

ROYAUME DU MAROC
*_**_*_*_*
HAUT COMMISSARIAT AU PLAN
*_**_*_*_*_*_*_*

INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE

INSEA



Projet de Fin d'Etudes

**Remises sur les garanties annexes en se basant sur la RC
Automobile.**

Préparé par

:

Mlle Layla ABOUKHERRAZ

Mlle Halima BENDARAG

Sous la direction de

:

M. Abderahim OULIDI (INSEA)
M. Achraf FTOUHI (Sanad 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. Abderahim OULIDI (INSEA)
- M.S.R NCIRI (INSEA)
- M. Achraf FTOUHI (Sanad Assurance)

Juillet 2011

RÉSUMÉ

Le but de ce projet de fin d'études est de proposer une tarification a **posteriori** correspondant au profil de chaque assuré dans le portefeuille automobile à usage tourisme de **SANAD**, Compagnie Marocaine d'Assurance, afin d'affecter des remises sur les garanties annexes pour les bons conducteurs. Cette tarification viendra dans un esprit de continuité de la recherche, corriger un tarif a **priori** effectué sur ce portefeuille. Pour accomplir cette tâche nous procéderons en deux étapes principales :

- La première étape consiste à élaborer un tarif technique, pour un portefeuille **RC-automobile** catégorie véhicules touristes. A cet effet, il a fallu, tout d'abord, constituer les différentes classes tarifaires en discrétisant et codifiant certaines variables qui étaient originellement continues, selon les résultats fournis par la **segmentation** du portefeuille d'étude. Après cela, nous avons adopté l'approche qui utilise les modèles **GLM** pour la **modélisation** de la fréquence des sinistres corporels et matériels ainsi que le montant des sinistres corporels et matériels. Par la suite, on a accompagné chaque modèle par une série de tests d'adéquation, de validation, ainsi par d'autres tests pour examiner la fiabilité des résultats obtenus.
- La deuxième étape consiste à proposer un système de personnalisation de la prime a posteriori relative à la garantie responsabilité civile en Automobile. En effet, la libéralisation de cette garantie en juillet 2006, a incité chaque compagnie d'assurance à mettre en place un système **Bonus-malus** adapté à leur portefeuille. Il s'agit d'augmenter la prime des assurés qui mettent des sinistres à la charge du groupe, et d'offrir **d'autant** une réduction sur les garanties annexes des bons conducteurs, sans toucher à l'encaissement global de la compagnie. Pour parvenir à ce but, nous avons justifié l'instauration d'un système Bonus-malus comme critère de tarification à posteriori, et ceci, en prouvant que l'hypothèse d'égalité des assurés devant le risque ne résiste pas à l'analyse statistique. Tester ensuite l'adéquation du modèle hétérogène sur notre portefeuille, à savoir : le modèle Binomial Négatif pour élaborer enfin les échelles Bonus Malus.

MOTS CLÉ

Assurance RC automobile, tarification a priori, segmentation, modèles linéaires généralisés, tarification a posteriori, Statistique Bayésienne, échelle Bonus Malus.

DÉDICACE

*À nos très chers parents, pour leurs soutiens et tous leurs
sacrifices*

À nos frères et sœurs pour leur amour et leur soutien

À toutes nos familles

À nos professeurs

À tous nos amis

À ceux qui nous ont soutenus tout au long de notre vie

Nous dédions ce modeste travail

Layla & Halima



REMERCIEMENT

Nous ne saurions commencer ce rapport sans remercier le tout puissant **ALLAH** miséricordieux, qui nous a permis d'effectuer notre stage au sein de Sanad Assurance dans de bonnes conditions et sans difficultés majeures.

Aussi nous remercions du fond du coeur **M. Abderahim OULIDI** notre encadrant interne de l'INSEA qui n'a ménagé aucun effort pour nous aider à mener à terme notre projet, nous remercions également **M. Achraf FTOUHI** notre encadrant externe du service actuariat de Sanad pour ces enseignements qui témoignent d'une grande expérience.

Nos remerciements s'en vont également à **Mlle. Loubna ELGHAZOUANI** et **M. Noureddine EL KABABI** pour la collaboration inestimable qu'ils nous ont accordée tout au long de notre période de stage.

Nous tenons à exprimer notre gratitude à notre Professeur **M. S.R. Nciri** qui nous a honorés en acceptant de juger notre travail. Veuillez trouver ici le témoignage de notre respect le plus profond.

Enfin, nous présentons notre grand respect au corps professoral de l'INSEA.

TABLE DES MATIÈRES

<u>RÉSUMÉ</u>	3
<u>MOTS CLÉ</u>	3
<u>DÉDICACE</u>	4
<u>REMERCIEMENT</u>	5
<u>TABLE DES MATIÈRES</u>	6
<u>LISTE DES ABRÉVIATIONS</u>	9
<u>LISTE DES TABLEAUX</u>	10
<u>LISTE DES FIGURES</u>	12
<u>INTRODUCTION GÉNÉRALE</u>	13
<u>PARTIE I : Tarification a priori</u>	15
<u>CHAPITRE I: LE MARCHE DE L'ASSURANCE AUMAROC</u>	17
<u>I.Description de la SANAD</u>	18
<u>II.Organisation et activité de la SANAD :</u>	18
<u>III.Problématique du marché de l'assurance automobile au Maroc :</u>	19
<u>IV.Principes de tarification et nécessité de segmenter :</u>	21
<u>V.Une innovation constante :</u>	23
<u>VI.Pré requis, limite et pertinence de la segmentation :</u>	24
<u>CHAPITRE II : ANALYSE DESCRIPTIVE ET PRESENTATION</u>	26
<u>I.Constitution de la base de données</u>	27
<u>II.Description et analyse des variables tarifaires :</u>	29
<u>II.1 La répartition par sexe :</u>	30
<u>II.2 La mesure de l'exposition au risque : la variable Véhicule année..</u>	31
<u>II.3 Présentation de l'algorithm CHAID :</u>	33
<u>II.4 Segmentation des variables :</u>	34
<u>II.4.1 Puissance véhicule et type moteur (PVM).</u>	34
<u>II.4.2 Ancienneté de véhicule:</u>	35
<u>II.4.3 L'âge et l'ancienneté du permis</u>	37
<u>II.4.4 Variable CRM (coefficient de réduction majoration).</u>	39
<u>II.4.5 La segmentation de la variable ville:</u>	40
<u>II.4 Discrétisation des variables</u>	44
<u>CHAPITRE III : MODELISATION DU NOMBRE DE SINISTRES CORPORELS ET MATERIELS</u>	47
<u>I .Le cadre général des modèles linéaires généralisés :</u>	48

I.1 Distributions discrètes :	48
I.2 Distributions continues :	49
II. Méthodes de sélection des variables explicatives	50
III. L'outil de base en RC Auto: la régression de Poisson	50
III.1 Description du modèle :	50
III.2 Estimation du maximum de vraisemblance	51
III.3 Illustration numérique :	52
III.4 Sur dispersion :	53
IV. GLM Binomiale négative	54
VI.1 Extension du modèle de Poisson pour palier au problème de sur dispersion :	54
V. Modèle de Régression logistique pour le nombre de sinistres:	56
V.1 Description du modèle :	56
V.2 Illustration numérique :	57
V.2.1 Modélisation de la fréquence matérielle	57
V.2.2 Modélisation de la fréquence corporelle	60
CHAPITRE VI : MODELISATION DU MONTANT DE SINISTRES CORPORELS ET MATERIELS	63
I. Modélisation du montant des sinistres corporels :	64
I.1. Descriptif et présentation :	64
I.2. L'écèlement des coûts :	65
I.3. Détermination du seuil d'écèlement :	65
I.4. Modélisation du montant de sinistres corporels :	67
I.5. Mesure d'influence :	72
II. Modélisation du montant des sinistres matériels :	73
III. Calcul de la prime commerciale :	77
PARTIE II : TARIFICATION A POSTERIORI	79
CHAPITRE I : Construction d'un système Bonus-Malus optimal	81
CHAPITRE II : Formulation des Modèles	87
I. Modèle Poissonien Portefeuille homogène :	86
II. Modèle binomial négatif Portefeuille hétérogène :	87
III. Modélisation par les processus Markoviens :	89
CHAPITRE III : Test d'adéquation des modèles homogènes et hétérogène	91
I. Le test χ^2 d'ajustement des lois :	92
II. Modèle poissonien de portefeuille homogène:	92
II.1 Sinistres corporels :	93

II.2 Sinistres matériels :	93
II.3 Sinistres RC :	94
III.Modèle Binomial négative de portefeuille hétérogène :	95
CHAPITRE IV : Elaboration des Echelles Bonus Malus	96
I .Construction d'un système bonus malus avec personnalisation à priori :	97
II.Par modèle Binomial négative :	97
III.Présentation Markovienne :	99
Application : Remises sur les garanties annexes en se basant sur la RC Auto.....	101
CONCLUSION GÉNÉRALE	108
ANNEXES	109
BIBLIOGRAPHIE	113

LISTE DES ABRÉVIATIONS

- GLM : modèles linéaires généralisés
- FMSAR : Fédération Marocaine des Sociétés d'Assurances et de Réassurance
- RC Auto : Responsabilité civile automobile
- CHAID : Chi-square Automatic Interaction Detected
- VBA : Visual Basic for Application
- DAPS : Direction des assurances et de prévoyance sociale

LISTE DES TABLEAUX

Tableau 1 : Tableau des statistiques simples	32
Tableau 2 : Tableau des corrélations entre les variables explicatives	33
Tableau 3 : Codage de la variable Puissance fiscale et combustible (PFC)	44
Tableau 4 : Codage de la variable Ancienneté de véhicule.....	45
Tableau 5 : Codage de la variable âge et l'ancienneté du permis	45
Tableau 6 : Codage de la variable coefficient de réduction majoration	45
Tableau 7 : Codage de la variable ville.....	46
Tableau 8 : Résultats de l'analyse de Type 1 pour le modèle avec les 5 variables explicatives.....	52
Tableau 9 : Résultats de la régression Poisson pour le nombre de sinistres matériels.	53
Tableau 10 : critère pour détecter la sur dispersion.....	54
Tableau 11 : Test de significativité des variables tarifaires (type 1).....	54
Tableau 12 : Résultat de la régression du modèle binomiale négative pour la fréquence corporelle .	55
Tableau 13 : Résultat de la régression du modèle binomiale négative pour la fréquence matérielle ..	56
Tableau 14: Résultats de la sélection séquentielle pour le nombre de sinistres matériels	58
Tableau 15 : Test de significativité des variables tarifaires (type 3)	59
Tableau 16 : Résultats de la régression LOGISTIQUE pour le nombre de sinistres matériels.	60
Tableau 17 : Résultats de la sélection séquentielle pour le nombre de sinistres corporels	60
Tableau 18 : Test de significativité des variables tarifaires (type 3).....	61
Tableau 19 : Résultats de la régression LOGISTIQUE pour le nombre de sinistres corporels.	62
Tableau 20 : Statistiques descriptives du montant des sinistres corporels	65
Tableau 21 : Les seuils d'écèlement.....	65
Tableau 22 : Les statistiques relatives à la déviance.....	67
Tableau 23 : Test de significativité des variables tarifaires (Type 1)	68
Tableau 24 : Test de significativité des variables tarifaires (Type 3)	68
Tableau 25 : Test de significativité des variables tarifaires (Type 1)	68
Tableau 26 : Test de significativité des variables tarifaires (Type 3)	69
Tableau 27 : Résultats de la régression Log-normal pour le montant de sinistres corporels	69
Tableau 28 : Test de significativité des variables tarifaires (Type 1)	70
Tableau 29 : Test de significativité des variables tarifaires (Type 3)	70
Tableau 30 : Test de significativité des variables tarifaires (Type 1)	70
Tableau 31 : Résultats de la régression Gamma pour le montant de sinistres corporels	71
Tableau 32 : La distance de Cook.....	72
Tableau 33 : Les coûts répétés ainsi que leurs poids dans l'échantillon.....	73
Tableau 34 : Statistiques descriptives du montant des sinistres matériels	73
Tableau 35 : Test de significativité des variables tarifaires (Type 1)	75
Tableau 36 : Test de significativité des variables tarifaires (Type 3)	75
Tableau 37 : Test de significativité des variables tarifaires (Type 1)	75
Tableau 38 : Test de significativité des variables tarifaires (Type 3)	76
Tableau 39 : Résultats de la régression Log-normal pour le montant de sinistres matériels.	76
Tableau 40 : La distance de Cook.....	76
Tableau 41 : sinistres corporels Poisson homogène.....	93

Tableau 42 : sinistres matériels Poisson homogène.....	94
Tableau 43 : tous les sinistres Poisson homogène	94
Tableau 44 : tous les sinistres BN hétérogène	95
Tableau 45 : estimation des paramètres du modèle Binomial Négative	98
Tableau 46 : Système Bonus malus obtenue par le modèle Binomiale Négative	99
Tableau 47 : formulation markovienne de l'échelle Binomiale Négative	100

LISTE DES FIGURES

Figure 1 : Organigramme de la SANAD	18
Figure 2 : Composition du chiffre d'affaire 2010 de la SANAD.....	19
Figure 3 : progression du marché d'assurances marocain.....	20
Figure 4 : classement des assureurs automobile en 2010 selon le chiffre d'affaires	21
Figure 5 : les catégories des véhicules existant dans notre portefeuille	29
Figure 6 : Distribution de la fréquence et du coût selon le sexe.....	30
Figure 7 : les périodes de couverture des polices	31
Figure 8 : la fréquence et le coût moyen de la nouvelle variable regroupée	35
Figure 9 : Arbre de décision de la variable âge véhicule	36
Figure 10 : la fréquence et le coût moyen corporel selon la variable âge du véhicule	37
Figure 12 : Arbre de décision de la variable âge conducteur	38
Figure 13 : La fréquence de la nouvelle variable âge-ancienneté permis	39
Figure 14 : la fréquence et le coût moyen de la variable Bonus.....	40
Figure 15 : diagramme arborescent segmentant la population par libellé de ville.....	41
Figure 16 : diagramme arborescent segmentant la population par code de ville.....	41
Figure 17 : la fréquence et le coût moyen de la variable région.....	44
Figure 19 : Ajustement Q-Q plot des montants des sinistres corporels écrêtés et leurs résidus	66
Figure 20 : histogramme des coûts matériels	74
Figure 21 : Ajustement Q-Q plot des montants des sinistres matériels écrêtés ainsi que leur résidu....	74
Figure 22 : Interface pour le calcul de la prime commerciale.....	77
Figure 23: La page d'accueil en VBA	102
Figure 24 : Interface de calcul de la prime commerciale	103
Figure 25 : le calcul de la prime commerciale	104
Figure 26 : détermination du type de client.....	105
Figure 27 : calcul de la remise pour un nouveau client	105
Figure 28 : calcul de la remise pour un client de Sanad.....	106
Figure 29 : Arbre de décision de toutes les variables explicatives	112

INTRODUCTION GÉNÉRALE

L'histoire de l'homme et de sa voiture a été l'une des plus grosses affaires du dernier siècle. La croissance exponentielle de la demande de véhicules à travers le monde, a constitué une considérable source de revenus pour les pays qui ont su prendre place parmi les grands producteurs. Seulement, cet accroissement de véhicule n'a fait qu'accroître des chiffres d'affaire. L'une des conséquences de cet accroissement a été aussi l'explosion de la mortalité par les accidents de la route. Le nombre de morts par accident de la route est estimé aujourd'hui à des millions de personnes par an à travers le monde. Cela met en ruine la couverture financière de beaucoup de familles car la possibilité de s'offrir une voiture n'est pas équivalente à la capacité de réparer tous les dommages que cette voiture est susceptible de causer. Les cas d'invalidités et d'orphelins sans couverture constituent une charge considérable pour les communautés sans compter les énormes dégâts matériels qui accompagnent ces phénomènes. Le Maroc n'a pas fait exception à cette réalité car son parc automobile a connu une croissance très forte pendant le dernier siècle constituant ainsi la plus grande part de marché en assurance. Parallèlement le taux de mortalité sur les routes du Maroc est devenu l'un des plus élevés au monde.

Les décideurs marocains ont donc rendu l'assurance de la responsabilité civile de tous les conducteurs obligatoire, dès 1969 pour faire face au problème. Cette forme d'assurance constitue la plus importante part en assurance automobile. Le caractère social que revêt un tel produit a poussé l'Etat à imposer des tarifs réglementés à l'abri des fluctuations que peuvent connaître les primes sur le marché. Les autres types de risques assurables en assurance automobile étaient régis par le même mode de tarification mais sont tous libéralisés de nos jours.

Dans les pays développés, les exigences de la libéralisation des marchés ont pris le dessus sur les justifications sociales, depuis les années quatre-vingt-dix, et des techniques ont été mises en place pour différencier les primes de manière considérablement poussée pour répondre aux exigences de la poussée du consumérisme dans ces sociétés. Le Maroc fera autant car l'imminence d'une libéralisation de l'assurance de la responsabilité civile tend plus vers une réalité qu'un mythe.

La libéralisation implique la liberté des prix. Ces prix peuvent être à la base de performantes stratégies commerciales mais peuvent être aussi à la base de la ruine du portefeuille, voir sa faillite s'ils sont mal évalués. Le débat sur la tarification est donc plus que jamais à l'ordre du jour et le présent travail traite d'une des multiples formes de cette tarification.

Le principe fondamental de l'assurance est de former des portefeuilles dans lesquels les assurés « diluent » leurs risques. Si un assuré apporte au portefeuille un risque considérablement différent de ceux des autres assurés, il est normal de lui proposer une prime

considérablement différente. Mais très souvent des caractéristiques très importantes du risque que nous apporte un nouvel assuré sont inobservables et ne peuvent pas être mesurés pour être introduits dans un tarif. Dans de pareilles situations une des procédures adéquates à adopter consiste à actualiser le profil de risque de l'assuré au fil du temps en tenant compte de sa sinistralité. Cette prise en compte ne doit pas pénaliser les sinistres causés par le simple hasard au risque d'enfreindre à la règle d'or de l'assurance. Elle ne doit non plus trop récompenser les non sinistrés pour ne pas voir le système s'affaiblir à cause de la « soif de récompense » qui a déjà mis en mal plusieurs systèmes à travers le monde.

Le but de ce projet de fin d'étude est d'élaborer un système de tarification personnalisée répondant aux différents profils de risque en coexistence dans le portefeuille Automobile à usage tourisme de la compagnie d'assurance **Sanad**. Pour parvenir à ce but, après avoir présenté succinctement la société d'accueil et parlé de certaines réalités du marché de l'assurance automobile au Maroc, nous avons élaboré un tarif technique pour un portefeuille RC-automobile où nous avons décrit les différentes variables tarifaires afin de les segmenter et par suite les codifier. Après codification, nous avons adopté l'approche qui utilise les modèles GLM pour la modélisation de la fréquence et le coût des sinistres corporels et matériels. Grâce à cette modélisation nous avons pu évaluer la prime pure pour chaque assuré. Cette prime, dite prime de base, sera corrigée par un système de personnalisation des primes a posteriori, ou, système Bonus-malus qui va tenir compte des facteurs difficilement observables. Ce système était basé sur l'approche bayésienne pour la formulation des deux modèles (Poisson et Binomiale Négative) que nous avons utilisés pour l'élaboration de l'échelle Bonus-Malus qui nous a permis ensuite de calculer les réductions sur les garanties annexes.

PARTIE I

**Tarification a priori : Segmentation et modélisation du
Tarif RC AUTO pour les véhicules de tourisme (Usage A)**

Introduction

La tarification est une problématique récurrente dans toute compagnie d'assurance. En effet, on ne peut appliquer un tarif uniforme à toute la population des assurés. Ainsi pour une prime donnée, seules les personnes qui présentent un niveau de risque identique ou supérieur à celui qui est associé au niveau de prime proposé vont souscrire au contrat. Cela peut être fatal pour un assureur ; si la prime est trop élevée, les bons risques finissent par sortir du contrat. L'assureur doit alors augmenter la prime en courant le risque d'une spirale d'éviction des contrats.

Dans le cas de la RC automobile, le phénomène est encore plus palpable. Du fait que la **RC-AUTO** est une assurance obligatoire, il est important de segmenter le portefeuille pour pouvoir attirer les bons clients et décourager les mauvais. La première étape de l'élaboration du tarif est donc la segmentation du portefeuille. Celle-ci se fait par le biais de variables jugées fortement corrélées avec la sinistralité. Une fois cette étape réalisée, on peut construire des modèles statistiques reliant les variables tarifaires aux variables de sinistralité.

La sinistralité en **RC-AUTO** est caractérisée par deux principales composantes : les sinistres matériels et les sinistres corporels. Il s'agira pour nous d'établir des modèles sur les deux composantes de ces risques - la fréquence d'occurrence et le coût moyen- afin de calculer la prime pure du tarif. On a opté pour un modèle collectif qui consiste à modéliser la charge engendrée par un assuré durant sa période de couverture, par la somme des coûts de sinistre occasionnés. Ce qui a nécessité l'établissement d'un modèle pour chacune des composantes aléatoires; le nombre de sinistres N et le coût d'un sinistre C .

Pour la fréquence des sinistres corporels nous avons adopté un modèle de « **GLM-Binomiale Négatif-lien identité** ». Pour la fréquence des sinistres matériels nous avons adopté un modèle de « **Régression logistique** ». Pour ce qui est du coût de sinistre corporel et matériel, nous avons utilisé deux modèles « **GLM-Normale-lien log** » et « **GLM-Gamma-lien identité** ».

Finalement on obtient la prime en combinant les modèles cités en haut. Mais avant toute chose présentons l'entreprise où ce travail a été réalisé.



L'objectif de cette partie est de donner un aperçu global sur la société d'accueil, le marché de l'assurance automobile au Maroc ainsi que les principes à la base de la tarification et la segmentation des risques comme un primordial à ne pas perdre de vue.

I. Description de la SANAD

La Compagnie Marocaine d'Assurance et de Réassurance (SANAD), Société Anonyme au Capital de 250 Millions de Dirhams, fondée en 1946, est une filiale HOLMARCOS Groupe Marocain multisectoriel d'envergure, et la Caisse de dépôt et de Gestion - CDG -, premier Groupe financier du royaume.

II. Organisation et activité de la SANAD :

La SANAD est constituée de plusieurs départements et services qui sont liés les uns aux autres pour la réussite et l'expansion de la compagnie.

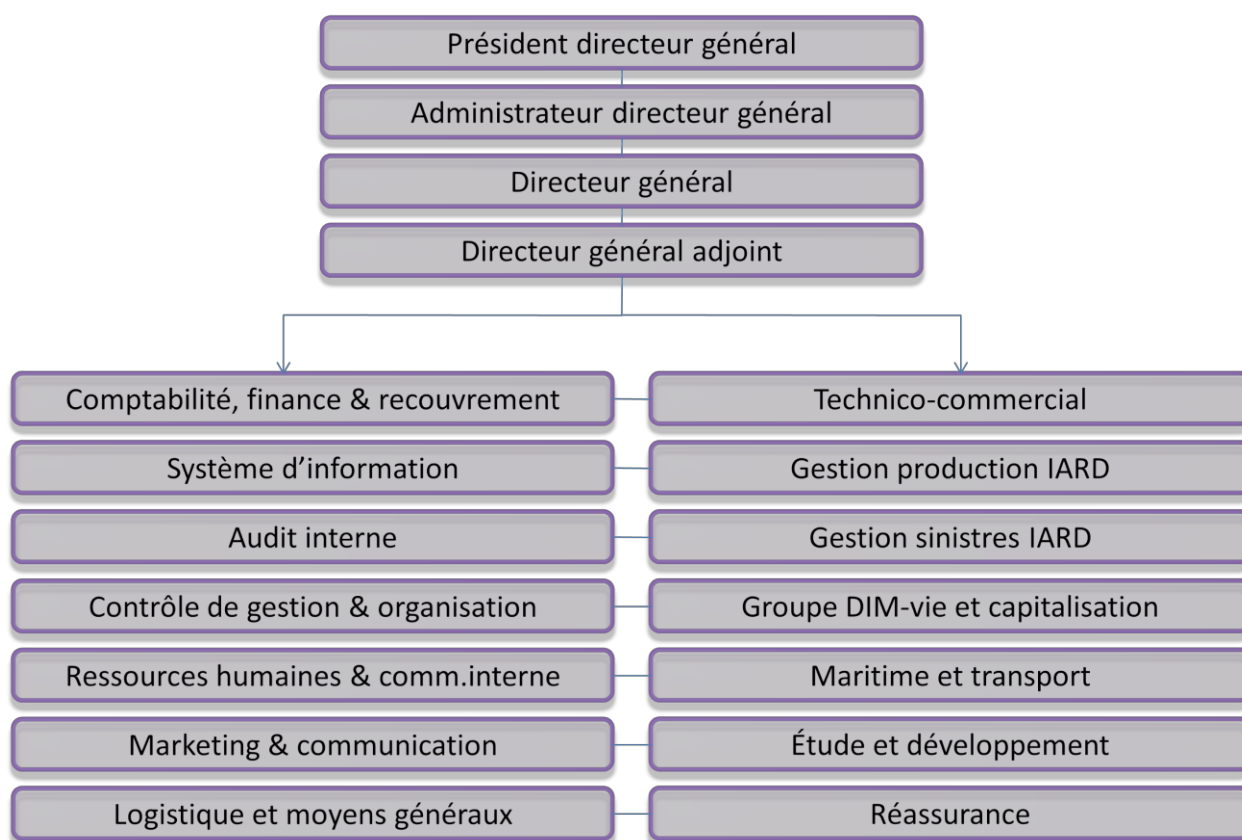


Figure 1 : Organigramme de la SANAD

La SANAD a réalisé en 2010, un chiffre d'affaires de 1.217.912.000 dirhams possédant 5.6% de part du marché et réalisant un taux de progression d'environ 2% par rapport à l'exercice de l'année précédente.

