



المندوبية السامية للتخطيط  
HAUT-COMMISSARIAT AU PLAN

ROYAUME DU MAROC  
\*\_.\_\*\_.\_\*  
HAUT COMMISSARIAT AU PLAN  
\*\_.\_\*\_.\_\*\_.\_\*\_.\_\*  
INSTITUT NATIONAL  
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE



**INSEA**

## Projet de Fin d'Etudes

\*\*\*\*\*

**Analyse actuarielle du portefeuille automobile usage  
tourisme de CNIA SAADA Assurance  
Segmentation à priori et correction de la prime pure  
à postériori (GLM, Bonus-Malus et Crédibilité)**

Préparé par : **M. Anass AHAYAN**  
**Mlle. Maha MZARI**

Sous la direction de : **M. Abdesselam FAZOUANE (INSEA)**  
**Mme. Mariam JAMALI (CNIA SAADA Assurance)**  
**M. Mohamed EL OUALI (CNIA SAADA Assurance)**

*Soutenu publiquement comme exigence partielle en vue de l'obtention du*

**Diplôme d'Ingénieur d'Etat**

**Option : ACTUARIAT-FINANCE**

*Devant le jury composé de :*

- **M. Abdesselam FAZOUANE (INSEA)**
- **M. Faissal GUEDIRA (INSEA)**
- **M. Mohamed EL OUALI (CNIA SAADA Assurance)**
- **Mme. Mariam JAMALI (CNIA SAADA Assurance)**



## Problématique

Dans la tarification à priori, les bonifications sont octroyées sur la base des observations faites par l'assureur dans un passé récent sur des assurés au profil identique. L'assuré est donc identifié à un collectif d'assurés qui lui sont indiscernables sur base des informations détenues par la compagnie, et se voit appliquer une tarification basée sur l'expérience relative à ce collectif. Bien évidemment, le groupe d'individus auquel est identifié chaque assuré ne lui est pas en tout point semblable ; il subsiste une hétérogénéité plus ou moins forte, selon l'étendue de l'information dont dispose l'assureur. A ce stade, deux questions méritent d'être posées :

1. de quelles informations dispose la compagnie à propos de ses assurés?

Clairement, si cette information est réduite, l'assuré sera identifié à des individus qui pourraient lui être très différents.

2. de combien d'assurés dispose la compagnie pour fonder ses décisions?

Une compagnie de taille modeste utilisant de nombreux facteurs de tarification en arrive à des classes de risque où ne se trouvent que quelques dizaines d'assurés, ce qui laisse planer un sérieux doute quant à la pertinence de ses conclusions.

Le principe de la tarification à priori consiste à déterminer une prime de risque dont la valeur dépend d'un système de segmentation prenant en compte une série de critères (âge, ancienneté de véhicule, ancienneté de permis de conduite, CRM... etc.)

Le problème de la tarification à priori est qu'elle ne fait aucunement intervenir le mérite de l'assuré; au contraire, la tarification d'expérience (tarification a posteriori) mutualise les risques entre les individus qui ont connu une sinistralité comparable. Sur la base de modèles actuariels, on décidera de la part de chaque sinistre due au hasard, et dont l'assuré n'a donc pas à subir les conséquences en vertu du principe fondamental de la mutualisation des risques, et de la part du sinistre qui révèle les caractéristiques non observés de l'assuré. La tarification a posteriori peut paraître plus équitable que la tarification à priori car elle se fonde sur des actes effectifs de l'assuré lui-même et non sur sa comparaison à un collectif d'individus indiscernables du point de vue de l'assureur.

Le processus de fixation de la prime est en fait un processus d'évaluation du risque; à tout instant, la prime doit être proportionnelle au risque. La superposition de la correction a posteriori à la tarification à priori est donc une méthode d'évaluation continue du risque. Il est clair que ces deux problèmes ne peuvent pas s'envisager indépendamment l'un de l'autre.

En utilisant l'information contenue dans les fichiers de notre assureur privé (CNIA SAADA Assurance), nous allons tenter dans un premier lieu de répondre à ces deux questions :

Quel est le lien statistique entre le nombre d'accidents d'un individu, ses propres caractéristiques et les caractéristiques de son automobile ?

Est ce que ce lien reflète vraiment la réalité concernant l'historique sinistre de l'assuré ?

## Résumé et mots clés

Nous avons entrevu à travers ce document une méthodologie de l'analyse de la segmentation, de la modélisation et de la tarification en assurance automobile.

Nous avons tout d'abord compris que dans un contexte hyperconcurrentiel des compagnies d'assurance, la maîtrise de la segmentation et de la tarification en découlant est primordiale pour préserver le portefeuille ou conquérir de nouveaux assurés. Nous avons également constaté que dans cet univers d'innovation continue, la segmentation des risques semble devenir de plus en plus poussée. A partir de ces constats, nous avons cherché à mettre en avant les différentes étapes de la modélisation du risque automobile. Nous avons ainsi entrevu de façon théorique et illustrée les notions d'extraction de bases de données, d'analyse descriptive, de la formation de classes de risques puis de modélisation linéaire généralisée, et enfin les notions fondamentales de la tarification dans ses deux formes à priori et à postériori.

Premièrement, à travers la logique sous jacente à la mise en place d'une extraction de données, nous avons pu comprendre et préciser l'importance de la détection d'erreurs et de l'épure des données à partir d'un périmètre d'étude clairement défini. L'étape suivante fut la modélisation linéaire généralisée. Nous avons pu mettre en avant l'ensemble des techniques classiques de la modélisation du risque automobile. Et pour finir, nous avons mis l'accent sur la correction de la prime calculée dans l'étude à priori et ce à travers la théorie de crédibilité et la construction d'un système Bonus-malus.

### **Mots clés :**

- Ratio S/P,
- Segmentation,
- Tarification à priori,
- Tarification à postériori,
- Bonus-malus,
- Crédibilité.

---

## Dédicace

---

*Je dédie ce travail,*

*A mes très chers parents,*

*Qui n'ont économisé aucun effort pour que je puisse être ce que je suis.*

*Aucun mot n'exprimera ma profonde gratitude pour tout l'amour,  
le soutien, les sacrifices, la confiance et surtout leurs prières qu'ils me  
font.*

*A mes adorables sœurs Sanae et Wiame.*

*A mon oncle Mustapha pour son soutien qu'il n'a cessé d'apporter au  
cours de ma formation.*

*A toute ma famille.*

*Et à tous mes amis avec qui j'ai partagé de meilleurs souvenirs.*

*Anass AHAYAN*

*A mon très cher père, pour son amour, son soutien moral et tout simplement pour avoir été là pour moi dans mes moments de joie et d'échec. Tu resteras toujours mon meilleur exemple dans la vie.*

*A ma tendre mère pour m'avoir éclairé le chemin et pour toutes ses prières.*

*A ma très chère sœur Fatin, mon amie et mon âme sœur. Un grand merci pour avoir été le garant d'une existence paisible et d'un avenir radieux,*

*A mes très chères frères Hakim et Ayoub.*

*A ma chère sœur Ghada.*

*A mon très cher Moad, merci d'avoir été la personne que j'ai toujours espérée.*

*A mon cher binôme Anass, pour son travail remarquable, son soutien, et surtout pour avoir été le meilleur collègue et ami qu'on pouvait avoir.*

*A ma très chère amie Ghizlane, merci de m'avoir aidé, encouragé et tout simplement d'avoir été ma meilleure amie pendant toutes ces années.*

*Maha MZARI*

## Remerciements

*Tout d'abord nous tenons à remercier notre cher professeur, directeur et encadrant M. Abdesselam FAZOUANE qui s'est montré à l'écoute et disponible, ainsi pour l'inspiration, l'aide et les conseils fructueux qu'il nous a prodigués tout au long de la réalisation de ce travail.*

*Qu'il nous soit permis, au terme de ce travail, d'exprimer notre gratitude et vifs remerciements à notre encadrante de stage Mme. Mariam JAMALI. Qu'elle trouve ici le témoignage de notre estime et de notre profonde reconnaissance pour sa disponibilité, ses directives précieuses, ses conseils, la qualité de son suivi et sa compétence qu'elle a su nous prodiguer tout au long de notre stage malgré ses occupations extrêmes.*

*On tient à remercier tout particulièrement M. Mohamed EL OUALI de nous avoir acceptés en tant que stagiaires, et d'avoir convenu de partager son professionnalisme avec nous débutants.*

*Un grand merci est adressé à M. Fouad Mari pour sa collaboration inestimable et ses directives précieuses.*

*On ne peut oublier la contribution significative de M. Pierre THÉROND que ce soit en terme de documents ou d'informations.*

*Nous ne saurions terminer ces remerciements sans rendre grâce à la gentillesse du personnel de la CNIA SAADA Assurance, en particulier M. Abdelhakim FATTY, Mme. Assia ABOULOULA et M. Khalil RZAMA qui ont rendu le milieu de travail plus aimable en faisant preuve d'un grand sens d'hospitalité.*

*Notre sincère reconnaissance va à tous ceux qui, de près ou de loin, ont contribué à l'aboutissement et au bon déroulement de ce modeste travail.*

## Table des matières

Problématique.....	3
Résumé et mots clés .....	4
Dédicace .....	5
Remerciements .....	7
Table des matières .....	8
Liste des abréviations .....	10
Liste des tableaux .....	11
Liste des figures .....	12
Introduction générale.....	13
Chapitre préliminaire : Aperçu sur le marché de l'assurance automobile au Maroc .....	15
Section I : L'organisme d'accueil .....	19
I.1 Généralités .....	20
I.2 Acquisitions .....	20
I.3 Rapprochement et fusion .....	21
I.4 Appartenance au Groupe SAHAM .....	21
I.5 Les activités de CNIA SAADA Assurance .....	22
I.6 Organigramme .....	23
Section II : Tarification à priori, segmentation et GLM .....	24
Chapitre I : Préliminaire à la modélisation du risque automobile.....	26
II.1.1 Introduction .....	26
II.1.2 Etude descriptive .....	27
II.1.3 La segmentation du portefeuille .....	35
Chapitre 2 : Modèle linéaire généralisé.....	41
II.2.1 Introduction .....	41
II.2.2 Analyse de la corrélation .....	42
II.2.3 Sélection des variables explicatives .....	43
II.2.4 Le modèle linéaire généralisé .....	44
II.2.5 Les classes de tarification .....	53
Conclusion.....	55
Section III : Tarification à postériori, élaboration d'un système Bonus-malus .....	56
Introduction .....	57
Chapitre 1 : Soubassements théoriques du système Bonus-malus .....	58
III.1.1 Introduction .....	58
III.1.2 Modèle poissonien : Portefeuille homogène .....	59

III.1.3 Modèle binomial négatif : Portefeuille hétérogène .....	59
III.1.4 Formulation du modèle Markovien à classes .....	61
Chapitre 2 : Modélisation du système Bonus-malus .....	62
III.2.1 Introduction .....	62
III.2.2 Le test $\chi^2$ d'ajustement des lois .....	63
III.2.3 Modèle poissonien : Portefeuille homogène .....	63
III.2.4 Modèle binomial négatif : Portefeuille hétérogène .....	64
Chapitre 3 : Élaboration des échelles Bonus-malus .....	66
III.3.1 Introduction .....	66
III.3.2 Préambule : <i>Principe adopté</i> .....	67
Conclusion.....	69
Section IV : Tarification à postériori, crédibilité .....	70
Introduction .....	71
Chapitre 1 : Description du modèle adopté, Bühlman-Straub .....	72
IV.1.1 Introduction .....	72
IV.1.2 Description du modèle .....	73
IV.1.3 Hypothèses du modèle .....	73
IV.1.4 Estimateurs des paramètres de structure .....	74
IV.1.5 Récapitulatif de la démarche .....	74
Chapitre 2 : Application aux données .....	75
IV.2.1 Introduction .....	75
IV.2.2 Commentaires et interprétations .....	76
IV.2.3 Estimation des paramètres et exemple d'illustration .....	77
Conclusion.....	78
Conclusion générale .....	79
Bibliographie .....	80
Annexes .....	81
Annexe I : Les statistiques après regroupement pour les segmentations retenues .....	81
Annexe II : Extraits des procédures SAS de la cartographie du risque automobile.....	82
Annexe III : Arbres de décision issues de SPSS ANSWER TREE .....	86
Annexe IV : Aperçu sur le système Bonus-malus marocain.....	88
Annexe V : Glossaire des termes clés de l'étude .....	89
Annexe VI : Extrait de la base PROD .....	90
Annexe VII : Extrait de la base Sinistre .....	91

## Liste des abréviations

CHAID	Chi-squared Automatic Interaction Detector
CIMR	La Caisse interprofessionnelle marocaine de retraite
CNIA	La Compagnie Nord-Africaine et Intercontinentale d'Assurance
CRM	Coefficient réduction majoration
CRT	Current Reality Tree
CSP	Catégorie socio professionnelle
DDL	Degré de liberté
FMSAR	Fédération Marocaine des Sociétés d'Assurances et de Réassurance
GLM	Modèles linéaires généralisés
INSEA	Institut National de Statistique et d'Economie Appliquée
MASI	Moroccan All Shares Index
MEC	Mise en circulation
PROD	Production
Q-Q plot	Quantile-Quantile plot
RC-auto	Responsabilité civile automobile
S/P	Ratio entre les coûts des sinistres d'un contrat et ses primes acquises
SAS	Statistical Analysis System
SBM	Système Bonus-malus
SM	Situation matrimonial

## Liste des tableaux

<b>Tableau 1</b>	Structure du chiffre d'affaire du secteur des assurances au Maroc.....	16
<b>Tableau 2</b>	Évolution du chiffre d'affaire des différentes branches du secteur des assurances au Maroc.....	17
<b>Tableau 3</b>	Informations sur la CNIA SAADA Assurance.....	20
<b>Tableau 4</b>	Les statistiques des valeurs manquantes des variables d'étude.....	31
<b>Tableau 5</b>	les quantités statistiques calculées sur l'ensemble des variables tarifaires.....	33
<b>Tableau 6</b>	Proportion des femmes et des hommes dans le portefeuille.....	34
<b>Tableau 7</b>	La sinistralité selon le sexe de l'assuré.....	34
<b>Tableau 8</b>	les quantités statistiques de la sinistralité selon la variable ancienneté de permis.....	39
<b>Tableau 9</b>	les quantités statistiques de la sinistralité du portefeuille selon l'âge de l'assuré.....	39
<b>Tableau 10</b>	les quantités statistiques de la sinistralité du portefeuille selon la situation matrimoniale.....	39
<b>Tableau 11</b>	les quantités statistiques de la sinistralité du portefeuille selon la variable ancienneté du permis.....	40
<b>Tableau 12</b>	les quantités statistiques de la sinistralité du portefeuille selon la variable énergie*puissance fiscale.....	40
<b>Tableau 13</b>	Les quantités statistiques de la sinistralité selon le CRM.....	41
<b>Tableau 14</b>	Les quantités statistiques de la sinistralité du portefeuille selon la variable région.....	41
<b>Tableau 15</b>	Corrélation entre les variables quantitatives.....	43
<b>Tableau 16</b>	Résultat du test khi2 d'indépendance entre la puissance fiscale et l'énergie.....	44
<b>Tableau 17</b>	Résultat du test khi2 d'indépendance entre le sexe et la région.....	44
<b>Tableau 18</b>	Résultats de l'analyse pour le modèle avec les 5 variables explicatives.....	45
<b>Tableau 19</b>	Lois classiques de la famille exponentielle.....	47
<b>Tableau 20</b>	Fonctions de lien pour les différentes lois.....	47
<b>Tableau 21</b>	Comparaison des qualités d'ajustement du GLM sur la fréquence.....	49
<b>Tableau 22</b>	Test de significativité de Wald pour le GLM avec la loi Binomiale Négative.....	49
<b>Tableau 23</b>	Test de significativité de Wald pour le GLM avec la loi Binomiale Négative après regroupement.....	50
<b>Tableau 24</b>	test de significativité de Wald pour le GLM avec la loi Binomiale Négative après le deuxième regroupement.....	51
<b>Tableau 25</b>	Comparaison des qualités d'ajustement du GLM sur le coût moyen des sinistres.....	51
<b>Tableau 26</b>	Test de significativité de Wald pour le GLM avec la loi Gamma.....	52
<b>Tableau 27</b>	Test de significativité de Wald pour le GLM avec la loi Gamma après le regroupement final.....	52
<b>Tableau 28</b>	Extrait du tableau exhaustif de l'ensemble des classes de tarifs.....	54
<b>Tableau 29</b>	La totalité des classes de tarifs a priori définitives.....	56
<b>Tableau 30</b>	Présentation d'une échelle Bonus- malus.....	62
<b>Tableau 31</b>	Ajustement de la sinistralité par la loi de Poisson.....	64
<b>Tableau 32</b>	Estimation des paramètres de la loi de Poisson.....	65
<b>Tableau 33</b>	Ajustement de la sinistralité par une binomiale négative.....	65
<b>Tableau 34</b>	Estimation des paramètres de la loi binomiale négative.....	66
<b>Tableau 35</b>	Ajustement de la sinistralité par une binomiale négative.....	68
<b>Tableau 36</b>	Échelle Bonus-malus dans le modèle binomiale négative à présentation Markovienne.....	69
<b>Tableau 37</b>	Les facteurs de crédibilité : minimum, maximum et moyen / Bühlman-Straub.....	77
<b>Tableau 38</b>	Les paramètres de structure estimés du modèle de Bühlman-Straub.....	78
<b>Tableau 39</b>	Échantillon de primes crédibilisées selon le modèle Bühlman-Straub.....	78

## Liste des figures

<b>Figure 1</b>	Logo du Groupe SAHAM.....	20
<b>Figure 2</b>	Logo de CNIA SAADA Assurance.....	20
<b>Figure 3</b>	L'organigramme du Groupe SAHAM.....	22
<b>Figure 4</b>	L'organigramme de CNIA SAADA Assurance.....	23
<b>Figure 5</b>	Variables de la base PROD.....	29
<b>Figure 6</b>	Variables de la base Sinistre.....	30
<b>Figure 7</b>	La fréquence et le coût des sinistres selon le sexe de l'assuré.....	35
<b>Figure 8</b>	La fréquence et le coût des sinistres selon l'énergie du véhicule.....	35
<b>Figure 9</b>	La fréquence et le coût moyen des sinistres selon l'ancienneté du véhicule.....	36
<b>Figure 10</b>	Cheminement de la construction d'un GLM.....	47
<b>Figure 11</b>	Q-Q plot pour la loi de Poisson.....	48
<b>Figure 12</b>	Q-Q plot pour la loi Gamma.....	49
<b>Figure 13</b>	La variation de la crédibilité des contrats individuels en fonction de leur prime.....	77

## Introduction générale

Le but essentiel de l'assurance consiste dans la formation d'une communauté dans la quelle les assurés mettent leurs risques en commun. L'assurance telle qu'elle est présentée et pratiquée par les sociétés commerciales n'est équitable que si la prime réclamée à chaque assuré correspond le plus exactement possible au risque réel que celui-ci fait peser sur le portefeuille.

Le principe essentiel de la tarification est la neutralité actuarielle (idéalement la prime pure doit couvrir le risque exact garanti par l'assureur), ce fonctionnement est altéré par des positions stratégiques et concurrentielles qui modulent le monde de l'assurance. Chaque compagnie s'intéresse un cœur de cible particulier, une clientèle spécifique qu'elle cherche à attirer, et cette clientèle ne correspond pas toujours à ce que l'on puisse appeler un **bon risque**. Les cibles visées par les assureurs sont bien souvent des populations importantes en besoin d'assurances et qui représentent par conséquent d'éventuelles parts de marché à conquérir.

La difficulté principale à laquelle est confronté un actuairé dans la construction d'un tarif technique réside dans la détermination du risque réel que représente individuellement chaque assuré. L'habileté à conduire une voiture ne dépend pas seulement de critères techniques observables à la conclusion du contrat, tels que l'ancienneté du véhicule et le sexe du conducteur, la puissance du véhicule par exemple, mais également de facteurs humains difficilement observables (comme le degré d'agressivité, l'abus d'alcool ou la connaissance du code de la route par exemple). Par conséquent, des assurés identiques du point de vue de l'assureur peuvent très bien s'avérer être des risques complètement incomparables. Nonobstant, on peut tenir compte a posteriori de ces facteurs non observables par le biais de l'historique des sinistres (nombre d'accidents) observé pour chaque assuré, afin de corriger la prime de base.

La tarification en assurance automobile peut généralement se décomposer en deux étapes. La première, dite tarification à priori, consiste à segmenter le portefeuille d'assurés en classes de risques homogènes en fonction de certains critères retenus dès la souscription (critères tarifaires). La deuxième, dite tarification a posteriori (tarification corrective et incitative), permet d'affiner la tarification en intégrant une dimension individuelle. Les systèmes de Bonus-malus sont une des méthodes possibles de la tarification corrective.

Le principe de ce système de réduction majoration est de placer l'assuré dans une échelle qui compte un certain nombre de degrés. A chacune de celles-ci est associé un pourcentage, et l'assuré doit acquitter un montant égal à ce pourcentage d'une prime de base librement fixée par la compagnie. Le fonctionnement du système Bonus-malus est simple : tout d'abord, un degré de l'échelle est prévu pour le nouvel assuré et, ensuite, un réajustement de la position de l'assuré sur cette échelle est effectué annuellement en fonction de la sinistralité de cet assuré et conformément aux règles du système. Le système Bonus-malus n'est applicable qu'aux véhicules relevant de l'usage **privé** ou **affaire**. La prime d'assurance varie à chaque échéance principale du contrat. Elle est déterminée en multipliant la prime de base pour la responsabilité civile par un coefficient de réduction ou de majoration fixé.

Nous limiterons notre propos à l'assurance de la responsabilité civile du conducteur d'un véhicule automoteur à l'égard des autres usagers de la route, mieux connue sous le nom de RC auto. Cette couverture revêt une importance sociale considérable de part son aspect

obligatoire, elle garantit aux citoyens la réparation du préjudice qu'ils auraient subi du fait d'un tiers en leur évitant d'avoir à pâtir de l'insolvabilité de celui-ci.

Notre travail va être enchaîné de la manière suivante :

Nous nous proposons, dans cette étude, de formaliser un modèle optimal de tarification basé sur les caractéristiques des assurés (un modèle à priori) et sur le nombre d'accidents passés des individus (un modèle à postériori), afin d'ajuster les primes individuelles selon le degré de risque intrinsèque à travers le temps, de sorte que chaque assuré paye une prime proportionnelle à sa fréquence d'accident et que l'assureur soit financièrement équilibré.

Les données sur lesquelles nous allons nous baser pour constituer notre modèle sont celles d'une compagnie d'assurances privée (CNIA SAADA Assurance) qui détenait plus de 8 % du marché d'assurance automobile en 2010 et dont la clientèle était distribuée dans toutes les régions du pays.

---

*« Sans les assurances,  
personne ne circulerait ... en  
voiture à travers les rues. Un  
bon chauffeur est conscient du  
risque qu'il court à chaque  
instant, le risque de renverser  
un piéton. »*

**Henry Ford**

---

## **CHAPITRE PRÉLIMINAIRE**

# APERÇU SUR LE MARCHÉ DE L'ASSURANCE AUTOMOBILE AU MAROC



### **I. Introduction**

Au Maroc, le secteur des assurances, de par son rôle essentiel en matière de collecte et d'injection des flux financiers dans le système économique, constitue un levier de développement et un stimulateur de la croissance.

Ce secteur ne cesse de prendre de l'importance au sein du marché marocain, que ce soit par les sommes importantes qu'il draine ou au niveau des tendances lourdes qui le caractérisent ces dernières années.

Il s'avère donc primordial avant d'entamer une étude de tarification d'un portefeuille en assurance automobile de s'intéresser à l'évolution de ce domaine au Maroc. Tout particulièrement en termes de son classement et sa position parmi ses concurrents internationaux, ainsi que son avancement en terme de chiffres d'affaires et de primes émises.

Cet aperçu va donc nous donner une idée sur la place de l'assurance non vie sous sa forme de l'assurance automobile et plus particulièrement l'assurance obligatoire de la responsabilité civile, sujet de notre projet, ainsi que son importance sur le développement de ce secteur.

## II. Évolution du secteur des assurances au Maroc

Le secteur marocain des assurances, classé deuxième après celui de l'Afrique du Sud, est l'un des secteurs les plus dynamiques de l'économie nationale, vu sa croissance significative et son important chiffre d'affaires annuel qui a atteint 26,03 milliards de dirhams (MMDH) en 2012, dont 8,84 MMDH en assurance vie et 17,19 MMDH en assurance non vie, selon la Fédération Marocaine des Sociétés d'Assurances et de Réassurance (FMSAR).

Avec l'amorce du nouveau millénaire, le secteur des assurances au Maroc à l'instar des autres pays de par le monde connaît de profondes mutations et se trouve ainsi confronté à de grands et importants défis (libéralisation, concentration, assurance maladie obligatoire, bancassurance) qui vont certainement affecter le processus de sa croissance, des défis qui une fois relevés, le secteur sortira certainement plus solide et plus apte à mener à bien sa principale mission, qui est celle de permettre à l'économie marocaine à mieux intégrer le nouveau siècle.

Parmi les secteurs de l'assurance qui ont vu une grande évolution, la branche non vie paraît comme l'acteur principal sur le podium de l'assurance.

Ce constat le témoigne le rapport annuel de la Fédération Marocaine des Sociétés d'Assurances et de Réassurance concernant le premier semestre de l'année 2012, qui montre que la branche de l'assurance automobile représente le pilier le plus important de l'assurance non vie.

