (voir annexe 3). D'après la circulaire sans tenir compte du système bonus-malus la prime RC automobile est de 214 DT.

À partir du modèle de fréquence et du coût moyen, cet exemple a été appliqué sur un échantillon de vingt assurés. Nous avons constaté que les primes estimées par le modèle sont supérieures aux primes administrées. L'écart moyen entre les deux montants de cet échantillon est de 31%. De ce fait, il est clair que les primes pratiquées doivent être majorées, ce qui signifie que la majoration de la prime RC est à peu près une certitude.

Toutefois résultats précédents ne peuvent pas être généralisés puisqu'ils sont tirés à partir d'un échantillon réduit. Cependant, le modèle que nous avons proposé offre une méthodologie pertinente pour un réexamen global de cette tarification avec plus de temps pour pouvoir analyser tous les détails et sur la totalité du portefeuille sur plusieurs années pour saisir la dynamique d'évolution.

Globalement, les résultats sont probants. Nous avons vérifié l'insuffisance en termes de valeur de la prime RC automobile et l'influence d'autres facteurs comme la région, la marque de véhicule et l'âge du véhicule dans la tarification de la garantie RC.

Au vu de nos résultats, il serait donc intéressant que le législateur tunisien permet aux compagnies de tarifier avec d'autres variables que celles retenues actuellement pour calculer la prime pure. Cela serait bénéfique pour les assureurs ainsi que pour les assurés.

CONCLUSION

Ce chapitre, a décrit d'une manière synthétique la situation actuelle du marché tunisien d'assurance automobile avec un focus sur la responsabilité civile et les enjeux règlementaire et les perspectives futures. Ensuite, nous avons présenté l'entreprise de parrainage GAT Assurances suivi d'une analyse univariée et bivariée des grandeurs endogènes et exogènes.

L'objectif central de ce chapitre, consiste à effectuer une application empirique relative à la détermination de la prime pure de la RC automobile de la GAT durant l'année 2019. Nous avons tarifié en utilisant le modèle GLM à l'aide du logiciel R, avec la loi Binomiale Négative pour la fréquence et la loi log-Normale pour le coût moyen. À la fin de ce chapitre, nous avons essayé d'illustrer nos résultats à travers un exemple simplifié avec une évaluation sur échantillon de taille réduite.

L'application d'un modèle multiplicatif nous a permis de dégager certains points importants concernant la tarification du risque automobile. Cette tarification a réparti les charges de sinistres d'une manière plus équitable entre les différents segments suivant leurs caractéristiques et les leurs risques associés.

La présente étude empirique offre une approche de tarification qui s'inscrit dans l'évolution constante des méthodes de tarification telles qu'elles sont de plus en plus pratiquées par les experts.

Ces méthodes convergent vers l'individualisation du risque automobile. Cette individualisation du risque vise à fidéliser les bons assurés puisque chaque assuré paie pour son risque et à faire face également à la pression de la concurrence par la détection des nouveaux segments plus rentables.

Cet effort de tarification de la prime doit être accompagné par une majoration de la prime RC automobile.

CONCLUSION GENERALE

Cette étude présente une méthodologie d'analyse de la segmentation et de la tarification en assurance automobile. Tout d'abord, nous avons compris que la maîtrise de ces deux notions est primordiale pour préserver le portefeuille et d'attirer des nouveaux assurés. Nous avons également constaté que la tarification en assurance automobile appartient à un univers d'innovation constant et la segmentation des risques semble devenir de plus en plus poussée allant jusqu'à l'individualisation de la prime.

À partir de ces constats d'ordre général, nous avons illustré ces notions à travers l'extraction d'une base de données du portefeuille automobile utilitaire de la compagnie de parrainage GAT Assurances de l'année 2019, dans laquelle nous avons saisi l'importance de l'étape d'épurement des données brutes. Cette approche a relevé l'importance du rôle de l'actuaire au sein d'une compagnie d'assurance, qui s'inscrit dans la logique de détermination de la prime de risque et d'évaluation des bases de données et de leurs fonctionnements en règle générale dans une optique de long terme.

Nous avons ensuite mis en avant les techniques de base concernant le regroupement des variables. Nous savons que d'un point de vue actuariel, le montant de prime est identique à chaque classe de risque homogène. Dans cette logique, il est nécessaire de regrouper un ensemble de modalités dans un même groupe ayant des caractéristiques homogènes afin de les rendre exploitables à la modélisation.

De plus, nous avons effectué une analyse univariée et bivariée entre les variables qui reste une étape préliminaire à ne pas négliger avant la modélisation du risque. En effet, cette dernière nous apporte un grand nombre d'informations utiles. Nous sommes, enfin, venus à la modélisation, dont laquelle nous avons appliqué les aspects théoriques de la modélisation du risque automobile.