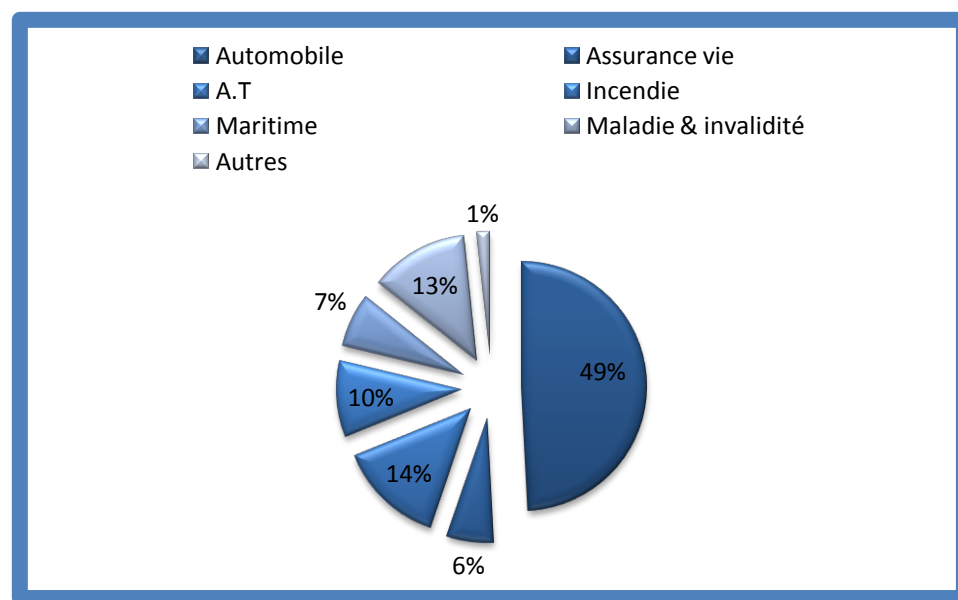


Figure 2 : Composition du chiffre d'affaire 2010 de la SANAD

Le diagramme nous montre que le chiffre d'Affaire de la SANAD, pour l'exercice 2010, révèle la prédominance de la branche automobile qui monopolise près de 49% du chiffre d'affaires. Ensuite, vient en seconde position les produits d'assurance de groupe vie (19%) dont l'assurance maladie fait partie, suivi des branches accident de travail (14%) et incendie (10%). Quant aux autres branches non vie, ils se départagent les 8% restant du chiffre d'affaire avec 1% pour autres.

III. Problématique du marché de l'assurance automobile au Maroc :

Le marché de l'assurance automobile au Maroc est l'un des plus importants, comme en témoigne le rapport annuel 2010 de la FMSAR (Fédération Marocaine des Sociétés d'Assurances et de Réassurance). En effet, il est le marché le plus développé de l'assurance Non Vie, le chiffre d'affaire en 2010 est composé à 32,3 % par l'assurance automobile. L'importance de cette branche d'assurance s'explique en grande partie par l'obligation d'assurance automobile minimale en responsabilité civile dite au tiers.

Structure du Chiffre d'Affaires		
	Chiffre d'Affaires	Part Marché
Automobile	7 075,8	32,3%
Assurances Vie et Capitalisation	6 590,9	30,1%
Accidents Corporels	2 726,8	12,5%
Accidents du Travail	1 894,3	8,7%
Incendie	1 032,6	4,7%
Transport	706,3	3,2%
Assistance - Crédit - Caution	699,3	3,2%
Responsabilité Civile Générale	458,2	2,1%
Risques Techniques	304,4	1,4%
Acceptations en réassurance	194,3	0,9%
Autres Opérations Non Vie	189,8	0,9%
Total	21 872,8	100%

En millions de dirhams

Source : Fédération Marocaine des Sociétés d'Assurance et de Réassurance-Mars 2011

En termes de progression, le marché ne cesse de grandir d'une année à une autre dont l'évolution la plus notable est à mettre au compte des produits d'assurance non vie. Le graphique ci-dessous nous montre l'évolution du chiffre d'affaire des branches vie et non-vie entre 2008 et 2010.

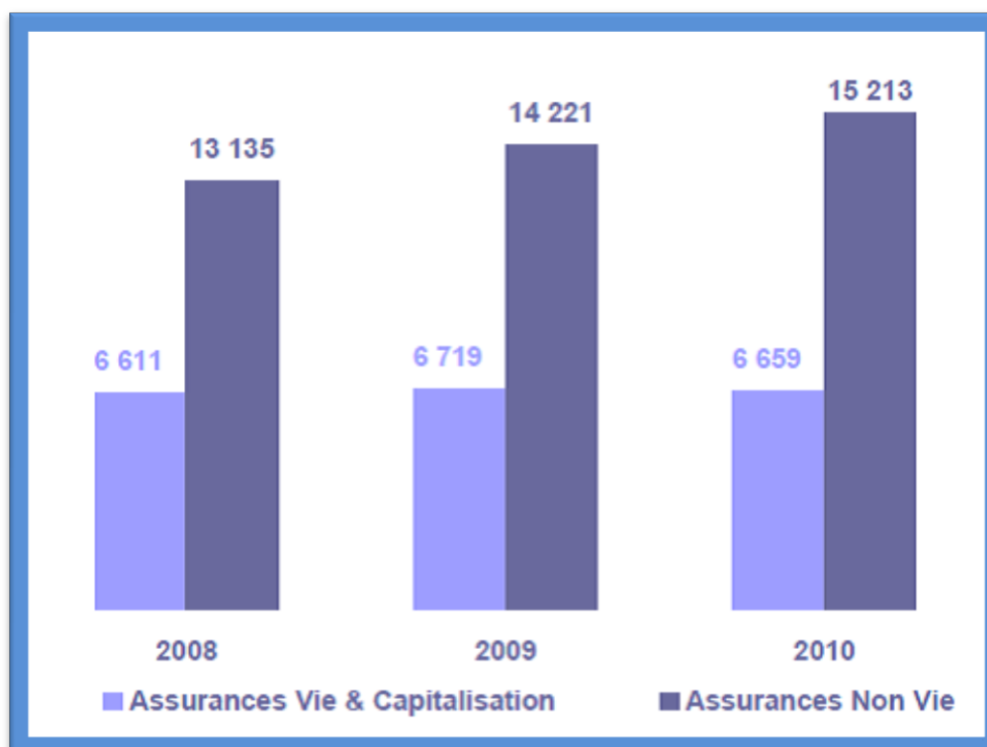


Figure 3 : progression du marché d'assurances marocain

Ce marché d'assurances de biens est certes le plus important, mais il est également le plus concurrentiel. De nombreux acteurs se disputent sans cesse les parts de marchés. Le marché de l'assurance est aujourd'hui qualifié d'hyperconcurrentiel. Les principaux acteurs du marché de l'assurance automobile, représentés ci-dessous dans le classement des assureurs automobile en 2010 (Fédération Marocaine des Sociétés d'Assurances et de Réassurance), doivent ainsi rivaliser d'une part en termes de prix, et d'autre part en termes d'inventivité.

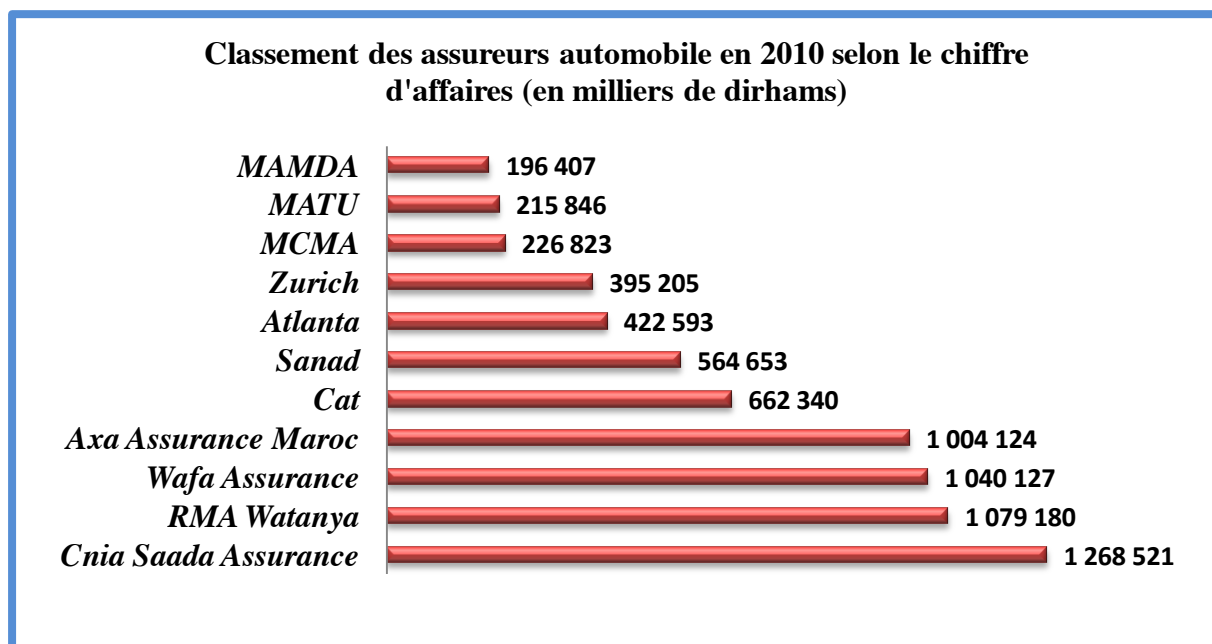


Figure 4 : classement des assureurs automobile en 2010 selon le chiffre d'affaires

Les acteurs habituels de ce marché que sont les sociétés d'assurances et les mutuelles doivent de plus faire face à l'arrivée sur le marché des bancassurances. Ces dernières peuvent utiliser un portefeuille de clients déjà important et source d'informations appropriées afin d'attirer de nombreux clients à partir d'offres intéressantes et de bas prix permis par un réseau d'agence important et l'absence d'intermédiaire, quand d'autres utilisent les services d'agents généraux et de courtiers.

IV. Principes de tarification et nécessité de segmenter :

Dans ce cadre, la législation marocaine impose la prise en compte par les assureurs d'un coefficient de réduction majoration (CRM) ou coefficient bonus malus. La mise en application n'est qu'à ses débuts, l'instauration du malus a été en 1995 et du bonus en 1998. La clause de « réduction-majoration » institue un système de diminution ou d'augmentation de la cotisation en fonction des accidents survenus.

- L'automobiliste qui ne cause pas d'accidents bénéficie d'un bonus : sa prime est réduite.
- L'automobiliste responsable d'un accident est pénalisé d'un malus : sa prime est majorée.

Lors de la première année d'assurance, le CRM est fixé à 1 (aucune modification de la prime d'assurance de base fixée par l'assureur).

Le CRM est transférable d'un assureur à un autre, il est propre à l'assuré en cas de changement d'assureur ou de changement de véhicule, mais il est assigné au véhicule durant la vie du contrat.

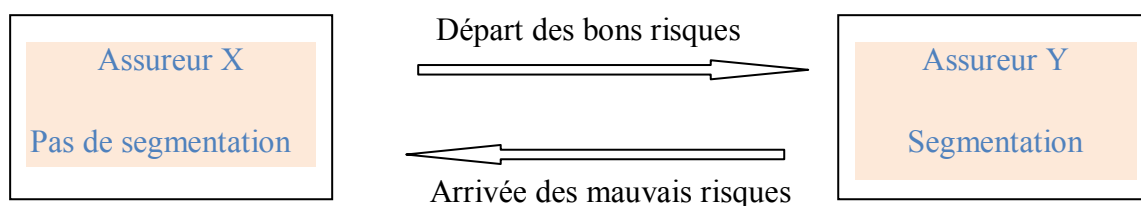
Les coefficients de réduction- majorations fixés par la réglementation sont comme suit :

- Un coefficient de 0,9 : si l'assuré n'a causé aucun sinistre engageant ou susceptible d'engager totalement ou partiellement sa responsabilité durant une période d'assurance de 24 mois consécutifs précédant la souscription ou le renouvellement du contrat. Il est toléré une seule interruption d'assurance ne dépassant pas 30 jours.
- Si l'assuré a causé un ou plusieurs sinistres engageant ou susceptible d'engager totalement ou partiellement sa responsabilité durant une période d'assurance de 12 mois précédant la souscription ou le renouvellement des contrat, ce coefficient, qui ne peut excéder 2,5, s'obtient en ajoutant à un pour chacun de ces sinistres pour l'usage «A» :
 - 20% si le sinistre est matériel.
 - 30% si le sinistre est corporel.

Au-delà de ce mécanisme de bonus malus, chaque assureur possède une tarification et une segmentation qui lui est propre, en adéquation avec le profil de son portefeuille d'assurés. Dans un contexte de marché très concurrentiel, la segmentation des risques est une nécessité. Celle-ci consiste à différencier les assurés et le risque qu'ils portent. On obtient ainsi différentes catégories de risques en fonction des caractéristiques de l'assuré et des garanties consenties par l'assureur. Chaque catégorie se verra ainsi attribuer un tarif qui lui sera propre, en adéquation avec le risque associé.

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. La segmentation permettra de prendre des mesures techniques à chacun des niveaux de segmentation, tant en tarification, qu'en souscription.

Prenons un exemple simple afin de comprendre la nécessité de segmenter, tout particulièrement en assurance automobile. L'assureur X décide de n'effectuer aucune segmentation. Il fait donc un profit avec les clients à faible probabilité de sinistre et une perte avec les assurés à forte probabilité de sinistre. L'assureur Y quant à lui, décide de faire payer plus cher les conducteurs à risques, déterminés à partir de variables telles que l'âge ou les antécédents de sinistres. En contrepartie, l'assureur Y diminuera la prime demandée aux assurés à risque faible. Suite à cette segmentation, les assurés prudents, ayant une faible probabilité d'avoir un sinistre et que l'on qualifiera de bons risques, décideront de s'assurer auprès de l'assureur Y, proposant des primes plus faibles. L'assureur X qui propose quant à lui un tarif unique, verra venir s'assurer chez lui la plupart des personnes ayant un comportement à risque, et donc une forte probabilité de sinistre. En effet, le tarif qui sera proposé à ces individus qualifiés de mauvais risques sera bien plus intéressant que celui de l'assureur Y. En conclusion, l'assureur X devra faire face à un risque de sinistre très élevé qu'il ne pourra pas absorber étant donné les primes demandées aux assurés. L'assureur X ne fera alors que des pertes, tandis que l'assureur Y ne fera que du profit.



Cet exemple simplifié exprime bien la nécessité pour l'assureur de segmenter. On peut également s'apercevoir à partir de cet exemple que la segmentation, si elle diminue la prime des conducteurs prudents, augmentera également sensiblement la prime des conducteurs à risque. Ce constat se révèle particulièrement problématique en ce qui concerne l'assurance automobile. En effet, les jeunes conducteurs, qui ne bénéficient pas d'ancienneté d'assurance, sont considérés comme une catégorie à risque et doivent souvent faire face soit au refus des assureurs, soit à des primes importantes. Conduire un véhicule peut donc devenir extrêmement coûteux pour un nouveau titulaire du permis de conduire.

On comprend à travers ces constats que la tarification, bien que nécessaire pour l'assureur, soulève un certain nombre de problématiques, et particulièrement en ce qui concerne la couverture des jeunes conducteurs. Il convient donc de rester prudent et de ne pas segmenter à l'extrême sans réflexion préalable, auquel cas l'assureur prendrait le risque de ne pas renouveler ses assurés, qui est bien souvent vieillissant dans les grandes compagnies d'assurance

V. Une innovation constante :

Le marché de l'assurance automobile affiche continuellement des progressions intéressantes en matière d'innovation, de nouvelles garanties telles que l'assistance ou la protection juridique ont apparues sur le marché. Ce marché qui a connu une libéralisation des tarifs de l'assurance responsabilité civile (RC) auto depuis juillet 2006, n'a pas assisté à une guerre des prix dans cette branche. Les compagnies d'assurances préfèrent livrer bataille sur les garanties annexes et non sur le niveau de la prime. La libéralisation s'est traduite par «une amélioration du contenu», soit le service complémentaire et la rapidité dans le règlement des sinistres. Quant au niveau des primes, il est presque le même partout. Il se base sur le tarif moyen calculé par la Direction des assurances et de la prévoyance sociale (DAPS). Pourtant, depuis un quatre ans, elles sont autorisées à calculer leurs propres tarifs, mais, pour cela, elles ont besoin de constituer une base de données. Chose qui demande du temps. Les assureurs hésitent à se lancer de craintes d'essuyer des déficits surtout que le marché est jugé étroit. Ce qui veut dire qu'aucune compagnie d'assurances ne peut établir un tarif technique fiable sur la base de son propre portefeuille. La branche automobile étant bénéficiaire, elles continuent donc à se baser sur le tarif moyen de la DAPS, lequel est calculé à titre indicatif. Les opérateurs s'accordent donc un délai de quelques années avant de passer à une véritable concurrence. Dans ce cas, le risque de dérapage sera contenu. Le Code des Assurances a mis en place des garde-fous en posant des conditions pour sauvegarder les intérêts des assurés et des assureurs. Tout changement tarifaire devrait être soumis à la tutelle qui garde un œil sur la solvabilité des entreprises. Les critères des tarifications sont également maintenus. Il s'agit en

l'occurrence de la puissance fiscale, du type de combustible et du mode d'utilisation du véhicule: transport de marchandises, tourisme ou de voyageurs.

La grande majorité des assureurs a profité de cette occasion pour personnaliser son offre à travers tout un panel de garanties d'assistance en partenariat avec des professionnels de ce métier. La protection juridique de l'assuré, est aujourd'hui une garantie qui ne surprend plus personne, de même que le remplacement du véhicule en cas de sinistre, autant de garanties annexes difficilement chiffrables par l'assuré, mais qui rassurent.

VI. Pré requis, limite et pertinence de la segmentation :

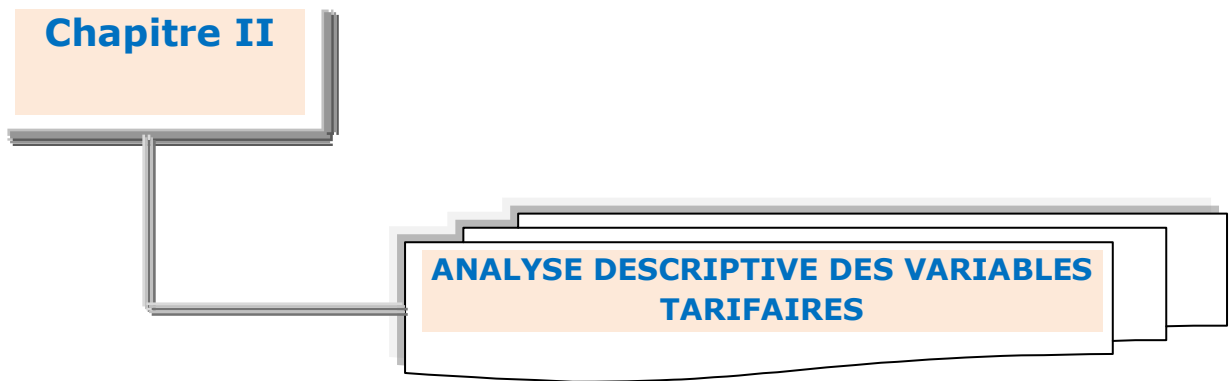
La mise en place ou la révision d'une segmentation, aussi méthodique et précise soit elle, est soumise d'une part à d'éventuels manques d'informations, et d'autres part à des enjeux marketing qu'il n'est pas toujours possible de modéliser mathématiquement parlant. De ce fait, la modélisation d'un risque tel que le risque automobile ne saurait se faire sans pré requis et intuitions préalables.

Il est tout d'abord nécessaire de prendre connaissance et de comprendre une tarification et donc une segmentation préexistante. Cette première étape primordiale permettra par la suite de pouvoir interpréter rapidement et clairement des résultats qui viendront alors s'inscrire dans une certaine logique. On pourra alors parler de logique actuarielle, mais également de logique d'entreprise ou d'enjeux marketing. En effet, si une tarification s'appuie essentiellement sur le principe de 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 entreprise possède 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 pourrait 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. Cette notion importante ne doit pas être perdue de vue lors de la modélisation ou tout du moins lors de l'interprétation de la segmentation ou des évolutions à mettre en place.

La modélisation, de même que l'analyse de donnée qui lui est préalable, connaît un certain nombre de limites. En effet, la qualité et la quantité des données utilisées doivent être contrôlées. A titre d'exemple, un portefeuille contenant une très faible quantité de véhicules d'une marque X auxquels sont associés une charge de sinistre très élevée ne signifiera pas que la marque X est à exclure du portefeuille. La cause de cette charge trop importante peut être liée à un unique véhicule, ou encore à la présence d'un sinistre grave ayant une probabilité d'occurrence très faible. On comprend donc que la modélisation d'un risque doit pouvoir s'asseoir sur une quantité suffisante de données afin d'être pertinente. La qualité des données est également primordiale. Une modélisation ne peut être correcte si elle s'appuie sur des données erronées. Toute modélisation doit donc faire l'objet d'une extraction de données rigoureuse ainsi que d'une épuration de la base de données utilisée. Enfin, une autre limite de la modélisation reste l'absence de données, ou tout du moins l'existence de données non accessibles. On évoquera tout d'abord les données non connues, soit par manque d'information, soit par la non collecte de ces informations qui occasionnerait un coût jugé trop élevé, soit enfin par l'impossibilité matérielle d'acquérir de telles données. L'absence de données peut également être la conséquence d'une franchise ou d'un écrêtage, voir même de l'utilisation de recours en assurance. Pour ce qui est de la franchise, il s'agit d'un montant, pas toujours fixe, qui reste à la charge de l'assuré en cas de sinistre. Dans ce cadre et plus

particulièrement dans l'assurance automobile, on comprend bien la volonté de l'assuré de ne pas déclarer d'éventuels sinistres dont le coût serait inférieur à la franchise. En effet, la déclaration d'un tel sinistre non seulement ne leur rapporterait pas un sou de la part de leur assureur, mais en plus pourrait faire augmenter leur coefficient bonus malus. Il existe ainsi des sinistres non déclarés, ce qui peut alors se révéler problématique si l'assureur souhaite modéliser le risque sous-jacent à une franchise inférieure. On parle alors de troncature à gauche, ce qui signifie que la variable Y reprenant le montant d'un sinistre (et donc l'existence d'un sinistre) n'est pas observable lorsqu'elle est inférieure à un seuil C (la franchise). L'écrêtage correspond quant à lui à un processus différent. L'assureur qui a la possibilité de se réassurer, a la possibilité de limiter la charge à laquelle il devra faire face en cas de sinistre important. De même, un écrêtage peut être effectué lors de la modélisation du risque, permettant ainsi de ne pas prendre en compte des sinistres exceptionnels, et donc peu nombreux, qui viendraient troubler la modélisation et seront traités séparément. On parle alors de censure à droite, ce qui signifie que la variable Y représentant le montant d'un sinistre ici n'est pas observable au-delà d'un seuil C fixe ou variable, mais que C est alors observable à la place. En clair, on n'observe plus Y , mais le minimum entre Y et C .

Un dernier point à souligner est l'historique de données. Une modélisation ayant pour but de créer une tarification devra s'appuyer sur un historique de données conséquent (tout en gardant à l'esprit que le risque ne doit pas avoir changé durant la période d'observation), tandis qu'une modélisation de « contrôle » de la tarification ou de suivi pourra éventuellement se concentrer sur une année d'exercice. Ce suivi de la tarification est nécessaire, d'autant plus lorsque l'historique utilisé lors de la création de la tarification en vigueur n'était pas assez important. L'apparition de nouvelles informations permet alors de combler les absences de la première étude, ou tout du moins d'apporter une information supplémentaire, ce qui permet ensuite une évolution des grilles tarifaires, voir même de la segmentation.



L'objectif de cette partie consiste à bien connaître le portefeuille sur lequel on travaille. Il faut donc prendre le temps de décrire en détail les différentes variables tarifaires et d'examiner la composition du portefeuille à analyser.

I. Constitution de la base de données :

La phase de constitution de la base de données est cruciale pour le tarif qui découlera de l'analyse : comment arriver à un tarif correct en utilisant des données erronées, incomplètes ou obsolètes ? Il est donc essentiel que l'actuaire se penche sur la constitution de la base de données, et ne délègue pas la tâche sans aucun contrôle au service informatique. 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 de 2 fichiers production et sinistres relatifs aux exercices 2007, 2008 et 2009 : un fichier de production et de sinistre, de l'année 2009, et l'autre concerne la sinistralité de ces 3 ans. Ces fichiers concernent les véhicules de tourisme dont le poids est inférieur à 3,5 tonnes (Usage A).

✓ Les données production

Les variables décrivant le fichier « Production »:

Il s'agit des variables suivantes :

- Année de référence
- Code usage
- N° de police
- Code intermédiaire
- Date d'effet
- Date d'échéance
- Code ville
- ville
- date naissance
- Sexe
- Date permis

- Coefficient de réduction majoration (Bonus/Malus)
- Combustion
- Puissance fiscale
- Âge du véhicule
- Montant de la prime acquise

Après la collecte des données vient la phase de l'épuration à l'aide des tests de fiabilité et aussi sous l'Excel. Cette phase consiste à effectuer une opération d'audit pour détecter l'ensemble des anomalies pouvant exister dans le fichier d'étude, par exemple :

- La répétition de la police sans l'existence d'un nouvel événement ;
- L'existence des polices hors l'année de garantie traitée ;
- L'existence de date de début ou de fin de garantie aberrante ;
- Information aberrante ou manquante concernant la date de naissance ou la date d'obtention du permis de conduire.

✓ **Les données sinistres**

Les variables décrivant le fichier « Sinistres »:

Il s'agit des variables suivantes :

- Année vue
- N° de police
- Montant des sinistres Corporels
- Nombre de sinistres Corporels
- Montant des sinistres Corporels Total
- Montant des sinistres Matériels
- Nombre de sinistres Matériels
- Montant des sinistres Matériels Total
- Nombre de sinistres Matériels Total

II. Description et analyse des variables tarifaires :

Notre échantillon d'assurance rassemble un nombre total de contrats de **50000**, et plus précisément **27464 contrats** de la catégorie des véhicules de tourisme, observés en 2009.

Le graphique ci-dessous montre les différentes catégories ainsi que leurs proportions dans notre portefeuille.