**Tableau 1 : Structure du chiffre d'affaire du secteur des assurances au Maroc**

### Structure du Chiffre d'Affaires

	Chiffre d'Affaires	Contribution
Automobile	4 614,5	32,1%
Assurances Vie et Capitalisation	4 183,7	29,1%
Accidents Corporels	1 597,8	11,1%
Accidents du Travail	1 321,9	9,2%
Incendie	889,9	6,2%
Assistance - Crédit - Caution	623,7	4,3%
Transport	365,7	2,5%
RC Générale	319,9	2,2%
Autres Opérations Non Vie	222,1	1,5%
Risques Techniques	171,0	1,2%
Acceptations en réassurance	57,3	0,4%
<b>Total</b>	<b>14 367,6</b>	<b>100%</b>

*En millions de dirhams*

*Source : Fédération marocaine des sociétés d'assurance et de réassurance (FMSAR)*

Selon la même source (FMSAR) ce pilier a présenté une évolution importante pendant les trois années 2010, 2011 et 2012.

**Tableau 2 : Évolution du chiffre d'affaire des différentes branches du secteur des assurances au Maroc**

Evolution du Chiffre d'Affaires

	1 <sup>er</sup> Sem. 2010	1 <sup>er</sup> Sem. 2011	1 <sup>er</sup> Sem. 2012	Evolution 2011/2012	Evolution 2010/2011
<b>Assurances Vie &amp; Capitalisation</b>	<b>3 053,3</b>	<b>3 604,7</b>	<b>4 211,7</b>	<b>16,84%</b>	<b>18,06%</b>
Individuelles	1 881,0	1 879,8	2 502,8	33,14%	-0,06%
Groupes	763,4	860,8	984,2	14,34%	12,75%
Capitalisation	222,9	590,8	537,5	-9,01%	165,07%
Contrats à Capital Variable	156,7	245,9	159,1	-35,32%	56,98%
Acceptations Vie	29,3	27,4	28,0	2,26%	-6,57%
<b>Assurances Non Vie</b>	<b>8 882,7</b>	<b>9 518,6</b>	<b>10 155,9</b>	<b>6,70%</b>	<b>7,16%</b>
Accidents Corporels	1 410,4	1 461,6	1 597,8	9,32%	3,63%
Accidents du Travail	1 201,4	1 298,7	1 321,9	1,78%	8,10%
Automobile	4 005,7	4 354,3	4 614,5	5,98%	8,70%
RC Générale	302,4	340,0	319,9	-5,89%	12,41%
Incendie	729,0	758,7	889,9	17,29%	4,08%
Risques Techniques	128,0	139,1	171,0	22,94%	8,67%
Transport	363,3	398,4	365,7	-8,21%	9,69%
Autres Opérations Non Vie	134,3	166,0	222,1	33,77%	23,59%
Assistance - Crédit - Caution	533,2	564,1	623,7	10,57%	5,79%
Acceptations Non Vie	74,9	37,7	29,3	-22,36%	-49,66%
<b>Total</b>	<b>11 936,0</b>	<b>13 123,3</b>	<b>14 367,6</b>	<b>9,48%</b>	<b>9,95%</b>

*Source : Fédération marocaine des sociétés d'assurance et de réassurance (FMSAR)*

A partir de ces deux tableaux, on constate que la part de la branche automobile est de 32,1 % et représente donc une tranche importante dans le chiffre d'affaire globale. Cette part de la branche automobile a connu une évolution de 8,70% entre les deux années 2010 et 2011 et a continué à évoluer en 2012 de 5,98% par rapport à l'an 2011.

La libéralisation de la responsabilité civile (RC) automobile n'avait pas un effet d'évolutif sur le tarif RC puisque ce dernier n'a pas subi une réelle modification et la concurrence reste limitée au niveau du tarif des garanties annexes. Il s'agit d'un marché en développement, dont le potentiel reste important au vu de la croissance continue du parc et le faible taux d'équipement des ménages en termes de véhicules.

La place importante qu'occupe cette branche d'automobile s'explique en grande partie par l'aspect obligatoire de l'assurance responsabilité civile, mais également par la volonté des assurés prudents considérés comme de bons risques de se couvrir au mieux contre ce risque quotidien.

### **III. La responsabilité civile au Maroc**

L'assurance de responsabilité civile a pour objet de garantir la dette de responsabilité d'une personne physique ou morale, suite à un dommage causé à un tiers par sa faute non intentionnelle, lorsqu'il est établi que ces dommages en sont la conséquence directe.

C'est dans ce cadre qu'intervient l'assurance Responsabilité civile automobile, qui garantit la responsabilité civile des auteurs de dommages corporels ou matériels, causés à des tiers par un véhicule terrestre à moteur non lié à une voie ferrée ou par ses remorques ou semi-remorques.

La responsabilité civile automobile, trouve son fondement dans les dispositions de l'article 78 et 88 du DOC de 1913 qui régissent respectivement la faute et la responsabilité du fait des choses.

Celle-ci a vu le jour en France par l'art 1384 alinéas 1 du code civil Français et a été confirmée par l'arrêt JAND HEUR du 13/02/30 de la cour de cassation, pour rattacher définitivement la responsabilité civile du fait de l'automobile à la RC du fait des choses.

Le législateur marocain a voulu mettre fin à plusieurs abus à l'égard des victimes des accidents de la circulation, et il lui est apparu indispensable de prendre en considération la situation de ces victimes et d'accélérer leur indemnisation. Pour ce faire, il a promulgué le Dahir du 02/10/1984 relatif à l'indemnisation des victimes des accidents de la circulation.

Selon la Fédération Marocaine des Sociétés d'assurance et de réassurance, tout propriétaire d'un véhicule terrestre à moteur non lié à une voie ferrée est tenu, au Maroc, de par la loi, de souscrire à une assurance dite « responsabilité civile automobile ».

Cette assurance permet de couvrir la responsabilité civile du souscripteur du contrat, du propriétaire du véhicule et de toute personne ayant, avec leur autorisation, la garde ou la conduite du véhicule.

Sont couverts par cette garantie:

- Les dommages matériels causés à d'autres véhicules ou à des biens appartenant à des tiers ;
- Les dommages corporels causés aux piétons, aux occupants des véhicules tiers ainsi qu'aux personnes transportées.

Sont exclus de la garantie :

- Le souscripteur du contrat, le propriétaire du véhicule assuré et de toute personne ayant, avec leur autorisation, la garde ou la conduite du véhicule;
- Le conducteur du véhicule ;
- Lorsqu'ils sont transportés dans le véhicule assuré, des représentants légaux de la personne morale propriétaire du véhicule assuré ;
- Pendant leur service, les salariés ou les préposés de l'assuré (ou du conducteur) responsable de l'accident.

# **SECTION I**

## **L'organisme d'accueil**

### **CNIA SAADA Assurance**

**Plan :**

- Généralités;
- Acquisitions ;
- Rapprochement et fusion ;
- Appartenance au Groupe SAHAM ;
- Les activités de CNIA SAADA Assurance ;
- Organigramme.

## I.1 Généralités

**Figure 1 : Logo du Groupe SAHAM**



**Figure 2 : Logo de CNIA SAADA Assurance**



CNIA SAADA Assurance est une société anonyme d'assurances et de réassurance au capital social de 411.687.400 de DH. Entreprise réglé par la loi n° 17-99 portant code des assurances. Elle est N°1 en assurance automobile et en santé. Elle dispose du 1er réseau d'agents exclusifs avec plus de 385 agents.

Depuis sa création en 1949, sous le nom de la Compagnie Nord-Africaine et Intercontinentale d'Assurance. CNIA Assurance s'est progressivement affirmée comme un des leaders du secteur des assurances au Maroc.

Créée en 1961, Es Saada a connu une progression significative depuis ces cinq dernières années, notamment grâce à l'obtention d'importants marchés auprès de nombreux institutionnels et grandes entreprises privées.

**Tableau 3: Informations sur la CNIA SAADA Assurance**

<b>Création</b>	2009
<b>Dates clés</b>	1949 : Création de CNIA
	2009 : Création de CNIA SAADA Assurance
<b>Personnages clés</b>	Moulay Hafid Elalamy (PDG)
<b>Forme juridique</b>	Société anonyme
<b>Action</b>	MASI : <b>CNIA SAADA</b>
<b>Slogan</b>	Vis ta vie
<b>Siège social</b>	CNIA SAADA Assurance, 216 Boulevard Zerktouni - Casablanca
<b>Direction</b>	Mehdi Tazi (DG)
<b>Actionnaires</b>	Groupe Saham (57 %)
	Sanam Holding (15 %)
<b>Activité</b>	Assurances et services financiers ("Protection financière")
<b>Produits</b>	Assurance vie, assurance dommages et gestionnaire d'actifs
<b>Effectif</b>	Entre 700 et 1000

*Source : CNIA SAADA Assurance*

## I.2 Acquisitions

En 1997 : L'Etat cède sa part dans le capital de la compagnie à un groupe d'assurance de premier ordre, Arab Insurance Group qui avec 67% des parts devient actionnaire

majoritaire dans le capital de la compagnie et ce qui a permis à cette dernière de bénéficier d'un savoir faire additionnel pour renforcer sa position sur le marché de l'assurance.

En 2005, le Groupe Saham acquiert CNIA Assurance, devenant l'actionnaire de référence. Cette acquisition est suivie par le déploiement d'un dispositif d'initiatives stratégiques visant une croissance rentable et un positionnement basé sur la qualité de service et de l'innovation continue. CNIA Assurance bénéficie ainsi d'une remarquable progression de ses performances commerciales, techniques et opérationnelles, confirmant son statut d'acteur incontournable du secteur.

En 2006, CNIA Assurance acquiert Es Saada, leader national du réseau de distribution depuis 50 ans. Cette nouvelle acquisition stratégique permet à CNIA Assurance d'offrir plus de proximité à sa clientèle et d'affirmer sa vocation d'entreprise "en action" dans un secteur de plus en plus dynamique. Plusieurs actions de mise à niveau sont entamées au niveau de Es Saada : assainissement des comptes, amélioration de la qualité de service, respect des engagements.

### **I.3 Rapprochement et fusion**

Présentes sur l'ensemble des segments d'assurance vie et d'assurance dommages incluant l'assurance automobile, l'assurance maladie, l'assurance vie et la retraite par capitalisation, CNIA Assurance et Es Saada entament un nouveau tournant de leur histoire en 2007, marqué par le rapprochement des deux compagnies. Le 24 juin 2009, les marques CNIA Assurances et Es Saada donnent naissances à CNIA SAADA Assurance.

La fusion de ces deux compagnies a été réalisée après l'élaboration d'un planning et d'un cahier des charges stricts respectant les standards internationaux en matière d'intégration, ont précisé les responsables des deux sociétés lors d'une conférence de presse tenue le 24 juin 2009.

Entamée depuis un an, cette fusion a pour objectifs de poursuivre le rapprochement des services de l'assurance des clients (particuliers et entreprises) et répondre à leurs préoccupations.

La conjugaison des forces des deux compagnies confèrera à la nouvelle entité tous les atouts nécessaires pour envisager l'avenir avec encore plus de professionnalisme et de sérénité, souligne un communiqué de presse.

### **I.4 Appartenance au Groupe SAHAM**

Avec le rachat de CNIA Assurance en 2005 et des Assurances Es Saada en 2006, puis suite à leur fusion en 2009, le Groupe Saham est devenu un acteur majeur du marché de l'assurance.

Actionnaire de référence de CNIA SAADA Assurance, le Groupe Saham détient 51% des parts et participe résolument au dispositif d'initiatives stratégiques, commerciales et techniques mis en place par la Compagnie.

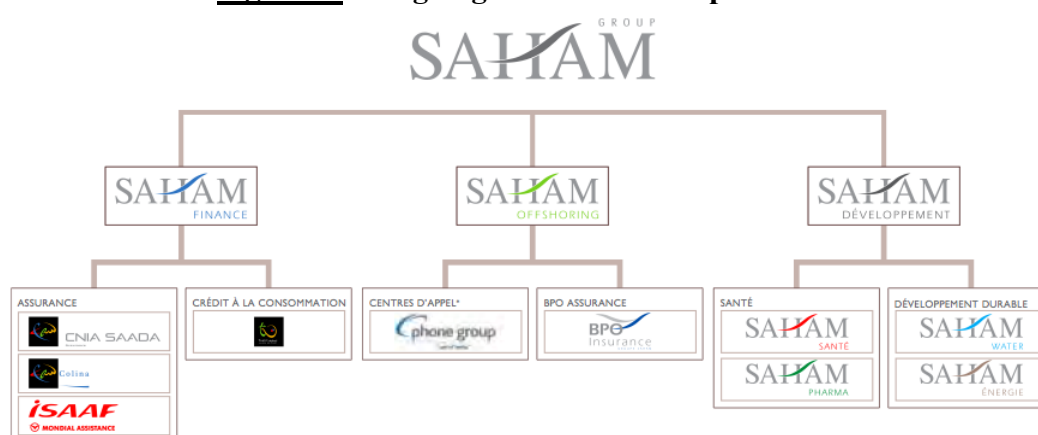
Au niveau national, le Groupe Saham se positionne comme un opérateur clé dans plusieurs secteurs d'activité et plus particulièrement dans les métiers de service où il développe des synergies autour de deux secteurs: l'Offshoring et les métiers de la Finance.

Doté d'une structure flexible et d'une équipe performante dédiée au service de

l'organisation et du développement, le Groupe Saham apporte à ses filiales un support et une assistance opérationnelle en matière de pilotage stratégique et budgétaire, de gestion de la trésorerie et d'optimisation des modèles de financement, d'audit et du contrôle interne. Enfin, les filiales du Groupe profitent d'une veille concurrentielle et d'un support commercial pour l'identification de projets, l'élaboration d'analyses sectorielles et la recherche d'opportunités commerciales.

Dans le cadre de sa stratégie de développement, privilégiant une cohérence plus affirmée et une intégration plus étendue de ses métiers, le Groupe est actuellement en phase de création de son pôle santé et de ses premières structures pour le développement du Business Process Outsourcing en assurance.

**Figure 3: L'organigramme du Groupe SAHAM**



*Source : Rapport annuel 2011 de CNIA SAADA Assurance*

Le Conseil d'Administration de CNIA SAADA compte sept Administrateurs dont deux Administrateurs dirigeants :

- Moulay Hafid Elalamy, Président Dirigeant ;
- Ghita Lahlou, Administrateur Dirigeant ;
- Saïd Alj, Administrateur Non Dirigeant ;
- Kofi Bucknor, Administrateur Non Dirigeant ;
- Alain Demissy, Administrateur Non Dirigeant ;
- SAHAM SA, Administrateur Non Dirigeant ;
- SAHAM Finances, Administrateur Non Dirigeant.

### **I.5 Les activités de CNIA SAADA Assurance**

CNIA SAADA est une société d'assurance toutes catégories, exerçant son activité dans les assurances de dommages et les assurances de personnes. Elle offre à ses clients une gamme de produits incontournables. Elle vise tout type de catégorie socio professionnelle entre autres les particuliers, les professionnels, les communautés et les Entreprises.

La compagnie offre un grand nombre de produits en : assurance automobile, assurance de personnes, CNIA SAADA est présente dans les catégories suivantes :

- Maladie – Maternité, Hospitalisation, Incapacité – Invalidité, Individuelle ;
- Décès ;
- Retraite (CIMR, autres produits de capitalisation, à titre individuel ou collectif) ;
- Epargne – Education.

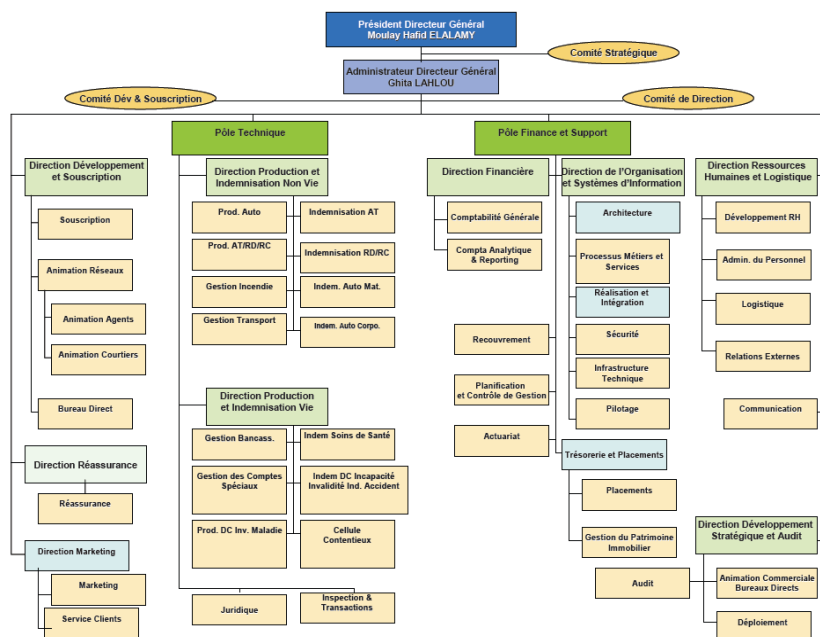
En ce qui concerne les assurances de dommages, la Compagnie opère dans les catégories suivantes :

- Automobile ;
- Accidents du travail ;
- Incendie et perte de bénéficiaires ;
- Maritime (corps et facultés) ;
- Responsabilité civile (générale, professionnelle, produits, chasse) ;
- Risques techniques (tous risques chantiers, informatique, montages, bris de machine, RC décennale) ;
- Multirisques (Habitation, Professions Libérales, Industrielles, Artisans-Commerçants) ;
- Autres Risques Divers (Vol, Bris de glace, Dégâts des eaux).

Ayant à son actif 13,7% des parts de marché, « CNIA SAADA Assurance » occupe la quatrième position à l'échelle nationale. Elle est le **premier assureur automobile**, le premier assureur **santé privé** du Royaume et le premier réseau d'agences évalué à plus de trois cents agents.

## I.6 Organigramme

**Figure 4: L'organigramme de CNIA SAADA Assurance**



Source : Rapport annuel 2011 de CNIA SAADA Assurance

# **SECTION II**

## **Tarification à priori**

### **Segmentation et GLM**

**Plan :**

- Introduction ;
- Préliminaire à la modélisation du risque automobile ;
- Modèle linéaire généralisé :  
Élaboration des tarifs actuariels ;
- Conclusion.

## **Introduction**

Dans un contexte de concurrence de secteur d'assurance automobile au Maroc, on comprend la nécessité de posséder une tarification précise et adaptée à chaque portefeuille. Cette tarification doit s'appuyer sur une segmentation plus ou moins poussée mais pertinente. Celle-ci consiste à différencier les assurés et le risque qu'ils portent. On obtient ainsi différentes classes de risques en fonction des caractéristiques de l'assuré et des quantités statistiques relatives à chacune d'elles. Chaque catégorie se verra ainsi attribuer un tarif qui lui sera propre, en adéquation avec le risque associé.

La tarification à priori repose sur une segmentation du portefeuille, on cherche à classer les assurés en fonction de leur risque potentiel. Il s'agit donc de sélectionner des critères de tarification qui soient pertinents et commercialement utilisables.

La segmentation consiste à analyser et contrôler l'adaptation des primes aux sinistres suivant des classes de risques homogènes, de façon à en tirer des conséquences du point de vue technique.

Ces critères permettent de classer tout nouvel assuré dans une catégorie uniforme de risque à laquelle est lié un niveau de prime sensé refléter les potentialités d'accident que représenter le nouvel assuré pour la compagnie d'assurance. L'objectif de cette tarification à priori est de réduire l'hétérogénéité qui préexiste dans un portefeuille d'assurés et ainsi de les classer selon le potentiel de risque qu'ils représentent pour l'assureur.

# **CHAPITRE 1**

## PRELIMINAIRE A LA MODELISATION DU RISQUE AUTOMOBILE



### **II.1.1 Introduction**

Avant chaque travail de tarification, une étude d'exploration de la base de données s'avère nécessaire afin de la rendre plus exploitable et adéquate. Cette étude va se passer en deux étapes : La première servant à faire adapter notre base aux buts de l'étude, et la deuxième consiste à vérifier la pertinence des variables de l'étude.

## **II.1.2 Etude descriptive**

### **II.1.2.1 Construction de la base de données**

Notre banque de données paraît intéressante, puisqu'elle nous permet d'avoir des statistiques individuelles sur toutes les caractéristiques, et de l'assuré et de sa voiture et ceci sur plusieurs années, nonobstant elle demeure brute avec des données qui peuvent être erronées, incomplètes ou obsolètes. Il est donc primordial pour arriver à un tarif correct de purifier notre base de données et de se pencher sur sa constitution.

De ce fait la phase d'extraction des données et leur nettoyage représente une part considérable du temps consacré à l'étude.

La plupart des compagnies sont à présent conscientes de la nécessité de disposer de données aussi nombreuses et de bonne qualité que possible. Les données servant à notre étude se composent en deux fichiers: Un fichier production considérant les années 2010 et 2011 et l'autre portant sur l'historique de sinistralité de cinq années consécutives, de 2007 à 2011.

#### **i) Variables de l'étude**

Dans l'assurance RC-auto, on peut choisir entre un grand nombre de critères de risque différents.

Notre base de données nous provient des fichiers « Émission & Sinistre » de la compagnie CNIA SAADA Assurance.

Pour chaque assuré, nous avons pu dégager :

*Les variables décrivant le fichier « Production » :*

**Figure 5 : Variables de la base PROD**

Nom de la variable

An\_Emission  
Annee\_effet  
CRM  
CYLINDREE  
CodeP\_client  
Code\_courtier  
Contrat\_resilie  
DATE\_MEC  
DATE\_PERMIS  
Date\_Effet  
Date\_Emission  
Date\_Fin  
Date\_First\_Ass  
Date\_Naissance  
Date\_Resil  
ENERGIE  
ID\_PRODUIT  
IMMATRIC  
IND\_FLOTTE  
Lieu\_Risque  
MARQUE  
MODEL  
NBR\_PLACES  
Nbr\_Conducteurs  
Num\_Avenant  
Num\_Pol  
Num\_attestation  
PRODUIT  
PUISS\_FISC  
RC  
Sexe  
Situation  
Situation\_Matrimoniale  
Type\_Echeance  
Type\_avenant  
Usage  
VILLE\_ZONE\_CIRCULATION  
Valeur\_Conv\_O\_N  
Valeur\_Conventionnelle  
Valeur\_neuve  
Valeur\_venale  
Ville\_Courtier

*Source : Elaborée par les auteurs, sortie SAS*

Ce fichier contient donc les informations relatives aux contrats des véhicules assurés par CNIA SAADA Assurance. Ces informations sont diverses et peuvent être ventilées ainsi :

- Les identificateurs : Numéro de police et numéro d'avenant.
- Les caractéristiques du véhicule : Puissance, énergie, ...
- Types de garanties : Responsabilité civile, dommage, ...
- Les primes 'RC' : Donne l'information sur le montant des primes collectées auprès des assurés.

Les variables décrivant le fichier « Sinistre » :

**Figure 6 : Variables de la base Sinistre**

Nom de la variable

---

Charge\_totale  
 Code\_Client  
 Code\_Garantie  
 Code\_Intermediaire  
 Code\_Mode\_Gestion  
 Code\_Nature\_Sinistre  
 Code\_Produit  
 Code\_Statut\_Sinistre  
 Date\_Cloture  
 Date\_declaration  
 Description\_Garantie  
 Immatriculation  
 Libelle\_Produit  
 Libelle\_usage  
 Lieu\_survenance  
 Nature\_Sinistre  
 Num\_Avenant\_Impact  
 Numero\_Contrat  
 Numero\_Sinistre  
 Prov\_Rec\_Emis  
 Prov\_Rec\_Subis  
 Provision\_Total  
 Rec\_Emis\_Reclames  
 Rec\_Payees  
 Rec\_Recuperes  
 Rec\_Subis\_Reclames  
 Reglements\_Total  
 Resp  
 Statut\_Sinistre  
 date\_ouverture  
 date\_survenance

*Source : Elaborée par les auteurs, sortie SAS*

Le fichier sinistre contient donc les informations liées à la sinistralité des contrats. On distingue :

- Les identificateurs : Numéro de contrat et Numéro d'avenant.
- Les informations liées aux sinistres : Date de survenance, nature du sinistre (matériel ou corporel), montant du sinistre, responsabilité de l'assuré (responsable, non responsable, partiellement responsable).

## ii) Les étapes de l'épuration de la base de données

Afin de nettoyer notre base de données nous avons eu recours au logiciel statistique SAS. Ce dernier nous a permis grâce aux différentes procédures (PROC SQL) de parvenir, à la fin de cette étape, à une base plus adéquate à notre étude.

### a. Les doublons et les valeurs non renseignées

Avant d'entamer notre étude, la première étape de nettoyage de notre base de données fut la connaissance des doublons ainsi que des valeurs manquantes et leur taux de remplissage dans chacune de nos variables.

Nous avons remarqué qu'aucun problème de doublon n'existe dans notre fichier d'étude.

Le deuxième problème fut les valeurs manquantes qui sont des informations non

déclarées par l'assuré ou non saisies par les gestionnaires de sinistres.

Le tableau suivant récapitule pour chaque variable jugée pertinente et essentielle à notre étude, le nombre exact des valeurs non renseignées.