La présente étude a pour objectif de proposer un modèle de tarification de l'assurance automobile intégrant la sinistralité des assurés comme un facteur déterminant de la prime pure de la garantie RC automobile. Pour ce faire, nous avons présenté une approche de tarification en se basant sur le principe de « Fréquence \times Coût Moyen » estimé par le modèle GLM. Nous avons ainsi utilisé plus particulièrement la loi Binomiale Négative et la loi Log-Normale, classiquement employées pour la modélisation des fréquences et des coûts moyens.

Les résultats d'une telle étude sont multiples. Nous avons vérifié l'insuffisance en termes de valeur de la prime RC automobile et que des variables autres que l'usage et la puissance des véhicules sont significatives pour la modélisation de la prime comme la région, la marque de véhicule et l'âge du véhicule.

Ces résultats permettent d'enrichir le débat sur les conditions éventuelles de la libéralisation des tarifs. Les expériences internationales ont montré les avantages d'une telle décision sur le secteur. Elles ont fourni aux assureurs des incitations en termes d'efficacité et d'innovation. Cette libéralisation entraînerait très probablement une forte concurrence entre les compagnies d'assurance. À cet effet, il est temps en Tunisie de revoir les décisions qui s'imposent en matière de tarification.

Cependant, nous tenons tout de même à relativiser nos conclusions. Vu les difficultés rencontrées étant donné que la richesse et le niveau de détails de la base de données exigent beaucoup de temps. Nous rappelons que nous avons travaillé sur un échantillon en nombre d'observations demeure relativement faible et pourrait nous conduire à des résultats moins précis. Une telle étude nécessite plus de temps pour pouvoir obtenir des résultats plus consistants et des conclusions générales.

En revanche, ce projet de fin d'études présente aux chercheurs plusieurs voix de développement possibles et qui peuvent être matérialisées par des nouvelles études telles que :

- Il s'avère très intéressant de réaliser la même étude sur la totalité du portefeuille pour l'ensemble des usages et sur plusieurs années.
- Introduire d'autres variables qui peuvent être utiles pour pouvoir expliquer la sinistralité des assurés telle que présentée dans le premier chapitre.
- Un autre axe de développement possible consiste à adopter une approche d'assurance comportementale, qui mettrait le comportement du conducteur au centre d'analyse de la tarification, ce qui exige des traitements statistiques et des technologies très développés. Cette approche exige aussi des nouveaux modèles de tarification, avec l'utilisation de nouvelles variables tarifaires et également de nouveaux principes.

BIBLIOGRAPHIE

❖ Articles et ouvrages

- Alain M. : Analyse Financière : Concepts et Méthodes Édition Dunod, Paris, 2007
- Bressand C. : A propos de la tarification de l'assurance automobile, Économie et Prévision, 75-96, 1993
- Charpentier A. & Denuit M. : Mathématiques de l'assurance non vie, tome 2, Tarification et Provisionnement, 2009
- Charpentier A. et Denuit M. : Mathématiques de l'assurance non vie, tome 1, Tarification et Provisionnement, 2009
- Collet I. et Chouchet M. : Etude sur la tarification de l'assurance automobile à travers le monde, Fixage, 2017
- Dionne, G. et Vanasse C. : Automobile Insurance Ratemaking in the Presence of Asymmetrical Information. Journal of Applied Econometrics, vol 7, 149-165, 1992
- Ernst & Young : Introducing pay how you drive insurance, 2016
- Ghali O. : Un modèle de tarification optimal pour l'assurance automobile dans le cadre d'un marché réglementé : application à la Tunisie, Cahier de recherche 01-09, Décembre 2001
- Ginglinger P. : Gestion financière de l'entreprise, Édition Dalloz, Paris, 1997
- Gourieroux C et al : Pseudo Maximum Likelihood Methods: Applications to Poisson Models, The Econometric Society, Vol. 52, No. 3, 1984
- Kahane Y. et Porat M. : Financial Analysis of Underwriting Results - A Corrected Approach to Loss Ratio Analysis, with a Special Reference to Inflation, The Journal of Risk and Insurance. 1994
- Kotler, P. et Keller, K. : Marketing Management. Englewood Cliffs, N.J., Prentice Hall, 2006
- La Presse Business, La libéralisation de l'assurance automobile : A chacun sa vérité, 18 au 24 Juillet 2016
- Lemaire J. : Bonus-Malus Systems in Automobile Insurance. Kluwer Academic Publishers, 1995