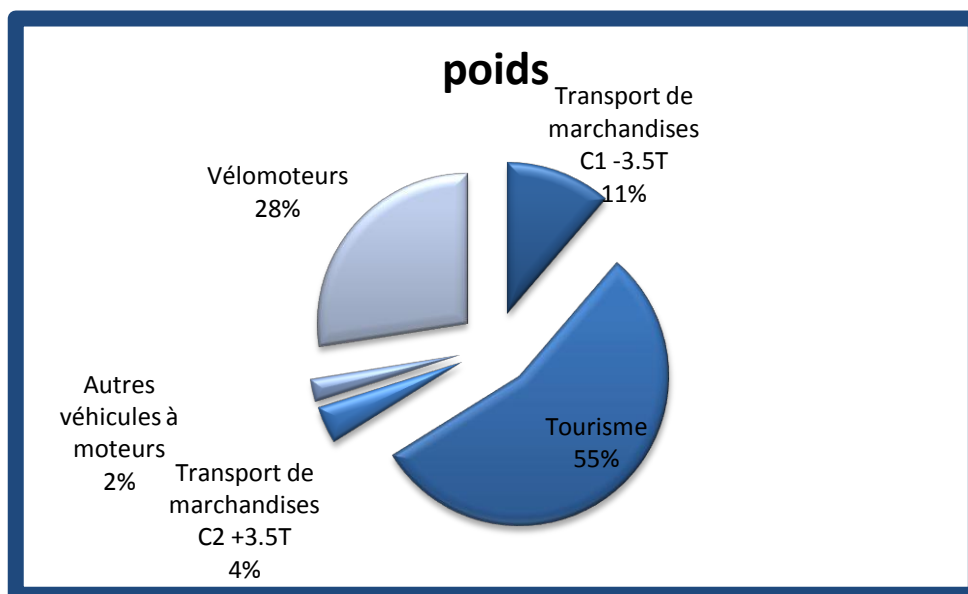


Figure 5 : les catégories des véhicules existant dans notre portefeuille

Les tableaux ci-dessous donnent le nombre de sinistres et des montants de sinistres par type de sinistre : matériel ou corporel.

		<u>2009</u>
Nombre total de sinistres	Corporels	252
		22,01%
	Matériels	893
		77,99%
	Total	1 145

		<u>2009</u>
Montant total de sinistres	Corporels	57678802,6
		75,38%
	Matériels	18836982,15
		24,62%
	Total	76515784,8

II.1 La répartition par sexe :

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.

En terme de proportion, les hommes représentent 98,39% du portefeuille, tandis que les femmes à peu près 1,6% . Les valeurs manquantes représentent près de 0,01% de la base de données (Voir tableau ci-dessous).

	Fréquence%	Coût Total
Hommes	4,00%	69623,10
Femmes	0,17%	38187,51

Face à la sinistralité la différence est grande entre hommes et femmes. Comme on peut le voir sur les graphiques ci-dessous les hommes sont plus « dangereux » que les femmes.

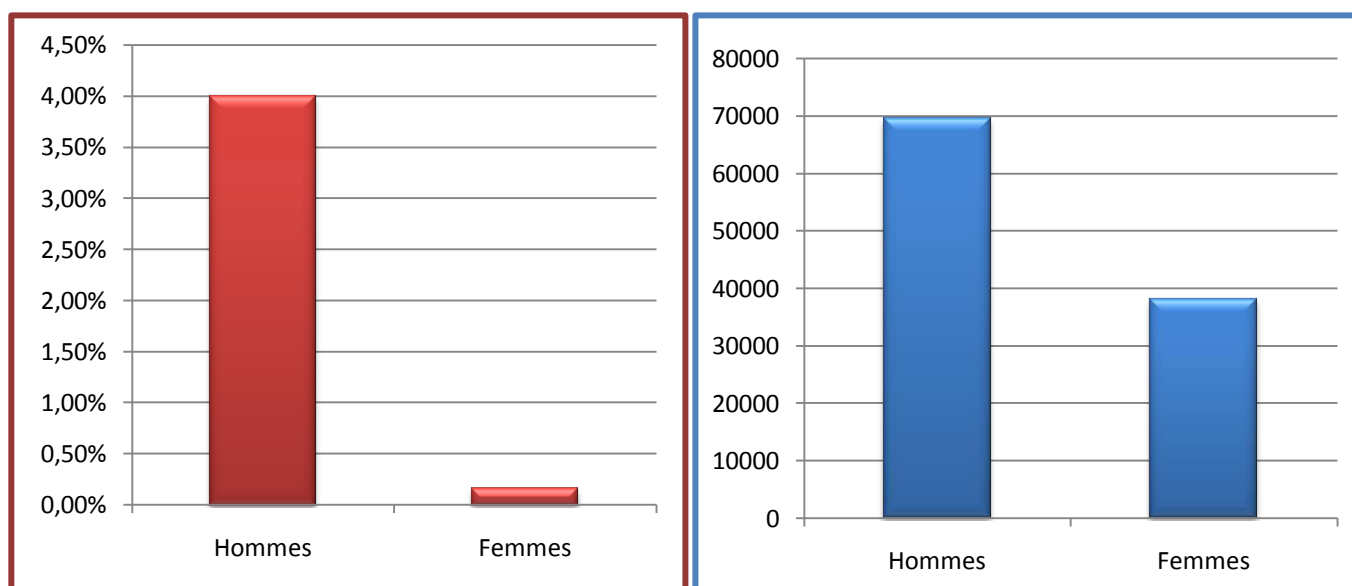


Figure 6 : Distribution de la fréquence et du coût selon le sexe

On remarque que :

- *Nombre de sinistres est une Donnée*
- *Charge totale est une Donnée*
- *Prime acquise est une Donnée*
- *Véhicule année : Durée d'exposition au risque calculée par une macro en Excel (voir Annexes)*

Donc on peut calculer facilement :

$$\text{Coût moyen} = \frac{\text{Charge totale}}{\text{Nombre de sinistres}}$$

$$\text{Fréquence} = \frac{\text{Nombre de sinistres}}{\text{Exposition}}$$

II.2 La mesure de l'exposition au risque : la variable Véhicule année

Il s'agit du nombre de jours où la police a été en vigueur durant l'année 2009. 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.

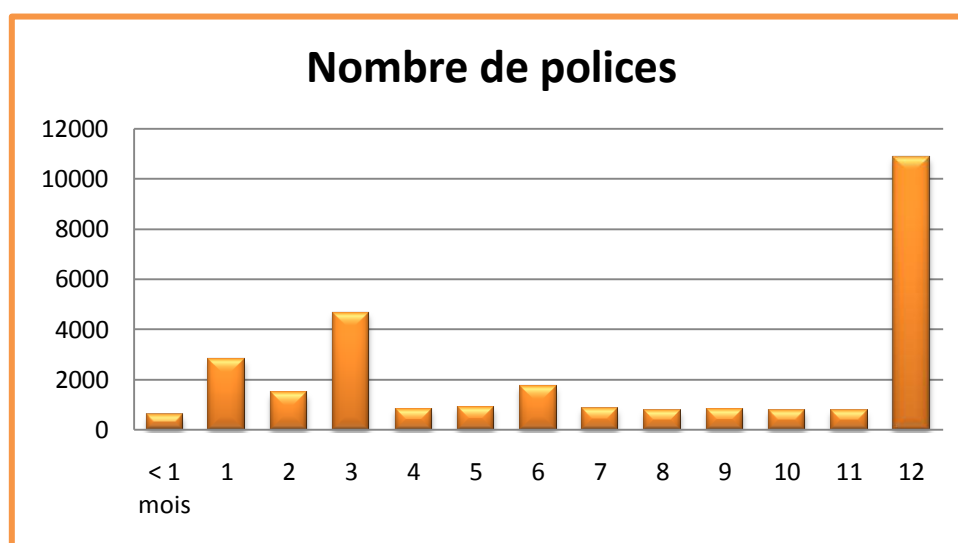


Figure 7 : les périodes de couverture des polices

La figure ci-dessus donne une idée des périodes de couverture des différentes polices en portefeuille. On constate une majorité de polices de couverture durant toute l'année, suivis par

des polices de durée de couverture trimestrielle et ensuite par celles de couverture mensuelle et semestrielle, avec aucune répartition particulière des autres polices de couvertures différentes de celles citées avant.

A. corrélation des variables explicatives avec la variable cible

Afin d'éliminer quelques variables nous allons étudier les corrélations entre les variables explicatives et les variables cibles, fréquence et charge de la sinistralité dommage.

S'il n'existe pas de corrélations entre ces variables, alors il n'est pas utile de les conserver car elles ne seront pas utilisées lors de la modélisation.

Cette étude a permis de supprimer quelques variables. En effet les informations apportées par certaines variables des véhicules, tels que le nombre de place, poids de véhicule....

B. corrélation des variables explicatives entre elles :

Nous allons étudier les corrélations entre les futures variables explicatives du modèle afin d'essayer de réduire leur nombre. Si deux variables sont extrêmement corrélées alors il faudra, malgré la perte d'informations, retirer la variable la moins intéressante ou les croiser afin d'obtenir une nouvelle variable.

La corrélation a été effectuée grâce à **la PROC CORR** voir Annexes.

Variable	Nb	Moyenne	Écart-type	Somme	Minimum	Maximum
age_condu	19779	45.86415	12.40643	907147	10.00000	92.00000
anc_permi	14389	14.14073	10.14354	203471	0	68.00000
anc_vehi	27502	13.73867	9.04042	377841	0	59.00000
Crn_Nive	27623	0.96449	0.05833	26642	0	1.70000
Codevill	27729	161.40052	219.68608	4475475	21.00000	999.00000
Puisvehi	27726	7.70923	1.97459	213746	1.00000	40.00000
Crn_Nive	27623	0.96449	0.05833	26642	0	1.70000

Tableau 1 : Tableau des statistiques simples

Coefficients de corrélation de Pearson							
Nombre d'observations							
	age_condu	anc_permi	anc_vehi	Crms_Nive	Codevill	Puisvehi	Crms_Nive
age_condu	1.00000	0.59041	-0.01569	-0.19078	-0.02381	0.01921	-0.19078
age_condu	19779	10813	19661	19727	19779	19779	19727
anc_permi	0.59041	1.00000	-0.09764	-0.20985	-0.03251	0.02094	-0.20985
anc_permi	10813	14389	14297	14344	14389	14389	14344
anc_vehi	-0.01569	-0.09764	1.00000	0.08786	0.04769	-0.03406	0.08786
anc_vehi	19661	14297	27502	27404	27502	27502	27404
Crms_Nive	-0.19078	-0.20985	0.08786	1.00000	0.01395	0.02034	1.00000
Crms_Nive	19727	14344	27404	27623	27623	27620	27623
Codevill	-0.02381	-0.03251	0.04769	0.01395	1.00000	-0.08466	0.01395
Codevill	19779	14389	27502	27623	27729	27726	27623
Puisvehi	0.01921	0.02094	-0.03406	0.02034	-0.08466	1.00000	0.02034
Puisvehi	19779	14389	27502	27620	27726	27726	27620
Crms_Nive	-0.19078	-0.20985	0.08786	1.00000	0.01395	0.02034	1.00000
Crms_Nive	19727	14344	27404	27623	27623	27620	27623

Tableau 2 : Tableau des corrélations entre les variables explicatives

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

- Age et l'ancienneté de permis sont deux variables corrélées.
- Toutes les variables explicatives semblent indépendantes.

II.3 Présentation de l'algorithme « CHAID »

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 nous utilisons 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 etc....

Chaid est l'acronyme de -Chi-square Automatic Interaction Detected-, il désigne comme on dit tantôt un algorithme de segmentation de variable. Il se base pour cela sur une variable dépendante qui lui servira de guide. L'algorithme subdivise la population initiale en plusieurs ensembles distincts en se basant sur les variables indépendantes les plus significatives, chacun des groupes ainsi obtenu 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. Le résumé de toutes ces démarches est donné sous forme d'un arbre de décision où chaque nœud représente un regroupement d'observations.

II.4 Segmentation des variables :

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

- ✓ puissance véhicule combinée au type moteur;
- ✓ l'âge de véhicule ;
- ✓ la région;
- ✓ le sexe du conducteur;
- ✓ l'âge du conducteur combiné à l'ancienneté du permis;
- ✓ bonus malus;

Pour ces variables explicatives retenues pour la tarification, la détermination d'un niveau de segmentation s'impose.

Si cette segmentation est implicite pour certaines variables comme le sexe (homme et femme), elle l'est moins pour des variables prenant un grand nombre de modalités, comme l'âge du conducteur ou l'ancienneté de permis.

Un traitement spécifique de ces variables doit donc être effectué.

II.3.1 Puissance véhicule et type moteur (PVM) :

Dans notre portefeuille, 17.783 polices concernant des véhicules diesel (soit 64,2%) et 9.937 des véhicules à essence (soit 35,8%). En effet, l'acquisition d'un véhicule diesel se justifie souvent par un usage fréquent.

La puissance fiscale est une variable dépendante du véhicule assuré, elle augmente en général le risque en croissant. Dans notre portefeuille, on constate que les véhicules de 6,7, et 8 chevaux sont majoritaires. Ils représentent à peu près les trois quarts du portefeuille. L'expérience a montré que la combinaison des deux variables –puissance fiscale et combustible – donnait de meilleurs résultats.

On a donc créé une nouvelle variable à partir de la combinaison des deux premières. Elle comporte 5 modalités : 2 pour l'essence et 3 pour le diesel.

CODEcompf	FR %	POIDS	cout moyen
ESSENCE>=8	0,60%	18,48	189923,30
ESSENCE<=7	0,65%	17,38	328393,93
DIESEL<=6	0,64%	12,41	369728,37
DIESEL>=8	1,43%	29,34	440585,82
DIESEL= 7	0,85%	22,38	403287,31

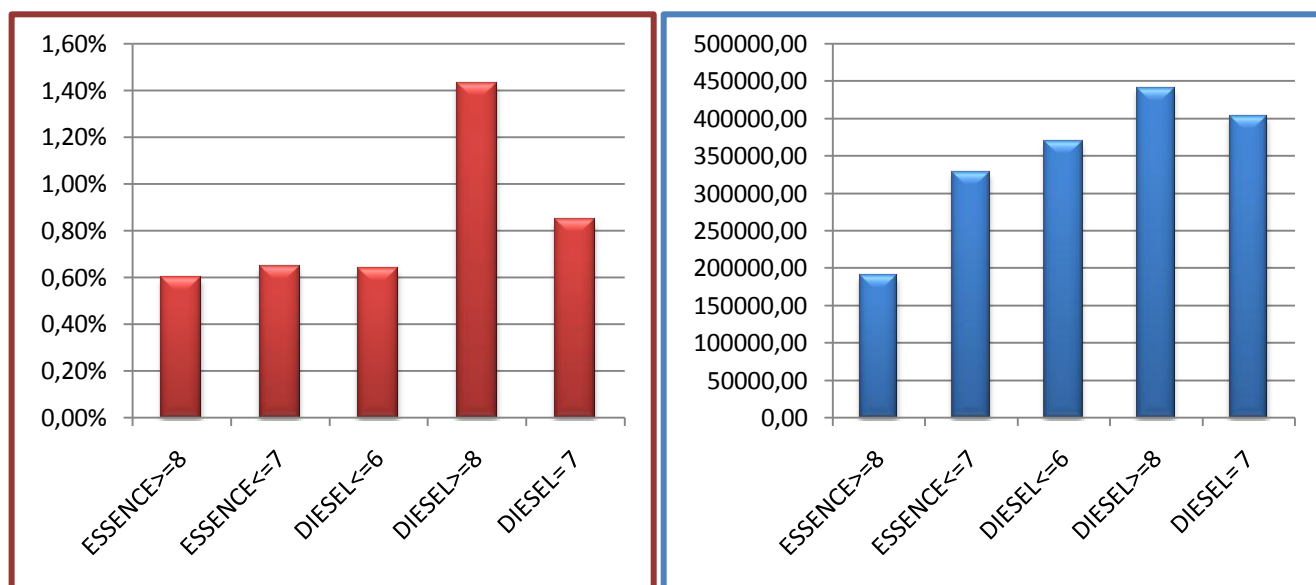


Figure 8 : la fréquence et le coût moyen de la nouvelle variable regroupée

Les graphiques révèlent une grande différence entre les modalités de la nouvelle variable tant en termes de coût qu'en termes de fréquence. Globalement, l'usage Essence a une fréquence plus faible que l'usage Diesel.

II.3.2 Ancienneté de 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 vieille la voiture cause des sinistres plus importants et des coûts moyens aussi plus élevés.

Elle se compose de cinq modalités selon l'algorithme **CHAID**:

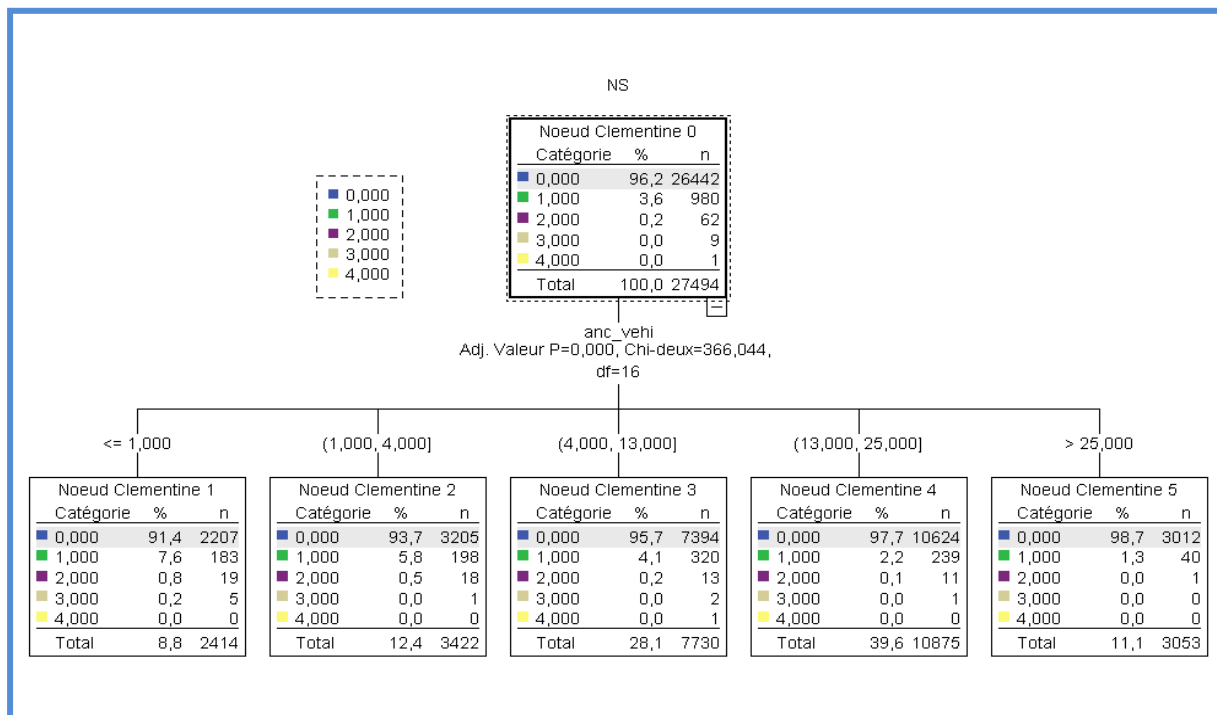
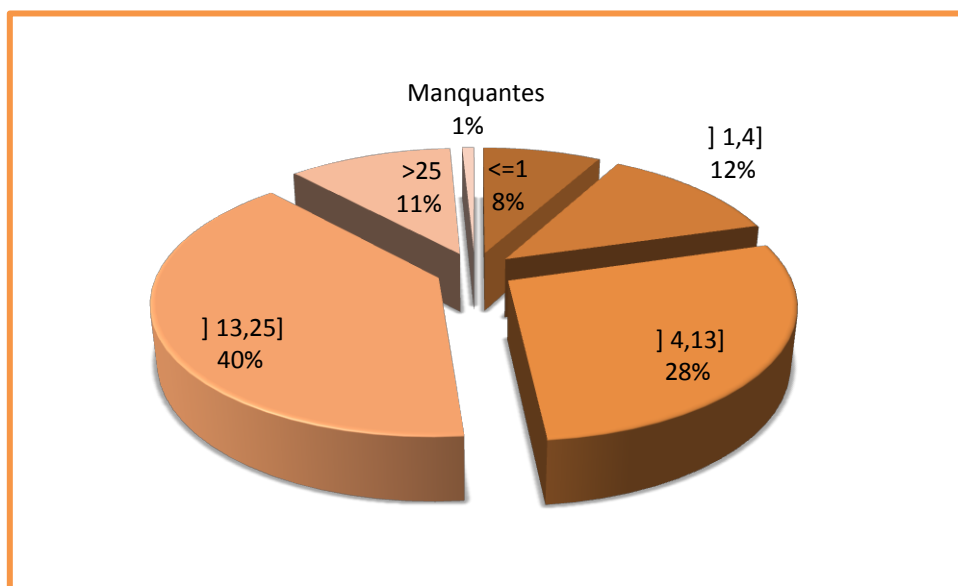


Figure 9 : Arbre de décision de la variable âge véhicule



Ce diagramme nous illustre le poids de chaque modalité de la classe âge véhicule dans notre portefeuille. On remarque une forte présence des véhicules anciens.

Les modalités se représentent comme suit :

variable	LA VARIABLE AGEVEHIC CODEE	FR%	cout moyen
codagevehi1	≤ 1	0,86%	282314,63
codagevehi2] 1,4]	0,92%	291524,6
codagevehi3] 4,13]	1,27%	200898,69
codagevehi4] 13,25]	0,95%	292044,98
codagevehi5	>25	0,21%	1174762,06
codagevehi6	Manquantes	0,07%	1053548,4

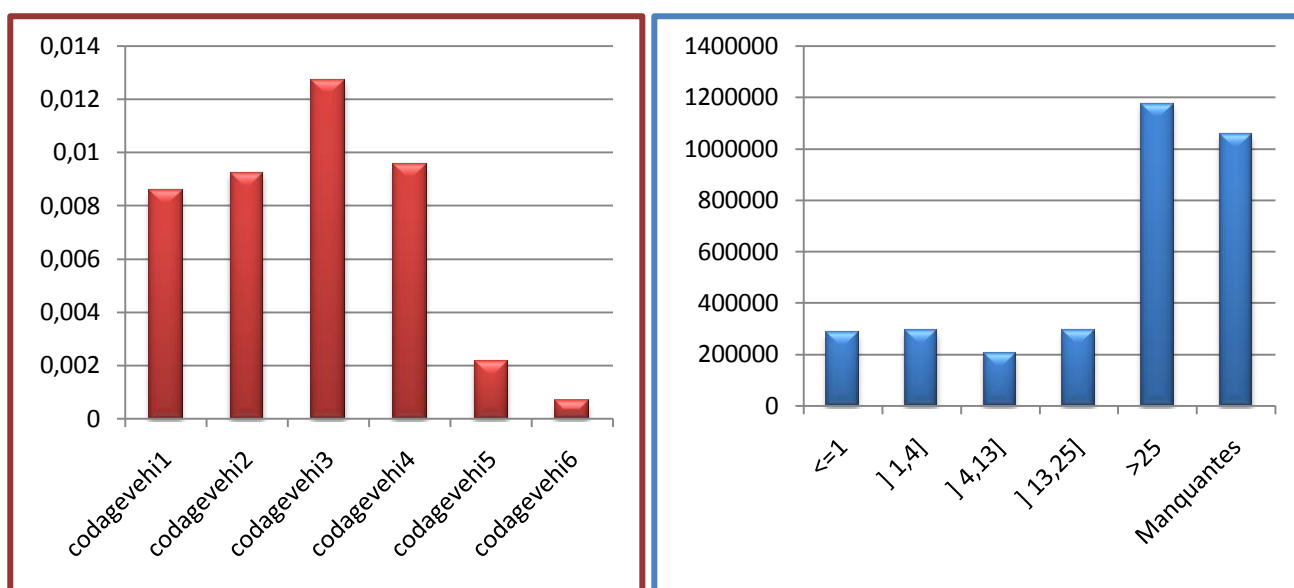


Figure 10 : la fréquence et le coût moyen corporel selon la variable âge du véhicule

On remarque que la fréquence est de plus en plus élevée que l'âge du véhicule augmente, jusqu'à l'âge de 13 ans avec un coût moyen relativement uniforme pour tout âge inférieur à 25 ans de l'ordre de 291524 à peu près et un coût moyen très élevé concernant les véhicules d'âge supérieur à 25 ans ce qui permet de dire que les véhicules les plus anciens causent des dégâts plus graves.