**Tableau 4 : Les statistiques des valeurs manquantes des variables d'étude**

Variables	Description	Nombre de Valeurs manquantes
Nature_sinistre	La nature du sinistre	2
Libelle_usage	Type de l'usage du véhicule	1
Date_déclaratiotion	La date de déclaration du sinistre	1918
Resp	La responsabilité de l'assuré	7
Reglements_total	Le règlement total du sinistre	39461
Code_intermediaire	Code du courtier	514
Date_cloture	La date du clôturation du contrat	1701
Date_ouverture	Date de l'ouverture du contrat	1866
Date_survenance	La date de survenance du sinistre	1826
Description_garantie	Le type de la garantie choisie par l'assuré	14
Ann-Emission	Année de l'émission du contrat	3
Date_effet	Date d'effet de la garantie	1909
Date_permis	Date de l'obtention du permis	19528
CRM	Une cotisation de l'assuré	21
Date_Naissance	La date de naissance de l'assuré	2
SM	La situation matrimoniale	5
Ville_courtier	la ville de l'intermédiaire de la compagnie	166
Marque	La marque du véhicule	395
Puiss_fisc	la puissance fiscale du véhicule	55
CSP	Catégorie socioprofessionnelle	22
Date_first_ass	Date du première assurance	5387
Date_MEC	Date de la mise en circulation	15460
Prime_RC	Prime	30471
Immatric	Immatriculation	622768
Année_effet	L'année de l'effet du contrat	15
<b>Total</b>		<b>743517</b>

*Source : Elaboré par les auteurs*

#### *b. La suppression de certaines variables*

L'étape suivante fut la suppression des variables jugées non essentielles à notre étude de tarification. Cette étape était effectuée sur les deux fichiers : Sinistre et PROD. Cette phase nous a permis de rendre notre base de données plus légère et objective.

Dans cette étape, on a éliminé des variables telles que le nombre de conducteurs, le nombre de places, le nombre d'enfants, la condition socioprofessionnelle, puisque ces variables n'auront aucun effet sur notre étude de tarification vu qu'elles ne jouent pas un rôle dans la survenance des sinistres dans un portefeuille d'assurés.

### c. Le calcul des variables

Pour effectuer notre étude on avait besoin d'un ensemble de variables qui n'étaient pas présentes sur les bases de données fournies par la compagnie, mais qu'on pouvait calculer directement à partir des autres données.

Parmi ces variables calculées on avait :

- La variable 'région'

C'est une variable qualitative à 18 modalités (16 régions territoriales plus une région 'Autres' pour les villes inconnues pour le découpage ainsi qu'une région 'NR' pour les champs non renseignés) qui nous renseigne sur la région où réside l'agent courtier de la compagnie.

En se basant sur un découpage administratif du Maroc de 2009, on a affecté chaque ville du courtier à la région correspondante.

Le choix de cette variable se justifie par le fait que travailler avec les villes des courtiers demeure plus condensé (167 modalités) et moins pratique, en revanche on ne dispose qu'un nombre limité de modalités régionales.

- La durée d'exposition au risque

Il s'agit du nombre de jours où la police a été mise en vigueur durant les années 2010 et 2011. On s'en servira pour mesurer l'exposition au risque. Elle permettra de tenir compte du fait qu'un sinistre déclaré par une police en vigueur durant un mois est un plus mauvais signe pour l'assureur qu'un sinistre relatif à une police en vigueur toute l'année.

Notons que l'exposition au risque devrait être mesurée par le nombre de kilomètres parcourus plutôt que par le nombre de jours où la police était en vigueur (le véhicule pourrait fort bien au garage pendant certaines périodes, n'étant donc pas soumis au risque). Le kilométrage annuel est cependant extrêmement difficile à mesurer.

L'étape suivante du calcul de la durée journalière d'exposition au risque fut suivie par rendre cette dernière annuelle en la divisant par 365.

- L'ancienneté du véhicule

C'est une variable pertinente pour décrire la sinistralité des assurés. En effet chaque fois que le véhicule de l'assuré est vieux, il cause des sinistres plus importants et des coûts moyens aussi plus élevés.

Cette variable a été calculée à partir de la différence entre la date d'effet et la date de mise en circulation du véhicule de l'assuré. Il s'agit donc d'une variable quantitative à valeurs entières décrivant l'âge de la voiture assuré.

- L'ancienneté du permis de conduire

Il s'agit d'une variable quantitative à valeurs entières donnant l'ancienneté du permis de conduire de l'assuré (en années révolues) le jour de la souscription du contrat pour une année donnée. Elle a été calculée à partir de la différence entre la date d'effet des contrats et la date de permis de conduire.

- L'âge des assurés

C'est une variable quantitative à valeurs entières donnant l'âge de l'assuré (en années

révolues). Elle a été calculée comme étant la différence entre la date de naissance de l'assuré et le jour de souscription du contrat.

Le tableau ci-dessous définit les quantités statistiques calculées sur l'ensemble des variables clés à notre étude de tarification :

**Tableau 5 : les quantités statistiques calculées sur l'ensemble des variables tarifaires**

<b>Nombre de sinistres</b>	Comptage du nombre de sinistre par rapport au numéro de sinistre au niveau de la table PROD_SINISTRE (combinée).
<b>Charge totale (pour un sinistre distinct)</b>	Provision_Total + Reglements_Total - Prov_Rec_Emis - Rec_Emis_Reclames - Rec_Recuperes + Prov_Rec_Subis + Rec_Subis_Reclames + Rec_Payes.
<b>Coût moyen</b>	Charge totale / Nombre de sinistres.
<b>Année risque</b>	Durée d'exposition au risque / 365.
<b>S/P</b>	Charge totale / Prime acquise.
<b>Fréquence</b>	Nombre de sinistres / Année risque.

*Source : Elaboré par les auteurs*

#### *d. Le tri des observations*

A l'aide du logiciel SAS, on a essayé de trier les données des deux fichiers PROD et sinistre et ce :

##### Selon la description de la garantie

Puisque notre étude s'intéresse à la garantie responsabilité civile (RC-auto), on a commencé par filtrer nos contrats d'assurance selon cette description pour ne garder que ceux correspondants à la garantie RC-auto à caractère individuel.

##### Selon l'âge

Après avoir calculé l'âge des assurés, nous avons filtré nos données pour ne garder que les assurés dont l'âge est de 18 ans et plus.

Nous avons constaté qu'il existe 157 observations pour laquelle l'âge du porteur d'assurance est aberrant (inférieur à 18 ans) y compris les non renseignés. Cela est dû soit à l'ignorance de la date de naissance du porteur de l'assurance, soit à une erreur de saisie. Les analyses nous ont permis de remarquer, heureusement que ces erreurs étaient aléatoires, c'est pourquoi nous avons procédé à leur élimination du fichier de l'étude.

##### Selon les années d'étude

Afin d'établir un seul fichier contenant toutes les variables de notre étude, on a besoin de joindre les deux bases PROD et Sinistre, pour ce faire on a restreint nos données aux deux années 2010 et 2011 concernant la date de survenance du sinistre (cette restriction concerne uniquement la partie à priori).

#### *e. La jointure des deux bases : PROD et Sinistre*

La base de données qui sera traitée tout au long de notre étude doit combiner les deux bases : PROD et Sinistre, ainsi chaque contrat sera identifié par la totalité des variables qui lui sont relatives et qui seront retenues pour notre étude. Ceci dit, on a donc besoin de joindre les deux bases (base Sinistre et base PROD), pour cela on a choisit pour clé primaire (clé commune entre les deux fichiers) :

Pour la première base {(Num\_avenant), (Num\_Pol)}

Pour la deuxième base {(Num\_avenant\_Impact), (Num\_contrat)}

Cette jointure nous a amené à calculer le nombre de sinistre par numéro de police distinct ainsi que la charge totale pour chaque contrat en les sommant.

Bref, à ce point, nous disposons d'une base de données nettoyée et modifiée donc exploitable et conforme aux contraintes de notre étude

### **II.1.2.2 Analyse descriptive des variables**

Malgré une épuration préalable de la base de données, il est possible que certaines variables ne soient pas pertinentes ou ne correspondent pas à la segmentation issue de la tarification en vigueur. Une première analyse descriptive peut permettre de le repérer.

Cette analyse va nous permettre de déterminer les caractéristiques d'un individu moyen afin de connaître la population des assurés de l'entreprise.

Par souci de lisibilité et tenant compte de la multitude des graphiques, nous avons choisi de présenter les résultats pour trois variables tarifaires (le sexe de l'assuré, la puissance fiscale du véhicule et son énergie ainsi que la variable décrivant l'ancienneté de la voiture).

#### **i) Répartition du portefeuille selon le sexe de l'assuré**

Le sexe est une variable importante pour expliquer la sinistralité. En effet le comportement des femmes et des hommes au volant n'est pas le même; les femmes sont supposées être plus prudentes en matière de conduite.

**Tableau 6 : Proportion des femmes et des hommes dans le portefeuille**

Sexe	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
F	12371	5.29	12371	5.29
M	221544	94.71	233915	100

*Source : Elaboré par les auteurs*

En terme de proportion, les hommes représentent 94.71% du portefeuille, tandis que les femmes occupent les 5.29% restant.

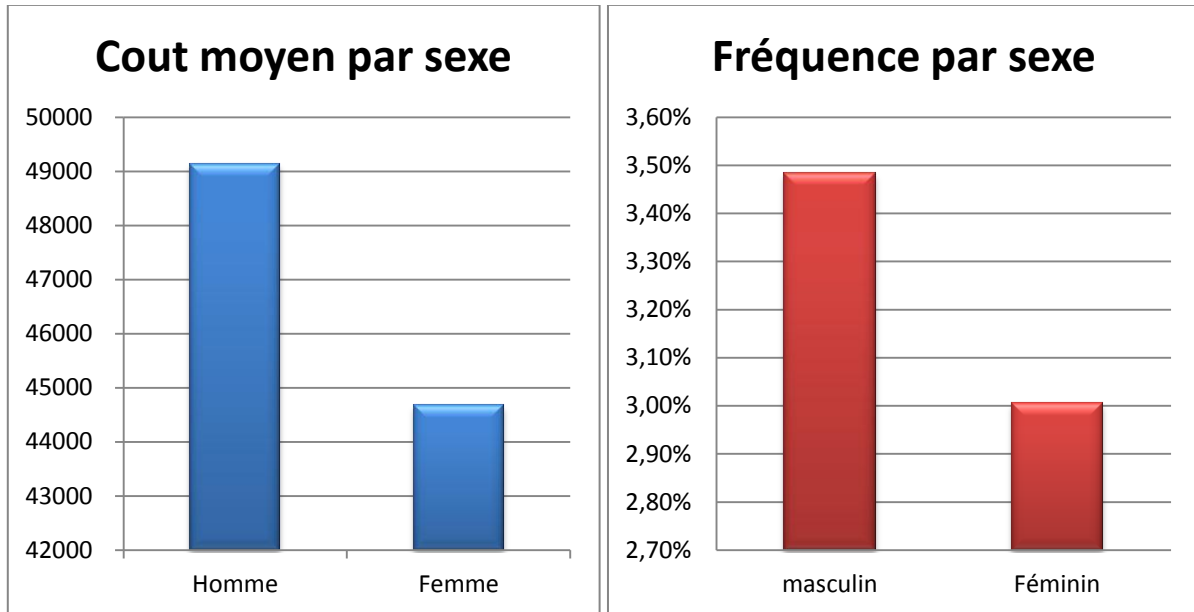
En terme de sinistralité, le tableau ci-dessus montre la répartition du coût et de la fréquence des sinistres entre les deux sexes.

**Tableau 7 : La sinistralité selon le sexe de l'assuré**

Sexe	Nombre d'observations	Variables	Total
F	12371	Coût	4706057
		Freq	6.3973
M	221544	Coût	75862579
		Freq	50.5076

*Source : Elaboré par les auteurs*

Face à la sinistralité, la différence est grande entre hommes et femmes. Comme on peut le voir plus clairement sur les graphiques ci-dessous les hommes sont plus « dangereux » que les femmes. Leur prime moyenne est de **34%** plus élevée que celle des femmes. Mais ce résultat est à prendre avec des pincettes du fait de la proportion des valeurs manquantes (1022274 assurés dont le sexe n'est pas renseigné).

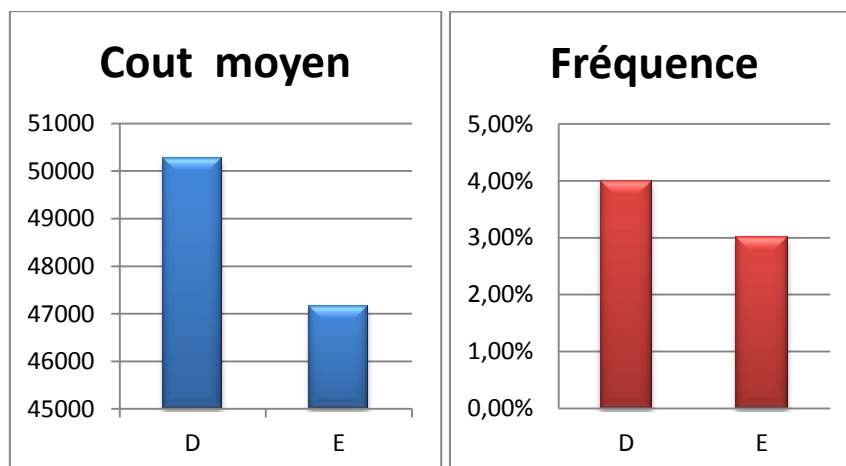
**Figure 7 : La fréquence et le coût des sinistres selon le sexe de l'assuré**

Source : Elaborée par les auteurs, sortie excel

## ii) Répartition du portefeuille selon la puissance fiscale et l'énergie du véhicule

La puissance fiscale et l'énergie du véhicule, pris séparément n'influencent pas clairement la sinistralité, mais on peut s'attendre à ce qu'elles soient fortement corrélées entre eux (ce qu'on va confirmer à l'aide de l'étude de corrélation). En effet, l'acquisition d'un véhicule Diesel se justifiant souvent par un usage plus fréquent, et pour de plus longues distances puisque le prix d'achat est plus élevé.

Les graphiques ci-dessus illustrent l'évolution du coût et de la fréquence des sinistres selon la variable énergie du véhicule assuré.

**Figure 8 : La fréquence et le coût des sinistres selon l'énergie du véhicule**

Source : Elaborée par les auteurs, sortie excel

En comparant les deux énergies, on remarque la fréquence de l'usage Diesel est de **4%** contre un pourcentage de **3%** pour ce qui est de l'usage Essence. Le coût moyen quant à lui paraît de l'ordre de **51 300 DH** pour l'énergie diesel donc une somme plus élevée que celle du type Essence qui est de l'ordre de **47 500 DH**. Ceci dit, cela peut très bien être expliqué par l'usage de chaque catégorie d'énergie.

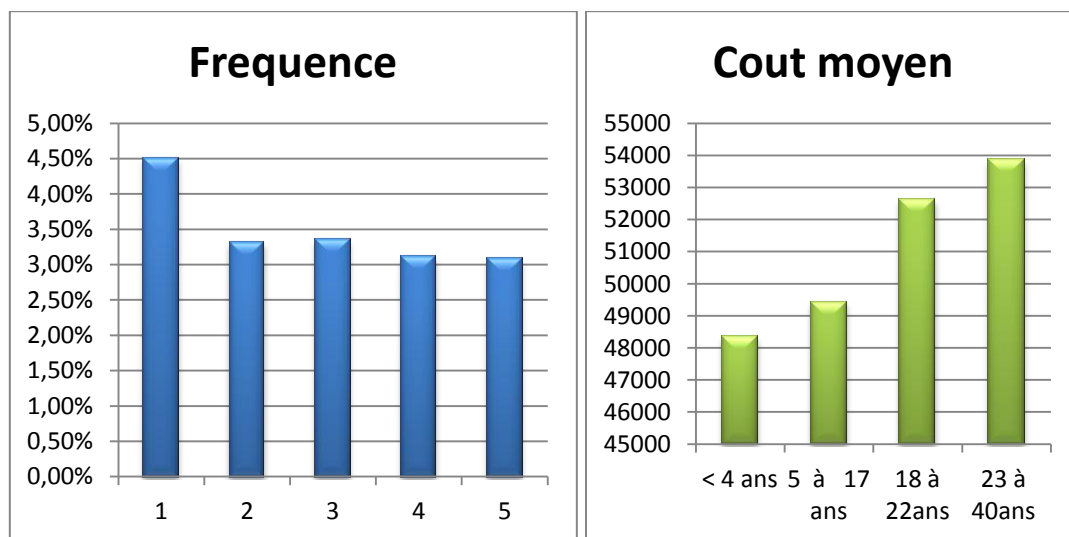
Pour ce qui est de la puissance fiscale on constate que les véhicules de 6,7, et 8 chevaux sont majoritaires. Ils représentent à eux près des  $\frac{3}{4}$  du portefeuille.

### iii) Répartition du portefeuille selon l'ancienneté du véhicule

Il s'agit de la période du temps pendant laquelle le véhicule est mis en circulation. C'est une variable pertinente pour décrire la sinistralité des assurés ; en effet chaque fois que le véhicule de l'assuré est vieux, il cause des sinistres plus importants et des coûts moyens aussi plus graves.

Les graphiques suivants donnent une idée sur l'effet de l'âge du véhicule sur la fréquence et le coût moyen des sinistres.

**Figure 9 : La fréquence et le coût moyen des sinistres selon l'ancienneté du véhicule**



*Source : Elaborée par les auteurs, sortie excel*

A partir de ces résultats, on remarque que le coût moyen des sinistres évolue dans le même sens que l'âge du véhicule. Alors que pour la fréquence des sinistres, elle demeure presque uniforme.

A partir de cette brève analyse, on a pu concevoir une idée sur la répartition de notre portefeuille selon quelques variables classiques à la tarification.

## II.1.3 La segmentation du portefeuille

### II.1.3.1 Introduction

La segmentation désigne le fait de découper un portefeuille en plusieurs sous-ensembles homogènes et distincts composés d'individus ayant des comportements communs.

Non seulement cet exercice de segmentation est important au niveau de l'équité entre assurés, mais aussi pour la santé financière de l'assureur qui réduit ainsi son risque d'anti-sélection.

Concrètement, l'étude sur la segmentation que nous avons effectuée est un moyen de différenciation de la population des assurés suivant les critères tarifaires relatives au client et au véhicule assuré, que nous retenons pour notre étude de tarification de ce portefeuille RC-auto. A l'issue de la segmentation, l'assureur dispose de classes homogènes d'assurés.

La segmentation des variables est une étape importante du processus de tarification. En effet elle nous permet de passer d'une infinité de classes de risques (car parmi ces variables nous utilisons aussi des variables continues) à un nombre fini de classes de risque facilement interprétables. Pour faire la segmentation nous allons utiliser les arbres de décision. Il existe plusieurs algorithmes de construction de ces arbres : l'algorithme CRT, l'algorithme CHAID ...

Donc avant d'entamer une étude de segmentation du portefeuille globale, on va d'ailleurs segmenter les différentes variables tarifaires en des sous ensembles tout en calculant les différentes quantités statistiques de la sinistralité pour chaque segment construit.

Afin d'effectuer cette étude de segmentation, on aura besoin d'un outil statistique. Pour notre projet on a choisit de travailler avec l'algorithme CHAID. L'application de cet algorithme sera effectuée à l'aide du logiciel ANSWER TREE développé par SPSS.

### **II.1.3.2 Représentation de l'algorithme de CHAID**

L'acronyme CHAID signifie en fait "Chi-squared Automatic Interaction Detector". Il s'agit de l'une des méthodes d'arbres de classification les plus anciennes, initialement proposée par Kass (1980 ; d'après Ripley, 1996, l'algorithme CHAID est une version modifiée de l'algorithme THAID développé par Morgan et Messenger, 1973). Ce nom **CHAID** provient de l'algorithme utilisé pour la construction d'arbres de décision (non-binaires), et qui, pour les problèmes de classification (lorsque la variable dépendante est de nature catégorielle) se sert d'un test du Chi-2 pour déterminer la meilleure division suivante à chaque étape; pour les problèmes de régression (variable dépendante continue), le programme va en fait calculer des tests F.

L'algorithme CHAID se déroule en trois étapes :

- a. Préparation des prédicteurs: tout d'abord, ANSWER TREE va créer des prédicteurs catégoriels à partir de chacun des prédicteurs continus, en répartissant la distribution des différents prédicteurs continus en un certain nombre de catégories d'effectifs sensiblement égaux. Pour les prédicteurs catégoriels, les catégories (classes) sont définies "naturellement".
- b. Fusion des classes: ANSWER TREE va ensuite examiner les prédicteurs afin de déterminer pour chacun, le couple de catégories (du prédicteur) les plus semblables (c'est-à-dire significativement moins différentes) par rapport à la variable dépendante ; pour les problèmes de classification (où la variable est également catégorielle), le programme va calculer un test du Chi-2 (Chi-deux de Pearson); pour les problèmes de régression (où la variable dépendante est continue), le programme va calculer des tests F. Si le test respectif, pour un couple donné de catégories du prédicteur, ne peut être considéré comme significatif eu égard à une valeur alpha-de-fusion, le programme va alors regrouper les catégories correspondantes du prédicteur et répéter ce processus (c'est-à-dire, rechercher le couple suivant de catégories, qui à présent peuvent être des catégories précédemment fusionnées). Si le couple respectif de catégories du prédicteur est statistiquement significatif (inférieur à la valeur correspondante du alpha-de-fusion), le programme va alors calculer (éventuellement) une valeur-p ajustée de Bonferroni pour l'ensemble des catégories du prédicteur respectif.

- c. Sélection de la variable de séparation: ANSWER TREE va ensuite choisir pour la division, la variable prédictive qui possède la plus faible valeur-p ajustée, c'est-à-dire la variable prédictive qui permet de produire la division la plus significative, si la plus petite valeur-p ajustée (Bonferroni) des prédicteurs est supérieure à une certaine valeur alpha-de-division, le processus de division prend fin, et le nœud respectif est un nœud terminal.

Dans cette dernière étape, l'algorithme utilise les probabilités pour estimer si une catégorie peut être divisée. Ce processus se poursuit jusqu'à ce qu'il ne soit plus possible de réaliser d'autres divisions (compte tenu des valeurs alpha-de-fusion et alpha-de-division).

L'algorithme de CHAID se base pour cela sur une variable dépendante qui lui servira de guide. L'algorithme subdivise par ailleurs la population initiale en plusieurs ensembles distincts en se basant sur les variables indépendantes les plus significatives, chacun des groupes ainsi obtenus est aussitôt subdivisé en deux sous groupes distincts ou plus en considérant les variables restantes. On répète ce processus jusqu'à la fin; soit toutes les variables sont utilisées, soit une condition spécifiée par l'utilisateur est réalisée (les conditions d'arrêt).

Le résumé de toutes ces démarches est donné sous forme d'un arbre de décision où chaque nœud représente un regroupement d'observations. Il s'agit donc de trouver un partitionnement des individus que l'on représente sous la forme d'un arbre de décision. L'objectif est de produire des groupes d'individus les plus homogènes possibles du point de vue de la variable à prédire (variable Target). Il est d'usage de représenter la distribution empirique de l'attribut à prédire sur chaque sommet (nœud) de l'arbre.

L'ensemble des arbres de décision concernant toutes quelques variables retenues pour notre étude sera affiché en annexe.

### **II.1.3.3 La segmentation des variables candidates à la tarification à priori**

Les variables de tarification usuelles utilisées pour expliquer la fréquence et le coût moyen des sinistres se rapportant au véhicule sont :

- ✓ La puissance fiscale ;
- ✓ L'ancienneté du véhicule ;
- ✓ l'ancienneté du permis ;
- ✓ la région ;
- ✓ le sexe du conducteur ;
- ✓ l'âge du conducteur ;
- ✓ CRM.

Ces variables sont jugées statistiquement pertinentes dans la modélisation de la fréquence et du coût des sinistres en assurance RC-auto.

Dans cette section, nous souhaitons donc appréhender le comportement des variables candidates à la tarification de notre portefeuille face aux critères de la fréquence et du coût moyen des sinistres ainsi que de la prime moyenne et du rapport S/P.

Nous créons ainsi, des classes plus ou moins homogènes en intra et hétérogènes en extra pour les variables quantitatives et des classes de modalités pour les variables qualitatives. On va tenter de calculer, pour chacune des variables et pour chacun de ses segments, l'ensemble

des quantités statistiques de la sinistralité.

**i) Variables caractérisant le conducteur assuré**

**a. L'ancienneté du permis**

La segmentation de la variable ancienneté du permis, a été réalisée grâce à l'algorithme présenté précédemment.

Nous avons pu limiter le nombre des classes pour l'ancienneté du permis de conduire à trois intervalles d'âges de permis.

**Tableau 8 : Les quantités statistiques de la sinistralité selon la variable ancienneté de permis**

Classes	Ancienneté de permis	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
1	[0,3]	18,8%	87 739	16,1%	136 422 471	118 203 431	6 483	18 233	7,4%	86,6%	1 347	1 555
2	]3,7]	12,2%	59 187	10,9%	99 263 166	70 514 227	4 363	16 162	7,4%	71,0%	1 191	1 677
3	]7,+[	69,1%	397 983	73,0%	785 830 809	342 380 365	25 240	13 565	6,3%	43,6%	860	1 975
		100%	544 909	100%	1 021 516 446	531 098 023	36 086	14 718	6,6%	52%	975	1 875

*Source : Elaboré par les auteurs, sortie excel*

L'examen de ces résultats nous a permis de conclure à propos de la sinistralité basée sur l'ancienneté de permis de l'assuré. En se basant sur le rapport S/P, on peut dire que les personnes dont l'âge du permis de conduire est inférieur à trois années sont les plus risquées alors que celles qui représentent une ancienneté au delà de sept ans sont les moins risquées.

**b. L'âge de l'assuré**

Nous avons, ici aussi limité le nombre de classes pour l'âge du conducteur à trois classes : l'intervalle le moins risqué, celui le plus risqué et enfin celui qui est moyennement risqué en terme de sinistralité.

**Tableau 9 : Les quantités statistiques de la sinistralité du portefeuille selon l'âge de l'assuré**

Classes	Age	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
1	32]	19,0%	88 007	16,2%	142 912 727	123 061 295	8 435	14 589	9,6%	86,1%	1 398	1 624
2	]32,42]	30,5%	154 245	28,3%	271 367 281	147 902 172	10 226	14 463	6,6%	54,5%	959	1 759
3	]42,+[	50,5%	302 657	55,5%	607 236 438	260 134 556	17 425	14 929	5,8%	42,8%	860	2 006
		100%	544 909	100,0%	1 021 516 446	531 098 023	36 086	14 718	6,6%	52%	975	1 875

*Source : Elaboré par les auteurs, sortie excel*

On remarque donc que les assurés relativement jeunes ( $\text{âge} \leq 32$ ) représentent la catégorie la plus risquée, avec une proportion de 19% de la population cible, alors que celle des moins risqués est représentée par les assurés qui sont plutôt « vieux » et qui représentent 50,5% du total de la population assurée.