- Lemaire J. : Assurance automobile : modèles actuariels. Huebner International Series sur le risque, l'assurance et la sécurité économique, 1985
- Noël F. : Économie et Organisation de l'Assurance, BTS Assurances, 2007
- Okba R., Abdelouhab L. et Riadh R. : Une évaluation empirique de la tarification de l'assurance automobile en Algérie, Volume 5, Numéro 9, 2015
- Paulin M. : Economie et Organisation de l'Assurance, Séfi édition, Québec, canada, 2007
- Ramage P. : Analyse et diagnostic financier, Edition d'organisation, paris, 2001
- Ray L : L'assurance des flottes automobiles, Souscription, tarification, gestion, éditions l'Argus de l'assurance, 2008
- Redjem N., Méthodes d'analyse financière, Édition Dar El Ouloum, 2005
- Riad M. : Modèle de tarification optimal en assurance automobile dans le cadre d'un marché réglementé, Université de Tipaza, 2015
- Rouach M. et Naulleau G. : Contrôle de gestion et stratégie dans la banque ,3ème édition, édition Dunod, 2000
- Trainar P. et Thourot P. : Gestion de l'entreprise d'assurance, édition DUNOD, 2 éd, paris, 2017
- Vernimmen P, Finance d'entreprise. Dalloz, 8ème édition, 2010

❖ **Thèses et mémoires**

- Aboukherraz L. et Bendarag M., remises sur les garanties annexes en se basant sur la RC automobile, Institut National de Statistique et d'Economie Appliquée, 2011
- Attia R. Mise en place de modèles de tarification alternatifs face à la suppression réglementaire d'une variable tarifaire en automobile Master Actuariat de Dauphine, 2016
- Azzousi Z. : Tarification en assurance automobile, Modélisation de la prime pure et individualisation du risque, IFID, 2017
- Bel Abria I. : Tarification en Assurance Maladie Collective Synthèse théorique avec application sur des données du GAT, IFID, 2020
- Collins T. : Pricing policy and profitability level of an organization, Centria University Of Applied Sciences, 2021
- Gonnet G. : Etude de la tarification et de la segmentation en assurance Automobile, Université Claude Bernard – Lyon 1, 2010

- Henriet D. et Rochet C. : La logique des systèmes bonus-malus en assurance automobile: une approche théorique, Annales d'Économie et de Statistique. 1986:133-52.
- Khichane D. et Kadri H, Tarification de la prime d'assurance automobile. Cas : de la Société d'assurance algérienne SAA, Faculté des Sciences Economiques, commerciales et sciences de gestion, 2020
- Magali R. : Elaboration d'un véhiculier en assurance automobile, Institut Supérieur d'Actuariat, 2016
- Marn, M. et Rosiello, R. : Managing price, gaining profit, Boston, MA: Harvard College, 1992
- Niny L., Le modèle mutualiste automobile : viabilité dans un univers concurrentiel, Centre d'Etudes Actuarielles, 2013
- Prévot P. Management de l'Innovation dans l'Assurance, École Polytechnique, 2005
- Rachel A. : Mise en place de modèles de tarification alternatifs face à la suppression réglementaire d'une variable tarifaire en automobile, 2016

❖ Cours et rapports

- Amira H., Cours d'assurance automobile : Souscription automobile et gestion des sinistres, IFID 2021
- Banque Mondiale, Motor Third-Party Liability Insurance in Developing Countries : Raising Awareness and Improving Safety, 2009
- CGA et Banque Mondiale, Réforme de l'assurance de la responsabilité civile automobile en Tunisie : état des lieux et recommandations, 2015
- CGA, Étude actuarielle du nouveau système d'assurance automobile en Tunisie, 15 septembre 2003
- Circulaire du Ministère des Finances : n°1/2017 du 28 février portant fixation du tarif de l'assurance de la responsabilité civile du fait de l'usage des véhicules terrestres à moteur
- Etude stratégique et actuarielle portant sur l'élaboration d'un contrat programme relatif au secteur des assurances en Tunisie, 2014
- Ktata R., Cours en actuariat non vie, IFID, 2021
- Rapport CGA 2018

- Rapport CGA 2019
- Rapport CGA 2020
- Rapport CGA 2021
- Zarrouk L. L'Assurance Automobile En Tunisie: De la Réforme engagée aux Défis à relever, 4 novembre 2015