II.3.3 L'âge et l'ancienneté du permis :

L'âge et l'ancienneté du permis sont deux variables corrélées (voir la matrice de corrélation partielle) très importantes dans l'analyse de la sinistralité. Malheureusement, elles comportent toutes les deux un taux élevé de valeurs manquantes. Ce constat nous impose la combinaison de ces deux variables en une seule, qui se compose de cinq classes. D'après l'algorithme de CHAID :

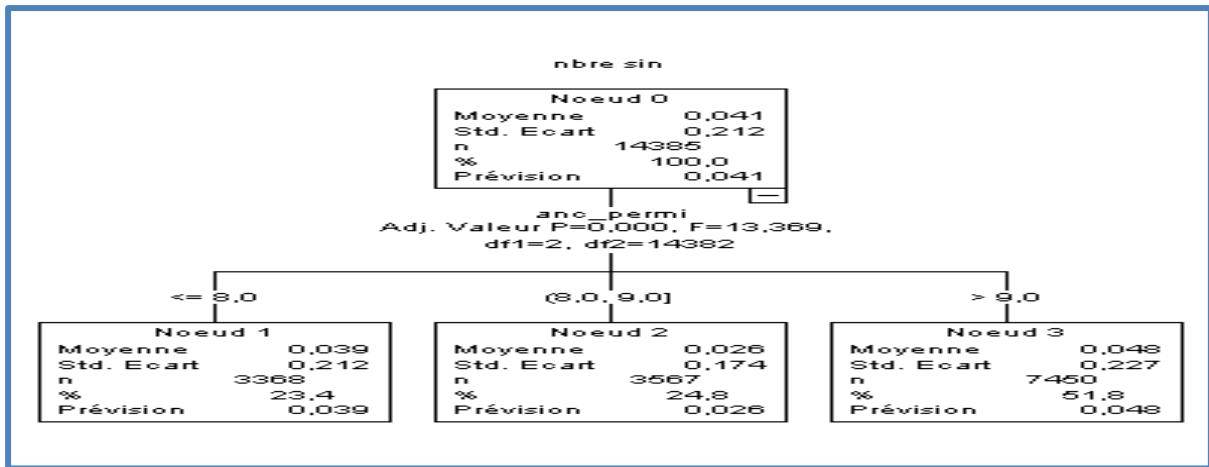


Figure 11 : Arbre de décision de la variable ancienneté de permis

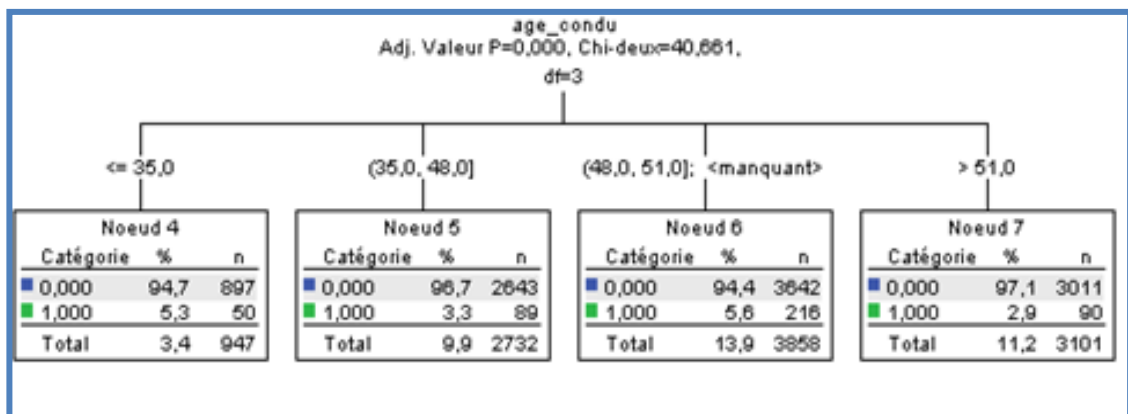


Figure 12 : Arbre de décision de la variable âge conducteur

Le tableau ci-dessous permet d’avoir un meilleur regard sur la composition des différentes classes de la nouvelle variable âge _ ancienneté permis :

CODEAGEPERMI	POIDS%	FR %	cout moyen
Manquants	61	2,71%	102843,80
âge >= 39 ans et une ancienneté de permis >=15 ans	55,66	2,62%	106124,53
39< âge >=62 et <=9 ancienneté de permis <= 14 ans	11,35	0,28%	712358,55
29=< âge >= 49 et 3=>ancienneté de permis <= 8 ans	3,26	0,13%	1626105,97
âge < 29 et une ancienneté de permis < 8	2,46	0,07%	3545928,03

Le graphique ci-dessous représente les 4 classes autres que les valeurs manquantes :

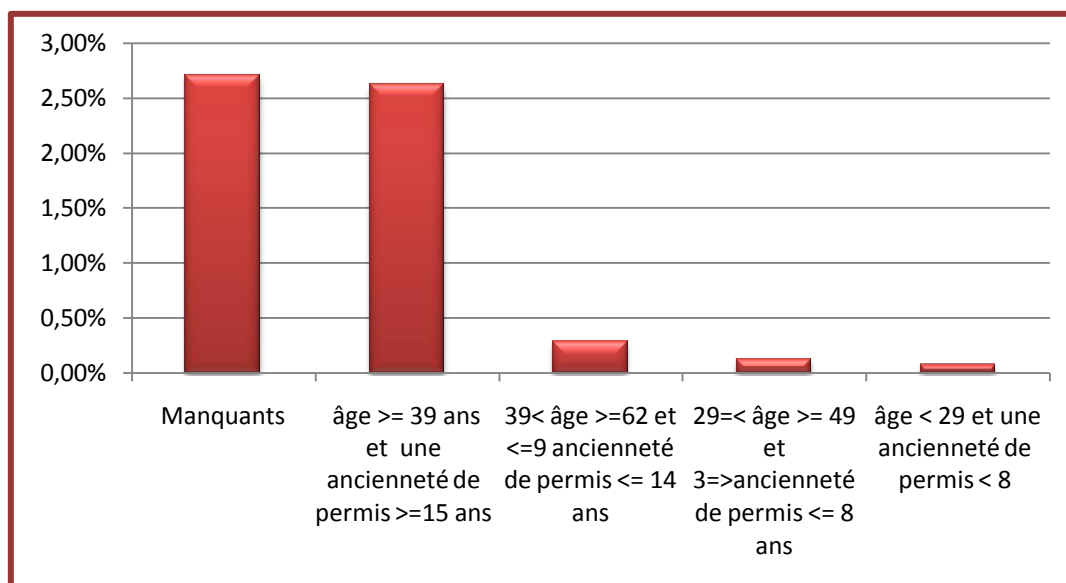


Figure 13 : La fréquence de la nouvelle variable âge-ancienneté permis

On remarque que la nouvelle variable souffre d'un fort taux de valeurs manquantes ce qui nous oblige à éliminer cette variable de notre étude.

II.3.4 *Variable CRM (coefficient de réduction majoration)*

Cette variable est généralement déterminante dans une tarification car elle rend compte de la probabilité que l'assuré a d'avoir un sinistre. Nous parlerons de fréquence de sinistres qui sera alors plus faible pour les assurés ayant un faible CRM, et inversement.

Notre portefeuille contient 16.552 (soit 60,26%) assurés qui ont un coefficient CRM de 100 (c.-à-d. paient la totalité du montant de la prime), 10.573 (soit 38,36%) avec un CRM de 90 (c.-à-d. une réduction de 10% du montant de la prime), et 339 (soit 1,23%) assurés avec un CRM compris entre 120 et 170.

LA VARIABLE BONUS CODEE	FR %	Coût moyen
BONIFIE	1,86%	143 808,399
NON BONIFIE	1,74%	159 740,678
MAJORE	0,56%	472 056,237

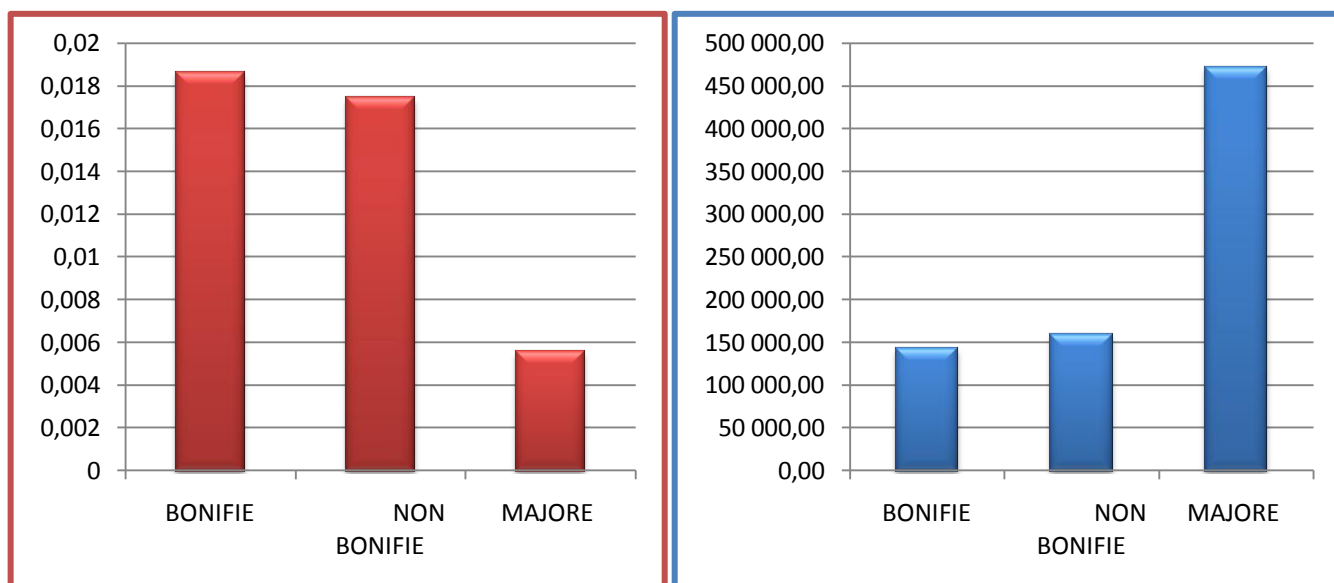


Figure 14 : la fréquence et le coût moyen de la variable Bonus

On constate que les assurés bonifiés et non bonifiés ont le même comportement au niveau de la fréquence et du coût moyen. Par contre le comportement des majorés est tout à fait différent. En effet, les majorés occasionnent peu de sinistres par rapport aux deux autres modalités mais avec un coût moyen très élevé ce qui signifie que ces sinistres sont souvent de type grave contrairement aux deux autres types (BONIFIE et NON BONIFIE) à caractère moins grave.

II.3.5 La segmentation de la variable ville:

Les études précédentes ont montré que cette variable est la plus pertinente. En effet, chaque ville est caractérisée par une sinistralité différente selon la fréquence des sinistres et le coût moyen.

L'algorithme CHAID du module « Answer Tree » de SPSS 19 utilisé dans notre étude, produit des diagrammes arborescents clairs qui segmentent une population en sous-groupes partageant les mêmes caractéristiques. Ces diagrammes sont les suivants :

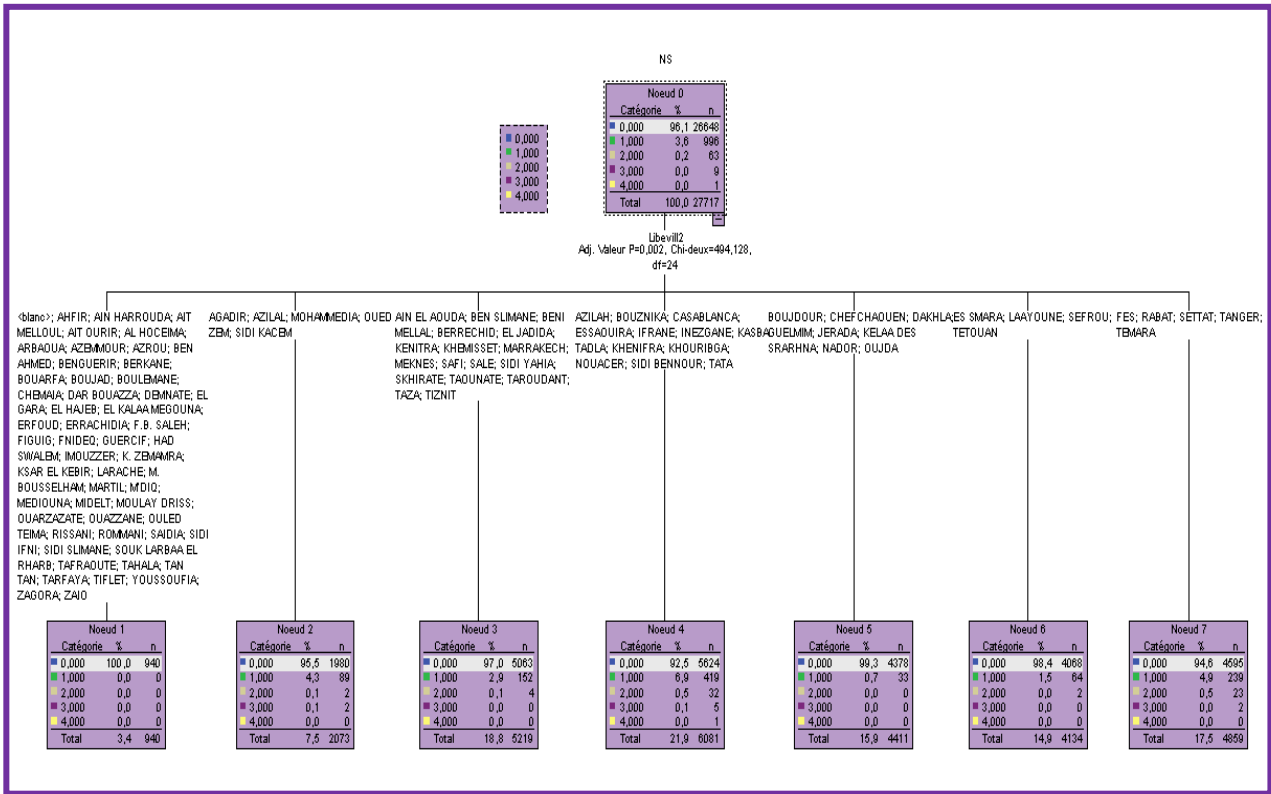


Figure 15 : diagramme arborescent segmentant la population par libellé de ville

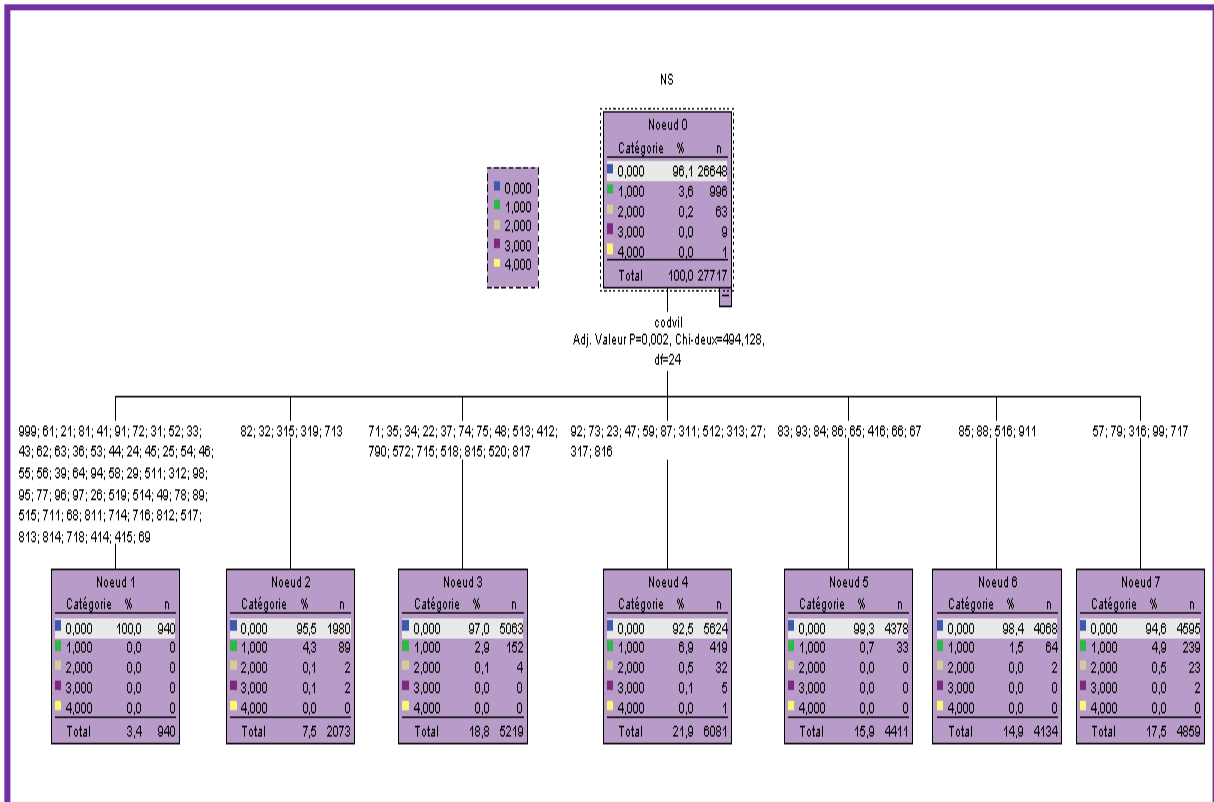


Figure 16 : diagramme arborescent segmentant la population par code de ville

Les groupes de villes sont définies comme suite :

Groupe 1 :

AHFIR ; AIN HARROUDA ; AIT MELLOUL ; AIT OURIR ; AL HOCEIMA ; ARBAOUA ; AZEMMOUR ; AZROU ; BEN AHMED ; BENGUERIR ; BERKANE ; BOUARFA ; BOUJAD ; BOULEMANE ; CHEMAIA ; DAR BOUAZZA ; DEMNATE ; EL GARA ; EL HAJEB ; EL KALAA MEGOUNA ; ERFOUD ; ERRACHIDIA ; FB SALEH ; FIGUIG ; FNIDEQ ; GUERCIF ; HAD SWALAM ; MARTIL ; MDIQ ; MEDIOUNA ; MIDELT ; MOULAY DRISS ; OURZAZATE ; OUZZANE ; OULED TEIMA ; RISSANI ; ROMMANI ; SOUK LARBAA ; TAFRAOUTE ; TAN TAN ; TAHALA ; TARFAYA ; TIFLET ; YOUSSEUFIA ; ZAGORA ; ZAIO

Groupe 2 :

AGADIR ; AZILAL ; MOHAMMEDIA ; OUED ZEM ; SIDI KACEM

Groupe 3 :

AIN EL AOUDA ; BEN SLIMANE ; BENI MELLAL ; BERRECHID ; EL JADIDA ; KENITRA ; KHMISSET ; MARRAKECH ; MEKNES ; SAFI ; SALE ; SIDI YAHIA ; SKHIRATE ; TAOUNATE ; TAROUDANT ; TAZA ; TIZNIT

Groupe 4 :

AZILAH ; BOUZNIKA ; CASABLANCA ; ESSAOUIRA ; IFRANE ; INZGANE ; KASBA TADLA ; KHNIFRA ; KHOURIBGA ; NOUACER ; SIDI BENNOUR ; TATA

Groupe 5 :

BOUJDOUR ; CHEFCHAOUEN ; DAKHLA ; GUELMIM ; JERADA ; KELAA DESSRAGHNA ; NADOR ; OUJDA

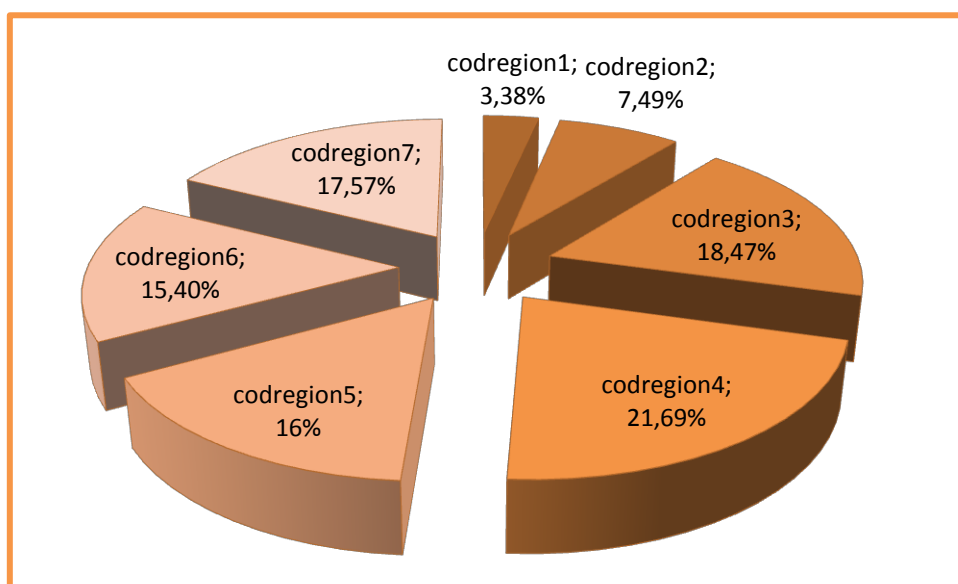
Groupe 6 :

ES SMARA ; LAAYOUN ; SEFROU ; TETOUAN

Groupe 7 :

FES ; RABAT ; SETTAT ; TANGER ; TEMARA

Le poids de chaque région dans notre portefeuille est illustré dans le diagramme suivant :



On remarque que la région la plus dominante dans notre portefeuille est 4, avec une proportion de 21,69%, suivie par la région 3 avec 18,47%.

	Fréquence%	Coût moyen
codregion1	0	0
codregion2	0,36%	746398,56
codregion3	0,55%	510105,23
codregion4	1,81%	133757,94
codregion5	0,11%	2315649,70
codregion6	0,22%	1207047,55
codregion7	1,08%	228010,73

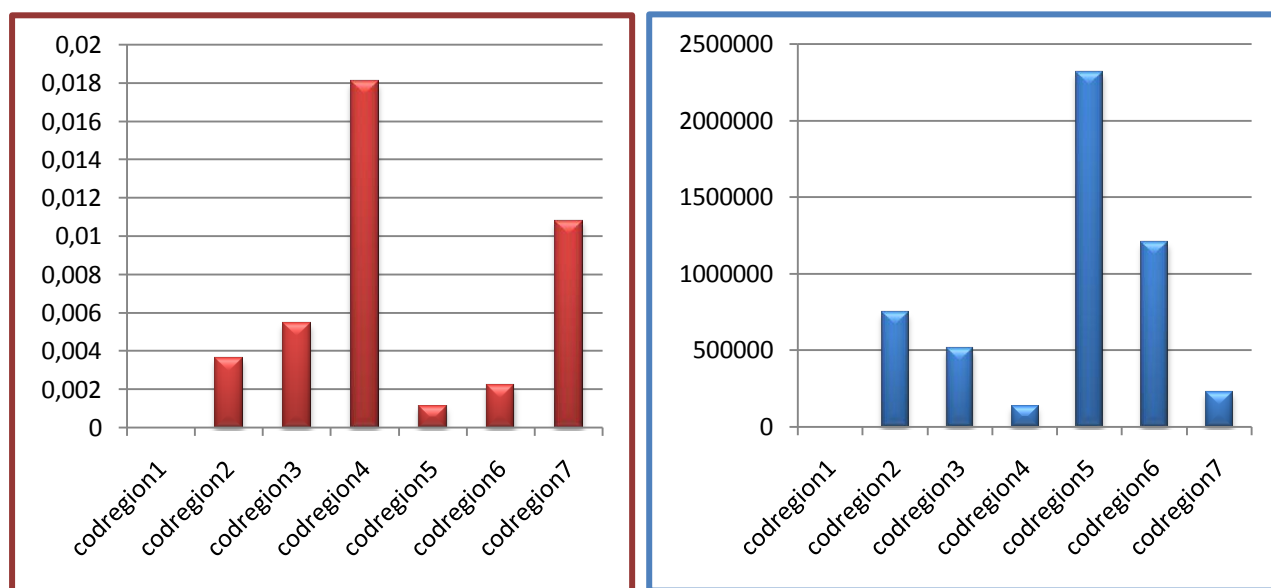


Figure 17 : la fréquence et le coût moyen de la variable région

On remarque qu'en termes de fréquence, la région 4 affiche la plus forte valeur de 1,81%. Cette région rassemble les villes de «AZILAH ; BOUZNICA ; CASABLANCA ; ESSAOUIRA ; IFRANE ; INZGANE ; KASBA TADLA ; KHNIFRA ; KHOURIBGA ; NOUACER ; SIDI BENNOUR ; TATA ». Ceci peut être expliqué par l'infrastructure routière dégradée au niveau de ces villes (KASBA TADLA ; KHNIFRA ; KHOURIBG, BOUZNICA, SIDI BENNOUR ; TATA). En ce qui concerne CASABLANCA, la capitale économique du pays, ceci peut être expliqué par la circulation intense que connaît cette ville. Malgré l'enregistrement de la plus grande valeur au niveau de fréquence, le coût moyen représente la plus petite valeur au niveau de toutes les régions ; ceci signifie que les sinistres réalisés sont de type moins grave.

II.5 Discretisation des variables

A l'importation, nous avons à la fois des variables qualitatives et des variables quantitatives. Dans un souci d'homogénéité (nécessaire pour les prochaines étapes), nous allons discrétiser ces variables.

variable	codtyppuis	on affecte la valeur
Codtyppuis1	ESSENCE>=8	1
codtyppuis2	ESSENCE<=7	2
Codtyppuis3	DIESEL<=6	3
Codtyppuis4	DIESEL>=8	4
Codtyppuis5	DIESEL= 7	5

Tableau 3 : Codage de la variable Puissance fiscale et combustible (PFC)

variable	LA VARIABLE AGEVEHIC CODEE	on affecte la valeur
codagevehi1	<=1	1
codagevehi2] 1,4]	2
codagevehi3] 4,13]	3
codagevehi4] 13,25]	4
codagevehi5	>25	5
codagevehi6	Manquant	6

Tableau 4 : Codage de la variable Ancienneté de véhicule

variable	CODEAGEPERMI	on affecte la valeur
codagpermi1	MANQUANTS	0
codagpermi2	Individus ayant un âge de 39 et plus et une ancienneté de permis 15 ans et plus	1
codagpermi3	Individus ayant un âge entre 39 et 62 et une ancienneté de permis 9 ans et 14 ans	2
codagpermi4	Individus ayant un âge entre 30 et 49 et une ancienneté de permis 3 ans et 8 ans	3
codagpermi5	Individus ayant un âge moins de 29 et une ancienneté de permis moins de 8 ans	4

Tableau 5 : Codage de la variable âge et l'ancienneté du permis

Variable	LA VARIABLE BONUS CODEE	on affecte la valeur
codbonu1	BONIFIE	1
codbonu2	NON BONIFIE	2
codbonu3	MAJORE	3

Tableau 6 : Codage de la variable coefficient de réduction majoration

Variable	Code ville	on affecte la valeur
codregion1	999.61.21.81.41.91.72.31.52.33.43.62.63.36.53.44.24.45.25.54.46.55.56.39.64.94.58.29.511.312.98.95.77. 96.97.26.519.514.49.78.89.515.711.68.811.714.716.812.517.813.814.718.414.415.69	1
codregion2	82.32.315.319.713	2
codregion3	71.35.34.22.37.74.75.48.513.412.790.572.715.518.815.520.817	3
codregion4	92.73.23.47.59.87.311.512.313.27.317.816	4
codregion5	83.93.84.86.65.416.66.67	5
codregion6	85.88.516.911	6
codregion7	57.79.316.99.717	7

Tableau 7 : Codage de la variable ville

Le travail le plus important c'était la construction et l'épuration de la base de données, car il influence toute l'étude de la tarification, surtout la segmentation, et la codification. Finalement, on a trouvé six variables codifiées qui vont expliquer les variables d'intérêts, à savoir fréquence corporelle, fréquence matérielle, coût corporel, coût matériel, ce qui va être explicité à partir de la partie suivante.



On a vu précédemment des généralités concernant l'assurance automobile ainsi que des techniques d'analyses descriptives. Ces notions constituent une première approche et un préliminaire à la réalisation de la modélisation du risque automobile. Nous nous concentrerons tout d'abord sur l'aspect théorique des modèles linéaires généralisés, on va essayer d'estimer les fréquences annuelles des sinistres matériels et corporels. Nous avons établi dans ce cadre les modèles suivants :

- ❖ *La régression de poisson*
- ❖ *GLM Binomiale négative*
- ❖ *La régression logistique*

I .Le cadre général des modèles linéaires généralisés :

Depuis quelques années, l'outil principal utilisé en tarification est le modèle linéaire généralisé, Dans cette section, nous allons présenter le cadre des GLM, avant de rentrer dans l'application en tarification dans les sections suivantes.