**c. La situation matrimoniale**

Pour cette variable, L'algorithme de CHAID nous a fourni trois classes correspondantes aux différentes modalités de ladite variable.

**Tableau 10 : Les quantités statistiques de la sinistralité du portefeuille selon la situation matrimoniale**

Classes	Situation matrimoniale	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
1	Célibataire	47,0%	248 359	45,6%	446 711 647	258 657 390	17 391	14 873	7,0%	57,9%	1 041	1 799
2	Marié(e); Pacs	52,8%	295 503	54,2%	572 741 813	271 591 350	18 603	14 599	6,3%	47,4%	919	1 938
3	Veuf(ve); Divorcé(e)	0,2%	1 047	0,2%	2 062 986	849 283	92	9 231	8,8%	41,2%	811	1 970
		100%	544 909	100%	1 021 516 446	531 098 023	36 086	14 718	6,6%	52%	975	1 875

*Source : Elaboré par les auteurs, sortie excel*

A partir de cette classification, on peut conclure en terme de la sinistralité décrite par le rapport S/P, que les célibataires qui représentent 47% de la population de l'étude demeurent les assurés les plus risqués.

## ii) Variables caractérisant la voiture de l'assuré

### a. L'ancienneté du véhicule

Pour cette variable qui désigne la période du temps pendant laquelle le véhicule est mis en circulation, on a restreint le nombre d'intervalles à deux : ceux qui ont une ancienneté inférieure à une année contre ceux dont l'ancienneté est supérieure à une année avec des rapports S/P similaires.

**Tableau 11 : Les quantités statistiques de la sinistralité du portefeuille selon la variable ancienneté du permis**

Classes	Ancienneté de véhicule	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
1	[0,1]	8,0%	56 395	10,3%	125885305,1	63 564 918	6 785	9 368	12,0%	50,5%	1 127	2 232
2	]1,+[	92,0%	488 515	89,7%	895631141,2	467 533 105	29 301	15 956	6,0%	52,2%	957	1 833
		100%	544 909	100%	1 021 516 446	531 098 023	36 086	14 718	6,6%	52%	975	1 875

Source : Elaboré par les auteurs, sortie excel

On déduit de l'ensemble de ces résultats que la majorité des assurés de notre portefeuille (92% de l'ensemble de la population) ont une ancienneté de permis supérieur à une année et représente un S/P de 52,2%.

### b. L'énergie croisée la puissance fiscale

Cette variable qui croise l'énergie de la voiture assurée et sa puissance fiscale a été segmentée en sept classes et ce, selon le paramètre S/P, et pour chacun de ces segments on a calculé l'ensemble des quantités statistiques qui décrivent la sinistralité.

**Tableau 12 : Les quantités statistiques de la sinistralité du portefeuille selon la variable énergie\*puissance fiscale**

Classes	Energie	Puissance fiscale	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
1	E	[0,6]	12,0%	62 973	11,6%	80 819 321	55 466 153	3 962	14 000	6,3%	68,6%	881	1 283
2	E	]6,8]	13,2%	67 982	12,5%	102 957 371	57 577 393	4 127	13 951	6,1%	55,9%	847	1 514
3	E	]8,9]	2,9%	14 169	2,6%	21 769 362	12 649 030	1 048	12 070	7,4%	58,1%	893	1 536
4	E	]9,+[	3,1%	15 542	2,9%	31 533 451	15 199 869	1 261	12 054	8,1%	48,2%	978	2 029
5	D	[0,5]	1,0%	5 746	1,1%	9 414 514	6 247 502	644	9 701	11,2%	66,4%	1 087	1 639
6	D	]5,8]	56,0%	315 985	58,0%	622 599 525	318 162 537	20 865	15 249	6,6%	51,1%	1 007	1 970
7	D	]8,+[	11,8%	62 513	11,5%	152 422 902	65 795 540	4 179	15 744	6,7%	43,2%	1 053	2 438
			100%	544 909	100%	1 021 516 446	531 098 023	36 086	14 718	6,6%	52%	975	1 875

Source : Elaboré par les auteurs, sortie excel

On constate à partir de ce tableau qu'une puissance fiscale  $\leq 6$  et croisée avec un moteur Essence correspond à la classe la plus risquée en terme du rapport S/P.

Alors que la classe qui regroupe la grande partie de la population est celle qui regroupe à la fois une énergie Diesel et une puissance fiscale comprise entre 5 et 8 chevaux fiscaux avec un S/P assez important (51.1%).

## iii) Autres variables tarifaires

### a. CRM

La segmentation au niveau du coefficient de réduction majoration a révélé une segmentation en trois niches que nous avons résumée dans le tableau ci-dessus. En effet, on

trouve premièrement la classe CRM égal à 100, qui représente un poids de la population totale de l'étude de 64.1% avec un rapport S/P très élevé suivit par l'intervalle des CRM supérieurs à 100 et finalement la classe des CRM inférieurs ou égal à 90.

**Tableau 13 : Les quantités statistiques de la sinistralité selon le CRM**

Classes	CRM	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
1	[0;90]	34,6%	243 266	44,6%	509 789 675	173 989 931	13 592	12 801	5,6%	34,1%	715	2 096
2	100	64,1%	293 389	53,8%	490 444 290	346 258 562	21 419	16 166	7,3%	70,6%	1 180	1 672
3	]100;250]	1,4%	8 254	1,5%	21 282 481	10 849 530	1 075	10 093	13,0%	51,0%	1 314	2 578
		100%	544 909	100%	1 021 516 446	531 098 023	36 086	14 718	6,6%	52%	975	1 875

Source : Elaboré par les auteurs, sortie excel

En terme de sinistralité décrite par le rapport S/P, la classe du coefficient de réduction majoration de 100 représente la catégorie la plus risquée.

#### b. La région

La région est une variable capital pour notre étude de tarification de ce portefeuille RC automobile. Elle fut segmentée par un ensemble de trois intervalles: le premier regroupement de régions qui représente un poids de 39.9% des assurés du portefeuille et un S/P de l'ordre de 77%, il s'agit donc d'une zone fortement sinistrée suivit par une classe d'un rapport moyen et finalement une classe d'un faible S/P.

**Tableau 14 : Les quantités statistiques de la sinistralité du portefeuille selon la variable région**

Région	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	Charge/Prime	Coût*Fréquence	Prime moyenne
Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	23,9%	122 976	22,6%	204 053 600	158 008 594	5 965	26 489	4,9%	77,4%	1 285	1 659
Grand casablanca; Marrakech - Tensift - Al Haouz; Rabat - Salé - Zemmour - Zaïr; Fès - Boulemane.	39,5%	231 814	42,5%	463 641 752	238 315 597	22 099	10 784	9,5%	51,4%	1 028	2 000
Tanger - Tétouan; Méknès - Tafilalet; Taza - Al Hoceima - Taounate; Oriental; Souss - Massa Darâa.	36,6%	190 119	34,9%	353 821 095	134 773 832	8 022	16 801	4,2%	38,1%	709	1 861
	100%	544 909	100%	1 021 516 446	531 098 023	36 086	28 308	6,6%	52%	1 875	1 875

Source : Elaboré par les auteurs, sortie excel

À ce stade on a pu, au niveau de chaque variable tarifaire, construire des intervalles repartis d'une manière uniforme selon le S/P tout en calculant l'ensemble des quantités statistiques de risque pour chacune de ces classes.

Ceci va être un point de départ pour construire les classes de risque qui croisent l'ensemble de toutes ces variables.

## **CHAPITRE 2**

# MODELE LINEAIRE GENERALISE ÉLABORATION DES TARIFS ACTUARIELS



### **II.2.1 Introduction**

Nous avons pu entrevoir précédemment des généralités concernant l'assurance automobile ainsi que des techniques d'analyse descriptive et de segmentation. Ces aspects constituaient une première approche et un préliminaire à la réalisation de la modélisation du risque automobile que nous allons à présent décrire.

L'enjeu de la modélisation est de prédire ou d'expliquer une variable  $Y$ , à partir de  $p$ -variables explicatives appelées également prédicteurs.

En premier lieu, nous s'assurons des corrélations entre les variables et ce en étudiant la corrélation pour les variables quantitatives et pour les variables qualitatives  $\hat{q}$  travers le test de Chi-deux.

Dans la seconde partie, nous analysons l'ajustement de la fréquence et du coût moyen aux différentes lois candidates. Il est d'usage en assurance automobile, de modéliser séparément le coût moyen et la fréquence de sinistre. La prime pure est ensuite calculée en multipliant ces deux derniers.

Dans la dernière partie on va essayer de construire les classes de tarifs à partir des variables retenues pour l'étude et ce, en se servant des tableaux de segmentation formés ultérieurement.

## II.2.2 Analyse de la corrélation

Avant de procéder à la modélisation de la fréquence et du coût moyen des sinistres, on entame une étude de corrélation entre les variables qualitatives et quantitatives séparément.

A partir de cette étude, on va essayer de réduire le nombre des variables. Si deux variables sont extrêmement corrélées alors il faudra, malgré la perte d'informations, retirer la variable la moins intéressante ou la croiser avec la variable présentant un certain niveau de corrélation.

### II.2.2.1 Entre les variables quantitatives

Les variables quantitatives mises en jeu dans notre étude sont : l'âge de l'assuré, l'âge du véhicule, l'âge du permis, la puissance fiscale et le CRM. Les corrélations sont mesurées par le coefficient de Pearson résumé sur la matrice de corrélation ci-dessous :

**Tableau 15: Corrélation entre les variables quantitatives**

Coefficients de corrélation de Pearson					
	CRM	Puissance fiscale	âge	Ancienneté du permis	Ancienneté du véhicule
CRM	1	0,01336	-0,19202	-0,21718	0,05783
Puissance fiscale	0,0134	1	0,05493	0,05397	0,00838
âge	-0,192	0,05493	1	<b>0,65021</b>	-0,01019
Ancienneté du permis	-0,2172	0,05397	<b>0,65021</b>	1	-0,06047
Ancienneté du véhicule	0,0578	0,00838	-0,01019	-0,06047	1

*Source : Elaboré par les auteurs, sortie SAS*

#### *Commentaire et interprétation :*

En étudiant les corrélations les plus élevées nous pouvons déjà obtenir quelques conclusions :

- Age et l'ancienneté du permis sont deux variables fortement corrélées, ce qui est normal puisque l'ancienneté du permis suppose l'avancement de son titulaire dans l'âge.
- Toutes les variables quantitatives restantes semblent indépendantes.

A partir de ces résultats, on a décidé de croiser l'âge et l'ancienneté de permis afin d'obtenir une nouvelle variable. Combiner les deux variables va donc nous aider à éviter d'une part le problème de manque de l'information et d'autre part la redondance.

### II.2.2.2 Entre les variables qualitatives

Les variables qualitatives objet de notre étude de tarification sont : La région, le sexe du conducteur et le carburant du véhicule.

Afin de mettre en évidence les corrélations existantes entre ces variables nous ferons appel au test de Chi-deux pour l'indépendance. Sous SAS ce test est donné par la procédure «**PROC FREQ**».

Les résultats sont comme suit :

- Chi-deux pour l'indépendance entre la puissance fiscale du véhicule et son énergie*

**Tableau 16 : Résultat du test khi2 d'indépendance entre la puissance fiscale et l'énergie**

Statistique	DF	Valeur	Proba
Chi-deux	2	39886.8904	< 0.00001
Test du rapport de vraisemblance	2	53303.8599	< 0.00001
Khi-2 de Mantel-Haenszel	1	39008.1841	< 0.00001

Source : Elaboré par les auteurs, sortie SAS

*Commentaire et interprétation :*

Le test met en évidence une corrélation entre la puissance fiscale du véhicule et son énergie, cette corrélation s'explique par l'usage du diesel par exemple comme carburant dans tout usage relatif à l'activité commerciale.

ii) *Khi2 pour l'indépendance entre la région et l'énergie du véhicule*

**Tableau 17 : Résultat du test khi2 d'indépendance entre le sexe et la région**

Statistique	DF	Valeur	Proba
Chi-deux	32	36613.5077	< 0.00001
Test du rapport de vraisemblance	32	40335.2122	< 0.00001
Khi-2 de Mantel-Haenszel	1	5001.8191	< 0.00001

Source : Elaboré par les auteurs, sortie SAS

*Commentaire et interprétation :*

Le test met en évidence une quasi corrélation entre la variable sexe et la variable région. Cette corrélation peut être expliquée par le constat que les femmes qui conduisent au Maroc sont généralement celle circulant dans les grandes régions du pays : Casablanca, Marrakech, Rabat et Tanger.

### **II.2.3 Sélection des variables explicatives**

La sélection de variables est une perspective naturelle à plus d'un titre. La principale raison est que certaines variables ne contribuent pas à l'explication de la variable dépendante. Aussi la présence des variables qui sont très corrélées apporte une redondance d'information. On a envie donc de les éliminer du modèle. Il faut bien noter que l'on cherche toujours à privilégier le modèle le plus simple possible permettant ainsi une interprétation facile et pouvant éviter à l'expérimentateur des coûts d'acquisition de certaines données.

Les méthodes les plus utilisées pour la sélection des variables explicatives sont dites "pas à pas" dans lesquelles les variables sont introduites ou supprimées dans la régression l'une après l'autre.

Il existe trois variantes de ces méthodes que nous décrivons ci-dessus.

*Ascendante:*

La procédure commence avec le terme constant  $\beta_0$ , soit le modèle nul:  $Y_i = \beta_0 + E_i$   
Ensuite, elle s'effectue en plusieurs étapes :

Etape1:

On choisit la variable  $X_{kl}$  parmi l'ensemble des variables de départ qui contribue le plus à expliquer Y, i.e. celle qui fait le plus augmenter le  $R^2$  ou encore telle que  $\rho(Y, X_{kl})$  est maximal.

Ensuite, on teste la nullité du coefficient de régression associé et la variable est retenue en cas de significativité du test.

#### Etape 2 :

On choisit la variable  $X_{k2}$  parmi l'ensemble des variables auquel on a retiré  $X_{k1}$  telle que  $\rho(Y, X_{k2} | X_{k1})$  est maximal.

C'est n'est donc pas la seconde variable la plus corrélée à  $Y$  mais c'est celle qui apporte le plus d'information en plus de  $X_{k1}$ . De la même façon que précédemment, le coefficient de régression est testé.

Il existe plusieurs tests d'arrêts de la procédure: en choisissant un nombre à priori de variables ou une valeur final de  $R^2$ , ou encore dès que le test de nullité de la dernière variable introduite n'est pas significatif. C'est cette dernière solution qui est faite dans le logiciel statistique SAS.

#### Descendante :

C'est la procédure symétrique de la précédente qui part du modèle complet et élimine à chaque étape la variable correspondante au plus petit coefficient de corrélation partielle.

#### Stepwise :

Cette procédure est semblable à l'ascendante avec remise en cause à chaque étape des variables déjà introduites. En effet, il arrive souvent que des variables introduites en tête, par le biais de leur liaison avec une ou plusieurs variables introduites ultérieurement, ne soient plus significatives.

La stratégie utilisée dans notre étude est la procédure Stepwise. Le modèle final retenu est :

**Tableau 18 : Résultats de l'analyse pour le modèle avec les 5 variables explicatives**

Stepwise Selection Summary					
Etape	Effet saisi	Nombre de params dans	R carré du modèle	R-carré ajusté	Pr > F
0	Intercept	1	0	0	1
1	Region_Segment	3	0,0023	0,0023	<.0001
2	Anc_Vehicule_Segment	4	0,0034	0,0034	<.0001
3	CRM_Segment	6	0,0074	0,0074	<.0001
4	Age_anciennete_permi	14	0,0094	0,0094	<.0001

*Source : Elaboré par les auteurs, sortie SAS*

L'examen des résultats de ce tableau nous permet de diminuer le nombre de variables explicatives. Les variables explicatives retenues sont : la région, l'ancienneté du véhicule, l'âge croisé avec l'ancienneté du permis, et le CRM.

## **II.2.4 Le modèle linéaire généralisé**

### **II.2.4.1 Le cadre général**

Les modèles linéaires généralisés ont été définis pour la première fois par John Nelder et Robert Wedderburn en 1972. Ils sont venus pallier à l'étroitesse des modèles linéaires.

En effet, la classe des GLM est une extension des modèles linéaires traditionnels qui permet à la moyenne d'une population de dépendre d'un estimateur linéaire via une fonction non linéaire dite *fonction lien*. Ainsi, le choix d'une fonction lien adéquate et d'une distribution pour la variable réponse sont inévitables à la construction d'un modèle linéaire généralisé. De surcroît, la modélisation GLM est paramétrique dans le sens où la variable expliquée est supposée munie d'une loi de probabilité. Plus exactement, elle est prétendue appartenir à la famille des lois exponentielle.

Les avantages précités et d'autres encore font des modèles linéaires généralisés l'un des choix de modélisation statistique les plus populaires, notamment en matière de tarification.

*i) L'hypothèse du modèle*

La modélisation en GLM se base sur l'hypothèse que l'échantillon est constitué de variables aléatoires respectant la normalité des résidus. Cette hypothèse est vérifiable grâce à une multitude de tests statistiques.

*ii) Les composantes du modèle*

La construction d'un modèle linéaire généralisé requiert la disposition puis l'assemblage de deux composantes:

- La partie linéaire du modèle qui se résume en l'estimateur linéaire ;
- La fonction lien notée  $g$  qui décrit l'obtention de, la valeur prévue de  $Y$ , à partir de l'estimateur linéaire.

*iii) La distribution de la variable dépendante et la famille de lois exponentielle*

En modélisation GLM, la distribution de la variable réponse  $Y$  doit appartenir à la famille de lois exponentielle. La forme générale de la fonction de masse pour cette famille de lois s'écrit :

$$f(y, \theta, \varphi) = \exp \left( \frac{\theta y - b(\theta)}{a(\varphi)} + c(y; \varphi) \right)$$

Les fonctions  $a$ ,  $b$  et  $c$  sont fixées ;

$\varphi$  est un paramètre de dispersion ou (nuisance paramètre).

- $a(\varphi) = \varphi$  pour des données individuelles,
- $a(\varphi) = \frac{\varphi}{\omega}$  pour des données de groupes ( $\omega$  effectif du groupe).

Exposons quelques modèles dont on aura besoin par la suite en spécifiant les caractéristiques et les propriétés spécifiques de chaque loi, on commence par la loi de poisson, ensuite on présente la loi binomiale Négative puis la loi Normale et enfin la loi Gamma :

Distributions discrètes :

- La distribution de probabilité d'une loi poisson  $P(\lambda)$  de paramètre  $\lambda$  est donnée par :

$$P(Y = y; \mu) = \frac{\mu^y e^{-\mu}}{y!}$$

L'égalité  $\text{Var}(Y) = E(Y)$  introduit une limitation forte dans l'application de ce modèle en assurance non-vie.

- Loi binomiale négative  $BN(r, r/r+\mu)$  :

$$P(Y = y; \pi) = \binom{m}{my} \pi^{my} (1 - \pi)^{m-my}$$

Distributions continues :

- Loi Gamma  $\gamma(v; v/\mu)$  :

$$f(y; v; \mu) = \frac{1}{\Gamma(v)y} \left(\frac{yv}{\mu}\right)^2 \exp(-yv/\mu)$$

En résumé voici une illustration des lois de la famille exponentielle classiquement utilisée en modélisation linéaire généralisée :

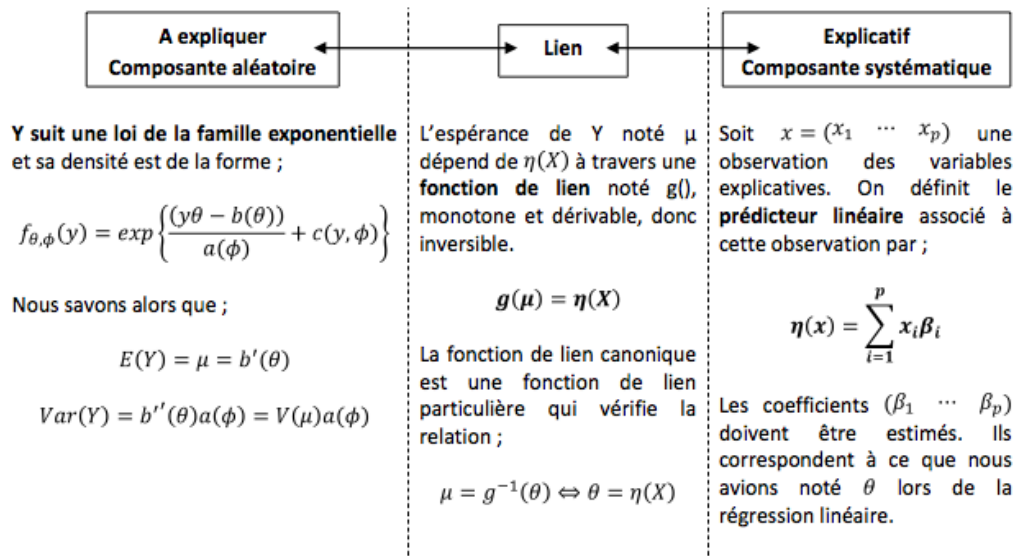
**Tableau 19 : Lois classiques de la famille exponentielle**

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

Source : Abderrahim LOULIDI. Cours de l'assurance non vie, 2011

Ci-dessus, on trouve l'ensemble des étapes qu'il faut suivre afin de construire un modèle linéaire généralisé :

**Figure 10 : Cheminement de la construction d'un GLM**



Source : Abderrahim LOULIDI. Cours de l'assurance non vie, 2011

Et pour accomplir le lien entre les composantes aléatoires et celles systématiques, voici les fonctions de lien à adopter pour chacune des lois usuelles :

**Tableau 20 : Fonctions de lien pour les différentes lois candidates**

Loi	Nom du lien	Fonction de lien
Bernouilli/Binomiale	lien logit	$g(\mu) = \text{logit}(\mu) = \log(\mu/1 - \mu)$
Poisson	lien log	$g(\mu) = \log(\mu)$
Normale	lien identité	$g(\mu) = \mu$
Gamma	lien réciproque	$g(\mu) = -1/\mu$

Source : Abderrahim LOULIDI. Cours de l'assurance non vie, 2011

Bref, nous avons donc vu ci-dessus qu'il est nécessaire d'effectuer deux choix majeurs pour construire un modèle linéaire généralisé. Le premier concerne la loi de la variable à expliquer. Comme nous l'avons signalé précédemment, ce choix peut être orienté par le type

de la variable et des connaissances préalables. Le deuxième choix porte sur la fonction de lien à travers les fonctions de liens classiquement utilisées.

Le choix de la densité peut alors dépendre de la loi. En effet, si  $Y$  est binaire, on préférera utiliser les liens *logit*, *probit* ou *cloglog*, si  $Y$  est un comptage, on utilisera classiquement le lien *log* (ce qui est le cas pour notre étude), et enfin si  $Y$  est continue, on pourra utiliser les liens canoniques des lois Normale et Gamma. Le choix du lien peut également être déterminé par l'existence d'études passées, par une connaissance préalable du problème.

#### **II.2.4.2 Etude des lois des variables dépendantes et tests d'adéquation**

Dans la modélisation des processus de comptage, ici de la fréquence des sinistres, deux sortes de modèle sont couramment mis en œuvre, le modèle de Poisson et le modèle binomial négatif. On trouve une littérature abondante sur l'utilisation de ces modèles : GREENE (1996), WOOLDRIDGE (1997), CAMERON et TRIVEDI (1998), WINKILMANN (2000) YAU et AI (2003), YANG et AI (2007).

Aussi en actuariat, pour le choix de la distribution du montant des sinistres (coût moyen des sinistres) on a généralement recours à une loi continue avec un support compris dans  $\mathbb{R}^+$  (ici on va tester les lois Gamma et Log-Normale). En effet, ces lois correspondent bien à une distribution continue, définie sur les réels positifs, et ayant une variabilité qui augmente avec la moyenne.

##### ***i) Pour la fréquence des sinistres :***

La première étape consiste à choisir la distribution de la variable fréquence des sinistres. La régression de Poisson et la régression Binomiale Négative mettent en évidence les mêmes variables explicatives de la fréquence des sinistres, avec des effets semblables.

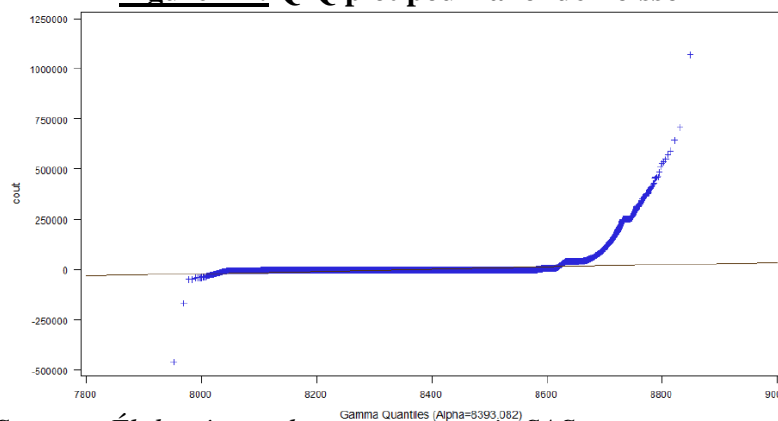
Dans un premier temps, on compare ces deux modèles. Pour choisir le meilleur modèle pour notre étude, on va avoir recours au logiciel SAS pour effectuer les tests statistiques. Plusieurs tests permettent de s'assurer de la pertinence de tel choix.