❖ Sites Web

- <https://bonne-assurance.com/automobile/actualites/2009/10/06/solly-azar-propose-easy-drive-aux-jeunes-conducteurs/>
- <https://www.atlas-mag.net/article/augmentation-du-nombre-d-accidents-de-la-route-entunisie#:~:text=L'Observatoire%20National%20de%20la,la%20m%C3%A9me%20p%C3%A9riode%20de%202020.>
- <https://www.larevuedudigital.com/payer-comme-vous-conduisez-chez-axa-direct-assurance-a-la-fin-de-lannee/>
- <https://www.swissre.com/institute/>

LISTE DES ANNEXES

| | |
|--|----|
| Annexe 1 : Les critères les plus fréquents de tarification a priori | 78 |
| Annexe 2 : Le système bonus-malus selon le pays..... | 79 |
| Annexe 3 : Le barème des tarifs de l'assurance responsabilité civile en Tunisie :..... | 80 |
| Annexe 4 : Le système bonus-malus Tunisien | 82 |
| Annexe 5 : Acronyme des variables et des modalités..... | 82 |
| Annexe 6 : Méthode de k-means pour la variable BM | 83 |
| Annexe 7 : Méthode de k-means pour la variable PV | 83 |
| Annexe 8 : Méthode de k-means pour la variable PTV | 83 |
| Annexe 9 : Méthode de k-means pour la variable CU..... | 83 |

LES ANNEXES

Annexe 1 : Les critères les plus fréquents de tarification a priori

| Critère | Sous-critère | Pays |
|-------------|---------------------------------------|---|
| Véhicule | L'âge du véhicule | Italie, France, Royaume-Uni, Grande Bretagne, Portugal, Canada, Norvège ; Chypre |
| | Poids du véhicule | Pays bas, Norvège, Portugal et Canada |
| | La marque | Norvège |
| | Le modèle | Québec, Liban, Chypre et en Italie ; |
| | Vitesse | France, Royaume-Uni |
| | Puissance | Liban, France, Royaume-Uni, Portugal, Canada et Norvège, Chypre et en Italie ; |
| | Volume moteur | France, Royaume-Uni, Portugal, Canada et Norvège |
| | Valeur de la voiture | Luxembourg |
| Lieu | Habitation | France (5 zones), Pays-Bas (3 zones), Italie (8 zones), Espagne (3 zones), Allemagne (8 zones), Grande-Bretagne (7 zones), Suède (7 zones), Norvège (3 zones), États-Unis (selon les Etats) et Canada (8 zones) |
| | Place de parking | R.-U., au Nigeria et en France ; |
| Utilisation | Distance parcourue | R.-U., en Norvège, en Italie, en Suède, au Canada et en France |
| | Activité professionnelle | n Italie, au Canada, à Chypre et au R.-U, France Allemagne, Canada |
| | Conduite de jour ou de nuit | R.-U |
| Conducteur | Age | en Belgique, en Suède et aux Pays-Bas, Belgique Espagne Suisse. Grande-Bretagne, Japon |
| | Genre | États-Unis et Québec. |
| | Nombre d'années de permis de conduire | France et Luxembourg |
| | Nationalité | Suisse |

Source : Bressand (1993)

Annexe 2 : Le système bonus-malus selon le pays

| Pays | Modèle |
|-----------------------------------|--|
| Liban | Absence d'un système de bonus en cas de non déclaration de sinistres |
| Maroc | Si l'assuré n'a aucune responsabilité totale ou partielle dans un accident pendant deux ans et il bénéficie de 10% de remise, sans effet mémoire |
| Royaume-Uni | Le système bonus est basé seulement sur le nombre d'années sans sinistre pour déterminer le bonus |
| Luxembourg | Pour chaque année sans sinistre un rapport un point de bonus et un sinistre entraîne un malus de 3 points. Le nombre de point du système bonus/malus pour tous les usagers sont compris entre 22 et -3 |
| Italie | Un système de bonus/malus n'est pas très efficace car la majorité des assurés se trouvent dans la meilleure classe. |
| Zone CIMA | Un système bonus/malus dépend du nombre d'années sans accidents responsables |
| Canada | Un système de bonus est basé sur le nombre d'accidents fautifs au cours des trois dernières années |
| Belgique | Descente rapide du bonus en cas d'accident et retour automatique au bonus maximum si 4 années sans accident |
| Les pays émergents | Un système du bonus/malus n'est pas beaucoup développé et n'incite pas l'assuré à améliorer son comportement |
| Espagne, Suède, Chypre et Vietnam | Un système de pur bonus |
| Etats-Unis | Un système de pur malus |

Source : Collet et Chouchet (2017)

Annexe 3 : Le barème des tarifs de l'assurance responsabilité civile en Tunisie :

1- L'usage privé :