Les modèles linéaires généralisés sont une généralisation du modèle linéaire Gaussien, Obtenu en autorisant d'autres lois (conditionnelles) que la loi Gaussienne. Les lois possibles doivent appartenir à la famille exponentielle, dont la densité (ou mesure de probabilité dans le cas discret) s'écrit :

$$f(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp \left\{ \frac{\boldsymbol{\theta} \mathbf{y} - \mathbf{b}(\boldsymbol{\theta})}{\phi} + \mathbf{c}(\mathbf{y}, \phi) \right\}, \mathbf{y} \in \mathbf{S}$$

Où $\boldsymbol{\theta}$ est un paramètre réel, appelé paramètre naturel, $\phi > 0$ (ϕ éventuellement donné) est un paramètre de dispersion. Les fonctions $\mathbf{b}(\cdot)$ et $\mathbf{c}(\cdot)$ sont spécifiques à la distribution, $\mathbf{b}(\cdot)$ étant « régulière ». Le support \mathbf{S} est un sous-ensemble de \mathbb{N} ou de \mathbb{R} .

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 :

I.1 Distributions discrètes :

- *Loi de Poisson $Poi(\lambda)$:*

La distribution de probabilité d'une loi poisson de paramètre λ est donnée par :

$$Poi(Y = y) = \exp(-\lambda) \frac{\lambda^y}{y!}$$

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

- Loi binomiale Négative BN $\left(r, \frac{r}{r+\mu}\right)$

Une introduction alternative de la sur dispersion reposerait sur la loi Binomiale Négative à deux paramètres r et μ . Cependant celle-ci n'appartient pas au type exponentiel des GLM. On a en effet :

$$P(Y = y) = \frac{\Gamma(r + y)}{\Gamma(r)y!} \frac{r^y \mu^y}{(r + \mu)^{r+y}} \text{ avec } y \in \mathbb{N}$$

I.2 Distributions continues :

- Loi Normale $N(\mu, \sigma^2)$:

Cette distribution rentre dans le type exponentiel car :

$$f(y) = \exp \left\{ \frac{1}{\sigma^2} \left(\mu y + \frac{\mu^2}{2} \right) + c(y, \sigma^2) \right\} \text{ avec } y \in \mathbb{R}$$

- Loi Gamma $\gamma\left(\nu, \frac{\nu}{\mu}\right)$

$$f(y) = \exp \left[\left(-\frac{y}{\mu} - \ln(\mu) \right) \nu + c(y, \nu) \right] \text{ avec } y > 0$$

Dans le cadre des modèles linéaires généralisés, pour chaque individu appartenant à une population donnée dont on dispose d'informations sous forme d'un ensemble de variables x_i^t , on cherche à déterminer des paramètres inconnues β tel que :

$$\mu_i = g^{-1}(x_i^t \beta)$$

g est appelée la fonction de lien. Pour chaque type de distribution, il y a une fonction de lien canonique.

Les fonctions les plus utilisés sont la fonction logarithmique, la fonction puissance et la fonction Identité.

II. Méthodes de sélection des variables explicatives

Trois approches existent dans la plupart des logiciels : forward , backward et stepwise. La procédure forward part d'un modèle sans variables explicatives (comportant uniquement un intercepte β_0 , donc supposant les observations identiquement distribuée) et incorpore un à un les facteurs de risque jugés les plus pertinents sur base de la comparaison des log-vraisemblances. Aucune variable n'est plus introduite lorsque la prise en compte de celles-ci ne rend pas le modèle significativement meilleur. Cette première approche est fort semblable à celle de la segmentation des portefeuilles selon un arbre de classification.

La procédure backward quant à elle part du portefeuille le plus segmenté et regroupe les classes définies à partir des facteurs les moins pertinents.

L'approche stepwise conjugue l'esprit des deux algorithmes précédents (elle peut se voir comme une procédure forward pour laquelle, après chaque inclusion de variable explicative, on se demande si une des variables entrées dans le modèle ne pourrait pas être supprimée.

la procédure GENMOD DE SAS, qui permet d'effectuer la régression de Poisson, n'offre pas les procédures décrites ci-dessus au contraire de logistique, mais bien l'analyse de type 1 à la même fonction que la procédure stepwise ,elle introduit une à une les variables dans le modèle, dans l'ordre dans lequel elles ont été spécifiées dans MODEL

III. L'outil de base en RC Auto: la régression de Poisson

III .1 Description du modèle :

L'analyse du nombre des sinistres ne peut en général pas se faire dans le cadre classique du modèle de régression linéaire. En effet, la moyenne de la variable aléatoire décrivant le nombre des sinistres déclarées par un assuré sur une période est en général proche de 0, la variable aléatoire elle-même ne prenant que quelques valeurs entières 0,1,2, ...n la forme dissymétrique de la densité de probabilité discrète associée est loin de ressembler à la courbe de Gauss et l'hypothèse de normalité est clairement inadéquate.

La régression de Poisson se place dans le contexte des modèles de comptage. Dorénavant, Y_{it} représente le nombre de sinistres déclarées par l'assuré i durant la période t , $i = 1, 2, \dots, n$, $t = 1, 2, \dots, T_i$, où T_i désigné le nombre de périodes d'observation pour l'assuré i . Nous noterons d_{it} la durée de la $t^{\text{ème}}$ période d'observation pour l'individu i . Lors de chaque modification des variables observables, un nouvel intervalle commence, de sorte que d_{it} peut être différent de 1.

Nous supposons que nous disposons par ailleurs d'autres variables X_{it} , connues au début de la période t , et pouvant servir de facteurs explicatifs. Ces variables peuvent comporter des facteurs purement individuels indépendants du temps, des facteurs purement temporels, des facteurs dépendant des deux indices voire même des valeurs passées $Y_{i, t-j}$, $j = 1, 2, \dots$ de la Variable d'intérêt (quoique l'inclusion de la sinistralité passée de l'assuré en composantes de régression ne soit pas très orthodoxe).

Le modèle de Poisson suppose que la loi conditionnelle de Y_{it} sachant X_{it} est de Poisson. Spécifions une fonction de lien de type logarithmique, sa moyenne s'écrit sous la forme :

$$E[N_i/x_i] = d_i \exp(\beta^t x_i) \quad , i=1, 2, \dots, n$$

β est le vecteur des coefficients de régression inconnus et qu'on va estimer sur la base de nos données.

Après la catégorisation des variables explicatives chaque assuré est représenté par un vecteur X_i . Par la suite, la fréquence annuelle des sinistres $\mu_i = \exp(\beta^t x_i)$ apparaît comme un produit de coefficient de majoration ou de réduction par rapport à la fréquence de l'individu de référence. Plus précisément on a :

$$\mu_i = \exp(\beta^t x_i)$$

$$\mu_i = \exp(\beta_0) \prod_{j=1}^p \exp(\beta_j x_{ij})$$

$$\mu_i = \exp(\beta_0) \prod_{j/x_{ij}=1}^p \exp(\beta_j)$$

Le terme $\exp(\beta_0)$ représente la fréquence annuelle de l'individu de référence. Chacun des facteurs $\exp(\beta_j)$ traduit l'influence d'un critère de segmentation. En effet, si $\beta_j > 0$ les assurés présentant cette caractéristique subiront une majoration de la prime, alors que si $\beta_j < 0$ une réduction de la prime.

III .2 Estimation du maximum de vraisemblance

L'estimation des paramètres du modèle se fait par maximisation de la vraisemblance associée aux observations dont l'expression vaut :

$$L(\beta/x) = \prod_{i=1}^n \exp[d_i \mu_i] \frac{[d_i \mu_i]^{n_i}}{n_i!} \quad \text{où } \mu_i = \exp(\beta^t x_i)$$

La maximisation de cette quantité permet de déterminer $\hat{\beta}$ l'estimateur de β qui s'accorde le mieux avec nos observations.

III.3 Illustration numérique :

La procédure GENMOD type1 de SAS permet de sélectionner les variables explicatives. (Pour le code voir Annexes). Les résultats sont présentés dans le tableau suivant :

Statistiques LR pour Analyse de Type 1				
Source	Écart	DF	Khi 2	Pr > Khi 2
Intercept	6287.2300			
sexe	6257.1718	2	30.06	<.0001
codepuis	6221.7819	5	35.39	<.0001
codeagevehi	5802.0085	6	419.77	<.0001
codregion	5675.3640	7	126.64	<.0001
codebonu	5319.8159	3	355.55	<.0001

Tableau 8 : Résultats de l'analyse de Type 1 pour le modèle avec les 5 variables explicatives.

L'examen des résultats de Tableau 10 nous permet de diminuer le nombre de variables explicatives. Nous constatons en effet que toutes les modalités de la variable CODEAGEPERMI n'est pas significative. Son omission n'affecte pas le modèle, Nous l'éliminons donc du modèle. Les variables les plus explicatives sont : la région ; l'âge du véhicule ; type moteur et la puissance fiscale ; le bonus et le sexe.

La procédure GENMOD de SAS permet de réaliser la régression de Poisson du nombre total de sinistres matériels sur les 5 variables explicatives présentées plus haut. Les résultats sont présentés ci-dessous. :

Number of Observations Read	27718
Number of Observations Used	27717
Missing Values	1

Informations sur le niveau de classe		
Classe	Niveaux	Valeurs
sexe	2	0 1
codepuis	5	1 2 3 4 5
codeagevehi	6	1 2 3 4 5 6
codregion	7	1 2 3 4 5 6 7
codebonu	3	1 2 3

Analyse des résultats estimés de paramètres							
Paramètre		DF	Estimation	Erreur standard	Wald 95Limites de confiance %	Khi 2	Pr > Khi 2
Intercept		1	-0.7131	0.2734	-1.2490	6.80	0.0091
sexe	0	1	0.4316	0.1671	0.1042	6.68	0.0098
sexe	1	0	0.0000	0.0000	0.0000	.	.
codepuis	1	1	0.2993	0.1195	0.0651	6.27	0.0123
codepuis	2	1	0.3871	0.1165	0.1589	11.05	0.0009
codepuis	3	1	0.3208	0.1173	0.0909	7.48	0.0062
codepuis	4	1	0.5553	0.0967	0.3658	32.98	<.0001
codepuis	5	0	0.0000	0.0000	0.0000	.	.
codeagevehi	1	1	0.3621	0.2546	-0.1369	2.02	0.1549
codeagevehi	2	1	-0.0038	0.2530	-0.4998	0.00	0.9879
codeagevehi	3	1	-0.4516	0.2530	-0.9474	3.19	0.0742
codeagevehi	4	1	-1.2854	0.2584	-1.7919	24.74	<.0001
codeagevehi	5	1	-1.9359	0.3305	-2.5836	34.32	<.0001
codeagevehi	6	0	0.0000	0.0000	0.0000	.	.
codregion	1	1	-0.6631	0.1137	-0.8860	34.00	<.0001
codregion	2	1	-3.0365	1.0018	-5.0000	9.19	0.0024
codregion	3	1	1.1341	0.3388	0.4701	11.21	0.0008
codregion	4	1	0.2589	0.1422	-0.0199	3.31	0.0687
codregion	5	1	-1.4567	0.3827	-2.2068	14.49	0.0001
codregion	6	1	-0.3678	0.0759	-0.5166	23.45	<.0001
codregion	7	0	0.0000	0.0000	0.0000	.	.
codebonu	1	1	-2.0478	0.1064	-2.2564	370.44	<.0001
codebonu	2	1	-2.5893	0.1102	-2.8053	552.03	<.0001
codebonu	3	0	0.0000	0.0000	0.0000	.	.

Tableau 9 : Résultats de la régression Poisson pour le nombre de sinistres matériels.

III.4 Sur dispersion :

Le modèle de Poisson impose des contraintes assez fortes sur la dépendance entre la variable de Comptage Y_{it} et les facteurs de risque X_{it} . Ceci revient donc à supposer l'égalité entre le nombre moyen de sinistre et la variabilité de ce nombre au sein de chaque classe de risque. L'équidispersion est rarement satisfaite en pratique, ce qui met en doute le modèle de Poisson.

Ce tableau nous permet de détecter une éventuelle sur dispersion :

Critère pour évaluer la qualité de l'ajustement			
Critère	DF	Valeur	Valeur/DF
Deviance	28000	40400.0655	1.4429

Tableau 10 : critère pour détecter la sur dispersion

On s'intéresse surtout à la quantité (VALEUR/DF) Doit être très proche de 1

Dans notre cas : D = 1.4429  **Problème de Sur dispersion**

IV. GLM Binomiale négative

Un des candidats les plus efficaces est la régression binomiale négative qui permet de prendre en compte cette sur dispersion en superposant un terme d'erreur aléatoire au prédicteur linéaire. En fait, ceci revient à reconnaître l'hétérogénéité des assurés au sein de chaque classe de tarif ; bien que ces individus soient identiques pour l'assureur, ils présentent néanmoins des profils de risque relativement différents.

IV.1 Extension du modèle de Poisson pour palier au problème de sur dispersion :

La procédure TYPE 1 de sélection des variables explicatives nous donne :

Source	Degrés de lib. num.	Valeur F	Pr > F	Khi 2	Pr > Khi 2
codregion	6	3266.26	<.0001	19597.6	<.0001
codepuis	4	6961.30	<.0001	27845.2	<.0001
codebonu	2	6307.08	<.0001	12614.2	<.0001

Tableau 11 : Test de significativité des variables tarifaires (type 1)

Les résultats de la régression sont présentés au tableau ci-dessous :

Paramètre		DF	Estimation	Erreur standard	Wald 95Limites de confiance %		Khi 2	Pr > Khi 2
Intercept		1	-1.8000	0.0481	-1.8943	-1.7058	1400.51	<.0001
codepuis	1	1	-0.0064	0.0197	-0.0450	0.0321	0.11	0.0441
codepuis	2	1	-0.0487	0.0194	-0.0868	-0.0107	6.32	0.0119
codepuis	3	1	-0.0645	0.0213	-0.1061	-0.0228	9.19	0.0024
codepuis	4	1	-0.0606	0.0170	-0.0938	-0.0273	12.73	0.0004
codepuis	5	0	0.0000	0.0000	0.0000	0.0000	.	.
codebonu	1	1	-0.5335	0.0465	-0.6247	-0.4424	131.59	<.0001
codebonu	2	1	-0.1728	0.0462	-0.2634	-0.0821	13.96	0.0002
codebonu	3	0	0.0000	0.0000	0.0000	0.0000	.	.
codregion	1	1	0.0761	0.0358	0.0060	0.1462	4.53	0.0334
codregion	2	1	-0.0607	0.0262	-0.1120	-0.0094	5.38	0.0204
codregion	3	1	-0.0381	0.0199	-0.0771	0.0010	3.66	0.0559
codregion	4	1	-0.0923	0.0192	-0.1300	-0.0546	23.02	<.0001
codregion	5	1	0.0525	0.0214	0.0105	0.0944	6.00	0.0143
codregion	6	1	-0.0999	0.0213	-0.1416	-0.0582	22.03	<.0001
codregion	7	0	0.0000	0.0000	0.0000	0.0000	.	.

Tableau 12 : Résultat de la régression du modèle binomiale négative pour la fréquence corporelle

Pour modéliser le nombre de sinistres matériels, on procède de la même manière que dans la partie précédente, les coefficients se présentent dans le tableau qui suit :

Analyse des résultats estimés de paramètres						
Paramètre		DF	Estimation	Erreur standard	Wald 95 Limites de confiance %	
Intercept		1	-1.7902	0.0566	-1.9012	-1.6793
codregion	1	1	0.0075	0.0491	-0.0886	0.1037
codregion	2	1	-0.0866	0.0350	-0.1552	-0.0181
codregion	3	1	-0.1262	0.0270	-0.1791	-0.0733
codregion	4	1	-0.0482	0.0252	-0.0976	0.0012
codregion	5	1	-0.0536	0.0293	-0.1111	0.0039
codregion	6	1	-0.2048	0.0292	-0.2620	-0.1477
codregion	7	0	0.0000	0.0000	0.0000	0.0000
codepuis	1	1	-0.0234	0.0267	-0.0757	0.0290
codepuis	2	1	-0.0644	0.0262	-0.1158	-0.0129
codepuis	3	1	-0.0561	0.0284	-0.1118	-0.0004
codepuis	4	1	-0.0685	0.0229	-0.1133	-0.0237
codepuis	5	0	0.0000	0.0000	0.0000	0.0000
codebonu	1	1	-0.8135	0.0545	-0.9203	-0.7067
codebonu	2	1	-0.5289	0.0542	-0.6352	-0.4226
codebonu	3	0	0.0000	0.0000	0.0000	0.0000

Tableau 13 : Résultat de la régression du modèle binomiale négative pour la fréquence matérielle

V. Modèle de Régression logistique pour le nombre de sinistres:

V.1 Description du modèle :

- **Spécification d'un modèle de probabilité**

Dans la régression logistique binaire (dichotomique), on considère une variable cible binaire $Y=0$ ou 1 , et J variables explicatives X_j continues, binaires ou qualitatives.

L'objectif de la régression logistique est celui de toute régression : modéliser l'espérance conditionnelle $E(Y_i / X=x)$ on veut connaître la valeur moyenne de Y_i pour toute valeur de X_i . Pour une valeur Y_i valant 0 ou 1. Cette valeur moyenne est la probabilité que $Y_i = 1$ (loi de Bernoulli). On a donc :

$$E(Y_i / X=x) = 1 * \text{Porba}(Y_i = 1 / X=x) + 0 * \text{Porba}(Y_i = 0 / X=x) = \text{Porba}(Y_i = 1 / X=x)$$

$$= p_i$$

Où P_i se calcule comme suit :

$$p_i = \text{Porba}(Y_i = 1 / X_i = x) = \frac{e^{\beta X_i}}{1 + e^{\beta X_i}}$$

$\beta = (\beta_0, \dots, \beta_j)$ sont les paramètres inconnus du modèle.

- **Évaluation et adéquation du modèle : Test de Wald**

Soit le test suivant : $H_0 : \beta_k = 0$ contre $H_1 : \beta_k \neq 0$.

Ainsi la statistique de Wald s'écrit :

$$W = \frac{\hat{\beta}_k}{s(\hat{\beta}_k)}$$

Avec $s(\hat{\beta}_k)$: l'écart type estimé de $\hat{\beta}_k$.

Sous H_0 : W suit approximativement la loi normale $N(0,1)$.

- **Règle de décision :**

Au niveau de signification α : on rejette H_0 si $W > \phi(1-\alpha)$ avec ϕ : la fonction de réciproque de la loi normale $N(0,1)$.

V.2 Illustration numérique :

Cette étude concerne l'assurance automobile usage A. L'échantillon est constitué de 27 464 assurés observés en 2009 et répartis en deux groupes :

- ❖ Les assurés qui n'ont eu aucun sinistre matériel dans l'année, qui sont au nombre de 26 319 (modalité « 0 sin » de la variable SINIST)
- ❖ Les assurés ayant eu au moins un sinistre dans l'année, qui sont au nombre de 893 (modalité « ≥ 1 sin » de la variable SINIST).

V.2.1 Modélisation de la fréquence matérielle

Sélection des variables par la procédure stepwise pour la modélisation de la fréquence matérielle:

- ✚ Après chaque inclusion de variable explicative, on se demande si une des variables entrées dans le modèle ne pourrait pas être supprimée :

Exemple de l'une des étapes de cette approche :**Step 1. Effect codebonu entered:**

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et
AIC	7370.158	6933.404
SC	7378.388	6958.093
-2 Log L	7368.158	6927.404

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	440.7549	2	<.0001
Score	1020.9202	2	<.0001
Wald	588.3232	2	<.0001

Test du Khi 2 résiduel		
Khi 2	DF	Pr > Khi 2
656.3393	16	<.0001

Les résultats récapitulatifs sont présentés au tableau suivant :

Récapitulatif sur la sélection séquentielle

Étape	Effet		DF	Nombre dans	Khi 2 du score	Pr > Khi 2	Libellé de variable
	Saisi	Supprimé					
1	codebonu		2	1	1020.9202	<.0001	codebonu
2	codregion		6	2	404.6298	<.0001	codregion
3	codeagevehi		5	3	189.2437	<.0001	codeagevehi
4	codepuis		4	4	28.6267	<.0001	codepuis

Tableau 14: Résultats de la sélection séquentielle pour le nombre de sinistres matériels.

L'examen des résultats de Tableau 15 nous permet de diminuer le nombre de variables explicatives. Nous constatons en effet que toutes les modalités de la variable âge_permis et sexe ne sont pas significative. Donc elles n'affectent pas le modèle. Nous les éliminons du modèle.

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
codepuis	4	28.2455	<.0001
codeagevehi	5	175.1130	<.0001
codregion	6	193.8604	<.0001
codebonu	2	424.8498	<.0001

Tableau 15 : Test de significativité des variables tarifaires (type 3)

La modélisation de la fréquence matérielle a été créée grâce à la proc **logistique** (voir Annexes).

Les résultats sont présentés au tableau qui suit :

		Profil de réponse		
Number of Observations Read	27718	Valeur ordonnée	SINIST_M	Fréquence totale
Number of Observations Used	27717	1	0	26900
		2	1	817

Probability modeled is SINIST_M=1.

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	7370.158	6484.026
SC	7378.388	6681.542
-2 Log L	7368.158	6436.026

Paramètre		DF	Estimation	Erreur std	Exp(Est)
Intercept		1	-4.9143	26.2901	0.007
codepuis	1	1	0.0705	0.0844	1.073
codepuis	2	1	0.0395	0.0820	1.040
codepuis	3	1	-0.0455	0.0848	0.955
codepuis	4	1	0.2424	0.0628	1.274
codeagevehi	1	1	0.7890	0.0939	2.201
codeagevehi	2	1	0.3883	0.0898	1.474
codeagevehi	3	1	0.0186	0.0846	1.019
codeagevehi	4	1	-0.6122	0.0945	0.542
codeagevehi	5	1	-1.0473	0.1943	0.351
codregion	1	1	-11.3249	157.7	0.000
codregion	2	1	2.4032	26.2902	11.059
codregion	3	1	1.8542	26.2901	6.386
codregion	4	1	2.8620	26.2900	17.497
codregion	5	1	0.5792	26.2908	1.785
codregion	6	1	1.1311	26.2905	3.099
codebonu	1	1	-0.6049	0.0587	0.546
codebonu	2	1	-1.1597	0.0615	0.314

Tableau 16 : Résultats de la régression LOGISTIQUE pour le nombre de sinistres matériels.

V.2.2 Modélisation de la fréquence corporelle

Pour modéliser le nombre de sinistres corporels, on procède de la même manière que dans la partie précédente :

Récapitulatif sur la sélection séquentielle

Étape	Effet		DF	Nombre dans	Khi 2 du score	Pr > Khi 2
	Saisi	Supprimé				
1	codebonu		3	1	539.5120	<.0001
2	codregion		6	2	28.7625	<.0001
3	codeagevehi		5	3	12.1342	0.0330

Tableau 17 : Résultats de la sélection séquentielle pour le nombre de sinistres corporels.

L'examen des résultats de Tableau 17 nous permet de diminuer le nombre de variables explicatives. Nous constatons en effet que toutes les modalités de la variable age_permi,

codepuis et sexe ne sont pas significatives. Donc elles n'affectent pas le modèle, Nous les éliminons du modèle. Nous en arrivons alors au modèle retenu, lequel est décrit au Tableau qui suit :

Analyse des effets Type 3			
Effet	DF	Khi 2 de Wald	Pr > Khi 2
codeagevehi	5	11.2790	0.0461
codregion	6	14.0174	0.0294
codebonu	2	270.4893	<.0001

Tableau 18 : Test de significativité des variables tarifaires (type 3)

La procédure logistique donne ces résultats :

Profil de réponse		
Valeur ordonnée	SINIST_C	Fréquence totale
1	0	27452
2	1	265

Statistiques d'ajustement du modèle		
Critère	Coordonnée à l'origine uniquement	Coordonnée à l'origine et covariables
AIC	2993.996	2819.184
SC	3002.226	2860.333
-2 Log L	2991.996	2809.184

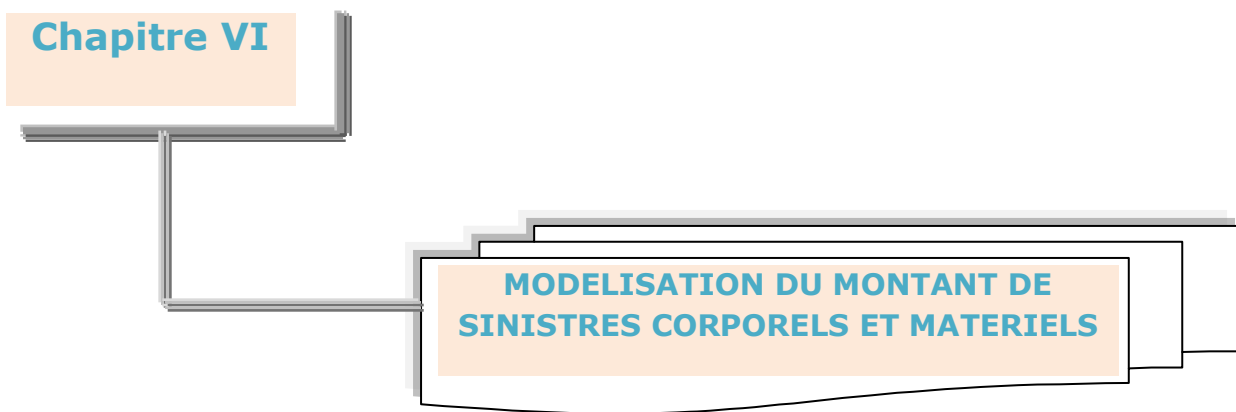
Test de l'hypothèse nulle globale : BETA=0			
Test	Khi 2	DF	Pr > Khi 2
Likelihood Ratio	182.8117	4	<.0001
Score	559.5789	4	<.0001
Wald	296.3999	4	<.0001

Les coefficients du modèle se présentent dans le tableau qui suit :

Paramètre		DF	Estimation	Erreur std	Exp(Est)
Intercept		1	-6.0155	40.4566	0.002
codeagevehi	1	1	0.5187	0.2294	1.680
codeagevehi	2	1	-0.3274	0.2531	0.721
codeagevehi	3	1	0.1835	0.2032	1.201
codeagevehi	4	1	0.3173	0.1989	1.373
codeagevehi	5	1	0.0713	0.2620	1.074
codregion	1	1	-11.5338	242.7	0.000
codregion	2	1	2.2187	40.4566	9.195
codregion	3	1	2.0672	40.4564	7.902
codregion	4	1	2.0674	40.4564	7.904
codregion	5	1	1.0418	40.4568	2.834
codregion	6	1	1.9387	40.4565	6.949
codebonu	1	1	-1.0357	0.0979	0.355
codebonu	2	1	-0.8852	0.0892	0.413

Tableau 19 : Résultats de la régression LOGISTIQUE pour le nombre de sinistres corporels.

Ainsi, la fréquence des sinistres est modélisée selon les modèles : logistique et GLM Binomiale négative, après avoir détecté une sur dispersion dans le présent portefeuille, ce qui a éliminé la GLM poisson du champ d'étude.