##### ***a) L'analyse à l'aide de Q-Q plot :***

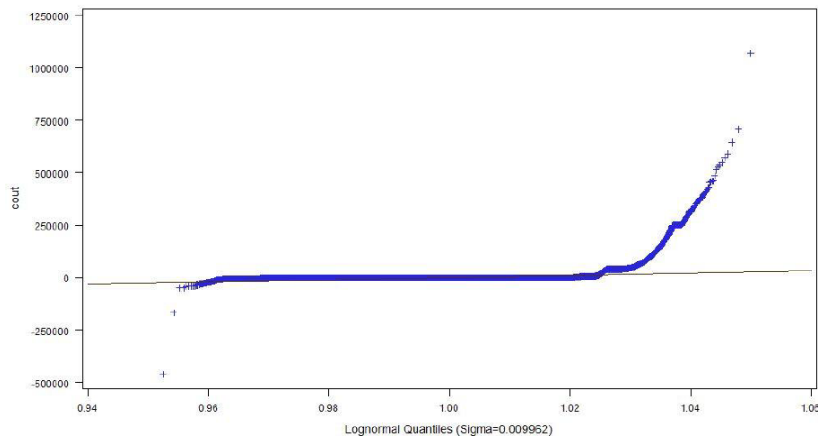
L'analyse à l'aide de l'outil graphique Q-Q plot n'est que descriptive. Elle permet d'avoir une première idée sur la meilleure distribution à choisir sans pour autant confirmer définitivement ce choix. Par la suite, et pour valider le modèle sélectionné, on doit impérativement avoir recours à des tests statistiques plus rigoureux.

Les graphes suivant sont des Q-Q Plot respectivement pour la loi Poisson, et pour la loi Gamma :

**Figure 11 : Q-Q plot pour la loi de Poisson**



*Source : Élaborée par les auteurs, sortie SAS*

**Figure 12 : Q-Q plot pour la loi Gamma**

*Source : Élaborée par les auteurs, sortie SAS*

On remarque que ces deux graphes ne nous fournissent pas une idée claire sur le choix du modèle à adopter puisque les deux figures sont quasiment similaires avec une différence près concernant les valeurs extrêmes.

*b) Qualité d'ajustement du modèle :*

Les distributions sur lesquelles nous nous sommes focalisé feront l'objet d'une analyse de la qualité d'ajustement avant d'en choisir la meilleure, pour cela, nous aurons recours à quelques concepts statistiques tel que la déviance.

La déviance est définie comme deux fois la différence entre le maximum possible de la log vraisemblance et le maximum atteint sur le modèle estimé.

Plus la valeur de D/DDL (déviance / degré de liberté) est faible, plus le modèle est crédible. Alors comparons ce rapport à la valeur 1.

**Tableau 21 : Comparaison des qualités d'ajustement du GLM sur la fréquence**

	Poisson	Binomiale négative
Déviance	751015,26	<b>100197,10</b>
Déviance/DDL	0,5979	<b>0,0798</b>

*Source : Élaboré par les auteurs, sortie SAS*

*Commentaire :*

Le test sur la qualité d'ajustement confirme que le modèle à distribution poisson n'est pas le bon. En effet le rapport en question dépasse celui de binomiale négative. Nous en concluons que le modèle à distribution binomiale négative est le plus susceptible d'être le modèle approprié. Par conséquent nous continuerons l'étude uniquement sur le modèle à cette même distribution afin de vérifier la significativité des variables explicatives.

*c) Test de significativité de Wald :*

**Tableau 22 : Test de significativité de Wald pour le GLM avec la loi Binomiale Négative**

Analyse des valeurs estimées du paramètre de vraisemblance maximum							
Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %	de Wald à	Pr > Khi-2
Intercept		1	-2,5924	0,0624	-2,7148	-2,47	1723,7 <.0001
CRM_Segment	1	1	-0,75	0,0619	-0,8714	-0,6286	146,62 <.0001

CRM_Segment	2	1	-0,7131	0,0615	-0,8336	-0,5926	134,49	<.0001
CRM_Segment	3	0	0	0	0	0	.	.
Age_anciennete_permi	1	1	0,4049	0,0301	0,3458	0,464	180,41	<.0001
Age_anciennete_permi	2	1	0,2783	0,0346	0,2104	0,3462	64,58	<.0001
Age_anciennete_permi	3	1	0,4221	0,0328	0,3579	0,4864	165,66	<.0001
Age_anciennete_permi	4	1	-0,0226	0,0351	-0,0913	0,0461	0,41	<b>0,05198</b>
Age_anciennete_permi	5	1	-0,0799	0,0414	-0,1609	0,0012	3,73	<b>0,0534</b>
Age_anciennete_permi	6	1	0,1419	0,0206	0,1015	0,1823	47,39	<.0001
Age_anciennete_permi	7	1	-0,0667	0,036	-0,1373	0,0039	3,43	<b>0,0642</b>
Age_anciennete_permi	8	1	-0,0079	0,0491	-0,1043	0,0884	0,03	<b>0,08716</b>
Age_anciennete_permi	9	0	0	0	0	0	.	.
Anc_Vehicule_Segment	1	1	0,7077	0,0268	0,6553	0,7602	699,05	<.0001
Anc_Vehicule_Segment	2	0	0	0	0	0	.	.
Region_Segment1	1	1	0,2138	0,0209	0,1729	0,2548	104,84	<.0001
Region_Segment1	2	1	0,8602	0,0177	0,8255	0,8948	2363,3	<.0001
Region_Segment1	3	0	0	0	0	0	.	.
Dispersion		1	56,575	0,4401	55,7124	57,438		

Source : *Élaboré par les auteurs, sortie SAS*

Commentaire :

On constate que le test ne rejette pas l'hypothèse de significativité des variables. On en conclut que les variables requises contribuent significativement à l'explication de la variable fréquence des sinistres. Néanmoins il met en évidence la nécessité de regrouper certaines modalités dans une seule modalité. Ceci fait, le résultat du test devient comme suit :

**Tableau 23 : Test de significativité de Wald pour le GLM avec la loi Binomiale Négative après regroupement**

Analyse des valeurs estimées du paramètre de vraisemblance maximum								
Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 Wald	de Pr > Khi-2
Intercept		1	-2,5988	0,0624	-2,7212	-2,4765	1734,04	<.0001
CRM_Segment	1	1	-0,7461	0,062	-0,8676	-0,6246	144,75	<.0001
CRM_Segment	2	1	-0,7186	0,0615	-0,8392	-0,598	136,36	<.0001
CRM_Segment	3	0	0	0	0	0	.	.
Age_anciennete_permi	1	1	0,4158	0,0297	0,3576	0,474	196,14	<.0001
Age_anciennete_permi	2	1	0,2885	0,0343	0,2213	0,3557	70,86	<.0001
Age_anciennete_permi	3	1	0,1694	0,0225	0,1254	0,2134	56,88	<.0001
Age_anciennete_permi	4	1	0,15	0,02	0,1107	0,1893	55,96	<.0001
Age_anciennete_permi	5	0	0	0	0	0	.	.
Anc_Vehicule_Segment	1	1	0,7092	0,0268	0,6567	0,7618	700,7	<.0001
Anc_Vehicule_Segment	2	0	0	0	0	0	.	.
Region_Segment1	1	1	0,2052	0,0208	0,1644	0,246	97,32	<.0001
Region_Segment1	2	1	0,865	0,0177	0,8303	0,8996	2387,99	<.0001
Region_Segment1	3	0	0	0	0	0	.	.

Dispersion	1	56,7835	0,4414	55,9184	57,6486		
------------	---	---------	--------	---------	---------	--	--

Source : *Élaboré par les auteurs, sortie SAS*

**Tableau 24 : Test de significativité de Wald pour le GLM avec la loi Binomiale Négative après le deuxième regroupement**

Analyse des valeurs estimées du paramètre de vraisemblance maximum								
Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		1	-2,596	0,0624	-2,7184	-2,4736	1729	<.0001
CRM_Segment	1	1	-0,7518	0,062	-0,8734	-0,6302	146,86	<.0001
CRM_Segment	2	1	-0,7173	0,0616	-0,838	-0,5966	135,76	<.0001
CRM_Segment	3	0	0	0	0	0	.	.
Age_anciennete_permi	1	1	0,2607	0,0184	0,2246	0,2968	200,13	<.0001
Age_anciennete_permi	2	1	0,1492	0,0201	0,1099	0,1885	55,34	<.0001
Age_anciennete_permi	3	0	0	0	0	0	.	.
Anc_Vehicule_Segment	1	1	0,7098	0,0268	0,6573	0,7624	701,5	<.0001
Anc_Vehicule_Segment	2	0	0	0	0	0	.	.
Region_Segment1	1	1	0,2029	0,0208	0,1622	0,2437	95,22	<.0001
Region_Segment1	2	1	0,8638	0,0177	0,8291	0,8985	2380,92	<.0001
Region_Segment1	3	0	0	0	0	0	.	.
Dispersion		1	56,8572	0,4419	55,9912	57,7233		

Source : *Élaboré par les auteurs, sortie SAS*

**ii) Pour le coût moyen des sinistres :**

Lors de la modélisation de la charge, deux lois candidates ont fait objet de l'étude, la loi Gamma et la loi Log-Normale. A présent, nous allons appliquer le modèle linéaire généralisé sur les deux distributions et vérifier lequel des deux modèles expliquera mieux le coût moyen des sinistres. Ce serait également une occasion pour fiabiliser ou remettre en cause notre choix de distribution.

**a) Qualité d'ajustement du modèle :**

On rappelle que la déviance mesure l'écart entre la vraisemblance théorique et celle de la distribution. Ainsi lors d'une comparaison, le modèle qui aura une déviance minimale sera jugé meilleur. Le résultat de notre comparaison est le suivant :

**Tableau 25 : Comparaison des qualités d'ajustement du GLM sur le coût moyen des sinistres**

	Gamma	Log-normale
Déviance	<b>43559,39</b>	43348474000000
Déviance/DDL	<b>1,6783</b>	34509324,837

Source : *Élaboré par les auteurs, sortie SAS*

*Commentaire :*

Le test sur la qualité d'ajustement confirme que le modèle à distribution Gamma n'est pas le bon. En effet le rapport en question dépasse amplement 1. Nous en concluons que le modèle à distribution log normale est le plus susceptible d'être le modèle approprié.

De même le test sur les deux déviations qualifie le modèle à distribution log normale au détriment de celui à distribution Gamma. Par conséquent nous continuerons l'étude uniquement sur le modèle à distribution log–normale pour vérifier la significativité des variables explicatives et tester la normalité des résidus.

*b) Test de significativité de Wald :***Tableau 26 : Test de significativité de Wald pour le GLM avec la loi Gamma**

Analyse des valeurs estimées du paramètre de vraisemblance maximum								
Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		1	9,812	0,0485	9,7168	9,9071	40870,9	<.0001
CRM_Segment	1	1	0,1258	0,0469	0,0339	0,2176	7,2	0,0073
CRM_Segment	2	1	0,2782	0,0461	0,1879	0,3684	36,46	<.0001
CRM_Segment	3	0	0	0	0	0	.	.
Age_anciennete_permi	1	1	0,0222	0,0271	-0,031	0,0754	0,67	<b>0,4131</b>
Age_anciennete_permi	2	1	0,0551	0,0321	-0,0077	0,1179	2,95	<b>0,0856</b>
Age_anciennete_permi	3	1	-0,1673	0,0295	-0,2252	-0,11	32,13	<b>&lt;.0001</b>
Age_anciennete_permi	4	1	0,1118	0,0359	0,0415	0,1821	9,71	0,0018
Age_anciennete_permi	5	1	0,0313	0,0409	-0,0489	0,1114	0,58	<b>0,4447</b>
Age_anciennete_permi	6	1	-0,0654	0,0202	-0,1049	-0,026	10,5	<b>0,0012</b>
Age_anciennete_permi	7	1	0,1617	0,0362	0,0907	0,2328	19,91	<.0001
Age_anciennete_permi	8	1	0,1441	0,0494	0,0473	0,2408	8,52	0,0035
Age_anciennete_permi	9	0	0	0	0	0	.	.
Anc_Vehicule_Segment	1	1	-0,4104	0,0198	-0,4493	-0,372	427,6	<.0001
Anc_Vehicule_Segment	2	0	0	0	0	0	.	.
Region_Segment1	1	1	0,3571	0,0228	0,3125	0,4017	246,31	<.0001
Region_Segment1	2	1	-0,2808	0,0181	-0,3162	-0,245	241,78	<.0001
Region_Segment1	3	0	0	0	0	0	.	.
Scale		1	0,718	0,0054	0,7076	0,7287		

*Source : Élaboré par les auteurs, sortie SAS*

*Commentaire :*

On constate que le test ne rejette pas l'hypothèse de significativité des variables. On en conclut que les variables requises contribuent significativement à l'explication de la variable coût moyen des sinistres. Néanmoins il met en évidence la nécessité de regrouper certaines modalités dans une seule modalité. Ceci fait, le résultat du test devient comme suit :

**Tableau 27 : Test de significativité de Wald pour le GLM avec la loi Gamma après le regroupement final**

Analyse des valeurs estimées du paramètre de vraisemblance maximum								
Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		1	9,8362	0,0483	9,7414	9,9309	41405,3	<.0001
CRM_Segment	1	1	0,1163	0,0469	0,0244	0,2082	6,15	0,0131
CRM_Segment	2	1	0,2874	0,0461	0,1971	0,3777	38,89	<.0001
CRM_Segment	3	0	0	0	0	0	.	.
Age_anciennete_permi	1	1	-0,0266	0,0176	-0,0611	0,0078	2,3	0,01296
Age_anciennete_permi	2	1	-0,0914	0,0196	-0,1299	-0,0529	21,67	<.0001
Age_anciennete_permi	3	0	0	0	0	0	.	.
Anc_Vehicule_Segment	1	1	-0,4138	0,0199	-0,4527	-0,3748	434,07	<.0001
Anc_Vehicule_Segment	2	0	0	0	0	0	.	.
Region_Segment1	1	1	0,3652	0,0227	0,3208	0,4097	259,09	<.0001
Region_Segment1	2	1	-0,2826	0,0181	-0,318	-0,2472	244,82	<.0001
Region_Segment1	3	0	0	0	0	0	.	.
Scale		1	0,7164	0,0054	0,706	0,727		

*Source : Élaboré par les auteurs, sortie SAS*

En guise de conclusion, les lois utilisées pour la modélisation de la fréquence et du coût moyen par le modèle linéaire généralisé sont respectivement la loi Binomiale Négative et la loi Gamma.

Si les modèles actuels sont suffisamment sophistiqués pour que l'on puisse les considérer comme des outils utiles et performants et non plus comme des curiosités théoriques, il ne faut pas oublier qu'un modèle a ses limites et ne donne qu'une image imparfaite de la réalité. Les modèles doivent être utilisés de façon souple, sans y croire complètement à la limite. L'approche devant être ouverte et multiforme, et en ce sens, il n'y a pas une seule méthode pour un problème.

Une modélisation économétrique permet donc un meilleur contrôle des risques et de leurs impacts sur l'entreprise en fournissant une information fiable et robuste pour les décideurs.

La prime de chaque segment sera, par ailleurs, calculée comme le produit du coût moyen et de la fréquence des sinistres. Ainsi, les modèles définitifs s'écrivent de la manière suivante :

- Pour le coût moyen :

$$\text{Coût moyen}_{i,j,k,l} = \exp(\beta_0 + \beta_{CRM_i} + \beta_{A*AP_j} + \beta_{AV_k} + \beta_{R_l})$$

Où

- $i$  prend les valeurs (1, 2, 3) pour les CRM.
- $j$  prend les valeurs (1, 2, 3) pour les classes des âge \* ancienneté du permis.
- $k$  prend les valeurs (1, 2) pour les classes des anciennetés de véhicules.
- $l$  prend les valeurs (1, 2, 3) pour les classes des régions.

- Pour la fréquence :

$$Fréquence_{i,j,k,l} = \exp(\beta_0' + \beta_{CRM_i}' + \beta_{A*AP_j}' + \beta_{AV_k}' + \beta_{R_l}')$$

Où

- $i$  prend les valeurs (1, 2, 3) pour les CRM.
- $j$  prend les valeurs (1, 2, 3) pour les classes des âge \* ancienneté du permis.
- $k$  prend les valeurs (1, 2) pour les classes des anciennetés de véhicules.
- $l$  prend les valeurs (1, 2, 3) pour les classes des régions.

➤ Le calcul de la prime s'écrit donc :

$$Prime\ a\ priori_{i,j,k,l} = Fréquence_{i,j,k,l} \times Coût\ moyen_{i,j,k,l}$$

Où

- $i$  prend les valeurs (1, 2, 3) pour les CRM.
- $j$  prend les valeurs (1, 2, 3) pour les classes des âge \* ancienneté du permis.
- $k$  prend les valeurs (1, 2) pour les classes des anciennetés de véhicules.
- $l$  prend les valeurs (1, 2, 3) pour les classes des régions.

## **II.2.5 Les classes de tarification**

La construction des classes de risque en assurance automobile est stratégique pour que le principe de mutualisation soit fonctionnel dans cet environnement concurrentiel.

Ces classes, constituées à partir de caractéristiques de l'assuré et de son véhicule, sont supposées homogènes en termes de sinistralité. A l'aide des variables tarifaires, l'assureur subdivise son portefeuille en classes d'assurés identiques du point de vue des critères qu'il a sélectionnés. Une prime sera calculée pour chaque classe et sera appliquée aux individus de cette dernière.

Ces classes, constituées à partir de caractéristiques de l'assuré et de son véhicule, sont supposées homogènes en terme de sinistralité. Cette homogénéité a été mesurée par des indicateurs de risque, à savoir le rapport S/P, la fréquence et le coût moyen des sinistres.

En suivant ce principe et en intégrant toutes les variables explicatives, on a pu former 54 classes tarifaires. Pour chacune de ces classes formées, on a calculé l'ensemble des quantités statistiques de la sinistralité (fréquence, coût moyen, prime moyenne, et le S/P ...).

Ensuite, à partir de la modélisation, de la fréquence et du coût moyen, on a pu en s'appuyant sur les procédures du logiciel SAS, calculer les valeurs estimées des paramètres du modèle (nombre de sinistre et la charge totale ...).

### **II.2.5.1 Les classes de tarif initiales**

En se servant des fonctions de liens adaptées au modèle jugé adéquat et pour le coût et pour la fréquence, on a calculé les valeurs estimées de ces deux derniers pour chacune des classes tarifaires.

**Tableau 28 : Extrait du tableau exhaustif de l'ensemble des classes de tarifs**

Classe	CRM	Région	Ancienneté du véhicule	Num classe Age*Anciennete_permis	Age	Ancienneté du permis	S/P	Prime à priori
1	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	1	32]	[0,3]	8,3%	159
2	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	2	32]	]3,7]	17,5%	324
3	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	3	32]	]7,+[	7,9%	153
4	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	4	]32,42]	[0,3]	136,8%	2 626
5	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	5	]32,42]	]3,7]	72,1%	1 408
6	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	6	]32,42]	]7,+[	53,3%	1 072
7	[0;90]	Laâyoune - Boujdour - Sakia El Hamra; Gharb - Chrarda - Béni Hssen; Chaouia - Ouardigha; Tadla - Azilal; Oued Eddahab - Lagouira; Doukkala - Abda; Autres; NR; Guelmim - Es-Smara.	[0,1]	7	]42,+[	[0,3]	36,4%	714

*Source : Élaboré par les auteurs, sortie excel*

A ce stade on va tenter, à partir de ces 54 classes, de réduire le nombre des classes de tarif et ce en se basant sur les quantités statistiques (S/P, poids en année risque et prime pure estimée).

Pour ce faire, on va essayer de construire plusieurs propositions de regroupement afin qu'ensuite pourvoir sélectionner la proposition optimale de nombre de classes qui seront prises comme les classes définitives de risque.

### **II.2.5.1 La classification définitive de tarif**

Pour choisir la classe finale de tarification, on s'est basé sur cinq propositions et ce en se focalisant à la fois sur la prime moyenne et sur le S/P moyen.

**Tableau 29 : La totalité des classes de tarifs à priori définitives**

Classe	Année Risque	Poids Année risque	S/P	Nombre de sinistre estimé	Charge estimée	Coût moyen estimé	Fréquence estimée	Prime estimée
1	60387	11,08%	23,7%	2 234	45 450 592	20 345	3,70%	<b>752,65</b>
2	263259	48,31%	39,2%	16 604	306 441 242	18 456	6,31%	<b>1 164,03</b>
3	86262	15,83%	58,8%	6 366	119 675 277	18 801	7,38%	<b>1 387,34</b>
4	111393	20,44%	92,7%	6 702	177 579 084	26 497	6,02%	<b>1 594,17</b>
5	23607	4,33%	142,3%	1 388	48 161 641	34 696	5,88%	<b>2 040,11</b>
	<b>544909</b>	<b>100,00%</b>	<b>52%</b>	<b>33 293</b>	<b>697 307 835</b>	<b>20 944</b>	<b>6,11%</b>	<b>1279,68</b>

*Source : Élaboré par les auteurs, sortie excel*

Ainsi pour sélectionner la meilleure classification parmi ces propositions, on s'est basé sur la conformité de ces classes tarifaires aux besoins du secteur marocain de l'assurance automobile ainsi que les besoins de la compagnie CNIA SAADA Assurance.

### **Conclusion**

Bref, les modèles convenables à la fois à la fréquence et au coût moyen des sinistres sont enfin bien adaptés et les critères ainsi que les classes de risque sont empiriquement bien choisis et formés, la société disposera dorénavant d'une base de données prête à être exploitée pour une analyse multidimensionnelle. Que ce soit pour classifier les nouveaux assurés ou pour les opérations des mouvements des assurés sur l'ensemble des classes. Ce présent constat va présenter un point de départ de notre prochaine section concernant l'étude corrective de la prime de base à priori.

# **SECTION III**

## **Tarification à postériori**

### **Elaboration d'un système Bonus-Malus**

**Plan :**

- Introduction ;
- Soubassements théoriques du système Bonus-malus ;
- Modélisation du Bonus-malus ;
- Élaboration des échelles Bonus-malus ;
- Conclusion.

## Introduction

Les systèmes de tarification à priori i.e exclusivement basés sur les variables tarifaires recueillies à la souscription (hors celles décrivant une sinistralité antérieure) ne permettent pas de cerner totalement le risque et ainsi constituer des classes homogènes.

Ils ne capturent en effet pas le phénomène de propension à la sinistralité : la sinistralité passée est un bon prédicateur de la sinistralité future, d'où l'idée de faire appel à un système de personnalisation des primes corrigées ou système Bonus-malus qui peut tenir compte a posteriori de ces facteurs non observables par le biais de la sinistralité observée pour chaque assuré, afin de corriger la prime de référence.

Le principe des clauses de bonification-majoration est d'une part d'offrir une diminution de prime aux assurés n'ayant déclaré aucun sinistre responsable, et d'autres parts de pénaliser les assurés accidentés en fonction du nombre de leurs sinistres. Ces systèmes de réduction-majoration sont une des méthodes de tarification a posteriori qui permettent d'affiner la tarification en intégrant une dimension individuelle.

On peut trouver deux justifications à l'emploi de tarifs de types **Bonus-malus** en assurance automobile. La première concerne la sélection des risques et la seconde fait référence à l'incitation à la prudence au volant.

La tarification à priori ne fait aucunement intervenir le mérite de l'assuré, au contraire, la tarification d'expérience mutualise les risques entre les individus qui ont connu une sinistralité comparable. Sur la base de modèles actuariels, on décidera de la part de chaque sinistre due au hasard, et dont l'assuré n'a donc pas à subir les conséquences en vertu du principe fondamental de la mutualisation des risques, et de la part du sinistre qui révèle les caractéristiques cachées de l'assuré. La tarification a posteriori peut paraître plus équitable que la tarification à priori car elle se fonde sur des actes de l'assuré lui-même et non sur son identification à un collectif d'individus indiscernables du point de vue de la compagnie.

Le processus de fixation de la prime est en fait un processus d'évaluation du risque; à tout instant, la prime doit être proportionnelle au risque. La superposition de la correction a posteriori à la tarification à priori est donc une méthode d'évaluation continue du risque. Il est clair que ces deux problèmes ne peuvent pas s'envisager indépendamment l'un de l'autre.

# **CHAPITRE 1**

## SOUBASSEMENTS THEORIQUES DU SYSTEME BONUS-MALUS



### **III.1.1 Introduction**

Comme cela se fait presque partout dans le monde. Notre système Bonus-malus sera basé uniquement sur la fréquence et non pas sur les coûts des sinistres. En d'autre terme, un sinistre de 10.000 Dirhams entrainera le même malus qu'un sinistre de 1.000.000 de Dirham.

Ce système ne sera efficace que si l'hétérogénéité du portefeuille ne se situe qu'au niveau des fréquences et non des montants.

### III.1.2 Modèle poissonien : Portefeuille homogène

Formulons les trois hypothèses intuitives suivantes. Soit  $N(t, t+\Delta t)$  le nombre d'accidents dans l'intervalle de temps  $(t, t+\Delta t)$  :

- $P[N(t, t+\Delta t)=1] = \lambda \Delta t + o(\Delta t)$
- $P[N(t, t+\Delta t)>1] = o(\Delta t)$
- Si on note  $\tau$  et  $\tau'$  deux intervalles de temps non disjoints. Alors  
 $P[N(\tau) = k \text{ et } N(\tau') = k'] = P[N(\tau) = k] \times P[N(\tau') = k']$ .

La première hypothèse stipule que la probabilité d'avoir un accident durant un petit intervalle  $(t, t + \Delta t)$  est, en négligeant les termes de puissance supérieure ou égale à 2, proportionnelle à la longueur de l'intervalle. En particulier, cette probabilité ne dépend pas du début de l'intervalle.