En DT

| Puissance fiscale du moteur | Tarif (Hors Taxes) |
|-----------------------------|--------------------|
| 02 Chevaux | 94 |
| De 3 à 4 chevaux | 110 |
| 5 à 6 chevaux | 140 |
| De 7 à 10 chevaux | 170 |
| De 11 à 14 chevaux | 220 |
| 15 chevaux et plus | 264 |

2- Transport de marchandises pour son propre compte:

A- Véhicules dont le poids brut n'excède pas 3,5 Tonnes:

En DT

| Puissance fiscale du moteur | Tarif (Hors Taxes) |
|-----------------------------|--------------------|
| 02 Chevaux | 145 |
| De 3 à 4 chevaux | 171 |
| 5 à 6 chevaux | 214 |
| De 7 à 10 chevaux | 262 |
| De 11 à 14 chevaux | 338 |
| 15 chevaux et plus | 405 |

B- Véhicules dont le poids total dépasse 3,5 tonnes:

En DT

| | Tarif (Hors Taxes) |
|----------------------|--|
| Prime de base | 257 |
| Prime supplémentaire | 21 pour chaque tonne supplémentaire dont le poids total dépasse 3,5 tonnes |

3- Transport de marchandises pour des Tiers:

A- Véhicules dont le poids brut n'excède pas 3,5 tonnes:

(Appliquer le tarif relatif aux véhicules destinés au transport de marchandises pour son propre compte visé au point A)

B- Véhicules dont le poids total dépasse 3,5 tonnes:

En DT

| | Tarif (Hors Taxes) |
|----------------------|--|
| Prime de base | 250 |
| Prime supplémentaire | 4 pour chaque tonne supplémentaire dont la charge utile dépasse 3,5 tonnes |

1- Les usages agricoles:

A- Véhicules dont le poids brut n'excède pas 3,5 tonnes:

En DT

| Puissance fiscale du moteur | Tarif (Hors Taxes) |
|-----------------------------|--------------------|
| 02 Chevaux | 84 |
| De 3 à 4 chevaux | 97 |
| 5 à 6 chevaux | 122 |
| De 7 à 10 chevaux | 150 |
| De 11 à 14 chevaux | 193 |
| 15 chevaux et plus | 232 |

B- Véhicules dont le poids total dépasse 3,5 tonnes:

En DT

| | Tarif (Hors Taxes) |
|----------------------|---|
| Prime de base | Le tarif relatif aux véhicules destinés à un usage agricole d'un poids brut n'excédant pas 3,5 tonnes |
| Prime supplémentaire | 13 pour chaque tonne supplémentaire dont la charge utile dépasse 3,5 tonnes |

1- Machines et tracteurs agricoles:

En DT

| Type de Véhicule | Tarif (Hors Taxes) |
|--------------------------|--------------------|
| Type I | |
| Avec roues en caoutchouc | 42 |
| Avec chenilles | 31 |
| Type II | 117 |
| Type III | |
| Avec roues en caoutchouc | 84 |
| Avec chenilles | 59 |
| Type IV | 177 |

2- Les 02 Roues:

En DT

| Type de Véhicule | Tarif (Hors Taxes) |
|----------------------------|--------------------|
| Inférieur ou égal à 125 CC | 62 |
| Supérieur à 125 CC | 168 |

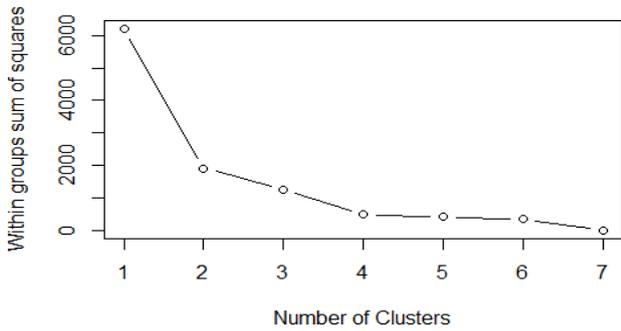
Annexe 4 : Le système bonus-malus Tunisien

| Pour l'usage affaire | | Pour les autres usages | |
|----------------------|-----------|------------------------|----------|
| 350 % | Classe 11 | 200 % | Classe 7 |
| 300 % | Classe 10 | 170 % | Classe 6 |
| 250 % | Classe 9 | 150 % | Classe 5 |
| 200 % | Classe 8 | 120 % | Classe 4 |
| 160 % | Classe 7 | 100 % | Classe 3 |
| 140 % | Classe 6 | 90 % | Classe 2 |
| 120 % | Classe 5 | 80 % | Classe 1 |
| 100 % | Classe 4 | | |
| 90 % | Classe 3 | | |
| 80 % | Classe 2 | | |
| 70 % | Classe 1 | | |