L'analyse des coûts de sinistres est sensiblement plus compliquée que celle des nombres. Avant toute chose, il est important de préciser que la variable à expliquer est en réalité la charge de sinistres, soit le montant réellement déboursé par l'assureur.

I. Modélisation du montant des sinistres corporels :

I.1. Descriptif et présentation :

Avant de commencer notre étude, expliquons pourquoi l'analyse des coûts de sinistres est sensiblement plus compliquée que celle des nombres.

Si toutes les polices du portefeuille peuvent servir à estimer la fréquence annuelle de sinistres, il est clair que seules les polices sinistrées peuvent être utilisées en vue d'étudier la loi des montants des sinistres. L'actuaire dispose donc d'un nombre assez limité d'observations à exploiter pour ajuster un modèle aux montants de sinistres. De plus les montants sont beaucoup plus difficiles à modéliser que les nombres, car le phénomène est beaucoup plus complexe.

Souvent, les sinistres de quelque gravité nécessitent des délais assez longs pour être clôturés. Pensez par exemples à un sinistre RC automobile avec dégâts corporels, où il faut attendre que l'état de la victime soit stabilisé avant de pouvoir déterminer le montant des indemnités. Pour cette raison, la compagnie ne dispose bien souvent que de prévisions des coûts dans ses fichiers, ce qui rend l'analyse hasardeuse.

Selon certains auteurs, la prise en compte des montants des sinistres dans la tarification RC automobile est contestable. En effet, les montants payés par l'assureur sont versés aux tiers lésés au titre d'indemnisation de leurs préjudices. Ainsi, un assuré qui renverserait un piéton exposerait son assureur à des dépenses très différentes selon que ce piéton est un vieillard malade seul au monde ou un jeune cadre dynamique père de deux enfants en bas âge. On voit mal comment les caractéristiques de l'assuré pourraient expliquer ce phénomène, même si on comprend bien que le profil de l'assuré puisse expliquer la probabilité de causer un accident avec dégâts corporels. Dans le même ordre d'idées, un assuré qui accrocherait un autre véhicule obligerait son assureur à payer des indemnités très différentes selon que le véhicule tiers est une limousine ou une voiture bas de gamme.

Dans notre portefeuille, 252 assurés ont déclaré au moins un sinistre corporel au cours de l'année 2009. La moyenne du montant de sinistres corporels est de l'ordre de 228884,14Dhs. On peut lire quelques statistiques descriptives au tableau ci-dessous.

Statistiques	Valeurs
Moyenne	228884,14
Médiane	174482,55
Mode	166650,00
Écart-type	196777,98
Variance	38721571744,78
Maximum	1875928,26
Minimum	3759,62
Aplatissement	21,63
Asymétrie	3,70
Étendue	1872168,64
somme	57678802,56

Tableau 20 : Statistiques descriptives du montant des sinistres corporels

Le coefficient d'asymétrie est strictement positif. Cela signifie que la distribution des coûts est décalée vers la droite. Quant au Kurtosis, il indique que la distribution est très aplatie.

I.2. L'écèlement des coûts :

La largeur de la plage des observations suggère une forte influence des montants extrêmes sur le coût moyen, ainsi, il serait injuste de modéliser les coûts tels quels sans écèlement. En effet, les polices qui ont subi ces événements extrêmes seront défavorisées. Il convient donc de partager la charge au-delà d'un certain seuil, appelé seuil d'écèlement, entre toutes les polices.

Mais la question qui se pose est : à partir de quel niveau la charge sinistre est elle considérée comme extrême ?

I.3. Détermination du seuil d'écèlement :

Pour choisir le seuil le plus convenable, on a sélectionné quatre seuils d'écèlement et nous nous sommes donnés comme condition que la loi des montants écélés se rapproche le plus possible de la loi choisie comme loi de modélisation.

Les quatre seuils choisis correspondent aux quatre quantiles suivant :

Quantiles	Montant de sinistres corporels
80%	268043,85
85%	326968,41
90%	411988,13
95%	626653,91

Tableau 21 : Les seuils d'écèlement.

Généralement, le montant des sinistres corporels ou leur gravité, est modélisé selon la loi log-normal ou Gamma, alors, pour vérifier cette affirmation, on a essayé de tracer le Q-Q Plot des montants de sinistres écrêtés, ainsi que celui des résidus.

L'ajustement Q-Q plot est une méthode graphique très simple qui fournit une idée de la distribution la plus adéquate d'une variable.

Elle s'appuie sur la relation linéaire existant entre la fonction de répartition et les quantiles. En effet, elle consiste à produire un nuage de points en plaçant en abscisse les quantiles observés, et en ordonnée les quantiles théoriques de la loi choisie.

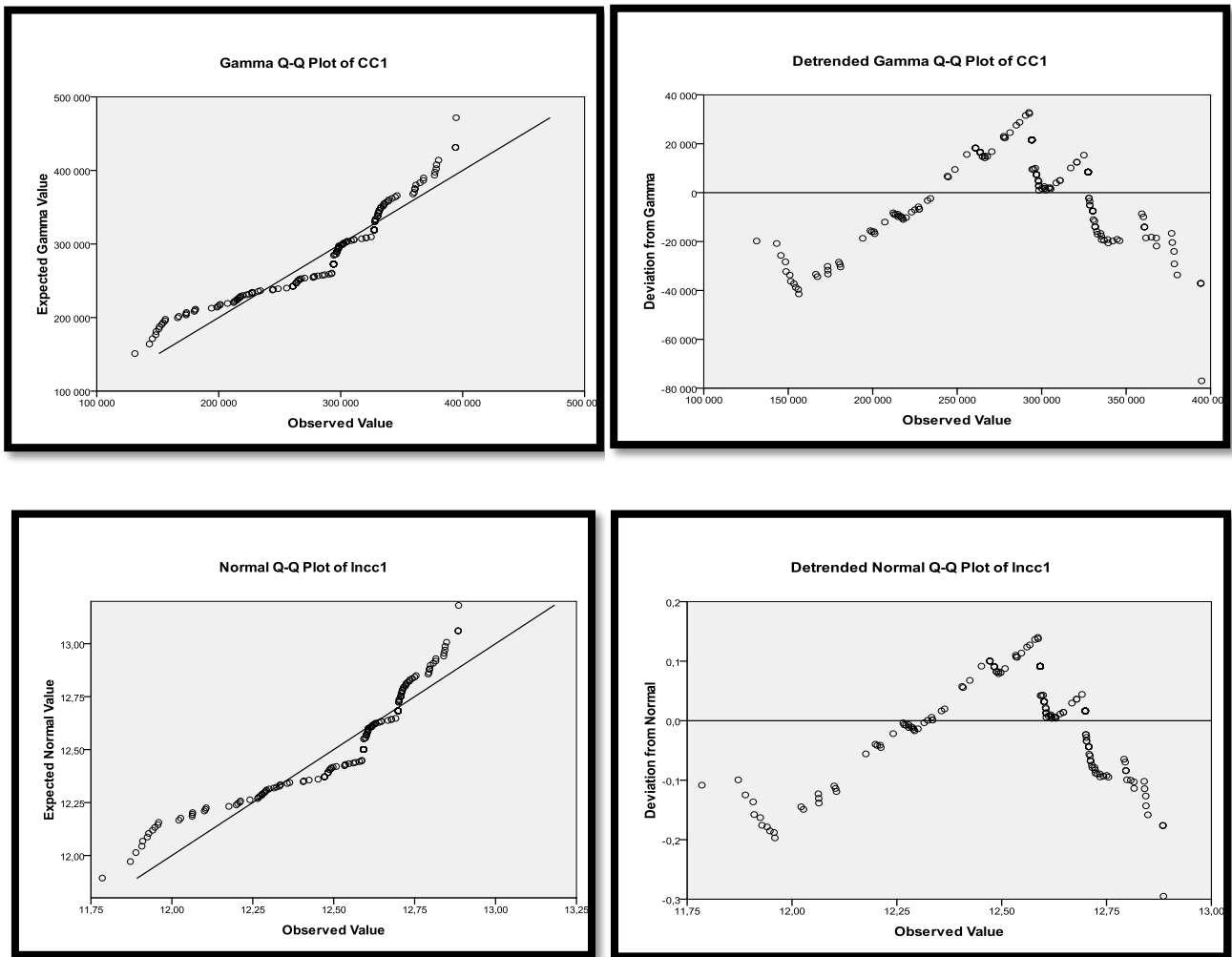


Figure 18 : Ajustement Q-Q plot des montants des sinistres corporels écrêtés ainsi que leurs résidus

Il apparaît de cette analyse Q-Q plot que les points forment presque une droite, on peut conclure que la distribution empirique est compatible avec les deux lois choisies.

Donc La modélisation du montant de sinistres corporels peut être opérée à partir d'une régression Log-normal ou Gamma

I.4. Modélisation du montant de sinistres corporels :

Les variables explicatives utilisées sont:

- **La Région ;**
- **Le Sexe ;**
- **La Puissance fiscale-Combustible ;**
- **L'Age-Ancienneté du permis ;**
- **L'Age du Véhicule ;**
- **Le Bonus.**

I.4.1 Modèle de régression LOG-NORMALE :

On estime dans un premier temps les paramètres de la loi Log-Normale. Une variable est dite Log-Normale si son logarithme suit une loi Normale. Elle est adaptée à la modélisation de coût puisqu'elle permet d'ajuster les phénomènes asymétriques. Nous allons commencer alors à réaliser une régression Log-normal. À l'aide du logiciel SAS, en utilisant la syntaxe (voir Annexes), nous obtenons les tables ci-dessous, ainsi que le tableau des paramètres estimés.

Le tableau ci-dessous illustre des statistiques relatives à la déviance, et ceci pour les quatre seuils du modèle log-normal.

	Valeur	DDL	Valeur/DDL
Déviance q(0,80)	22.5128	180	0.1251
Déviance q(0,85)	29.8149	193	0.1545
Déviance q(0,90)	42,3441	205	0,2076
Déviance q(0,95)	64,8395	216	0,3016

Tableau 22 : Les statistiques relatives à la déviance

Nous obtenons ici que le rapport de la déviance standardisée sur le nombre de degrés de liberté est inférieur à 1 pour les quatre seuils. Néanmoins, la série des coûts qui correspond au seuil égal au quantile 0,8 enregistre une valeur minimale. Nous pouvons donc admettre que le modèle de régression Log-normal avec ce seuil est le plus pertinent (on rappelle que le modèle est acceptable si le rapport de la déviance standardisée sur les degrés de liberté est très petit et n'est pas grand devant 1). Ainsi on écrête au seuil 268043,8580Dhs qui correspond au quantile d'ordre 0,80.

Reste la question de savoir quoi faire de la charge écrêtée (ou sur crête) ? Une solution est de l'ignorer entièrement. Néanmoins la charge sur crête correspond à un cout réel que l'assureur devra payer. Cette méthode n'est donc pas satisfaisante. Une autre approche consiste à répartir les coûts entre tous les assurés. L'idée est que ces très gros sinistres sont largement liés à la malchance et qu'ils auraient tout aussi bien pu arriver à quelqu'un d'autre. Il est donc logique que tous les assurés supportent une part de la charge.

Ainsi, on mutualise la charge sur crête des sinistres entre les contrats en répartissant équitablement la sur crête entre tous les assurés.

Illustration numérique :

Source	DF	Khi 2	Pr > Khi 2
Intercept			
codtyppuis	4	1.84	0.7657
codagevehi	5	7.53	0.1841
codagepermi	5	8.54	0.1287
codbonifi	2	17.72	0.0001
codregion	5	20.99	0.0008

Tableau 23 : Test de significativité des variables tarifaires (Type 1)

Analyse de Type 3			
Source	DF	Khi 2	Pr > Khi 2
codtyppuis	4	0.62	0.9606
codagevehi	5	9.21	0.1008
codagepermi	5	7.19	0.2067
codbonifi	2	17.30	0.0002
codregion	5	20.99	0.0008

Tableau 24 : Test de significativité des variables tarifaires (Type 3)

Nous constatons ici que le rajout des variables : Puissance fiscale-Combustible ; âge-ancienneté permis et âge Véhicule au modèle n'est pas pertinent. Donc, on est amené à rejeter ces variables et ne garder que les variables bonus et région qui sont significatives au seuil de 5%.

Le nouveau modèle donne comme résultats: (Voir code dans Annexes)

Source	DF	Khi 2	Pr > Khi 2
Intercept			
codbonifi	2	23.33	<.0001
codregion	5	17.15	0.0042

Tableau 25 : Test de significativité des variables tarifaires (Type 1)

Analyse de Type 3			
Source	DF	Khi 2	Pr > Khi 2
codbonifi	2	22.59	<.0001
codregion	5	17.15	0.0042

Tableau 26 : Test de significativité des variables tarifaires (Type 3)

On constate que toutes les variables contribuent à l'explication du montant de sinistres corporels, au seuil de 5%.

Les paramètres du modèle Log-normal sont présentés dans le tableau suivant :

Analyse des résultats estimés de paramètres						
Paramètre		DF	Estimation	Erreur standard	Wald 95Limites de confiance %	
Intercept		1	11.6669	0.0691	11.5316	11.8023
codbonifi	1	1	0.0456	0.0715	-0.0946	0.1858
codbonifi	2	1	0.2696	0.0679	0.1364	0.4027
codbonifi	3	0	0.0000	0.0000	0.0000	0.0000
codregion	2	1	-0.0236	0.1057	-0.2307	0.1834
codregion	3	1	0.0113	0.0774	-0.1404	0.1629
codregion	4	1	-0.2090	0.0658	-0.3380	-0.0799
codregion	5	1	0.0992	0.1030	-0.1026	0.3011
codregion	6	1	-0.0993	0.0846	-0.2651	0.0664
codregion	7	0	0.0000	0.0000	0.0000	0.0000

Tableau 27 : Résultats de la régression Log-normal pour le montant de sinistres corporels.

I.4.2 *Modèle de régression Gamma :*

De la même manière que la loi Log-Normale, nous reprenons ici uniquement les grands points de cette analyse.

La syntaxe utilisée sur SAS pour générer les tables ci-dessous, ainsi que d'autres paramètres qui vont nous servir dans cette partie, est dans l'Annexe.

Source	DF	Khi 2	Pr > Khi 2
Intercept			
Codtyppuis	4	4.54	0.3379
Codagevehi	5	12.26	0.3002
Codagepermi	5	5.61	0.3466
Codbonifi	2	2.41	0.0493
Codregion	5	23.74	0.0002

Tableau 28 : Test de significativité des variables tarifaires (Type 1)

Analyse de Type 3			
Source	DF	Khi 2	Pr > Khi 2
codtyppuis	4	5.34	0.2543
codagevehi	5	13.83	0.3357
codagepermi	5	5.45	0.3634
codbonifi	2	2.18	0.0167
codregion	5	23.74	0.0002

Tableau 29 : Test de significativité des variables tarifaires (Type 3)

Nous constatons ici que le rajout des variables : Puissance fiscale-Combustible ; âge-ancienneté permis et âge véhicule au modèle n'est pas pertinent. Donc, on est amené à rejeter ces variables et ne garder que les variables Bonus et région qui sont significatives au seuil de 5%.

Il est à signaler que le rajout des deux variables Bonus et région est pertinent dans les deux modèles, Log-normal et Gamma.

Le nouveau modèle donne comme résultats : (pour code voir Annexes)

L'estimation des paramètres de ce modèle a donné les résultats ci-dessous :

Source	DF	Khi 2	Pr > Khi 2
Intercept			
codbonifi	2	14.75	0.0123
codregion	5	23.67	0.0003

Tableau 30 : Test de significativité des variables tarifaires (Type 1)

On constate que toutes les variables contribuent à l'explication du montant de sinistres corporels, au seuil de 5%.

Les paramètres du modèle Gamma sont présentés dans le tableau suivant :

Analyse des résultats estimés de paramètres						
Paramètre		DF	Estimation	Erreur standard	Wald 95 Limites de confiance %	
Intercept		1	2.6144	1.0200	0.6153	4.6135
codbonifi	1	1	0.5240	0.3889	-0.2382	1.2861
codbonifi	2	1	0.2599	0.2021	0.1362	0.6560
codbonifi	3	0	0.0000	0.0000	0.0000	0.0000
codregion	2	1	-0.4708	0.1869	-0.8371	-0.1046
codregion	3	1	0.2110	1.0064	-1.7616	2.1835
codregion	4	1	0.1403	0.2802	-0.4089	0.6894
codregion	5	1	-0.6822	0.4235	-1.5123	0.1478
codregion	6	1	-0.3608	0.1421	-0.6392	-0.0823
codregion	7	0	0.0000	0.0000	0.0000	0.0000

Tableau 31 : Résultats de la régression Gamma pour le montant de sinistres corporels.

Il nous reste à comparer entre les deux modèles. Pour cela il suffit de comparer juste leurs déviations.

En principe, si le modèle ajuste bien les données observées, la déviance suit approximativement la loi $\chi^2_{(n-p; 1-\alpha)}$; cela suppose asymptotiquement que n soit grand et ce qui est vérifié par notre base de données. Une valeur de déviance très grande indique une qualité mauvaise du modèle.

Mais en pratique, pour juger de la qualité d'un modèle on compare la déviance avec le nombre de degrés de liberté. Plus le rapport est plus petit ou proche de l'unité, le modèle peut être jugé satisfaisant.

	DF	Valeur	Valeur/DF
Déviance de la loi Gamma	211	137,15	0.6500
Déviance de la loi Log-normal	194	24.3858	0.1257

Nous notons un rapport Déviance/(n - p) qui est minimal pour la loi Log-normal, donc ce modèle est jugé le plus adéquat.

I.5. Mesure d'influence :

Parfois l'existence de quelques observations particulières dites aussi points influents, dans un modèle de régression, peut nuire à la qualité de celui-ci.

On dit qu'une observation C_i a de l'influence sur le modèle considéré, lorsqu'une petite modification de C_i ou l'omission de celle-ci conduit à des estimations fort différentes des paramètres du modèle.

Ainsi, ces observations particulières peuvent avoir une influence considérable sur les estimations des paramètres et il est important de recommencer la phase d'estimation après les avoir enlevées, afin de s'assurer de la stabilité des résultats obtenus.

Parmi les tests qui nous permettent de détecter ces observations on trouve : la mesure de Cook. Cette mesure nous renseigne sur l'influence de chaque observation sur le vecteur β des paramètres de la régression.

Pour une régression de :

$$Y = X\beta + e \quad e \propto N(0, \sigma^2 I_n)$$

$$\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$$

La distance de Cook de la $i^{\text{ème}}$ observation, noté D_i et définie de la manière suivante :

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(-i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(-i)})}{ps^2}$$

Tel que : $s^2 = \sum r_i^2 / (n - p) = \hat{\sigma}^2$

$r_i = y_i - \hat{y}_i$ Le $i^{\text{ème}}$ résidu ;

$\hat{\beta}$ Les prédictions avec la $i^{\text{ème}}$ observation ;

$\hat{\beta}_{(-i)}$ Les prédictions sans la $i^{\text{ème}}$ observation ;

p est le nombre de variables explicatives.

Le seuil de tolérance associé à cette distance de Cook le plus souvent employé est la valeur 1. En règle générale, toute observation pour laquelle la distance de Cook est élevée sera retirée de l'étude. La distance de Cook mesure donc le caractère aberrant d'une observation.

	Min	Max	Moyenne	Ecart-type	Nombre d'observations
Cook's Distance Gamma	,000	,075	,005	,009	202
Cook's Distance Log-normal	,000	,065	,005	,009	202

Tableau 32 : La distance de Cook

Nous remarquons que toutes les observations de l'étude sont de plus convenables quand il s'agit de la loi Log-normal.

Le modèle Log-normal est le modèle qui ajuste le mieux la distribution des montants des sinistres corporels.

Les variables retenues dans le modèle final sont : l'âge du véhicule et la région.

II. Modélisation du montant des sinistres matériels :

En analysant la base de données, on a remarqué la répétition de quelques coûts, il s'agit des réserves que l'assureur garde provisoirement. Ainsi pour éviter le risque d'avoir une modélisation biaisée, on a proposé tout simplement de les éliminer dans la phase de la modélisation mais on va les réintégrer dans le calcul du coût moyen en les pondérant par leur poids dans l'échantillon.

coûts(Dhs)	nombre de répétition	poids
2499,75	68	8,19%
2759,72	94	11,33%
4999,5	15	1,81%
5259,47	38	4,58%
9999	14	1,69%
19998	161	19,40%
22757,72	42	5,06%

Tableau 33 : Les coûts répétés ainsi que leurs poids dans l'échantillon

Dans notre portefeuille, 830 assurés ont déclaré au moins un sinistre matériel au cours de l'année 2009. Après l'élimination des coûts répétés il ne reste que 389 observations à modéliser. La moyenne du montant de sinistres corporels est de l'ordre de 34716,6957Dhs. On peut lire quelques statistiques descriptives au tableau ci-dessous.

Statistiques	Valeurs
Moyenne	34716,6957
Médiane	24083,2050
Mode	12758,72
Écart-type	34165,67660
Variance	1,167E9
Maximum	219458,05
Minimum	799,92
Aplatissement	9,084
Asymétrie	2,695
Étendue	218658,13
somme	13817244,87

Tableau 34 : Statistiques descriptives du montant des sinistres matériels

Le coefficient d'asymétrie est strictement positif. Cela signifie que la distribution des coûts est décalée vers la droite. Quant au Kurtosis, il indique que la distribution est très aplatie.

Graphiquement on trouve :

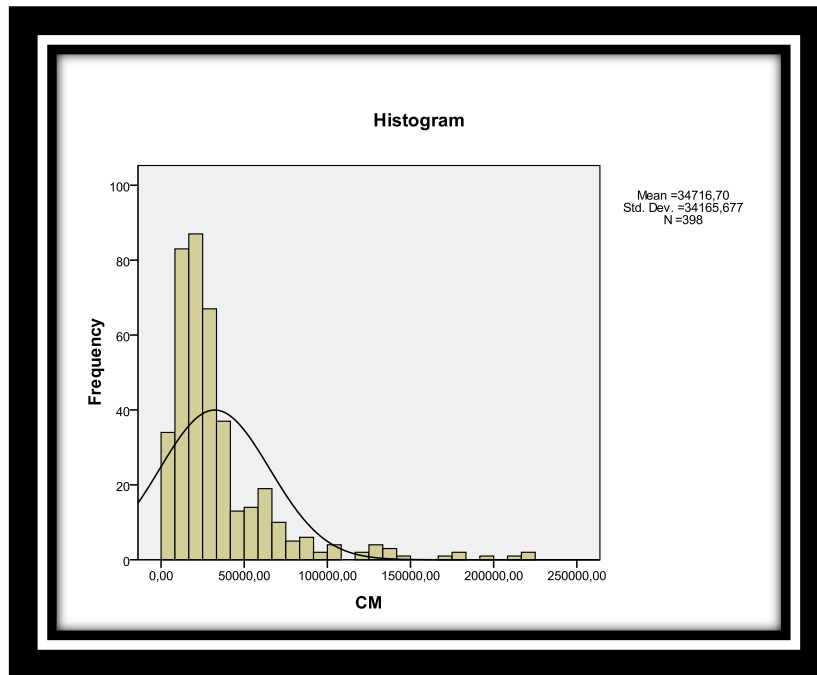


Figure 19 : histogramme des coûts matériels

Comme pour la modélisation du montant de sinistres corporels, nous reprenons ici les grandes lignes de notre analyse pour le montant de sinistres matériels.

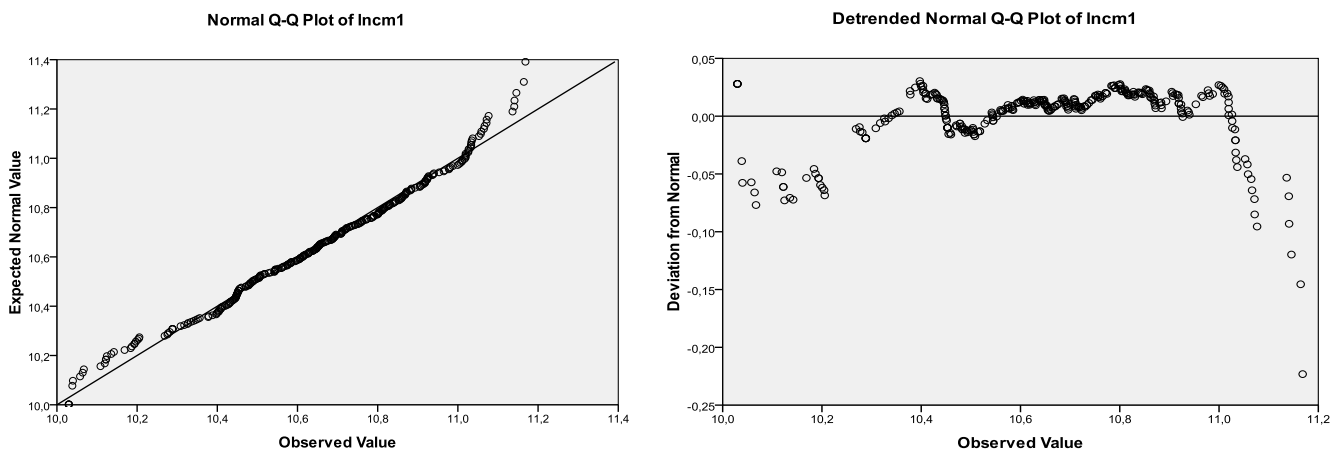


Figure 20 : Ajustement Q-Q plot des montants des sinistres matériels écartés ainsi que leurs résidus

Il apparaît de l'analyse Q-Q plot que la modélisation du montant de sinistres matériels peut être opérée à partir d'une régression Log-normal. On a écriété avec un seuil de l'ordre 49058,6340Dhs qui correspond à la plus petite déviance.

Illustration numérique :

Source	DF	Khi 2	Pr > Khi 2
Intercept			
codtyppuis	4	1.25	0.8691
codagevehi	5	0.64	0.9859
codagepermi	5	8.31	0.1398
codbonifi	2	16.36	0.0003
codregion	5	10.51	0.0321

Tableau 35 : Test de significativité des variables tarifaires (Type 1)

Analyse de Type 3			
Source	DF	Khi 2	Pr > Khi 2
codtyppuis	4	1.41	0.8424
codagevehi	5	2.97	0.7047
codagepermi	5	4.92	0.4257
codbonifi	2	18.20	0.0001
codregion	5	10.51	0.0321

Tableau 36 : Test de significativité des variables tarifaires (Type 3)

Nous constatons ici que le rajout des variables : Puissance fiscale-Combustible ; âge-ancienneté permis et âge Véhicule au modèle n'est pas pertinent. Donc, on est amené à rejeter ces variables et ne garder que les variables bonus et région qui sont significatives au seuil de 5%.