La deuxième veut dire que la probabilité d'avoir deux accidents ou plus dans cet intervalle est négligeable.

La troisième stipule que les nombres de sinistres occasionnés dans des intervalles de temps disjoints sont indépendants.

Il est aisé de démontrer qu'une loi de poisson satisfait à ces trois hypothèses. Réciproquement, on peut montrer que ces trois hypothèses impliquent que la distribution du nombre de sinistres  $P_k(t) = P[N(0,t) = k]$ ,  $k = 0, 1, 2$  est une loi de poisson de paramètre  $\lambda$ .

On a :  $P_k(t+\Delta t) = P_k(t) [1 - \lambda \Delta t] + P_{k-1}(t) \lambda \Delta t + o(\Delta t)$

Pour  $k = 0, 1, 2 \dots$  Avec l'hypothèse supplémentaire que  $P_{-1}(t) = 0$ . Par conséquent :

$$\frac{P_k(t + \Delta t) - P_k(t)}{\Delta t} = -\lambda P_k(t) + \lambda P_{k-1}(t) + \frac{o(\Delta t)}{\Delta t}$$

En passant à la limite pour  $\Delta t \rightarrow 0$ , on obtient la suite d'équations différentielles :

$$\begin{aligned} P_k'(t) &= -\lambda P_k(t) + \lambda P_{k-1}(t) & k = 1, 2, \dots \\ P_0'(t) &= -\lambda P_0(t) & k = 0. \end{aligned}$$

En résolvant de manière récursive ces équations différentielles, avec les conditions initiales  $P_0(0) = 1$  et  $P_k(0) = 0$  si  $k > 0$ .

Nous obtenons :

$$P_k(t) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$

En pratique, on estime  $\lambda$  par l'estimateur du maximum de vraisemblance  $\tilde{\lambda} = \frac{1}{n} \sum_{k=1}^{k_{max}} k n_k$  où  $n_k$  désigne le nombre de polices du portefeuille pour lesquelles on a observé  $k$  sinistres durant la période de référence,  $k_{max}$  est le nombre maximum observé de sinistres touchant une même police du portefeuille et, par convention nous notons  $n$  la somme des  $n_k$ ,  $n = \sum_{k=1}^{k_{max}} n_k$ .

### III.1.3 Modèle binomial négatif : Portefeuille hétérogène

Lorsqu'il s'agit de décrire le nombre de sinistres occasionnés par un conducteur, la loi de poisson dite loi des *événements rares* est un modèle probabiliste très important. Mais lorsqu'on s'intéresse au portefeuille, à cause de l'hétérogénéité des profils de risque des différents assurés, la loi de poisson devient peu adéquate pour la description du nombre de sinistres. D'où l'idée d'utiliser un mélange de poisson qui revient à supposer que pour chaque

assuré, le nombre de sinistre engendré au cours de l'année suit une loi de poisson de paramètres  $\lambda$ .

$\lambda$  varie d'un assuré à l'autre suivant une loi que nous supposons être de fonction de densité  $u$  et de répartition  $U$ . Pour chaque assuré la fonction de densité du nombre de sinistres annuels s'écrit :

$$g_0(k) = \frac{e^{-\lambda} \lambda^k}{k!}; \quad k = 1, 2, \dots$$

Choisissons comme distribution à priori de  $\lambda$  une  $\Gamma$  de paramètre  $a$  et  $\tau$ , dont la fonction de fréquence est encore appelée fonction de structure :

$$u(\lambda) = \frac{\tau^a e^{-\tau \lambda} \lambda^{a-1}}{\Gamma(a)}$$

$$\text{Avec } a, \tau > 0 \text{ et } \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$$

La loi de probabilité du nombre  $N$  de sinistres causés par un assuré du portefeuille est alors donnée par :

$$P(N = k) = \frac{\Gamma(k + a)}{\Gamma(k + 1) \Gamma(a)} \times \frac{\tau^a}{(1 + \tau)^{k+a}}$$

Nous obtenons une distribution Binomiale négative, dont la moyenne et la variance sont comme suite :

$$\bar{x} = \frac{a}{\tau}$$

$$S^2 = \frac{a}{\tau} \left( 1 + \frac{1}{\tau} \right)$$

Les estimateurs de  $a$  et  $\tau$  sont obtenus en résolvant le système précédent :

$$\tilde{a} = \frac{\bar{x}^2}{S^2 - \bar{x}}$$

$$\tilde{\tau} = \frac{\bar{x}}{S^2 - \bar{x}}$$

Montrons que la distribution a posteriori est également une  $\Gamma$ , mais de paramètres  $\tau' = \tau + t$  et  $a' = a + k$ , et où  $k = \sum_{i=1}^t k_i$  est le nombre total de sinistres.

Etant donné les hypothèses du modèle :

$$P(k_1, \dots, k_t / \lambda) = \frac{\lambda^k e^{-\lambda t}}{\prod_{j=1}^t (k_j!)}$$

Par conséquent, la fréquence des sinistres a posteriori du groupe des assurés ayant subi l'historique  $(k_1, \dots, k_t)$  est la moyenne de la distribution  $\Gamma$  de paramètres  $a'$  et  $\tau'$ , à savoir :

$$\lambda_{t+1}(k_1, \dots, k_t) = \frac{a'}{\tau'} = \frac{a + k}{\tau + t}$$

### III.1.4 Formulation du modèle Markovien à classes

Considérons une échelle Bonus-malus disposant de  $s$  degrés. A un moment donné, chaque assuré se trouve à un degré  $i$  de notre échelle. Notons  $C_i$  la classe composée de tous les assurés de degré  $i$ . Supposons que la classe d'un assuré au cours d'une année donnée est complètement déterminée par sa classe au cours de l'année précédente et le nombre de sinistres engendrés au cours de cette année. C'est une propriété importante qui nous permet de modéliser l'évolution d'une police donnée par un processus de Markov qui, il faut le rappeler, est un processus *sans mémoire*. Dans ces conditions, notre système Bonus-malus est déterminé par quatre éléments :

- i- Le nombre de classe :  $s$  ;
- ii- Le vecteur de niveaux de primes  $b = (b_1, \dots, b_s)$  où la prime des assurés de la classe  $i$  est exprimée en pourcentage de la prime de référence ;
- iii- La classe d'entrée dans le système :  $i_0$  ;
- iv- Les règles de transition d'une classe à l'autre sous la forme de fonctions  $T_k$  ( $k \geq 0$ ), indicées par le nombre  $k$  de sinistres responsables de la période précédente :

$$T_k: \{1, \dots, s\} \longrightarrow \{1, \dots, s\}$$

$$i \longrightarrow T_k(i)$$

où  $T_k(i)$  est la classe dans laquelle est placé un assuré issu de la classe  $i$ , pour  $k$  sinistres l'année précédente.

Tous ces éléments sont résumés dans le tableau suivant :

**Tableau 30 : Présentation d'une échelle Bonus-malus**

Classes	Niveaux de primes	Classes après ... $k$ sinistres
$s$	$b_s$	.
.	.	.
.	.	.
.	.	.
Classe d'accès => $i_0$	$b_{i_0}$	.
.	.	.
.	.	.
.	.	.
$i$	$b_i$	.
.	.	.
.	.	... $T_k(i)$
1	$b_1$	.

*Source : Christian PARTRAT, Jean-Luc BESSON. Assurance non-vie : Modélisation et simulation, Édition Economica, décembre 2004, pages 258 et 259*

Les classes sont rangées par niveaux de primes décroissants de  $s$  à 1 :  $b_s > \dots > b_1$ .  
Si  $b_s < 100$ , le système est qualifié de bonus.

## **CHAPITRE 2**

# MODELISATION DU SYSTEME BONUS-MALUS



### **III.2.1 Introduction**

Le but de ce chapitre est de fournir un ensemble de tests pour justifier l'inadéquation du modèle homogène de distribution du nombre de sinistres et pour choisir un modèle hétérogène et adéquat pour approcher la distribution du nombre de sinistres. Pour le test d'ajustement, nous utiliserons le test  $\chi^2$  d'adéquation que nous décrivons ci-dessous.

### III.2.2 Le test $\chi^2$ d'ajustement des lois

Le but de cette partie est de fournir un ensemble de tests pour justifier l'inadéquation du modèle homogène de distribution du nombre de sinistres et pour choisir un modèle hétérogène et adéquat pour approcher la distribution du nombre de sinistres. Pour le test d'ajustement, nous utiliserons le test  $\chi^2$  d'adéquation.

Considérons un échantillon  $X_1, X_2, \dots, X_n$  d'une population dans laquelle on étudie un caractère  $X$  dont la distribution est inconnue. On cherche à déterminer si une distribution théorique qui peut représenter adéquatement le comportement probabiliste de la variable étudiée en se basant sur les valeurs observées.

Il s'agit en d'autre terme de tester les hypothèses :

$H_0$  :  $X$  est distribuée suivant une distribution spécifiée  $f_0(X, \theta)$ ,  $\theta$  pouvant être connu ou inconnu.

$H_1$  :  $X$  n'est pas distribuée suivant  $f_0(X, \theta)$ .

Considérons le cas des variables discrètes (qui nous préoccupent ici) dont les paramètres sont connus et désignons par  $p$  la fonction de masse inconnue du caractère  $X$  étudié dans la population. Supposons que les réalisations de l'échantillon observé sont réparties en  $k$  valeurs :  $x_1, x_2, \dots, x_k$ .

Soit  $N_i$  le nombre de fois que  $X = x_i$  avec  $i = 1, 2, \dots, k$ ,  $N_i$  suit une  $B(n, p_i)$  où  $p_i = P(X = x_i)$  inconnue.

$$E[N_i] = n p_i$$

$$V[N_i] = n p_i q_i$$

Si  $n$  est grand de sorte que  $n p_i \geq 5$ , on peut montrer que la variable aléatoire :

$$X^2 = \sum_{i=1}^k \frac{(N_i - n p_i)^2}{n p_i} \rightsquigarrow \chi^2_{k-1}$$

On rejette  $H_0$  au seuil  $\alpha$  si  $X^2 > C_\alpha$  : où  $\alpha$  est la valeur critique que l'on aura choisie de façon à ce que  $P_{H_0}(X^2 > C_\alpha) = \alpha$ .

### III.2.3 Modèle poissonien : Portefeuille homogène

Nous allons justifier l'instauration d'un système Bonus-malus comme critère de tarification à postériori, et ceci, en prouvant que l'hypothèse d'égalité des assurés devant le risque ne résiste pas à l'analyse statistique.

Le tableau ci-dessous présente la distribution du nombre de sinistres matériels du portefeuille observé. Sa moyenne vaut 0,104 et sa variance  $S^2 = 0,114$ , ainsi la méthode du maximum de vraisemblance et la méthode des moments nous conduisent à estimer le paramètre  $\lambda$  de la distribution par la moyenne observée. En ajustant la distribution observée par une distribution de Poisson de paramètre  $\tilde{\lambda} = 0,104$ , nous obtenons des effectifs théoriques significativement différents des effectifs observés. Visiblement, l'ajustement n'est pas bon.

**Tableau 31 : Ajustement de la sinistralité par la loi de Poisson**

Nombre de sinistres $k$	Nombre de fréquence de polices observées $n_k$	$P_k$	Effectifs théoriques $n * P_k$	Statistique de Khi-deux
0	110 513	0,90	109 990	2
1	10 468	0,09	11 456	85

2	1 010	0,00	597	287
3	63	0,00	21	86
4	9	0,00	1	133
<b>Total</b>	122 063	-	-	<b>593,262</b>

Source : Elaboré par les auteurs, sortie excel

Estimation des paramètres :

**Tableau 32 : Estimation des paramètres de la loi de Poisson**

<b>Moyenne</b>	0,104
<b>Variance</b>	0,114
$\tilde{\lambda}$	0,104
$\chi^2_{4;0.01}$	13,277
$\chi^2_{\text{calculé}}$	<b>593,262</b>

Source : Elaboré par les auteurs, sortie excel

Le test  $\chi^2_{\text{calculé}} = 593,26 \gg \chi^2_{4;0.01} = 13,277$  confirme la très mauvaise qualité de l'ajustement par le modèle poissonien, ce qui nous amène à rejeter le modèle.

L'hypothèse d'homogénéité du portefeuille, constituée uniquement des sinistres matériels, ne résiste pas à l'analyse statistique.

### III.2.4 Modèle binomial négatif : Portefeuille hétérogène

Le modèle le plus fréquent utilisé pour la description de l'hétérogénéité des portefeuilles est le modèle binomial négatif. Nous allons essayer de décrire notre portefeuille par un tel modèle et faire le test d'ajustement entre les valeurs observées et les valeurs théoriques d'une binomiale dont les paramètres sont estimés par les observations. Les procédures d'estimation ont été décrites ci-dessus. Nous utiliserons les estimateurs de moments donc nous considérons qu'ils ne font pas perdre de degré de liberté à la statistique de Khi deux.

L'estimation des paramètres du modèle donnent comme résultats :

$$\tilde{\alpha} = 0.47363$$

$$\tilde{\tau} = 4.77776$$

Ainsi, ci-dessous le tableau présentant les résultats du test de Chi-deux :

**Tableau 33 : Ajustement de la sinistralité par une binomiale négative**

Nombre de sinistres $k$	Nombre de fréquence de polices observées $n_k$	$P_k$	Effectifs théoriques $n \cdot P_k$	Statistique de Khi-deux
0	110 513	0,9052	110 492	0
1	10 468	0,0863	10 529	0
2	1 010	0,0078	949	3,26
3	63	0,0007	84	5,25

4	9	0,0001	7	0
<b>Total</b>	122 063	-	-	<b>9,234</b>

*Source : Elaboré par les auteurs, sortie excel*

Estimation des paramètres :

**Tableau 34 : Estimation des paramètres de la loi binomiale négative**

<b>Moyenne</b>	0,104
<b>Variance</b>	0,114
$\tilde{a}$	1,120
$\tilde{\tau}$	10,756
$\chi^2_{4;0.01}$	13,277
$\chi^2_{\text{calculé}}$	<b>9,232</b>

*Source : Elaboré par les auteurs, sortie excel*

On a trouvé  $\chi^2_{\text{calculé}} = 9,23 < \chi^2_{4;0.01} = 13,277$

Donc, on accepte le test d'ajustement par une binomiale négative hétérogène. On remarque d'ailleurs la proximité entre les valeurs théoriques et les valeurs observées. Nous avons donc une preuve statistique de la nécessité de considérer une tarification personnalisée dans tout le portefeuille par le modèle binomial négatif.

## **CHAPITRE 3**

# ÉLABORATION DES ECHELLES BONUS-MALUS



### **III.3.1 Introduction**

A travers le chapitre suivant, on va essayer d'élaborer notre échelle Bonus-malus basée uniquement sur le nombre d'accidents passés.

Notre but principal fut toujours l'ajustement des primes individuelles à travers le temps. C'est ce qu'on va tenter d'appliquer à partir du système construit afin de suivre l'évolution des assurés sur cette échelle ainsi que la manière dont cette dernière varie selon la sinistralité.

### III.3.2 Préambule : *Principe adopté*

L'assureur peut faire trois choix concernant l'échelle Bonus-malus qu'il souhaite mettre en œuvre. Tout d'abord il peut choisir le nombre de classes, disons  $s$ . A chaque classe est associé un niveau de prime, en pourcentage de la prime de départ, disons  $C_m; m = 1, 2, \dots, s$ . Ensuite il peut choisir le niveau d'entrée d'un nouveau conducteur. Enfin il choisit les règles de transition à l'intérieur du système Bonus-malus.

*Formalisons ces règles de transition :*

Soit  $T_k(m) = n$  la règle qui envoie un assuré qui était au niveau  $m$  vers le niveau  $n$  lorsque  $k$  sinistres ont été déclarés durant la période d'assurance.

Un exemple de règle de transition est le suivant :

$$T_0(m) = \max(1; m - 1)$$

$$T_k(m) = \min(s; m + 1); k > 0$$

Cette règle de transition indique qu'une année sans sinistre implique une chute d'un échelon (Et l'on ne peut naturellement pas descendre sous le niveau 1) tandis que déclarer  $k$  sinistres implique une montée d'un échelon par sinistre déclaré (et l'on ne peut naturellement pas dépasser le niveau  $s$ ). Le système est markovien car la position d'un assuré ne dépend que de sa position précédente (et pas de son historique plus ancien, c'est-à-dire la manière dont il arrive à cette position précédente) et du nombre de sinistres qu'il a commis dans la dernière période d'observation.

#### Estimation des paramètres

La pénalisation que subit un assuré qui a causé  $k$  sinistres en  $t$  années est :  $\frac{a+k/\tau+t}{a/\tau} = \frac{\tau(a+k)}{a(t+\tau)}$  exprimée en pourcentage de la prime de référence, par exemple si la prime à priori proposée à un assuré est 1000 Dhs, la prime à postériori ainsi vaudra :  $1000 \times \frac{a+k/\tau+t}{a/\tau}$  pour un assuré après la  $t$ -ième année avec  $k$  sinistres.

Illustration numérique :

$$\tilde{a} = 0.47363$$

$$\tilde{\tau} = 4.77776$$

Les années sont représentées en lignes et le nombre de sinistres est représenté en colonnes.

**Tableau 35 : Ajustement de la sinistralité par une binomiale négative**

$t/k$	0	1	2	3	4	5	6
0	100	-	-	-	-	-	-
1	91	173	255	337	418	500	582
2	84	160	235	310	385	461	536
3	78	148	218	288	357	427	497

4	73	138	203	268	333	398	463
5	68	129	190	251	312	373	434
6	64	121	179	236	293	351	408
7	61	115	169	223	277	331	385
8	57	109	160	211	262	313	364
9	54	103	152	200	249	297	346
10	52	98	144	191	237	283	329

*Source : Elaboré par les auteurs, sortie excel*

De cette présentation, l'échelle Bonus-malus obtenue par le modèle Binomial Négative avec les paramètres estimés ci-dessus est représentée dans le tableau ci-dessus, nous nous sommes limités à 6 sinistres et la plus longue durée considérée est de 10 ans. On déduit que le rabais est de presque 10% sur une année sans sinistres et une majoration de 20% avec un seul sinistre.

L'inconvénient de cette présentation est qu'elle est infinie et ne tient pas compte de la répartition de la sinistralité dans le temps passé, conformément à ce qui est appliqué, en pratique, dans tous les systèmes Bonus-malus du monde. De plus, les malus qui devraient sanctionner les conducteurs ayant causé des sinistres sont difficilement, voire même non imposables du point de vue commercial.

Certes, les modèles actuariels aboutissent à des malus élevés, mais, il est à remarquer que ceux auxquels nous avons aboutit sont exorbitants. L'explication possible est que le phénomène de la soif du bonus pourrait être bien développé. Dans ce cas, il y a une sous estimation du nombre de sinistres servant à la modélisation du système Bonus-malus.

Il est proposé de corriger les résultats par l'estimation des sinistres, qui ne sont pas déclarés du fait de la soif de bonus. C'est ce qui pourrait faire l'objet d'un prochain projet de fin d'étude.

Le deuxième élément est que l'on ne tient pas compte dans cette étude des réalités commerciales.

#### ➤ Présentation Markovienne :

La présentation précédente est simple mais pas très pratique pour affecter des garanties annexes gratuites pour les gens qui méritent en se basant sur la RC-auto, c'est pour cette raison que nous allons simplifier l'échelle et la représenter en un nombre fini de classe sans la faire perdre ses principales caractéristiques. Nous nous limiterons à 9 classes et les règles de transition sont obtenues à partir des sauts de prime effectués d'une situation à une autre.

Les deux valeurs extrêmes de prime sont : 78% (minimum) et 173% (maximum) de la prime de base qui correspond respectivement au niveau 0 et 8 de l'échelle. Ainsi :

- Tout nouvel assuré intègre le système au niveau 4 de l'échelle qui correspond à un tarif de 100% de la prime de base,
- Chaque année sans sinistre fait descendre l'assuré d'un degré de l'échelle,
- Chaque sinistre fait grimper de 1 degré.

**Tableau 36 : Échelle Bonus-malus dans le modèle binomiale négative à présentation**

**Markovienne**

CRM	Degré	Degré après 0 sinistre	Degré après 1 sinistre ou plus
	8	7	8
<b>173</b>	7	6	8
<b>160</b>	6	5	7
<b>148</b>	5	4	6
<b>138</b>	4	3	5
<b>100</b>	3	2	4
<b>91</b>	2	1	3
<b>84</b>	1	0	2
<b>78</b>	0	0	1
<b>73</b>			

*Source : Elaboré par les auteurs, sortie excel*

## Conclusion

Cette étude a permis, non seulement de couvrir une panoplie de techniques pour ajuster le nombre de sinistres et modéliser le système Bonus-malus, mais aussi d'analyser les statistiques afin de mieux comprendre la composition du portefeuille et le comportement des assurés.

Les résultats ont révélés une tarification exorbitante du fait que les assureurs font état d'un décalage entre le nombre de sinistres réellement survenus et le nombre de déclarations. En effet, les conducteurs préfèrent prendre en charge les petits sinistres, plutôt que de subir un malus. Ce phénomène fait que le modèle théorique est trop sévère et loin d'être accepté du point de vue commercial. Toutefois, l'étude renvoie à une méthode de correction des données, sans pouvoir l'effectuer faute de temps. Il s'agit de l'estimation du nombre de sinistres non déclarés ou non enregistrés de la part des assureurs.

# **SECTION IV**

## **Tarification à postériori**

**Crédibilité sur le portefeuille  
automobile usage tourisme**

**Plan :**

- Introduction ;
- Description du modèle adopté :  
Bühlman-Straub ;
- Application aux données ;
- Conclusion.

## Introduction

Si les assureurs n'étaient autorisés à utiliser qu'une seule variable de tarification, celle-ci devrait être sous la forme d'un système de mérite. En effet, le meilleur prédicateur du nombre de réclamations n'est pas l'âge ou le sexe du conducteur, mais bien son expérience passée de sinistres.

Au cours des années, les assureurs ont essayé d'accumuler des informations longitudinales sur leurs assurés, motivant ainsi les analyses a posteriori de la correction de la prime. Conséquemment, lorsqu'un assuré accumule un historique de perte, sa prime annuelle est modifiée en fonction de son expérience. Contrairement aux idées populaires, il ne s'agit pas d'un mécanisme visant à ce que l'assuré rembourse ce que sa compagnie a dû payer pour ses sinistres passés, mais bien d'une façon de mieux évaluer le risque couvert par la compagnie d'assurance.

La révision de la prime se justifie de plusieurs manières : l'expérience passée donne une meilleure idée du risque réel de l'assuré, mais aussi parce qu'un accident d'automobile peut engendrer une contagion en modifiant les perceptions de conduite.

Dans cette partie, nous cherchons à évaluer la crédibilité des contrats individuels du portefeuille automobile usage tourisme matérialisée par la prime.

A chaque contrat on associe un facteur  $\alpha_j$  appelé crédibilité qui mesure la fiabilité de son expérience propre ( $0 \leq \alpha_j \leq 1$ ). La crédibilité d'un contrat donné s'écrit alors comme suit :

$$P_{\text{credibilisee}} = \alpha_j \times X_j + (1 - \alpha_j) \times P_{\text{a priori}}$$

Ainsi, si  $\alpha_j = 0$ , l'information du contrat est jugée comme absolument non fiable, par conséquent il faut tenir compte uniquement des résultats du groupe, si  $\alpha_j = 1$ , l'expérience du groupe est jugée parfaitement non fiable alors on se base uniquement sur l'expérience propre.

Il reste à savoir donc comment déterminer le facteur  $\alpha_j$ , plusieurs critères paraissent naturels :

- Plus l'expérience d'un contrat est grande et plus il paraît crédible,
- Certains contrats ont des résultats plus stables que d'autres et doivent donc être d'avantage crédibilisés.

## **CHAPITRE 1**

# DESCRIPTION DU MODELE ADOPTE BÜHLMAN-STRAUB



### **IV.1.1 Introduction**

La théorie de crédibilité va être mise en œuvre par le modèle de Bühlmann-Straub. En 1967, Bühlmann se penche sur le problème et développe un modèle de crédibilité où la prime d'assurance d'un assuré est exprimée linéairement en fonction d'une moyenne pondérée entre son expérience passée et sa prime à priori.

### IV.1.2 Description du modèle

Soit un portefeuille de  $N$  contrats observés dans une période de  $T$  années. On note  $X_{jt}$  la grandeur (objet de l'étude) observée au niveau du  $j^{\text{ème}}$  contrat, cette grandeur peut éventuellement être le nombre de sinistres, le coût des sinistres, le S/P, ou encore la prime pure.

Ici on s'intéresse à la prime pure. On note  $\omega_{jt}$  le poids de l'observation  $X_{jt}$ . Le poids  $\omega_{jt}$  peut être déterminé selon plusieurs critères liés à la taille du contrat, à la sinistralité ou à l'exposition. De plus on associe à chaque contrat un paramètre de risque  $\theta$ , la loi de  $X_{jt}$  est alors déterminée par  $\theta$  et  $\omega$ , néanmoins le paramètre  $\theta$  est évidemment inobservable d'où la nécessité de supposer que le paramètre  $X_{jt}$  est indépendant du temps.

### IV.1.3 Hypothèses du modèle

Le modèle de BUHLMAN-STRAUB repose sur un ensemble d'hypothèses, on les résume comme suit :

**Hyp 1** : Les variables  $X_{jt}$  sont de carrés intégrables (i.e. à variances finies).

**Hyp 2** : Les contrats du portefeuille sont deux à deux indépendants, mathématiquement cela est traduit par : Les vecteurs  $(\theta, X_{jt} \ t \geq 0)$  sont indépendants

**Hyp 3** : A priori, tous Les contrats ont le même risque, mathématiquement : Les vecteurs  $(\theta, X_{jt} \ t \geq 0)$  ont la même loi de distribution.