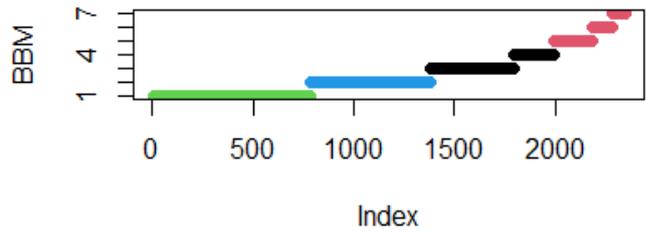
Annexe 5 : Acronyme des variables et des modalités

| Variable | Signification | Modalité |
|------------|--------------------------|---|
| M | La marque de véhicule | M1 : Allemande, Française et Italienne M2 : Américaine et Britannique M3 : Japonaise et Coréenne M4 : Chinoise, indienne et Roumanie |
| BM | Système bonus-malus | BM1 : Classe 1 BM2 : Classe 2 BM3 : Classe 3 et 4 BM4 : Classe 5 et plus |
| PV | Puissance du véhicule | PV1 : 5 CV PV2 : 6 et 7 CV PV3 : de 8 à 17 CV |
| PTV | Poids totale du véhicule | PTV1 : de 255 à 2000 PTV2 : de 2001 à 2600 PTV3 : de 2600 à 3000 PTV4 : de 3000 à 3500 |
| CU | Charge utile | CU1 : de 20 à 750 CU2 : de 751 à 900 CU3 : 901 à 1100 CU4 : 1101 à 3500 |
| A | Age du véhicule | A1 : de 0 à 1 A2 : de 2 à 3 A3 : de 4 à 6 A4 : 7 et plus |
| R | La région | R1 : Cap Bon R2 : Grand Tunis R3 : Centre R4 Nord-Ouest R5 : Sud |

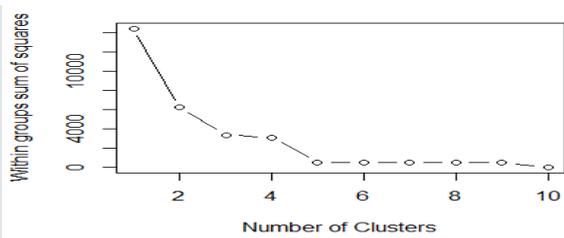
Annexe 6 : Méthode de k-means pour la variable BM



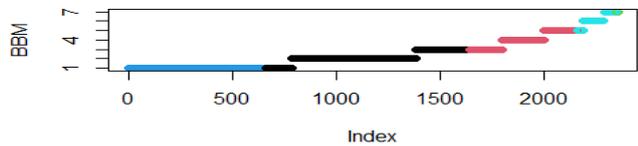
Partitionnement du BM par K-Means



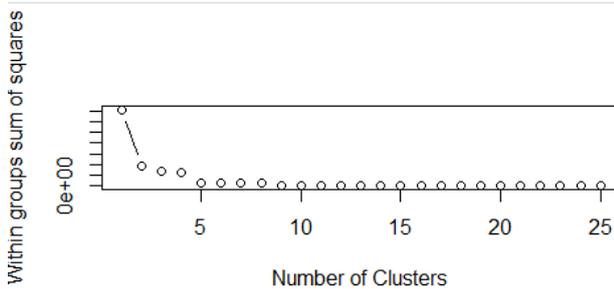
Annexe 7 : Méthode de k-means pour la variable PV



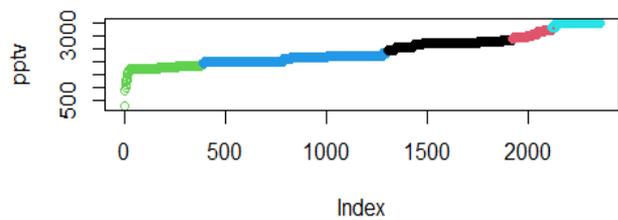
Partitionnement du PV par K-Means



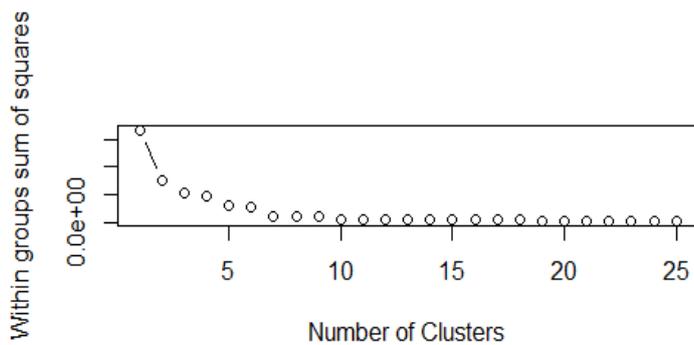
Annexe 8 : Méthode de k-means pour la variable PTV