Après élimination des variables non significatives on refait la modélisation, on obtient comme résultats les tableaux ci-dessous :

Source	DF	Khi 2	Pr > Khi 2
Intercept			
codbonifi	2	18.46	<.0001
codregion	5	10.59	0.0601

Tableau 37 : Test de significativité des variables tarifaires (Type 1)

Analyse de Type 3			
Source	DF	Khi 2	Pr > Khi 2
codbonifi	2	19.54	<.0001
codregion	5	10.59	0.0601

Tableau 38 : Test de significativité des variables tarifaires (Type 3)

On constate que toutes les variables contribuent à l'explication du montant de sinistres corporels, au seuil de 5%.

Les paramètres du modèle Log-normal sont présentés dans le tableau suivant :

Analyse des résultats estimés de paramètres						
Paramètre		DF	Estimation	Erreur standard	Wald 95Limites de confiance %	
Intercept		1	9.7563	0.0790	9.6014	9.9112
codbonifi	1	1	-0.0625	0.0757	-0.2109	0.0859
codbonifi	2	1	0.1256	0.0790	-0.0293	0.2805
codbonifi	3	0	0.0000	0.0000	0.0000	0.0000
codregion	2	1	-0.0671	0.0716	-0.2075	0.0733
codregion	3	1	0.0875	0.0746	-0.0587	0.2337
codregion	4	1	0.0532	0.0458	-0.0365	0.1428
codregion	5	1	0.0273	0.1453	-0.2576	0.3121
codregion	6	1	-0.2740	0.1265	-0.5220	-0.0260
codregion	7	0	0.0000	0.0000	0.0000	0.0000

Tableau 39 : Résultats de la régression Log-normal pour le montant de sinistres matériels.

Les mesures sur les influences individuelles de Cook sont rapportées dans le tableau ci-après :

	Min	Max	Moyenne	Ecart-type	Nombre d'observations
Cook's Distance Log-normal	,000	,029	,003	,004	319,000

Tableau 40 : La distance de Cook

Il ressort des résultats du tableau ci-dessus que le $\max(D_i) = 0,029 < 1$, par conséquent aucune observation n'a effet sur l'ensemble et par suite n'influence pas l'estimation des paramètres du modèle.

III. Calcul de la prime commerciale :

Après la finalisation des modèles, on peut maintenant calculer la prime pure, en utilisant la formule suivante :

$$\text{Prime pure} = (\text{FR}_{\text{CORP}} * \text{CM}_{\text{CORP}}) + (\text{FR}_{\text{MAT}} * \text{CM}_{\text{MAT}})$$

Et puisque les chargements employés par Sanad est de 30%, alors on obtient la prime commerciale.

$$\text{Prime commerciale} = 1,3 * \text{Prime pure}$$

Pour un assuré ayant le scénario suivant :

VILLE	CASABLANCA	REGION	4
AGE VEHICULE	10		
PUISSANCE VEHICULE	7		
TYPE MOTEUR	<input checked="" type="radio"/> E <input type="radio"/> D		
CRM	1,2		
LA PRIME	2422,23483		

<< >>

Figure 21 : Interface pour le calcul de la prime commerciale

Sa prime commerciale est de valeur : 2422,234 Dhs .

Conclusion

La tarification dans le secteur des assurances, est le moyen le plus approprié pour mener une réelle politique commerciale bien adaptée aux besoins du marché. Toutefois, elle présente aussi un enjeu actuariel quant à la modélisation et l'évaluation de la prime pure. A cet effet, différents modèles ont été élaborés. L'établissement de ces modèles consistait à lier de la meilleure manière possible des variables disponibles sur les assurés aux variables servant à calculer la prime pure. Globalement, les résultats auxquels nous avons aboutis sont probants. Ainsi la variable la plus importante est la « VILLE ». Elle est suivie de près par le « BONUS ». La puissance fiscale qui est la variable retenue par les organismes publics pour la tarification n'est donc pas la plus importante. Au vu de ces résultats, il serait donc intéressant que le législateur permette aux entreprises de tarifier avec d'autres variables que celles retenues actuellement. Cela serait bénéfique pour les assureurs ainsi que pour les assurés.

PARTIE II

TARIFICATION A POSTERIORI : SYSTEME BONUS MALUS

Introduction

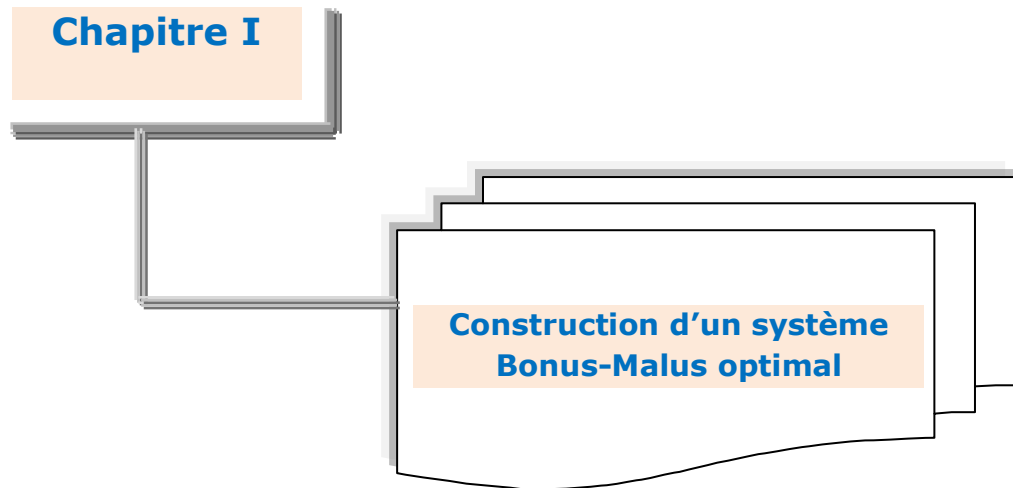
Le principe fondamental de l'assurance est la formation d'une communauté dans laquelle les assurés mettent leurs risques en commun. L'assurance telle qu'elle est pratiquée par les sociétés commerciales n'est équitable que si la prime réclamée à chaque assuré correspond le plus exactement possible au risque réel que celui-ci fait peser sur le portefeuille. En effet, dans ce type d'assurance, seule la mutualisation du risque est souhaitée par les assurés, sans la moindre solidarité. Si l'assuré recherche une certaine solidarité entre les conducteurs, en plus de la mutualisation des risques, il peut alors se tourner vers d'autres acteurs du marché, mutuelles en tête. Ces dernières n'ont pas d'actionnariat à rémunérer et sont donc mues par des considérations plus sociales. Quelle que soit la forme de la société pour laquelle il travaille, la difficulté principale à laquelle est confronté un actuaire 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'expérience et le sexe du conducteur, la puissance du véhicule, le lieu d'habitation par exemple, mais également de facteurs humains difficilement observables (comme le degré d'agressivité, les réflexes, l'abus d'alcool ou la connaissance du code de la route par exemple). Par conséquent, des assurés identiques du point de vue de la compagnie peuvent très bien s'avérer être des risques complètement différents. D'où l'idée de faire appel à un système de personnalisation des primes a posteriori, 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 base.

L'instauration d'un système Bonus-malus poursuit trois buts :

En application d'un tel système, les assurés ont un intérêt objectif à prendre autant de précautions que possible au volant pour éviter des accidents, sinon ils voient leur primes augmenter l'année suivante. A ce titre, le système Bonus-malus est un mécanisme permettant de combattre le risque moral (c'est à dire la tendance naturelle des assurés à prendre moins de précautions lorsqu'ils savent couverts par une police d'assurance). Toutefois, pour faire réussir ce système et pour combattre la concurrence déloyale, chaque assureur doit demander à chaque nouveau assuré, le nombre et la nature des sinistres RC enregistrés chez le dernier assureur, pour fixer la classe correspondante. Ainsi, les mauvais conducteurs n'échapperont pas à ce système de personnalisation des primes.

Ensuite, il s'agit de corriger le montant de la prime afin que celui-ci reflète le risque réel que l'assuré met à la charge des individus qui lui sont identiques sur bases des variables tarifaires.

En fin, il s'agit de diminuer la prime pour les bons conducteurs et de l'augmenter pour les mauvais sans toucher à l'encaissement global de l'assureur. Beaucoup d'assurés ne veulent plus supporter les conséquences de la mauvaise conduite de leurs semblables.



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.

Considérons un assuré observé pendant t années, et notons k_j le nombre d'accidents en tort déclarés au cours de l'année j . L'information dont nous disposons concernant l'assuré est donc un vecteur (k_1, \dots, k_t) .

Les k_j sont les réalisations de variables aléatoires K_j , supposées indépendantes. A tout ensemble d'observations k_1, \dots, k_t nous devons associer un nombre $\lambda_{t+1}(k_1, \dots, k_t)$ qui est le meilleur estimateur de λ à l'instant $t+1$.

Le problème de décision peut donc se formuler comme suit : étant donné une suite de variables aléatoires indépendante K_1, \dots, K_t, \dots , déterminer un ensemble de fonctions $\lambda_{t+1}(k_1, \dots, k_t)$ $t = 0, \dots, \infty$, qui estiment séquentiellement et optimalement λ .

Cette construction d'un système Bonus-Malus peut se présenter comme une suite de jeux statistiques Γ_t opposant la nature et l'actuaire.

Chaque jeu est défini par le triplet $\Gamma_{t+1} = (\Lambda, D_{t+1}, R_{t+1})$ où :

- ✓ Λ : l'espace des stratégies de la nature,
- ✓ D_{t+1} : l'espace des stratégies de l'actuaire à l'époque $t+1$, est une classe de fonctions de décisions $\lambda_{t+1}(k_1, \dots, k_t)$, qui associe à tout vecteur d'observations (k_1, \dots, k_t) un point $\lambda_{t+1} \in \Lambda$;
- ✓ $R_{t+1} = R_{t+1}(\lambda_{t+1}, \lambda)$ est la fonction de risque de l'actuaire à l'instant $t+1$; c'est l'espérance mathématique de la perte $F_{t+1}(\lambda_{t+1}, \lambda)$ qu'il encourt lorsqu'il prend la décision λ_{t+1} alors que la nature est dans l'état λ .

$F_{t+1}(\lambda_{t+1}, \lambda)$ est une fonction non négative de la différence entre λ_{t+1} et λ .

Il vient :

$$R_{t+1}(\lambda_{t+1}, \lambda) = E [F_{t+1}(\lambda_{t+1}, \lambda)] = \sum F_{t+1}(\lambda_{t+1}, \lambda) P(k_1, \dots, k_t \setminus \lambda)$$

en notant \sum la sommation portant sur tous les historiques sinistres, et $P(k_1, \dots, k_t \setminus \lambda)$ la distribution à t dimensions du nombre de sinistres pour un assuré caractérisé par sa fréquence des sinistres λ .

L'ensemble des Γ_t ($t=1, \dots, \infty$) forme le jeu statistique $\Gamma = (\Lambda, D, R)$, où :

- ✓ $D = D_1 \times \dots \times D_t \times \dots$ est le produit cartésien des D_t ,

✓ $R = R(\lambda_1, \dots, \lambda_t, \dots, \lambda) = \sum_{t=1}^{\infty} R_t(\lambda_t, \lambda) = \sum_{t=1}^{\infty} E[F_t(\lambda_t, \lambda)]$
est la perte totale de l'actuaire.

On dit qu'une suite $(\lambda^*_1, \dots, \lambda^*_t, \dots)$ est uniformément optimale si :

$R(\lambda^*_1, \dots, \lambda^*_t, \dots, \lambda) \leq R(\lambda_1, \dots, \lambda_t, \dots, \lambda)$ Pour toute valeur de λ et pour toute suite $(\lambda_1, \dots, \lambda_t, \dots)$.

En règle générale une telle suite n'existe pas. C'est pourquoi l'on décide d'appliquer le critère de Bayes. Soit $u(\lambda)$ la fonction de densité de la fréquence λ . Nous allons minimiser le risque moyen de l'actuaire :

$$R(\lambda_1, \dots, \lambda_t, \dots) = \int_0^{\infty} R(\lambda_1, \dots, \lambda_t, \dots, \lambda) dU(\lambda)$$

Par le théorème de Bayes on définit :

$$dU(\lambda \setminus k_1, \dots, k_t) = \frac{P(k_1, \dots, k_t \setminus \lambda) dU(\lambda)}{\int_0^{\infty} P(k_1, \dots, k_t \setminus \lambda) dU(\lambda)}$$

comme étant la distribution des sinistres pendant les t années d'observation dans le portefeuille.

Nous devons minimiser

$$\begin{aligned} R(\lambda_1, \dots, \lambda_t, \dots) &= \sum_{t=0}^{\infty} \int_0^{\infty} \sum F_t + 1(\lambda_{t+1}, \lambda) P(k_1, \dots, k_t \setminus \lambda) dU(\lambda) \\ &= \sum \sum_{t=0}^{\infty} \int_0^{\infty} F_t + 1(\lambda_{t+1}, \lambda) dU(\lambda \setminus k_1, \dots, k_t) \left(\int_0^{\infty} P(k_1, \dots, k_t \setminus \lambda) dU(\lambda) \right) \end{aligned}$$

Comme la fonction de perte est non négative, il suffit de minimiser pour tout (k_1, \dots, k_t) l'expression :

$$\int_0^{\infty} F_t + 1(\lambda_{t+1}, \lambda) dU(\lambda \setminus k_1, \dots, k_t)$$

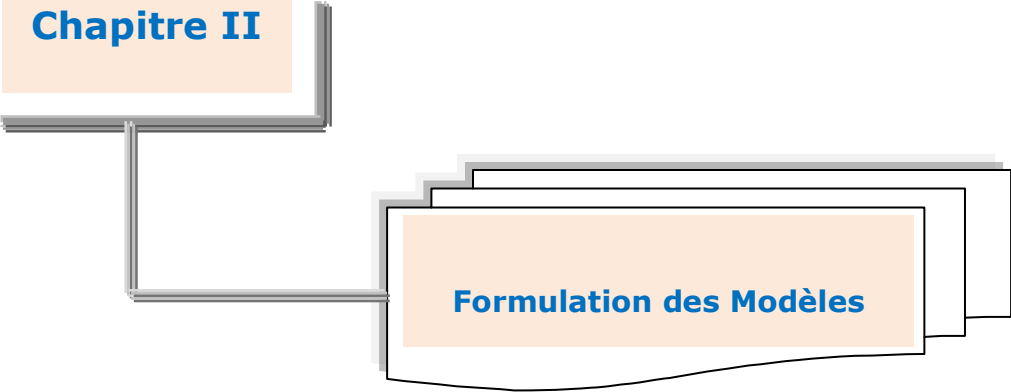
qui n'est rien d'autre que le risque a posteriori de λ .

Si l'on suppose la fonction de perte quadratique,

Le $\min \int_0^\infty (\lambda_{t+1} - \lambda)^2 dU(\lambda \mid k_1, \dots, k_t)$ conduit au système :

$$\lambda_{t+1}(k_1, \dots, k_t) = \int_0^\infty \lambda dU(\lambda \mid k_1, \dots, k_t)$$

C'est l'espérance mathématique a posteriori de λ .



Nous allons essayer dans ce chapitre de donner une formalisation des deux modèles que nous utiliserons dans ce projet de fin d'étude. Le premier est le modèle Poissonien adapté au portefeuille homogène, nous supposons que tous les assurés sont égaux devant le risque ; la survenance des accidents constitue un événement aléatoire et il n'y a pas lieu de pénaliser les assurés responsables de sinistres, ce qui veut dire que l'introduction d'un système Bonus-Malus n'est pas justifié.

Le deuxième est le modèle binomial négatif adapté au portefeuille hétérogène, nous supposons ici que les assurés ne sont pas tous égaux devant le risque, donc l'introduction d'un système Bonus-Malus est justifiée afin que chaque assuré paie, à la longue, une prime correspondante à sa fréquence propre.

I. 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 τ et τ' deux intervalles de temps non disjoints. Alors

$$P[N(\tau) = k \text{ et } N(\tau') = k'] = P[N(\tau) = k] * P[N(\tau') = k']$$

→ 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 poisson de paramètre λ .

On a :

$$\begin{aligned} P_k(t+\Delta t) &= P_k(t) \cdot P[N(t, t+\Delta t) = 0] + P_{k-1}(t) \cdot P[N(t, t+\Delta t) = 1] + \sum_{i=2}^k P_{k-i}(t) \cdot P[N(t, t+\Delta t) = i] \\ &= P_k(t) [1 - \lambda \Delta t + o(\Delta t)] + P_{k-1}(t) [\lambda \Delta t + o(\Delta t)] + \sum_{i=2}^k P_{k-i}(t) \cdot o(\Delta t) \\ &= P_k(t) [1 - \lambda \Delta t] + P_{k-1}(t) \lambda \Delta t + o(\Delta t) \end{aligned}$$

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 :

$$P_k'(t) = -\lambda P_k(t) + \lambda P_{k-1}(t) \quad k = 1, 2, \dots$$

$$P_0'(t) = -\lambda P_0(t) \quad k = 0.$$

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 :

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

En pratique, on estime λ par l'estimateur du maximum de vraisemblance

$\hat{\lambda} = \frac{1}{n} \sum_{k=1}^{k_{\max}} k n_k$ 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 , $n = \sum_{k=1}^{k_{\max}} n_k$

II. 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 sinistres engendrés au cours de l'année suit une loi de poisson de paramètres λ .

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

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

Choisissons comme distribution a priori de λ une Γ de paramètre a et τ , dont la fonction de fréquence est encore appelée fonction de structure :

$$u(\lambda) = \frac{\tau^a \exp(-\tau\lambda)\lambda^{a-1}}{\Gamma(a)} \quad \text{avec}$$

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

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

$$\begin{aligned} P[N=k] &= \int_{\lambda=0}^{\infty} P[N = k \mid \lambda] u(\lambda) d\lambda \\ &= \int_0^\infty \frac{\exp(-\lambda)\lambda^k}{k!} \frac{\tau^a \exp(-\tau\lambda)\lambda^{a-1}}{\Gamma(a)} d\lambda \\ &= \frac{\tau^a}{k! \Gamma(a) (1+\tau)^{(k+a)}} \int_0^\infty \exp(-\lambda(1+\tau)) [\lambda(1+\tau)]^{(k+a-1)} d[\lambda(1+\tau)] \\ &= \frac{\Gamma(k+a)}{\Gamma(k+1)\Gamma(a)} \frac{\tau^a}{(1+\tau)^{(k+a)}} \end{aligned}$$

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

$$\left\{ E[N] = \frac{a}{\tau} \quad \text{et} \quad \text{Var}[N] = \frac{a}{\tau} \left(1 + \frac{1}{\tau}\right) \right\}$$

Les estimateurs de a et τ sont obtenus en résolvant le système suivant :

$$\left\{ \begin{array}{l} \bar{x} = \frac{\tilde{a}}{\tilde{\tau}} \quad \text{et} \quad S^2 = \frac{\tilde{a}}{\tilde{\tau}} \left(1 + \frac{1}{\tilde{\tau}}\right) \\ \text{Donc} \quad \tilde{\tau} = \frac{\bar{x}}{(S^2 - \bar{x})} \quad \text{et} \quad \tilde{a} = \frac{\bar{x}^2}{(S^2 - \bar{x}^2)} \end{array} \right.$$

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

Etant donné les hypothèses du modèle :

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

Le théorème de Bayes s'écrit alors :

$$\begin{aligned} dU(\lambda \mid k_1, \dots, k_t) &= \frac{P(k_1, \dots, k_t \mid \lambda) dU(\lambda)}{\int_0^\infty P(k_1, \dots, k_t \mid \lambda) dU(\lambda)} \\ &= \frac{\frac{\lambda^k \exp(-\lambda t) \tau^a \exp(-\tau \lambda) \lambda^{a-1}}{\prod_{j=1}^t (k_j!) \Gamma(a)} d\lambda}{\frac{\tau^a}{\prod_{j=1}^t (k_j!) \Gamma(a)} \int_0^\infty \lambda^{k+a-1} \exp(-t+\tau)\lambda d\lambda} \\ &= \frac{\lambda^{k+a-1} \exp(-t+\tau)\lambda d\lambda}{\int_0^\infty \lambda^{k+a-1} \exp(-t+\tau)\lambda d\lambda} \\ &= \frac{\tau'^{a'} \lambda^{a'-1} \exp(-\tau \lambda) d\lambda}{\int_0^\infty (\lambda \tau')^{a'-1} \exp(-\tau \lambda) d(\tau \lambda)} \\ &= \frac{\tau'^{a'} \lambda^{a'-1} \exp(-\tau \lambda)}{\Gamma(a')} d\lambda \end{aligned}$$

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 Γ de paramètres a' et τ' , à savoir

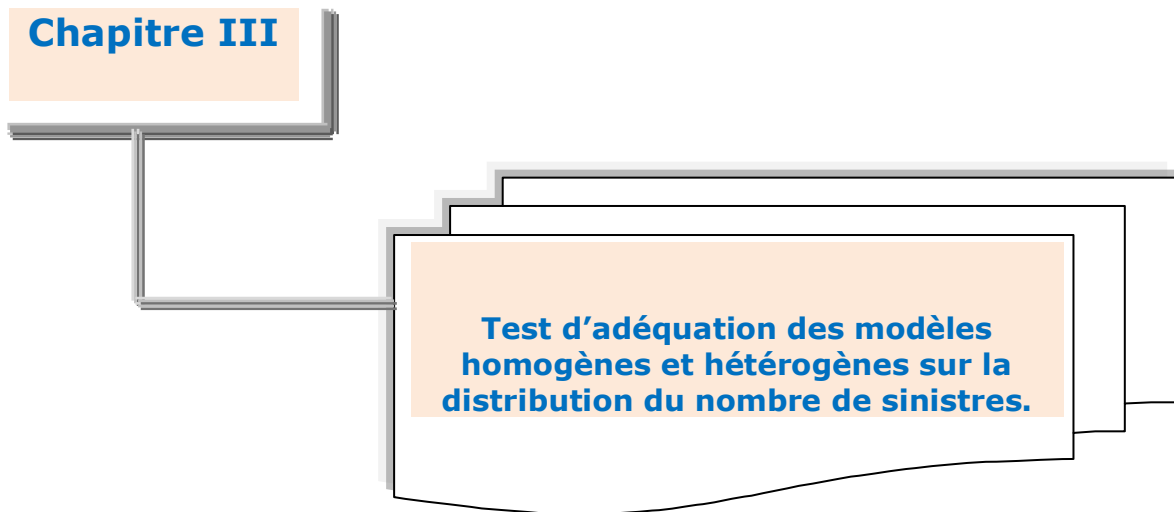
$$\lambda_{\tau+1}(k_1, \dots, k_t) = \frac{a'}{\tau'} = \frac{a+K}{t+\tau}$$

III. Modélisation par les processus Markoviens :

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 trois éléments :

- i. Un tableau des primes b_1, b_2, \dots, b_s (exprimées en pourcentage de la prime de base qui est calculée a priori) qui correspondent aux différents degrés de l'échelle Bonus-Malus.
- ii. le degré initial i_0 , affecté à un nouvel assuré lors de son entrée dans le système, et
- iii. la loi régissant le passage d'un niveau à un autre dans l'échelle. C'est la règle de transition. (qui sera détaillée dans le chapitre 'Elaboration des échelles Bonus-Malus')



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 χ^2 d'adéquation que nous décrivons ci-dessous.

I. Le test χ^2 d'ajustement des lois :

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 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 $f_0(X, \theta)$, θ 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 $X = x_i$ $i = 1, 2, \dots, k$.

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 :

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

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

II. Modèle poissonien de portefeuille homogène:

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

Nous utiliserons le test d'adéquation χ^2 pour tester l'ajustement du modèle Poissonien sur notre portefeuille qui est divisé en trois natures de sinistre, à savoir :

- Sinistres matériels.
- Sinistres corporels.
- Sinistres RC (sans distinction entre la nature de sinistre).

II.1 Sinistres corporels :

Le tableau suivant présente la distribution du nombre de sinistres matériels du portefeuille observé. Sa moyenne vaut **0,0319** et sa variance **$S^2 = 0,036$**

Nombre de sinistres corporels K	Effectifs observés nk	Probabilité pk	Effectifs théoriques n*pk	Test CHI 2
0	27452	0,990341709	27449	0,000265356
1	262	0,009611499	266	0,072735574
2	2	4,66E-05	1	0,386933779
3	1	1,51E-07	0	237,116409
>3	0			
Total	27717			237,5763437

Tableau 41 : sinistres corporels Poisson homogène

La méthode du maximum de vraisemblance et la méthode des moments nous conduisent à estimer le paramètre λ de la distribution par la moyenne observée. En ajustant la distribution observée par une distribution de Poisson de paramètre $\hat{\lambda} = 0,009705235$, nous obtenons des effectifs théoriques différents des effectifs observés. Visiblement, l'ajustement n'est pas bon.

Le test ($\chi^2_{\text{calc}} = 237,57 > \chi^2_{3,0.05} = 7,81$) confirme la très mauvaise qualité de l'ajustement, ce qui nous amène à rejeter le modèle. L'hypothèse d'homogénéité du portefeuille, constitué uniquement des sinistres matériels, ne résiste pas à l'analyse statistique.

Une autre manière de tester l'hypothèse d'homogénéité du portefeuille est de procéder à une comparaison entre la moyenne et la variance du portefeuille. En effet, la moyenne de notre portefeuille vaut 0,0319 qui est inférieure à la variance $S^2 = 0,036$, ce qui veut dire une sur dispersion des données, traduisant ainsi le rejet du modèle de Poisson dans le cas des sinistres matériels.

II.2 Sinistres matériels :

Dans le cas des sinistres matériels on a :

$$\chi^2_{\text{calc}} = 431,58 > \chi^2_{3,0.05} = 7,81$$

Ce qui confirme, là aussi, la mauvaise qualité de l'ajustement par un modèle de Poisson.

$$\lambda = 0,031893784$$

Nombre de sinistres matériels K	Effectifs observés nk	Probabilité pk	Effectifs théoriques n*pk	Test CHI 2
0	26900	0,968609459	26847	0,10483409
1	759	0,03089262	856	11,04549164
2	50	0,000492641	14	96,74384926
3	7	5,24E-06	0	323,6921811
>3	0			
Total	27717			431,5863561

Tableau 42 : sinistres matériels Poisson homogène

II.3 Sinistres RC :

Nous avons vu grâce aux deux précédentes analyses que le fait de considérer que la sinistralité matérielle ou corporelle d'un assuré est due au simple fait du hasard et par conséquent ne doit pas être sanctionné, ne résiste pas aux tests statistiques.