**Hyp 4** : La valeur moyenne ne dépend que du paramètre de risque :  $E(X_{jt} / \theta_j = \theta)$  est indépendante de  $j$  et de  $t$ .

**Hyp 5** : La variance du risque  $\theta$  à fixé est inversement proportionnelle au poids :  $V(X_{jt} / \theta_j = \theta) = \frac{\sigma(\theta)^2}{\omega_{jt}}$  ;  $\sigma(\theta)^2$  est indépendant du temps.

**Hyp 6** : Conditionnellement au paramètre de risque, les différentes observations d'un même contrat sont indépendantes, autrement : Conditionnellement à  $\theta_j$ , les variables  $X_{jt} ; t \geq 0$  sont non corrélés.

Avant d'énoncer le théorème de Bühlman-Straub, il convient d'introduire un ensemble de notations qui nous seront utiles lors de l'étude qui suit.

- $\mu_0 = E(X_{jt})$ . C'est le la prime à priori, autrement dit c'est la prime de base que l'on doit attendre d'un contrat dont on ne dispose pas de données.
- $\mu(\theta) = E(X_{jt} | \theta_j = \theta)$ . C'est la prime probable d'un contrat de paramètre  $\theta$ . Néanmoins on ne connaît pas en pratique le paramètre, par conséquent, on ne connaît pas non plus  $\mu(\theta)$ . On cherche donc à l'estimer.
- $\tau = V(E(X_{jt} | \theta_j = \theta))$  C'est la variance inter contrats. Elle mesure l'hétérogénéité du portefeuille. En matière de crédibilité, plus cette variance est importante plus les contrats risquent de s'éloigner nettement de la moyenne du portefeuille, plus il faudrait donc se fier au statistiques propres du contrat.
- $\sigma^2 = E(E(X_{jt} | \theta_j = \theta))$ . C'est la variance intra contrat Elle mesure la fluctuation de la prime dans le temps. Ainsi, Plus  $S^2$  est important, plus les primes individuelles sont

volatiles. L'expérience individuelle paraît donc peu fiable et on doit se fier plus à l'information collective.

#### IV.1.4 Estimateurs des paramètres de structure

On cherche à travers la méthode de crédibilité à déterminer le meilleur estimateur  $X_j$  possible prime de chaque contrat  $j$  donc le meilleur estimateur de  $\mu(\theta_j)$ .

Pour ce faire deux approches sont possibles :

- La crédibilité homogène : consiste à chercher l'estimateur sous forme d'une fonction linéaire des observations c'est-à-dire du type :  $\sum \beta_{it} X_{it}$
- La crédibilité non homogène : vise à déterminer l'estimateur comme fonction affine de la forme :  $\beta_0 + \sum \beta_{it} X_{it}$

Sous les hypothèses du modèle de Bühlman-Straub, ces estimateurs prennent une forme plus simple, c'est l'objet du théorème qui suit :

##### Théorème : Estimateurs de Bühlman-Straub

Sous les hypothèses (**Hyp 1**) à (**Hyp 6**), les estimateurs de Bühlman-Straub sont donnés par :

$$\text{Cas homogène : } \hat{X}_j = \alpha_j \bar{X}_j + (1 - \alpha_j) \mu_0$$

$$\text{Cas non homogène : } \hat{X}_j = \alpha_j \bar{X}_j + (1 - \alpha_j) \bar{X}$$

Un problème se pose : les estimateurs dépendent des paramètres structuraux  $\hat{\sigma}^2$  et  $\hat{\tau}^2$ , qui sont inconnus. Par conséquent, Il faut les estimer tout d'abord. Plusieurs estimations sont possibles. Ici on choisit l'estimateur classique de VYLDER donné comme suit :

$$\hat{\sigma}^2 = \frac{1}{I(n-1)} \sum_{i=1}^I \sum_{j=1}^n w_{ij} (X_{ij} - X_i)^2$$

$$\hat{\tau}^2 = \frac{w_{..}}{w_{..}^2 - \sum_{i=1}^I w_i^2} \left\{ \sum_{i=1}^I w_i (X_i - \bar{X})^2 - (I-1) \hat{\sigma}^2 \right\}$$

#### IV.1.5 Récapitulatif de la démarche

L'estimation paraît simple, par contre celle des paramètres  $\sigma^2$  et  $\tau^2$  n'est pas évidente vu qu'elle fait intervenir les coefficients de crédibilité qui eux même dépendent des paramètres  $\hat{\sigma}^2$  et  $w_{ij}$ . Ainsi une procédure itérative est envisageable pour résoudre le problème. Les étapes de la procédure se résument comme suit :

- Détermination des poids  $w_{ij}$  ;
- Calcul des  $X_i = \frac{S_i}{w_i}$  avec  $S_i$  est la charge du contrat  $i$  ;
- Estimation de  $\sigma^2$  par  $\hat{\sigma}^2$  ;
- Estimation de  $\tau^2$  par  $\hat{\tau}^2$  ;
- Estimation des  $\alpha_i$  par  $\hat{\alpha}_i = \frac{w_i}{w_{..}^2 + \frac{\hat{\sigma}^2}{\hat{\tau}^2}}$  ;
- Estimation de  $\mu_0$  par  $\hat{\mu}_0$  ;
- Calcul des  $\mu(\theta_1)^{\text{empirique}} = \hat{\alpha}_i \times X_i + (1 - \hat{\alpha}_i) \times \hat{\mu}_0$  .

## **CHAPITRE 2**

# APPLICATION AUX DONNEES



### **IV.2.1 Introduction**

Le présent chapitre va nous servir à mettre en œuvre la partie théorique citée auparavant, l'application du modèle a été réalisée sous le logiciel Excel après en s'appuyant sur les informations ramenées concernant l'historique du sinistre des assurés à partir du logiciel SAS. Le choix de l'outil servant à appliquer notre étude est dû à sa flexibilité et sa facilité pour traiter les données, en premier lieu on calcule les crédibilités des contrats individuels dans le portefeuille automobile et ce en se référant à l'année 2011.

### IV.2.2 Commentaires et interprétations

Nous rappelons que le nombre de contrats individuels présent dans le portefeuille s'élève à 709 045. De ce fait nous ne présentons pas la totalité des facteurs de crédibilité calculés par le modèle car on a procédé de manière individuelle.

Afin de pouvoir commenter les résultats, on présente ci-dessus quelques statistiques sur les crédibilités des contrats.

**Tableau 37 : Les facteurs de crédibilité : Minimum, maximum et moyen**

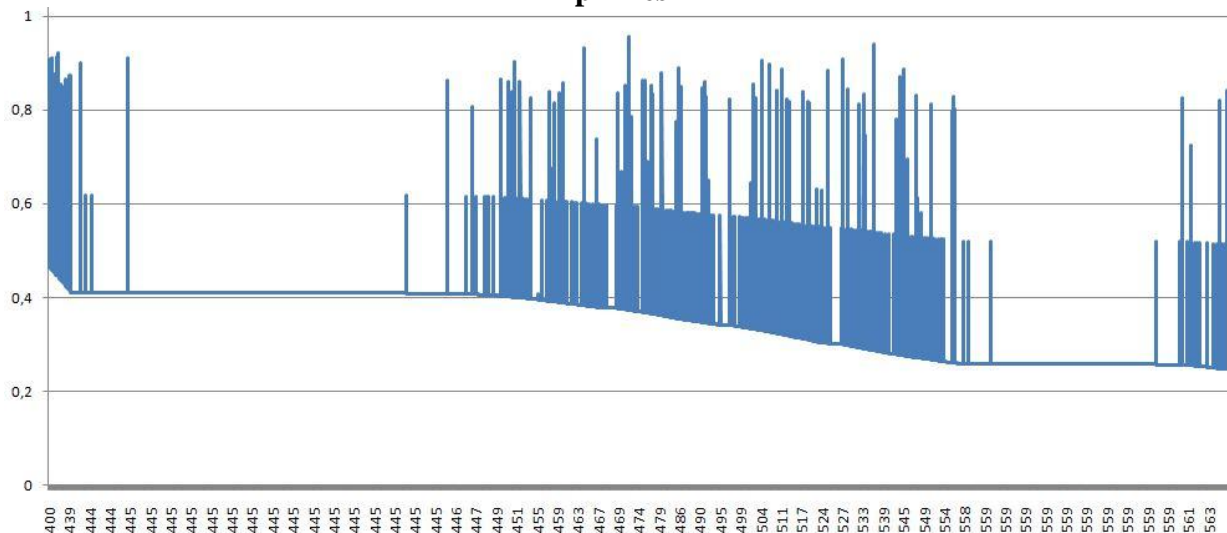
<b>Crédibilité minimal (<math>\alpha_{\min}</math>)</b>	0,000945
<b>Crédibilité maximal (<math>\alpha_{\max}</math>)</b>	0,957402
<b>Crédibilité moyenne (<math>\alpha_{\text{moyen}}</math>)</b>	0,181710

*Source : Elaboré par les auteurs, sortie excel*

Les contrats paraissent moyennement crédibles et l'étendue de la variable  $\alpha$  est très grande. Cela met en évidence l'existence d'un paramètre qui différencie les contrats et agit sur leur crédibilité. Lors du choix des poids, la durée d'exposition au risque et la charge des sinistres sont jugés comme étant les facteurs majeurs qui différencient plus entre les contrats en terme de leurs primes crédibilisées. De ce fait on propose de visualiser la variation des facteurs de crédibilité en fonction du poids des contrats (matérialisée par leur durée d'exposition au risque) afin de trouver une explication à la grande étendue de la variable  $\alpha$ .

Le graphique suivant représente la variation de la crédibilité des contrats en fonction de leur prime.

**Figure 13 : La variation de la crédibilité des contrats individuels en fonction de leurs primes**



*Source : Elaborée par les auteurs, sortie excel*

On constate que la crédibilité des contrats individuels diminue avec la prime, elle dépasse 80% pour les contrats individuels de basse prime. Ce qui explique l'étendue importante de la variable  $\alpha$ . Les résultats obtenus par le modèle de Bühlmann-Straub mettent en évidence la nécessité de subdiviser le portefeuille automobile usage tourisme à caractère individuel.

### IV.2.3 Estimation des paramètres et exemple d'illustration

L'application du modèle de BHULMAN STRAUB a été réalisée à l'aide des données formées dans la 1ère partie de la tarification à priori et ce en suivant les étapes citées auparavant passant par l'estimation des paramètres jusqu'au calcul des primes crédibilisées pour chaque contrat de manière individuelle.

i) *L'estimation des paramètres de structure :*

**Tableau 38 : Les paramètres de structure estimés du modèle de Bühlman-Straub**

$w$	0,7685
$\bar{X}$	1196,8488
$I$	709045
$\hat{\sigma}^2$	482692877,2956
$\widehat{\tau^2}$	166728521,1340

*Source : Elaboré par les auteurs, sortie excel*

i) *Exemple réel d'illustration :*

Après avoir estimé les paramètres du modèle, on peut à ce stade donner les nouvelles primes pour chaque numéro de contrat.  
Voici ci-joint un extrait du tableau donnant l'ensemble des primes crédibilisées.

**Tableau 39 : Échantillon de primes crédibilisées selon le modèle de Bühlman-Straub**

Num de police	Prime commerciale	Charge totale	Nombre sinistre	Année risque	Tarif à priori	Coefficient de crédibilité	Prime crédibilisée
AU03381360T	1 885	40 000	1	1,367	1 594	32,08%	10 468
20002310003	2 238	0	0	2,022	1 164	41,12%	685
20002310054	3 141	0	0	2,008	1 164	40,96%	687
20002310057	2 014	0	0	2,005	753	40,92%	445
20002310058	3 141	0	0	2,005	1 164	40,92%	688
AU03385031Y	2 429	40 400	1	0,866	1 594	23,02%	11 969
AU03389548N	3 141	5 000	1	0,874	1 164	23,19%	2 221
AU03390376P	2 429	6 000	1	0,866	1 164	23,02%	2 491
AU03390380P	2 014	10 000	1	1,874	753	39,29%	2 554

*Source : Elaboré par les auteurs, sortie excel*

On remarque que pour les assurés n'ayant causé aucun sinistre dans l'année, subissent une diminution de leur prime de base avec un abattement variant selon plusieurs critères (prime à priori, durée d'exposition au risque ...).

## **Conclusion**

À ce stade, on a pu effectuer une tarification corrective sur notre prime tarifaire et ce à travers une crédibilisation de la prime pure à priori déjà calculée au niveau de la section II.

Malgré son importance et sa valeur ajoutée à notre tarification, afin que cette étude de crédibilité soit plus exploitable au niveau de la compagnie, une prolongation sur moult années s'avère nécessaire pour donner des résultats plus fiables. Ceci dit, la contrainte de temps était obstacle devant une telle application, du coup on s'est référé à l'historique des sinistres relatifs aux années 2010 et 2011 pour crédibiliser la prime à payer pour 2012.

Aussi intéressante qu'elle soit, cette étude prolongée sur une multitude d'années fera un excellent sujet d'un nouveau projet.

## Conclusion générale

Tout au long de ce projet, on a pu traiter les deux tarifications majeures appliquées en pratique assurantielle automobile :

- Une tarification à priori dont l'idée fut de séparer les contrats (les assurés) en plusieurs catégories de telle façon qu'à l'intérieur d'une catégorie, les risques puissent être homogènes. Les classes étaient formées à partir d'information disponible sur l'assuré à priori.
- Une tarification a posteriori ou par expérience qui repose sur le fait que la survenance des sinistres est révélatrice de la qualité du risque et apporte une information supplémentaire par rapport aux informations demandées lors de la souscription.

Les deux formes de tarification utilisées furent complémentaires. Lorsqu'un tarif a été mis en place, on a pu l'affiner à l'aide de la tarification a posteriori par deux méthodes, et ce avec le biais de deux systèmes Bonus-malus et crédibilité qui tiennent compte des sinistres déclarés par les assurés pour le premier et la charge pour le deuxième.

Ainsi, notre assureur privé peut décider du degré de segmentation à priori qu'elle retient en choisissant un certain nombre et un certain type de variables explicatives et ceci va le conduire à proposer à ses assurés une tarification a posteriori adéquate.

La tarification de l'assurance automobile comporte donc une prime de base commune par catégories de classement à priori, à laquelle s'applique pour tous les assurés un mécanisme de prise en compte de l'expérience, non limitée dans le temps. Ce principe nous a permis de calculer une prime correspondant exactement à la sinistralité réelle de chaque assuré.

Afin d'affiner le système de tarification de notre assureur privé nous recommandons les propositions suivantes :

- ✚ L'utilisation du critère de l'historique des sinistres de chaque assuré est d'une importance capitale pour la concurrence sur le marché de l'assurance RC-auto donc notre assureur privé doit mener plusieurs études à postérieur sur le portefeuille du fait que le critère de l'historique des sinistres à la réputation d'être un critère *juste* puisque l'historique des sinistres passés est fortement influencé par le comportement de l'assuré, et ce d'une manière parfaitement *objective* sans que cela dépende de la possibilité de mesurer ou d'observer certaines caractéristiques.
- ✚ Un système de tarification vise à faire contribuer équitablement chaque assuré au financement des réclamations. Cependant, si les malus sont trop élevés, le système n'encourage pas les assurés à faire une réclamation et les empêche ainsi de bénéficier de la couverture à laquelle ils ont droit. De plus, il peut encourager les assurés à s'enfuir après avoir causé des dommages à des tiers. Pour éviter ces problèmes, les changements de prime entraînés par les bonus et les malus doivent être raisonnables, sans quoi le système ne transfère pas adéquatement le risque.
- ✚ Les systèmes de réduction majoration en vigueur devaient être conçus de telle sorte qu'ils doivent comporter un plancher et un plafond. Par ce principe les assureurs privés seraient obligés de respecter un seuil à ne pas dépasser sur l'échelle de Bonus-malus.

## Bibliographie

### ❖ Ouvrages

- [1] Arthur CHARPENTIER. Statistique de l'assurance : Assurance non vie tarification & provisionnement, 2007.
- [2] Michel DENUIT, Sandra PITREBOIS et Jean-François WALHIN. Méthodes de construction des systèmes Bonus-malus en RC-Auto, Juin 2007.
- [3] Christian PARTRAT, Jean-Luc BESSON. Assurance non-vie : Modélisation et simulation, Édition Économica, décembre 2004.

### ❖ Documents

- [1] Kamal BENCHEKROUN. Cours de la théorie de risque, 2011.
- [2] Abderrahim LOULIDI. Cours de l'assurance non vie, 2011.
- [3] Pierre THEROND. Théorie de crédibilité, Institut de Science Financière et d'Assurances - Université de Lyon, 2012.
- [4] Sébastien RINGUEDÉ. SAS version 9.2, introduction au décisionnel : méthode et maîtrise du langage, 2012.
- [5] Ministère de l'Intérieur, Direction Générale des Collectivités Locales. Découpage administratif marocain de 2009.

### ❖ Mémoires

- [1] Romain BOYER CHAMMARD. Processus de surveillance et de majoration des contrats flottes d'entreprises d'AXA France, mémoire d'Actuariat à l'université Paris Dauphine, novembre 2008. Consultable sur : <http://www.ressources-actuarielles.net/memoires>.
- [2] Guillaume GONNET. Étude de la tarification et de la segmentation en assurance automobile, Juin 2007. Consultable sur : <http://www.ressources-actuarielles.net/memoires>.

### ❖ Articles

- [1] Freddy CORLIER, Jean LEMAIRE, Dunia MUHOKOLO. Simulation of an Automobile Portfolio. Consultable sur : <http://www.sciencedirect.com>.
- [2] E. Gómez, A. Hernández, J. M Pérez, F.J Vázquez-Polo. Measuring sensitivity in a Bonus-malus system, April 2002. Consultable sur : <http://www.sciencedirect.com>.

### ❖ Wébographie

- [1] Portail intranet de CNIA SAADA Assurance : <http://www.cniasaada.ma>.
- [2] Portail intranet de la Fédération Marocaine des Sociétés d'Assurance et de Réassurance: <http://www.fmsar.org.ma>.

## Annexes

### Annexe I : Les statistiques après regroupement pour les segmentations retenues

- *Puissance fiscale* :

Classes	Situation matrimoniale	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	Coût moyen	Fréquence	S/P
1	Célibataire	47,0%	248 359	45,6%	446 711 647	258 657 390	17391	14873	7,0%	57,9%
2	Marié(e); Pacs	52,8%	295 503	54,2%	572 741 813	271 591 350	18603	14599	6,3%	47,4%
3	Veuf(ve); Divorcé(e)	0,2%	1 047	0,2%	2 062 986	849 283	92	9231	8,8%	41,2%
Total		<b>100%</b>	<b>544 909</b>	<b>100%</b>	<b>1 021 516 446</b>	<b>531 098 023</b>	<b>36086</b>	<b>14718</b>	<b>6,6%</b>	<b>52%</b>

*Source* : Elaboré par les auteurs, sortie excel

- *Ancienneté du permis croisée avec l'âge de l'assuré* :

Classes	Age	Ancienneté du permis	Age2	Ancienneté du permis2	Poids	Année Risque	Poids Année risque	Prime	Charge	Nombre de sinistre	S/P
1	[18,32]	-	]32,42]	[0,7]	28,7%	132 866	<b>24,4%</b>	212778231	173972907	11304	81,8%
2	]32,42]	]7,+[	-	-	20,7%	109 386	<b>20,1%</b>	201501776	96990560	7357	48,1%
3	]42,+[	-	-	-	50,5%	302 657	<b>55,5%</b>	607236438	260134555	17425	42,8%
Total					<b>100%</b>	<b>544 909</b>	<b>100%</b>	<b>1 021 516 446</b>	<b>531 098 023</b>	<b>36 086</b>	<b>52%</b>

*Source* : Elaboré par les auteurs, sortie excel

## Annexe II : Extraits des procédures SAS de la cartographie du risque automobile

- *Codification des variables :*

```
DATA tmp1.prod_sinistre; SET tmp1.prod_sinistre;
FORMAT CRM_Segment BEST12.;
  if CRM<=90 then CRM_Segment=1;
  if CRM=100 then CRM_Segment=2;
  if CRM>100 then CRM_Segment=3;
```

```
RUN;
```

```
DATA tmp1.prod_sinistre; SET tmp1.prod_sinistre;
FORMAT Anc_Vehicule_Segment BEST12.;
if anciennete_vehicule>=0 and anciennete_vehicule<=1 then
Anc_Vehicule_Segment=1;
if anciennete_vehicule>1 then Anc_Vehicule_Segment=2;
```

```
RUN;
```

```
DATA tmp1.prod_sinistrel; SET tmp1.prod_sinistrel;
FORMAT code_region BEST12.;
if region ='Autres'
  then code_region=0;
if region ='Chaouia - Ouardigha'
  then code_region=1;
if region ='Doukkala - Abda'
  then code_region=2;
if region ='Fès - Boulemane'
  then code_region=3;
if region ='Gharb - Chrarda - Béni Hssen'
  then code_region=4;
if region ='Grand casablanca'
  then code_region=5;
if region ='Guelmim - Es-Smara'
  then code_region=6;
if region ='Laâyoune - Boujdour - Sakia El Hamra'
  then code_region=7;
if region ='Marrakech - Tensift - Al Haouz'
  then code_region=8;
if region ='Méknès - Tafilalet'
  then code_region=9;
if region ='NR'
  then code_region=10;
if region ='Oriental'
  then code_region=11;
if region ='Oued Eddahab - Lagouira'
  then code_region=12;
if region ='Rabat - Salé - Zemmour - Zaïr'
  then code_region=13;
if region ='Souss - Massa Darâa'
  then code_region=14;
if region ='Tadla - Azilal'
  then code_region=15;
if region ='Tanger - Tétouan'
  then code_region=16;
if region ='Taza - Al Hoceima - Taounate'
  then code_region=17;
```

```
RUN;
```

```

DATA tmp1.prod_sinistre; SET tmp1.prod_sinistre;

FORMAT Region_Segment BEST12.;
if code_region=7 or code_region=4 or code_region=1 or code_region=15
  or code_region=12 or code_region=2 or code_region=10 or code_region=0
  or code_region=6 then Region_Segment=1;
if code_region=5 or code_region=8 or code_region=13 or code_region=3
  then Region_Segment=2;
if code_region=16 or code_region=9 or code_region=17 or code_region=11
  or code_region=14 then Region_Segment=3;

RUN;

```

```

DATA tmp1.prod_sinistre; SET tmp1.prod_sinistre;

FORMAT Age_anciennete_permis BEST12.;
if age<=32 or (age>32 and age<=42 and anciennete_permis<=7) then
Age_anciennete_permis=1;
if age>32 and age<=42 and anciennete_permis>7 then
Age_anciennete_permis=2;
if age>42 then Age_anciennete_permis=3;

RUN;

```

- *Calcul de la durée d'exposition au risque :*

```

DATA tmp1.prod; SET tmp1.prod;

FORMAT DER_2010 BEST12.;
if (date_effet < DEBUT) and (date_fin < DEBUT)
  then DER_2010=0;

if (date_effet > FIN) and (date_fin > FIN)
  then DER_2010=0;

if (date_effet <= DEBUT) and (date_fin >= FIN)
  then DER_2010=(FIN- DEBUT)+1;

if (date_effet < DEBUT) and ((date_fin >= DEBUT) and (date_fin<=FIN))
  then DER_2010=(date_fin - DEBUT)+1;

if (date_effet >= DEBUT) and (date_fin <= FIN)
  then DER_2010=(date_fin-date_effet)+1;

if ((Date_effet>=DEBUT) and (Date_effet <= FIN)) and (date_fin > FIN)
  then DER_2010=(FIN-Date_effet)+1;

RUN;

```

- *Jointure à gauche des tables PROD et Sinistre :*

```

PROC SQL;
CREATE TABLE tmp1.prod_sinistre AS SELECT prod_region.An_Emission
FORMAT=BEST12.,
prod_region.PRODUIT FORMAT=$24.,
prod_region.ID_PRODUIT FORMAT=BEST12.,
prod_region.Situation FORMAT=$8.,
prod_region.IND_FLOTTE FORMAT=$10.,
prod_region.Contrat_resilie FORMAT=$3.,
prod_region.Nbr_Conducteurs FORMAT=BEST12.,
prod_region.Type_Echeance FORMAT=$12.,
prod_region.Type_avenant FORMAT=$15.,
prod_region.Num_Pol FORMAT=$14.,
prod_region.Num_Avenant FORMAT=BEST12.,
prod_region.Date_Emission FORMAT=DDMMYY10.,
prod_region.Date_Effet FORMAT=DDMMYY10.,
prod_region.Annee_effet FORMAT=BEST12.,
prod_region.Date_Resil FORMAT=MMDDYY10.,
prod_region.Valeur_neuve FORMAT=BEST12.,
prod_region.Valeur_venale FORMAT=BEST12.,
prod_region.MARQUE FORMAT=$9.,
prod_region.MODEL FORMAT=$9.,
prod_region.IMMATRIC FORMAT=$10.,
prod_region.ENERGIE FORMAT=$1.,
prod_region.PUISS_FISC FORMAT=BEST12.,
prod_region.CYLINDREE FORMAT=BEST12.,
prod_region.DATE_MEC FORMAT=DDMMYY10.,
prod_region.Usage FORMAT=$10.,
prod_region.Num_attestation FORMAT=BEST12.,
prod_region.VILLE_ZONE_CIRCULATION FORMAT=$10.,
prod_region.CodeP_client FORMAT=BEST12.,
prod_region.Lieu_Risque FORMAT=$10.,
prod_region.CRM FORMAT=BEST12.,
prod_region.Sexe FORMAT=$1.,
prod_region.Date_Naissance FORMAT=DDMMYY10.,
prod_region.DATE_PERMIS FORMAT=DDMMYY10.,
prod_region.Situation_Matrimoniale FORMAT=$11.,
prod_region.Date_First_Ass FORMAT=DDMMYY10.,
prod_region.Code_courtier FORMAT=BEST12.,
prod_region.Ville_Courtier FORMAT=$20.,
prod_region.Region FORMAT=$37.,
prod_region.Valeur_Conventionnelle FORMAT=BEST12.,
prod_region.Valeur_Conv_O_N FORMAT=$1.,
prod_region.Date_Fin FORMAT=DDMMYY10.,
prod_region.RC FORMAT=BEST12.,
prod_region.age,
prod_region.anciennete_permis,
prod_region.anciennete_vehicule,
prod_region.DER_2010 FORMAT=BEST12.,
prod_region.DER_2011 FORMAT=BEST12.,
prod_region.DER FORMAT=BEST12.,
nbr_sinistre.Nombre_sinistre,
nbr_sinistre.charge_totale_sinistre
FROM tmp1.prod_region AS prod_region_der
LEFT JOIN tmp1.NBR_SINISTRE AS nbr_sinistre ON (prod_region.Num_Pol
= nbr_sinistre.Numero_Contrat) AND (prod_region.Num_Avenant =
nbr_sinistre.Num_Avenant_Impact);
QUIT;