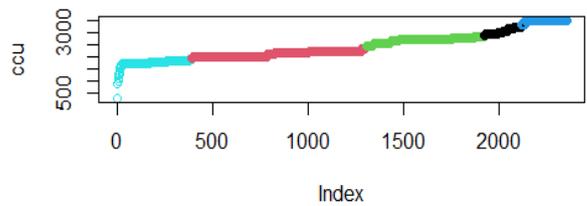
Partitionnement du PTV par K-Means



Annexe 9 : Méthode de k-means pour la variable CU



Partitionnement du PV par K-Means



LES CODES UTILISES SOUS R

✓ Les codes d'import

```
library(readxl)
auto <- read_excel("C:/Users/ASUS/Desktop/25092022/auto.xlsx")
View(auto)
```

✓ Les codes de changement des variables des caractères en numériques

```
auto$Prime<-as.numeric(as.character(auto$Prime))
auto$Mt_S<-as.numeric(as.character(auto$Mt_S))
auto$Fréq<-as.numeric(as.character(auto$Fréq))
auto $Exposition<-as.numeric(as.numeric(auto$Exposition))
auto$Nb_S<-as.numeric(as.character(auto$Nb_S))
```

✓ Les codes pour les statistiques descriptives

```
table(auto$M)
tabM <-table(auto$M)
barplot(tabM)
pie(tabM)
table(auto$A)
tabA <-table(auto$A)
barplot(tabA)
pie(tabA)
table(auto$R)
tabR <-table(auto$R)
barplot(tabA)
pie(tabR)
summary(auto$BM)
hist(auto$BM)
summary(auto$PV)
hist(auto$PV)
```

```
summary(auto$PTV)
```

```
hist(auto$PV)
```

```
summary(auto$CU)
```

```
hist(auto$CU)
```

```
summary(auto$Nb_S)
```

```
hist(auto$Nb_S)
```

```
summary(auto$Mt_S)
```

```
hist(auto$Mt_S)
```

✓ Les codes de regroupement par la méthode de K-means

```
library(ggplot2)
```

```
library(ggfortify)
```

```
library(dplyr)
```

```
library(stats)
```

```
wssplot <- function(auto, nc=15, seed=1234)
```

```
{
```

```
  wss <- (nrow(auto)-1)*sum(apply(auto,2,var))
```

```
  for (i in 2:nc){
```

```
    set.seed(seed)
```

```
    wss[i] <- sum(kmeans(data, centers=i)$withinss)}
```

```
  plot(1:nc, wss, type="b", xlab="Number of Clusters",
```

```
       ylab="Within groups sum of squares")
```

```
}
```

```
library(tidyverse) # data manipulation
```

```
library(cluster) # clustering algorithms
```

```
library(factoextra) # clustering algorithms & visual
```

```
library(NbClust) #use zip file to install it
```

```
##### A
```

```
wssplot(A,nc=18, seed=1234)
```

```
AA=matrix(c(sort(A$A)), ncol=1)
```

```
cl <- kmeans(AA, 5, nstart = 25)
```

```

plot(AA,main="Partitionnement d'age des véhicules par K-Means", xlab=
      "Indexation des contrats par valeurs croissante d'age", ylab="Age du véhicule", col = cl$cluster)
summary(A$A)
####BM
summary(BM)
wssplot(BM,nc=7, seed=1234)
BBM=matrix(c(sort(BM$BM)), ncol=1)
cl <- kmeans(BBM, 4, nstart = 25)
plot(BBM, main= "Partitionnement du BM par K-Means", col = cl$cluster)
summary(A$A)
#### PV
PP=matrix(c(sort(P$P)), ncol=1)
wssplot(PP,nc=10, seed=1234)
cl <- kmeans(PP, 5, nstart = 25)
plot(PP, main= "Partitionnement du PV par K-Means", col = cl$cluster)
#### PTV
wssplot(PTV,nc=25, seed=1234)
pptv=matrix(c(sort(PTV$PTV)), ncol=1)
cl <- kmeans(pptv, 5, nstart = 25)
####CU
wssplot(CU,nc=25, seed=1234)
ccu=matrix(c(sort(PTV$PTV)), ncol=1)
cl <- kmeans(ccu, 5, nstart = 25)
plot(ccu, main= "Partitionnement du PV par K-Means", col = cl$cluster)