Faisons donc un tel test pour tout le portefeuille sans tenir compte de la catégorie de sinistres et voyons quelle conclusion nous pouvons tirer quant à l'homogénéité des comportements face au risque dans notre portefeuille.

La distribution de sinistres dans le portefeuille et son ajustement par une loi de poisson dont le paramètre est estimé par la fréquence moyenne peuvent être résumés par le tableau suivant :

$$\lambda = 0,041599019$$

Nombre de sinistres K	Effectifs observés nk	Probabilité pk	Effectifs théoriques n*pk	Test CHI 2
0	26648	0,959254347	26588	0,13697312
1	996	0,039904039	1106	10,94415571
2	63	0,000829984	23	69,53479941
3	9	1,15E-05	0	236,2448994
>4	0			
Total	27717			316,8608276

Tableau 43 : tous les sinistres Poisson homogène

$$\chi^2_{\text{calc}} = 316,86 > \chi^2_{3,0.05} = 7,81$$

On voit qu'au sein du portefeuille, la qualité de l'ajustement par une poisson Homogène est très mauvaise. Le test Khi deux rejette l'hypothèse. Les résultats de ces tests constituent

une preuve que le modèle de distribution homogène n'est pas conforme aux réalités du portefeuille. Ceci étant, pour justifier l'utilisation d'une tarification personnalisée, il faut plus.

C'est pour cette raison que nous nous proposons dans ce qui suit de rechercher des modèles de distribution hétérogènes qui résistent à l'analyse statistique et modélisent avec un niveau de confiance considérable la distribution de sinistres de notre portefeuille.

III. Modèle Binomial négative de portefeuille hétérogène :

Le modèle le plus fréquent utilisé pour la description de l'hétérogénéité des portefeuilles est le modèle binomial négative. 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.

➤ Cas de tous les sinistres

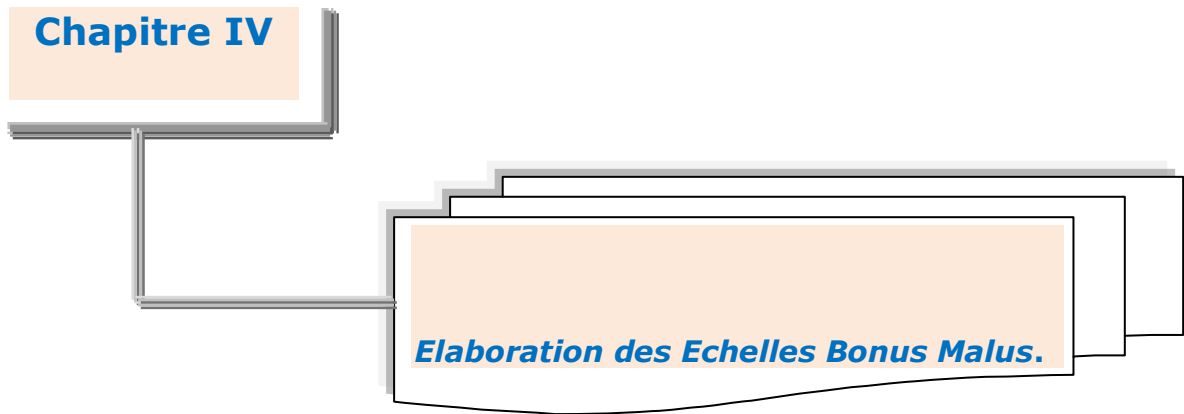
L'observation et les valeurs prédites peuvent être résumées dans le tableau suivant :

Nombre de sinistres K	Effectifs observés nk	Probabilité pk	Effectifs théoriques n*pk	Test CHI 2
0	26648	0,95253265	26401,34746	2,304332215
1	996	0,04412347	1222,970218	0,012324969
2	63	0,0020262	56,1601854	0,833029012
3	9	9,264E-05	2,567703048	2,113406921
>4	0			
Total	27717			5,263093117

Tableau 44 : tous les sinistres BN hétérogène

$$\chi^2_{\text{calc}} = 5,263 < \chi^2_{3,0.05} = 7,81$$

Donc dans le cas de tout le portefeuille on accepte le test d'ajustement par une binomiale négative hétérogène. On voit d'ailleurs la proximité entre les valeurs prédites 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 négative.



I .Construction d'un système bonus malus avec personnalisation a priori :

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 peut choisir 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:

$$\begin{aligned} T_0(m) &= \max(1; m - 1) \\ T_k(m) &= \min(s; m + 1) ; k > 0 \end{aligned}$$

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 est arrivé à cette position précédente) et du nombre de sinistres qu'il a commis dans la dernière période d'observation.

II.Par modèle Binomial négative :

On commence par estimer les paramètres a et τ de la loi Binomiale Négative sur base du portefeuille pour lequel on désire créer une échelle Bonus-Malus. Ensuite, la pénalisation que

subit un assuré qui a causé k sinistres en t années est : $\frac{\frac{a+k}{\tau+t}}{\frac{a}{\tau}}$ exprimée comme un

pourcentage de la prime de base $\mu = \frac{a}{\tau}$. Si la prime proposée à un nouvel assuré est par

exemple de 100 Dirhams, la prime a posteriori ainsi normée vaudra : $\frac{\frac{a+k}{\tau+t} 100}{\frac{a}{\tau}} = 100$

$\frac{\tau(a+k)}{a(t+\tau)}$ pour un assuré après la t ième année avec k sinistres.

En remplaçant a et τ par leurs estimations, nous obtenons un système Bonus-Malus optimal adapté au comportement des automobilistes assurés chez notre compagnie. Remarquons que dans tous les cas, la prime a posteriori dépend uniquement du nombre total des sinistres. Elle ne tient pas compte de la répartition des sinistres au fur des années. Sur le plan pratique ce n'est pas souvent le cas et la plus part des systèmes Bonus Malus du monde tiennent compte de cette répartition.

Illustration numérique :

Les paramètres estimés à partir des observations dans le portefeuille peuvent être résumé dans le tableau suivant : on observe que la fréquence moyenne 0,0416 est de et la variance 0,047

a	1,038227997
τ	7,703703704
moyenne	0,0416
variance	0,047

Tableau 45 : estimation des paramètres du modèle Binomial Négative

L'échelle Bonus Malus obtenue par le modèle Binomial Négative avec les paramètres estimés ci-dessus est représentée dans le tableau suivant : nous nous sommes limités à 6 sinistres et la plus longue durée considérée est de 11 ans. La sévérité de cette échelle en cas de sinistre est frappante, par rapport à ce qui se passe en réalité. Un sinistre entraîne une majoration de 74% tandis que le tarif actuel applique au maximum 30%. Les bonifications sont aussi plus importantes car pour une durée de deux ans sans sinistre l'assuré gagne 11,5% de réduction la première année et presque 21% la deuxième année, tandis que le système actuel accorde 10% pour les deux ans sans bonifications pour une seule année sans sinistres. Donc ce système est certes sévère mais cette sévérité peut être justifiée par deux éléments : le premier est qu'une grande partie de cette sévérité st due à la bonification qu'il propose aux bons assurés. Le deuxième élément est que l'on ne tient pas compte dans cette étude des réalités commerciales.

t/K	0	1	2	3	4	5	6
0	100						
1	88,51	173,76	259,01	344,27	429,52	514,77	600,02
2	79,39	155,86	232,32	308,79	385,25	461,72	538,19
3	71,97	141,29	210,62	279,94	349,26	418,58	487,91
4	65,82	129,22	192,62	256,02	319,42	382,82	446,22
5	60,64	119,05	177,46	235,87	294,28	352,68	411,09
6	56,22	110,36	164,51	218,66	272,80	326,95	381,09
7	52,39	102,86	153,32	203,78	254,25	304,71	355,18
8	49,06	96,31	143,56	190,81	238,06	285,31	332,56
9	46,12	90,54	134,96	179,38	223,81	268,23	312,65
10	43,51	85,43	127,34	169,25	211,16	253,08	294,99
11	41,19	80,86	120,53	160,20	199,87	239,55	279,22

Tableau 46 : Système Bonus malus obtenue par le modèle Binomiale Négative

III. Présentation Markovienne :

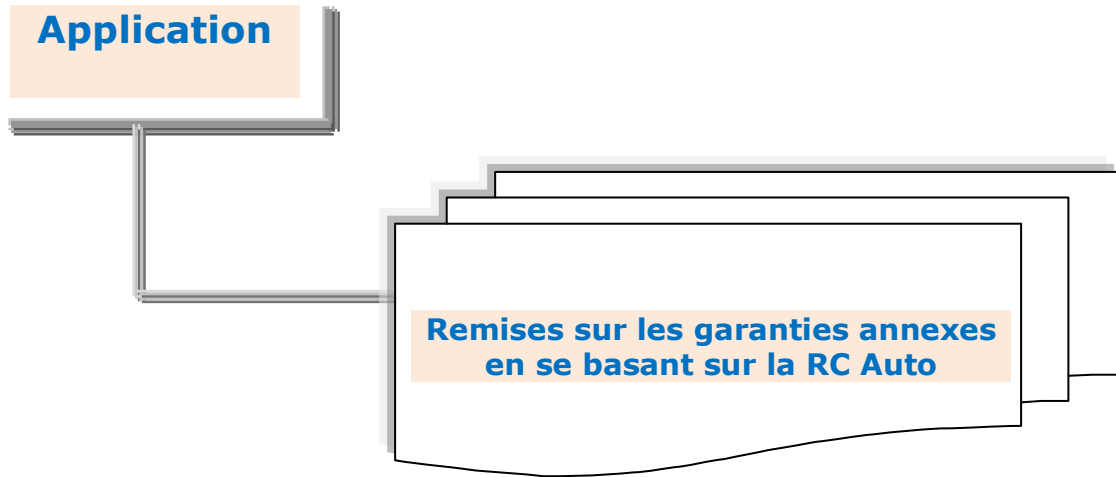
La présentation précédente est simple mais pas très pratique pour affecter des garanties annexes gratuites pour les gens qui méritent en se basant sur la RC, c'est pour cette raison que nous allons simplifier l'échelle et la représenter en un nombre fini de classe sans la faire perdre ses principales caractéristiques. Nous nous limiterons à 7 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 : 85% (minimum) et 119% (maximum) de la prime de base qui correspond respectivement au niveau 1 et 7 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.

prime(%)	Degré	Après 0 sinistre	1 sinistre
119	7	6	7
110	6	5	7
103	5	4	7
100	4	3	6
96	3	2	5
91	2	1	4
85	1	0	3

Tableau 47 : formulation markovienne de l'échelle Binomiale Négative



L'application que nous avons développée a deux objectifs principaux : calcul de la prime commerciale et le calcul de la remise sur les garanties annexes.

La figure suivante présente l'interface utilisée dans la première page :



Figure 22: La page d'accueil en VB

Alors pour commencer, il faut cliquer sur la commande « **vo**tre simulation », une simulation s'affiche automatiquement au premier lancement :

REGION	4
AGE VEHICULE	10
PUISSANCE VEHICULE	7
TYPE MOTEUR	E
CRM	1,2
La prime	2422,23483679899

>>

Figure 23 : Interface de calcul de la prime commerciale

Un clique sur le bouton permet de passer à une deuxième interface où il faut remplir la fiche d'information suivante :

VILLE	CASABLANCA	REGION	4
AGE VEHICULE	10		
PUISSANCE VEHICULE	7		
TYPE MOTEUR	<input checked="" type="radio"/> E <input type="radio"/> D		
CRM	1,2		
LA PRIME	2422,23483		

<< >>

Figure 24 : le calcul de la prime commerciale

Une fois que tous les champs de l'écran précédent sont remplis, on presse le bouton « **La prime** » pour obtenir la prime commerciale. Pour le calcul de la remise sur les garanties annexe, on presse le bouton on obtient une nouvelle fenêtre dans laquelle on doit préciser s'il s'agit d'un nouveau assuré ou tout simplement d'un assuré existant déjà dans le portefeuille de la compagnie.

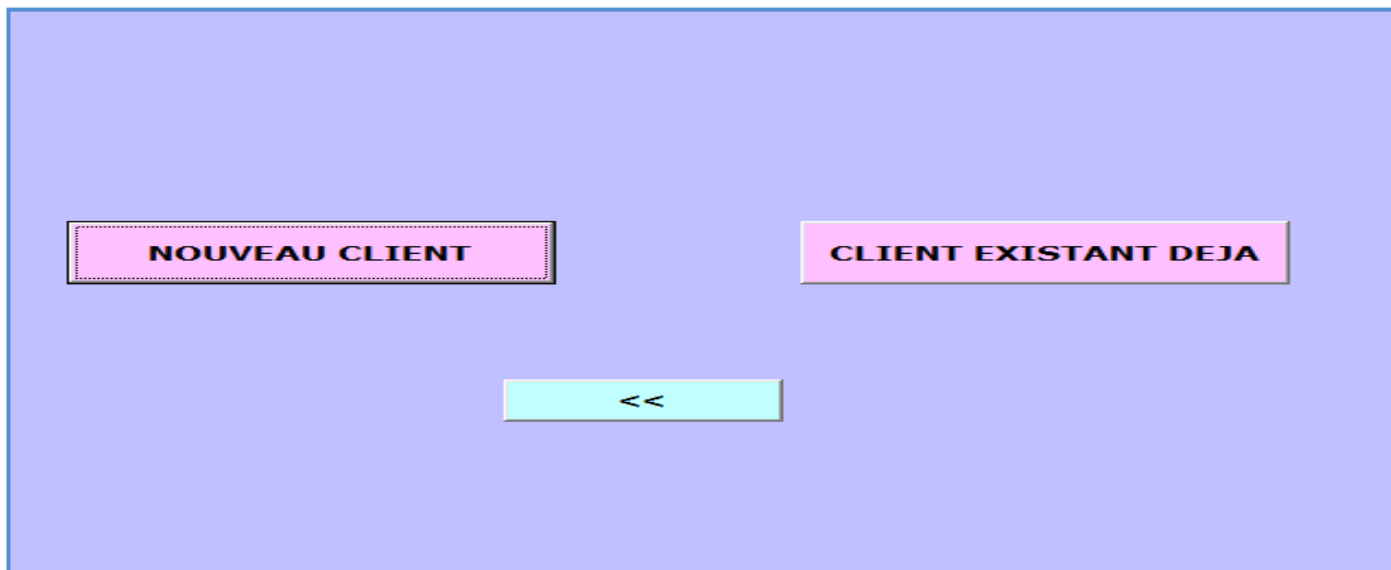


Figure 25 : détermination du type de client

Pour le cas d'un « nouveau client » :

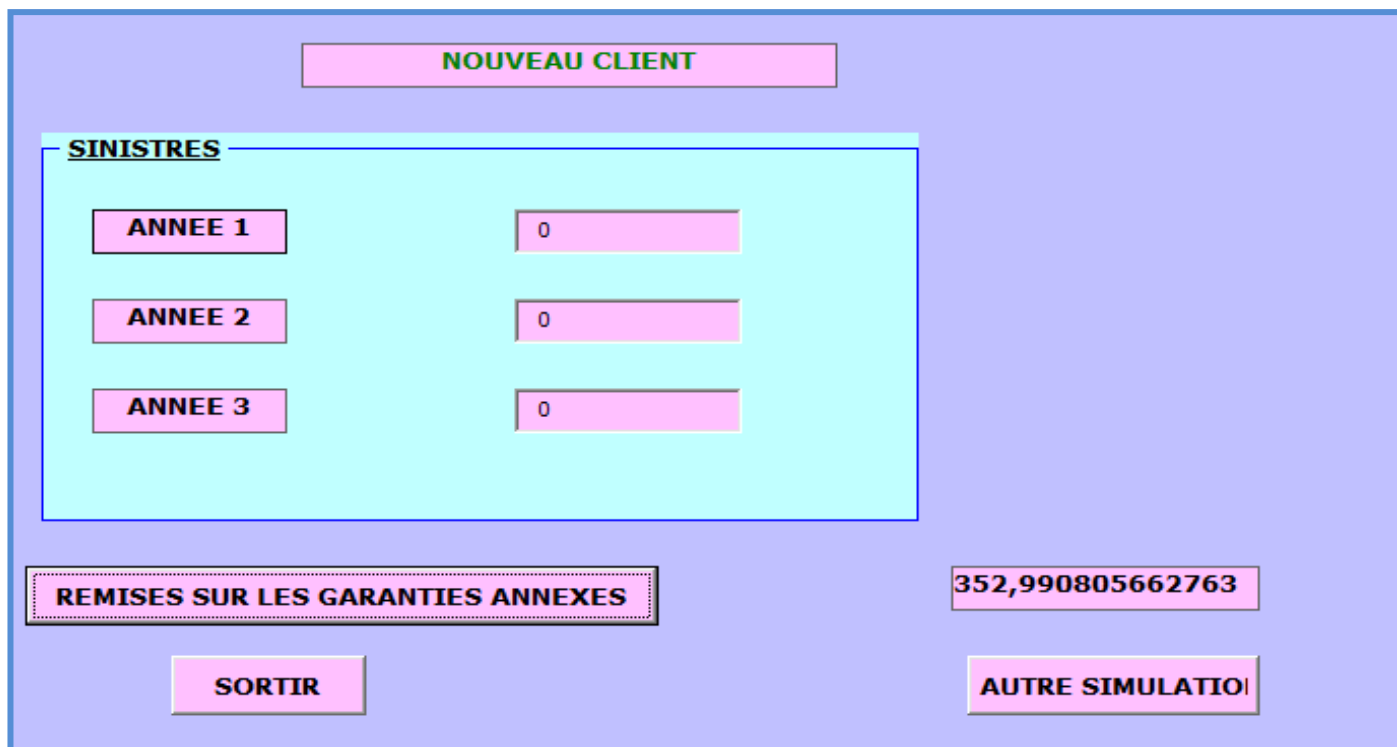


Figure 26 : calcul de la remise pour un nouveau client

Il suffit de remplir les cases représentant son historique des sinistres durant trois années successives et cliquer ensuite sur le bouton «remises sur les garanties annexes» pour obtenir la remise propre à ce client.

Pour le cas d'un « client de Sanad » :

The screenshot displays a simulation interface for a Sanad client. At the top, a yellow box contains the text "CLIENT SANAD". Below this, there are four rows of input fields and their corresponding values:

NUMERO CLIENT	368505
CODE INTERMEDIARE	548
SINISTRES	0
REMISES SUR LES GARANTIES	352,990805662763

At the bottom of the interface, there are two buttons: "SORTIR" on the left and "AUTRE SIMULATION" on the right.

Figure 27 : calcul de la remise pour un client de Sanad

Il suffit d'indiquer le numéro du client ainsi que le code de l'intermédiaire pour déterminer le degré de l'échelle Bonus-Malus de l'année précédente, dès qu'on remplit la case des sinistres effectués cette année, le simulateur détermine automatiquement le nouveau degré affecté au client. A la base de ce dernier, nous pouvons calculer la réduction que nous offrirons au client au niveau des garanties annexes.

CONCLUSION

Avec la mise en place d'un système Bonus-malus basé uniquement sur le nombre de sinistre (RC auto), notre compagnie d'assurance peut procéder à une correction de la prime selon le risque réel que représente individuellement chaque assuré.

Il a été en particulier possible de proposer une approche théorique du problème qui constitue un support indispensable pour l'application pratique.

Il a été aussi possible de prouver l'hétérogénéité du portefeuille et de proposer un modèle de distribution hétérogène s'ajustant de manière excellente à la distribution de sinistres dans le portefeuille.

Il nous a été permis d'élaborer un système Bonus-malus pour toute nature de sinistre, basé sur le modèle Binomial Négatif. Nous avons choisi ce modèle pour sa simplicité et son utilisation par les professionnels des assurances dans le monde.

Nous nous sommes basés dans ce projet sur la garantie Responsabilité Civile et nous avons affecté des remises au niveau des garanties contractuelles (Vol, Incendie, Bris de glace, Tiers, ...etc) pour les bons conducteurs, comme le montrait l'application présentée dans le dernier chapitre.

CONCLUSION GÉNÉRALE

Au début de ce projet nous nous sommes fixé comme objectif de proposer un système de tarification a posteriori pour une composante du portefeuille automobile de la compagnie *Sanad*. Ce système devrait proposer une mesure corrective adéquate à l'hétérogénéité suspectée par la segmentation et la tarification a priori effectuée précédemment sur le portefeuille.

Nous avons proposé des tarifs personnalisés pour les assurés du portefeuille et pour les nouveaux assurés en se basant sur un historique de trois ans. Malgré ces importantes réalisations, notre travail s'achève sans avoir parcouru toutes les pistes possibles, ni avoir proposé toutes les solutions adéquates au pertinent problème que constitue la tarification personnalisée. Il reste des modèles de distribution hétérogènes dont nous pouvons tester l'adéquation avec la distribution de sinistre dans notre portefeuille comme le modèle Good-Risk / Bad-Risk ainsi que d'autres modèles utilisés fréquemment par les professionnels des assurances.

En ce qui concerne la base de données nous pensons que la taille était largement suffisante pour que nos résultats soient fiables. Il est donc envisageable de procéder à une adaptation commerciale du système. Il faut enfin souligner l'importance de l'étude de l'influence de l'introduction d'un système, tel que celui qu'on a proposé dans ce rapport, sur les assurés. Cette étude devrait être une continuité de notre travail et être en mesure de déterminer la stratégie optimale des assurés en terme de rétention. En effet, avec un système ne tenant pas compte des montants de sinistres, il faut s'attendre à un développement de la « soif de Bonus » qui empêcherait aux assurés de déclarer les petits sinistres. Il serait alors, commercialement, plus ingénieux de proposer franchise.

ANNEXES

La macro pour calculer la durée d'exposition au risque :

```
Public Function Macro1(d1 As Date, d2 As Date) As String
Dim r As Integer
Dim dDebut As Date
Dim dFin As Date

dDebut = #1/1/2009#
dFin = #12/31/2009#

If (DateDiff("d", dDebut, d2) < 0) Or (DateDiff("d", d1, dFin) < 0) Then
    Macro1 = ""
Else
    If (DateDiff("d", d1, dDebut) > 0) Then d1 = dDebut
    If (DateDiff("d", dFin, d2) > 0) Then d2 = dFin

    Macro1 = DateDiff("d", d1, d2) + 1
    'Macro1 = d1
End If
End Function
```

La corrélation a été effectuée grâce a la PROC CORR suivante :

```
PROC SORT DATA=moi.base out=moi.SANAD;
BY sexeindi;
RUN;

proc corr data=moi.SANAD noprob plots=matrix;
var age_condu anc_permi anc_vehi anc_vehi Crm_Nive Codevill Puisvehi
;
```

Modélisation de nombres de sinistres matériels en utilisant la loi poisson :

```
proc genmod data=sanad;
CLASS sexe codepuis codeagevehi codregion codebonu codeagepermi ;
model NSM= sexe codepuis codeagevehi codregion codebonu
/ link=LOG dist=POISSON type1 ;
run;
```

Modélisation de nombres de sinistres matériels en utilisant la loi Binomiale négative :

```
proc genmod data=sanad;
class sexe codepuis codeagevehi codregion codebonu;
model NSC = sexe codepuis codeagevehi codregion codebonu
/ link=LOG dist=NEGBIN PSCALE OFFSET=dur_exp ;
run;
```

Sélection des variables par la procedure stepwise :

```
proc stepwise data= sanad;
class sexe codepuis codeagevehi codregion codebonu codeagepermi;
model SINIST_M(event="1") sexe codepuis codeagevehi codregion codebonu;
/ rule=single expb; run;
```

La modélisation de la fréquence matérielle a été créée grâce à la proc logistique :

```
proc logistic data=sanadbase;
class codepuis codeagevehi codregion codebonu;
model SINIST_M(event='1')= SINIST_M(event="1") sexe codepuis codeagevehi
codregion codebonu;
/ link=glogit;
run;
```

Modélisation du coût corporel en utilisant la loi log-normal :

```
proc genmod data =sasuser.CC1;
class codtyppuis codagevehi codagepermi codbonifi codregion ;
model LNCC1 = codtyppuis codagevehi codagepermi codbonifi codregion
/ dist = normal link = identity offset = dexpo type1 type3;
run;
quit;
```

Le nouveau modèle est :

```
proc genmod data =sasuser.CC1;
class   codbonifi codregion ;
model LNCC1 = codbonifi codregion
/ dist = normal link = identity offset = dexpo type1 type3;
run;
quit;
```

Modélisation du coût corporel en utilisant la loi Gamma :

```
proc genmod data =sasuser.CC1;
class   codtyppuis codagevehi codagepermi codbonifi codregion ;
model CC1 = codtyppuis codagevehi codagepermi codbonifi codregion
/ dist = gamma link = identity offset = dexpo type1 type3;
run;
quit;
```

Le nouveau modèle est :

```
proc genmod data =sasuser.CC1;
class   codagevehi codregion ;
model CC1 = codagevehi codregion
/ dist = gamma link = identity offset = dexpo type1 type3;
run;
quit;
```


BIBLIOGRAPHIE

Michel DENUIT (2001) : « Assurance Non-vie »

Arthur CHARPENTIER « Actuariat Non-vie », cours 2003-2004

Jean LEMAIRE (1995) « Bonus Malus Systems in Automobile Insurance»

Dominique HERRIET , Jean-Charles ROCHET : « La logique des systèmes bonus -malus en assurance automobile»

Claude Fluet : « Tarification et inter financement dans les régimes obligatoires d'assurance automobile »

MICHEL DENUIT, SANDRA PITREBOIS et JEAN-FRANC «Méthodes de construction de systèmes bonus-malus en RC Auto»

Lemaire J. 1995. «Bonus- Malus Systems in Automobile Insurance »

Richaudeau D. 1996. « Modélisation du risque d'accident automobile »

Sandra Pitrebois , Michel Denuit Jean-François Walhin« Marketing et Systèmes Bonus-Malus»

Olfa N. Ghali : « Un modèle de tarification optimal pour l'assurance automobile dans le cadre d'un marché réglementé : application à la Tunisie»

GEORGES DIONNE and CHARLES VANASSEA : « GENERALIZATION OF AUTOMOBILE INSURANCE RATING MODELS: THE NEGATIVE BINOMIAL DISTRIBUTION WITH A REGRESSION COMPONENT»

Jean Lemaire : « BONUS-MALUS SYSTEMS: THE EUROPEAN AND ASIAN APPROACH TO MERIT-RATING».

P. J. DELAPORTE « LES MATHEMATIQUES DE L'ASSURANCE AUTOMOBILE»