```

- *Procédures GENMOD pour la modélisation de la fréquence et du coût moyen des sinistres :*

- Pour le coût moyen des sinistres :

```

ODS TAGSETS.EXCELXP
file='C:\Users\ProBook\Desktop\PFE\sortie.xls';

proc genmod data=tmp1.prod_sinistre;
CLASS CRM_Segment Matri_Sinistre_Segment Puissance_fiscale_Segment
Energie_Puiss_Segment Age_Segment Anc_permis_Segment
Anc_Vehicule_Segment
Region_Segment sexe_code;
model cout= CRM_Segment Matri_Sinistre_Segment
Puissance_fiscale_Segment
Energie_Puiss_Segment Age_Anciennete_permis Anc_Vehicule_Segment
Region_Segment sexe_code
/dist=NORMAL link=identity type1;
run;

ODS TAGSETS.EXCELXP close;

```

- Pour la fréquence des sinistres :

```

ODS TAGSETS.EXCELXP
file='C:\Users\ProBook\Desktop\PFE\sortie.xls';

proc genmod data=tmp1.prod_sinistre;
CLASS CRM_Segment Matri_Sinistre_Segment Puissance_fiscale_Segment
Energie_Puiss_Segment Age_Segment Anc_permis_Segment
Anc_Vehicule_Segment
Region_Segment sexe;
model freq= CRM_Segment Matri_Sinistre_Segment
Puissance_fiscale_Segment
Energie_Puiss_Segment Age_Segment Anc_permis_Segment
Anc_Vehicule_Segment
Region_Segment1 sexe
/ link=LOG dist=negbin type1;
run;

ODS TAGSETS.EXCELXP close;

```

- *Calcul de la charge totale des sinistres :*

```

DATA tmp1.base_sinistre2010_2011; SET tmp1.base_sinistre2010_2011;

FORMAT charge BEST12.;

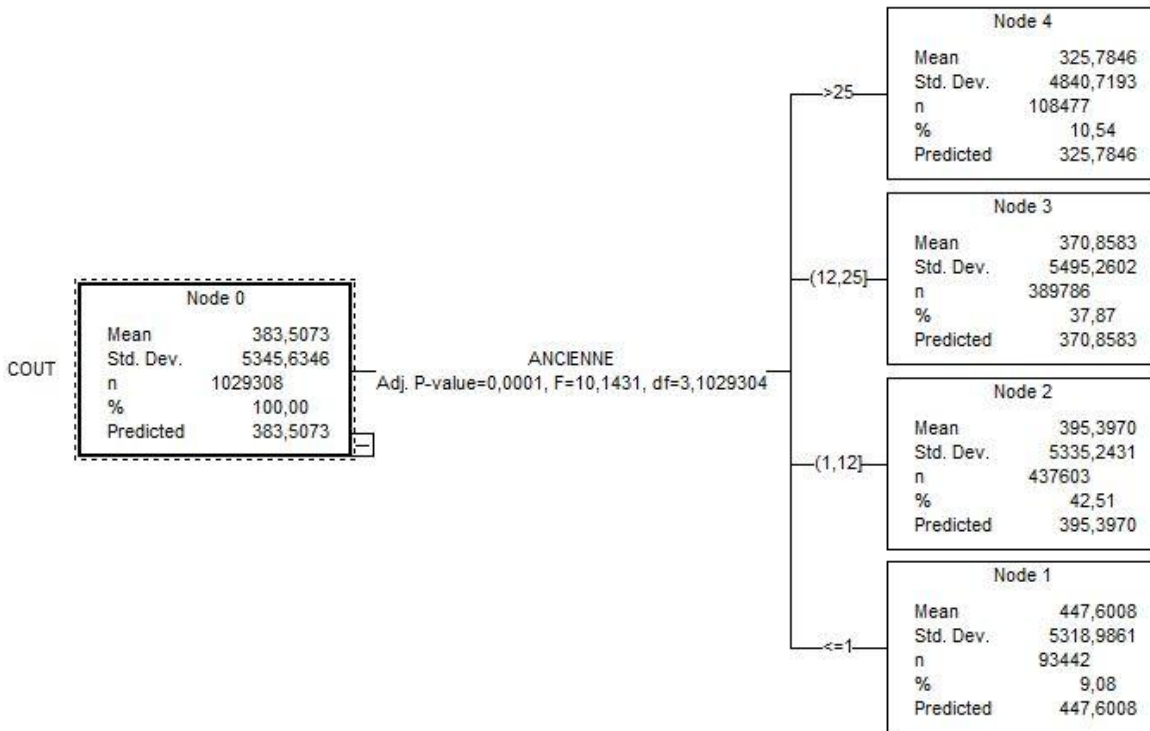
if charge_totale eq . then charge = 0;
else charge = Provision_total + Reglements_Total - Prov_Rec_Emis -
Rec_Emis_Reclames - Rec_Recuperes + Prov_Rec_Subis +
Rec_Subis_Reclames + Rec_Payes;

Run;

```

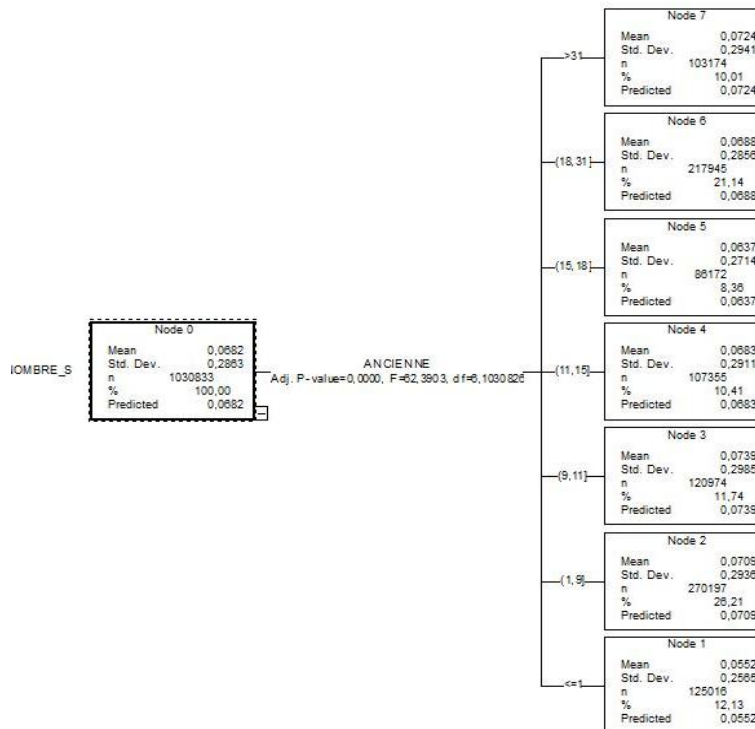
**Annexe III : Arbres de décision issues de SPSS ANSWER TREE**

- *Arbre de décision de la variable ancienneté du véhicule :*



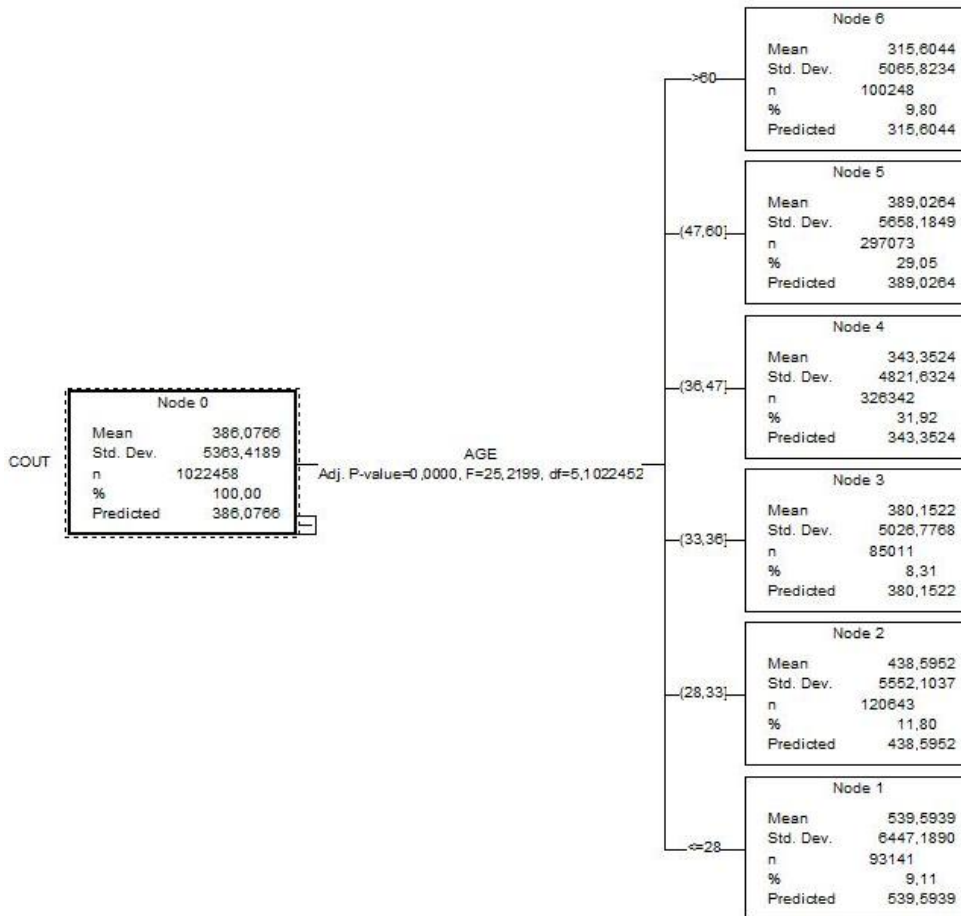
*Source : Elaborée par les auteurs, sortie SPSS ANSWER TREE*

- *Arbre de décision de la variable ancienneté du permis :*



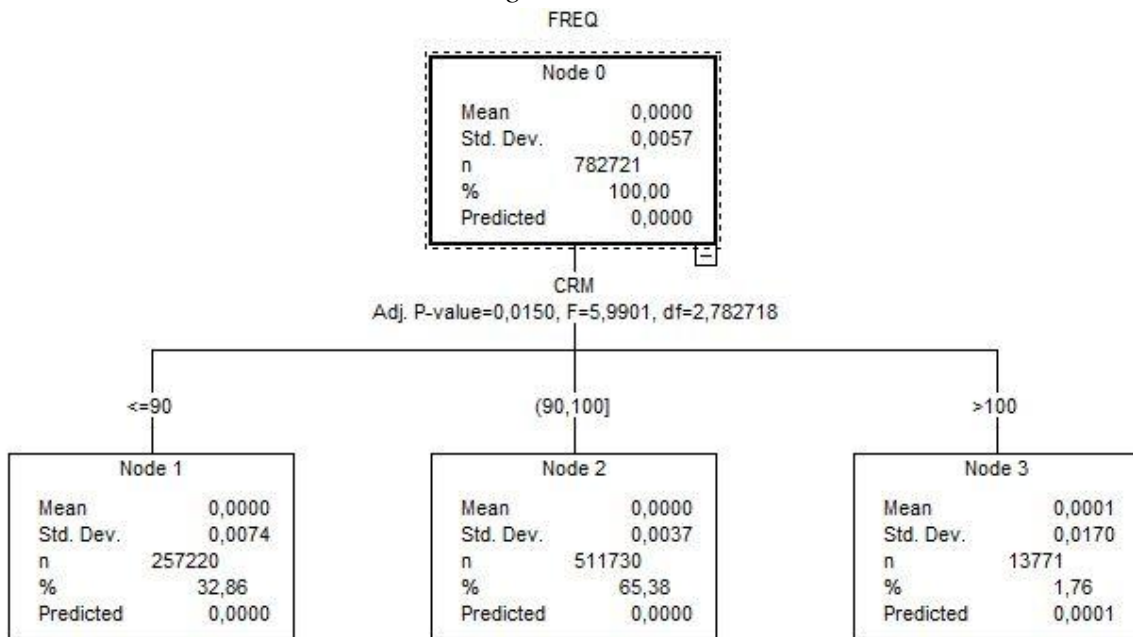
*Source : Elaborée par les auteurs, sortie SPSS ANSWER TREE*

- *Arbre de décision de la variable âge :*



*Source : Elaborée par les auteurs, sortie SPSS ANSWER TREE*

- *Arbre de décision de la variable âge :*



*Source : Elaborée par les auteurs, sortie SPSS ANSWER TREE*

#### **Annexe IV : Aperçu sur le système Bonus-malus marocain**

Après avoir été abandonnée vers la fin des années quatre vingt du siècle dernier, la clause de bonification-majoration a été réintroduite au Maroc par l'arrêté du ministère des finances et des investissements n°22977/97 du 6 Rabia I 1415 (15 août 1994). Dans moins d'un an, le système de tarification des primes de l'assurance automobile qui intègre cette clause, et qui est entré en vigueur à partir du premier septembre 1994, aura fait une décennie et beaucoup de malheureux. Il a toujours fait l'ordinaire du débat quotidien entre assureurs et assurés. Il serait sans doute opportun d'en faire une évaluation.

La clause de bonification-majoration dite clause de Bonus-malus insérée au contrat d'assurance automobile stipule que l'assuré s'oblige à payer une augmentation de prime en cas de sinistre : Majoration pour sinistre, mais peut dans le cas contraire bénéficier d'une réduction de prime : bonification pour non sinistre.

Quelques années après l'abandon de l'ancien système de Bonus-malus instauré par le tarif automobile de 1978, les assurés avaient fini par comprendre qu'à la suite d'un accident, ils ne seraient pas soumis à une majoration de prime. Profitant de cette faveur, certains d'entre eux se sont donnés à cœur joie dans la confection de faux constats amiables. Ces pratiques illicites ont trouvé un terrain fertile dû au manque de moyens de contrôle des compagnies, qui de surcroît préfèrent ne pas porter plainte systématiquement contre tous les auteurs des faux sinistres ; et ont coïncidé à l'époque (1990-1994) avec les promotions que faisait la douane marocaine en matière de dédouanement des voitures d'occasion.

Depuis le 06 juillet 2006, un nouveau système appelé "Coefficient Réduction Majoration" a été mis en place en remplacement du système Bonus-malus. Ce système agit sur la prime d'assurance au moyen d'un coefficient qui fonctionne comme suit :

- Réduction de 10% de la prime si vous n'avez causé aucun accident engageant ou susceptible d'engager totalement ou partiellement votre responsabilité durant une période d'assurance de vingt-quatre (24) mois consécutifs précédant la souscription ou le renouvellement de votre contrat.
- Si vous avez causé un ou plusieurs accidents engageant ou susceptibles d'engager totalement ou partiellement votre responsabilité durant les douze (12) mois précédant la souscription ou le renouvellement de votre contrat, votre prime d'assurance est majorée de 20% pour chaque accident matériel et de 30% pour chaque accident corporel sans pour autant pouvoir dépasser 250% de la prime de base.

Ces taux sont respectivement de 15% et de 20% si vous exploitez un véhicule destiné au transport public de voyageurs ou si vous êtes souscripteur d'une responsabilité civile garagiste.

*Source : Fédération Marocaine des Sociétés d'Assurance et de Réassurance (FMSAR)*

## **Annexe V : Glossaire des termes clés de l'étude**

**Rapport charges/primes** : Le rapport charges/primes : désigne le rapport entre les charges de sinistres et les encaissements bruts de primes pendant une année. On y prend en général pour charges de sinistres d'une année le montant devant être payé par la compagnie d'assurance pour les sinistres survenus pendant ladite année (indépendamment de l'année au cours de laquelle le règlement du sinistre, et donc la prestation réelle, a lieu). Dans certains cas, on considère comme charges de sinistres d'une année les prestations réalisées pendant l'année (pour des sinistres survenus pendant ladite année ou lors d'années précédentes).

**Prime tarifaire** : La prime tarifaire est la prime fixée par la compagnie d'assurance lors du calcul du tarif pour différents risques. Elle constitue la base utilisée pour la prise en considération de critères tarifaires supplémentaires (majorations et rabais, par ex.), comme dans le système Bonus-malus par exemple.

**La segmentation** : Elle désigne le fait de découper un portefeuille en plusieurs sous-ensembles homogènes et distincts composés d'individus ayant des comportements communs et ce selon différents critères.

**Les critères de segmentation** : Sont les caractéristiques choisies pour effectuer l'étude de segmentation. Elles doivent permettre d'obtenir des segments de population homogènes, de taille suffisante et opérationnelle.

**Tarifification à priori** : Elle repose sur la segmentation du portefeuille. C'est donc une tarification basée sur les caractéristiques de l'assuré (l'âge, le sexe...) et de son véhicule (l'ancienneté du véhicule, l'ancienneté du permis...). Elle consiste à créer des classes de risque au sein desquelles les assurés paient la même prime.

**Tarifification a posteriori** : Elle est basée essentiellement sur le nombre de sinistres déclarés par l'assuré et dépendant de la segmentation à priori pratiquée. Il s'agit donc d'une tarification corrective appliquée sur la prime de base correspondante à chaque classe de risque.

**Système Bonus-malus** : Système par lequel les assurés se voient accorder des réductions ou des majorations en fonction des sinistres qu'ils causent. Ce coefficient désigne le système de réduction ou de majoration des tarifs accordés aux assurés en fonction des accidents responsables. Il minore ou majore leur prime d'assurance de base.

**La théorie de la crédibilité** : La théorie de la crédibilité, comme l'une des formes de la tarification a posteriori, repose sur les principes de l'inférence bayésienne et a notamment pour objectif la mise à jour des primes pures à partir de l'historique des sinistres des contrats.

**Avenant** : Modification apportée au contrat, justifiée par un changement de nature du risque ou d'un changement des garanties souscrites. Un avenant peut générer soit une prime additionnelle (comptant) soit un remboursement (ristourne) soit aucun changement de prime (comptant nul).

## Annexe VI : Extrait de la base PROD

Num_Pol	MARQUE	IMMATRIC	PUISS_FISC	DATE_MEC	VILLE_ZONE CIRCULATION	CRM	Date de Naissance	DATE de PERMIS	Situation Matrimoniale	Region
AU07309891N	MERCEDES	565-A-44	8	30/11/1979	Tétouan	100	12/12/1965	31/10/1991	Célibataire	Tanger - Tétouan
AU06701582N	RENAULT	11494-A-44	7	01/02/2002	Ouezzane	100	01/01/1978	02/01/2005	Célibataire	Tanger - Tétouan
AU06461325X	MERCEDES	1913-A-44	10	01/03/1984	Tétouan	100	18/02/1976	27/06/2002	Marifé(e)	Tanger - Tétouan
AU06449577N	VOLKSWAGE	31145-A-44	8	26/04/1999	Chefchaoue	100	01/01/1986	02/01/2005	Célibataire	Tanger - Tétouan
AU06461325X	MERCEDES	1913-A-44	10	01/03/1984	Tétouan	100	18/02/1976	27/06/2002	Marifé(e)	Tanger - Tétouan
AU08959143Z	MERCEDES	1913-A-44	10	01/03/1984	Tétouan	100	18/02/1976	27/06/2002	Célibataire	Tanger - Tétouan
AU08959143Z	MERCEDES	1913-A-44	10	01/03/1984	Tétouan	100	18/02/1976	27/06/2002	Célibataire	Tanger - Tétouan
AU08959143Z	MERCEDES	1913-A-44	10	01/03/1984	Tétouan	100	18/02/1976	27/06/2002	Célibataire	Tanger - Tétouan
AU08129787L	.	30461-A-44	7	25/01/2008	Tétouan	100	01/01/1978	09/08/2001	Célibataire	Tanger - Tétouan
AU08067377R	VOLKSWAGE	30184-A-44	7	24/01/2008	Tétouan	100	11/05/1980	23/08/2001	Célibataire	Tanger - Tétouan
AU08354605L	MERCEDES	8998-A-44	8	10/02/1978	Tanger	100	01/02/1971	01/02/2008	Célibataire	Tanger - Tétouan
AU07839736L	MERCEDES	37608-A-44	9	07/06/2007	Martil	100	18/11/1972	27/12/1990	Célibataire	Tanger - Tétouan
AU07789599S	PEUGEOT	27722-A-44	6	14/05/2007	Tétouan	90	01/01/1946	30/06/1976	Marifé(e)	Tanger - Tétouan
AU07768157R	.	36183-A-44	6	10/02/2010	Tétouan	100	01/01/1969	02/01/2002	Célibataire	Tanger - Tétouan
AU10714736M	MAZDA	40252-A-44	11	05/01/2010	Tétouan	100	01/01/1950	02/01/1976	Marifé(e)	Tanger - Tétouan
AU10599226R	MERCEDES	8998-A-44	8	10/02/1978	Tanger	100	01/01/1977	07/11/2002	Célibataire	Tanger - Tétouan
AU10337716L	CHERY	37060-A-44	11	12/02/2009	Tétouan	100	08/05/1978	19/08/1999	Célibataire	Tanger - Tétouan
AU10166014R	DACIA	31199-A-44	8	11/07/2008	Tétouan	100	26/03/1981	02/01/2008	Célibataire	Tanger - Tétouan
AU10060111R	VOLKSWAGE	26232-A-44	8	09/06/2006	Chefchaoue	100	01/03/1985	01/11/2010	.	Tanger - Tétouan
AU10072347P	MITSUBISH	24069-A-44	13	01/09/2003	Tétouan	100	10/05/1973	01/08/2002	Célibataire	Tanger - Tétouan
AU09343738M	FORD	22420-A-44	10	21/11/2001	Tanger	100	01/01/1975	29/06/2000	Célibataire	Tanger - Tétouan

## Annexe VII : Extrait de la base Sinistre

Numero Sinistre	Numero Contrat	Nature Sinistre	Date survenance	Lieu survenance	Description Garantie	Resp	charge
2007YX06540	2007231698197F	Corporel	16/06/2007	93000 Tétouan	RC - Dommages corporels causés aux tiers	50	185118,06
2008YX14640	ES00060225S	Matériel	14/12/2008	93000 Tétouan	Dommages au véhicule lors d'un choc avec tiers	100	23541,78
AU200805-02516	AU00207251N	Corporel	03/11/2007	93000 Tétouan	RC - Dommages corporels causés aux tiers	66	12663,91
AU200805-08567	AU01059700Y	Matériel	31/05/2008	93000 Tétouan	Dommages au véhicule lors d'un choc sans tiers	0	0
AU200805-11553	AU00211461M	Matériel	05/08/2008	93000 Tétouan	RC - Dommages matériels causés au tiers	0	720
AU200905-19000	AU01931962Y	Matériel	08/12/2008	93000 Tétouan	Dommages au véhicule lors d'un choc sans tiers	0	32642,3
AU200905-28716	AU01757844M	Matériel	12/07/2009	93000 Tétouan	RC - Dommages matériels causés au tiers	100	6000
2007YX15983	2006231837892F	Corporel	07/11/2007	93000 Tétouan	RC - Dommages matériels causés au tiers	66	0
AU200805-07085	AU00457043N	Matériel	08/05/2008	93000 Tétouan	Collision	0	5704,12
AU201005-85585	AU02758150X	Corporel	06/09/2010	93000 Tétouan	RC - Dommages corporels causés aux tiers	100	5000
AU200905-27411	AU00500528L	Matériel	15/06/2009	93000 Tétouan	Dommages au véhicule lors d'un choc sans tiers	.	8302,28
AU200905-22762	424352070	Matériel	06/03/2009	93000 Tétouan	Bris de glaces du véhicule	0	900
AU201005-78975	AU04151212R	Corporel	31/08/2010	93000 Tétouan	RC - Dommages corporels causés aux tiers	100	34400
AU200805-12807	AU00017178Z	Matériel	05/09/2008	93000 Tétouan	Dommages au véhicule lors d'un choc sans tiers	0	14908,93
AU200905-26228	424308245	Matériel	16/05/2009	93000 Tétouan	Bris de glaces du véhicule	0	5369,51
2007YX15989	2007231012648F	Corporel	22/11/2007	93000 Tétouan	RC - Dommages corporels causés aux tiers	66	76585,31
AU200905-20159	424034960	Corporel	04/07/2008	93000 Tétouan	RC - Dommages corporels causés aux tiers	0	0
AU200905-19699	AU01925919Y	Matériel	12/12/2008	93000 Tétouan	Bris de glaces du véhicule	0	2432
AU200905-23719	AU01879884M	Matériel	06/03/2009	93000 Tétouan	Dommages au véhicule lors d'un choc avec tiers	50	1122,38
AU200805-05296	AU00124478R	Corporel	25/08/2007	93000 Tétouan	RC - Dommages corporels causés aux tiers	66	18381,29
2007YX08105	2006231671958F	Matériel	25/06/2007	93000 Tétouan	Bris de glaces du véhicule	0	4498,03
AU200905-21200	424360809	Matériel	08/02/2009	93000 Tétouan	Dommages au véhicule lors d'un choc sans tiers	0	5556,96