```

✓ [Les codes pour le regroupement par la méthode de CAH](#)

```

####M
summary(auto$Fréq)
b=tapply(auto$Fréq,auto$M, mean)
d=dist(b, method="euclidean")
hc=hclust(d, method="ward.D2")

```

```

plot(hc, hang = -1)
rect.hclust(hc, k=4, border="red")
bb=tapply(auto$CM,auto$M, mean)
dd=dist(bb, method="euclidean")
hcc=hclust(dd, method="ward.D2")
plot(hcc, hang = -1)
rect.hclust(hcc, k=4, border="red")

```

✓ Les codes des tests du Khi-deux d'indépendance

```

chisq.test(auto_$M,auto_$BM)
chisq.test(auto_$M,auto_$PTV)
chisq.test(auto_$BM,auto_$PTV)
chisq.test(auto_$M,auto_$CU)
chisq.test(auto_$BM,auto_$CU)
chisq.test(auto_$PTV,auto_$CU)
chisq.test(auto_$M,auto_$PV)
chisq.test(auto_$BM,auto_$PV)
chisq.test(auto_$PTV,auto_$PV)
chisq.test(auto_$CU,auto_$PV)
chisq.test(auto_$M,auto_$A)
chisq.test(auto_$BM,auto_$A)
chisq.test(auto_$PTV,auto_$A)
chisq.test(auto_$CU,auto_$A)
chisq.test(auto_$PV,auto_$A)
chisq.test(auto_$M,auto_$R)
chisq.test(auto_$BM,auto_$R)
chisq.test(auto_$PTV,auto_$R)
chisq.test(auto_$CU,auto_$R)
chisq.test(auto_$PV,auto_$R)
chisq.test(auto_$A,auto_$R)

```

✓ Les codes du test de Kruskal

```

kruskal.test(auto_$Freq~auto_$M)
kruskal.test(auto_$Freq~auto_$BM)
kruskal.test(auto_$Freq~auto_$PV)
kruskal.test(auto_$Freq~auto_$A)
kruskal.test(auto_$Freq~auto_$R)
kruskal.test(auto_$CM~auto_$M)
kruskal.test(auto_$CM~auto_$BM)
kruskal.test(auto_$CM~auto_$PV)
kruskal.test(auto_$CM~auto_$A)
kruskal.test(auto_$CM~auto_$R)

```

✓ Les codes pour le traitement des sinistres graves

```

library(ggplot.multistats)
library(ggplot2)
plot(auto$`Mt S`,col="blue")
summary(auto$`Mt S`)

```

✓ Les codes de choix de la loi pour la fréquence

```

library(vcd)
gf=goodfit(auto$Fréq,type="poisson",method="MinChisq")
plot(gf)
summary(gf)
gf=goodfit(auto$Fréq,type="nbinomial",method="MinChisq")
plot(gf)
summary(gf)

```

✓ Les codes pour l'estimation des coefficients de régression de la fréquence

```

library(tidyverse)
library(sjPlot)
library(lme4)
library(MASS)
regBNlog=glm.nb(auto$Nb_S~auto$M+auto$BM+auto$PV+auto$A+auto$R+
offset(log(auto$Exposition)))

```

```
summary(regBNlog)
```

```
fréquence<-regBNlog
```

✓ Les codes pour sélectionner les variables pour le modèle de fréquence

```
library(backward)
```

```
step=step(regBNlog,dir="backward")
```

```
step<-stepAIC(fréquence,direction = "backward")
```

```
summary(fréquence)
```

```
summary(step)
```

✓ Les codes de choix de la loi pour le modèle du coût

```
library(MASS)
```

```
library(fitdistrplus)
```

```
library(ggplot2)
```

```
library(stats)
```

```
fit1<-fitdist(auto_C$CM, distr = "gamma", method = "mle")
```

```
summary(fit1)
```

```
plot(fit1)
```

```
fit2<-fitdist(auto_C$CM,"lnorm")
```

```
summary(fit2)
```

```
plot(fit2)
```

✓ Les codes pour l'estimation des coefficients de régression du coût moyen

```
library(tidyverse)
```

```
library(sjPlot)
```

```
library(lme4)
```

```
library(MASS)
```

```
logcout <- glm(auto_C$logCM~auto_C$M+auto_C$BM+auto_C$PV+auto_C$A+auto_C$R+  
offset(log(Exposition)), family = gaussian (link = "identity"), data = auto_C)
```

```
summary(logcout)
```

✓ Les codes pour sélectionner les variables pour le modèle de coût moyen

```
CM<-reglogcout
```

```
library(backward)
```

```
step1=step(reglogcut="backward")  
step<-stepAIC(CM,direction = "backward")  
summary(CM)  
summary(step1
```