



المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN

ROYAUME DU MAROC

*_*_*_*_*

HAUT COMMISSARIAT AU PLAN

*_*_*_*_*_*_*_*

INSTITUT NATIONAL DE STATISTIQUE ET D'ECONOMIE APPLIQUEE



INSEA

Projet de Fin d'Etudes

Tarification D'un Produit Assurance Santé

Préparé par : **M. Assem BERROHO**
M. Said MOHIB

Sous la direction de : **M. Khalil SAID (INSEA)**
Mme. Fatima Zahra TALHAOUI (MCI CARE)

***Soutenu comme exigence partielle en vue de l'obtention du
Diplôme d'Ingénieur d'Etat***

Option : Actuariat Finance

Devant le jury composé de :

M. Khalil SAID (INSEA)
M. Fouad MARRI (INSEA)
Mme. Fatima Zahra TALHAOUI (MCI CARE)

N° PFE : 04

Année Universitaire 2019/2020

Remerciements

Nous tenons tout d'abord à remercier chaleureusement notre encadrant à l'Institut National de Statistique et d'Economie Appliquée, Monsieur Khalil SAID pour son soutien, son encadrement et ses conseils judicieux et précieux. Sans lui ce projet n'aurait pas pu se concrétiser. Nous remercions également Mme Fatima Zahra TALHAOUI, notre encadrante à la MCI CARE, pour son accompagnement durant notre projet de fin d'études et pour son aide bienveillante, Il sera opportun de remercier l'ensemble du corps professoral de l'Institut National de Statistique et d'Economie Appliquée, pour l'intérêt qu'ils manifestent dans la formation des futurs ingénieurs. En particulier Monsieur Fouad EL MARRI pour ses efforts énormes pour assurer des diverses formations au profit des étudiants actuels de l'INSEA. Que tous ceux et celles qui ont contribué de près ou de loin à l'accomplissement de ce travail trouvent l'expression de nous remercier les plus chaleureux.

Dédicaces Assem

Je dédie ce travail, comme preuve de respect, de gratitude et de reconnaissance à :

- *A mes chers parents, qui n'ont économisé aucun effort afin que je puisse être l'homme que je suis, qui m'ont toujours aimé et soutenu.*
- *A mon frère Nouh et mes sœurs Chaimae et Imane, pour leurs temps, leurs encouragements et leurs soutiens.*
- *A tout mes amis*

Dédicaces Said

Remerciement à Dieu, tout puissant, miséricordieux et clément. A mes chers parents Pour tous leurs sacrifices,

- *A toute ma famille Pour les encouragements,*
- *A mes amis Pour tous les moments de bonheur que nous avons passé ensemble,*
- *Je vous dédie cet humble travail*

Résumé

L'actuaire a pour mission de mettre en place des techniques de tarification des contrats d'assurance. Ce projet de fin d'études présente les éléments pour la tarification de la branche d'assurance maladie de base d'une compagnie privée d'assurance. Nous présenterons de manière globale les fondements de la branche de l'assurance maladie de base ainsi que les spécificités de l'environnement de tarification rencontré par l'actuaire. Ce projet se focalise sur la description d'une méthodologie pour la tarification des assurés de la branche assurance maladie. Contenant quatre chapitres, ce travail met l'accent dans sa première partie sur les enjeux de l'assurance maladie au Mali, le cadre économique du marché des assurances, ainsi qu'une présentation de la compagnie d'assurance dans laquelle le stage a été effectué. Le second chapitre est consacré à une présentation du portefeuille maladie, ses caractéristiques et spécificités, ainsi qu'une analyse descriptive de données selon plusieurs facteurs de sinistralité. Le troisième chapitre présente le cadre théorique ainsi que l'approche adoptée pour modéliser la prime de l'assurance maladie. Le dernier chapitre se focalise sur la modélisation de la fréquence et des frais moyens pour aboutir à un calcul du tarif de l'assurance maladie pour enfin analyser les résultats.

Table des matières

Résumé	5
Liste des abréviations	13
INTRODUCTION GENERALE	14
1 PRESENTATION DE L'ORGANISME D'ACCUEIL ET D'ENVIRONNEMENT POTENTIEL DU PROJET	16
1.1 Présentation générale de l'organisme d'accueil	16
1.1.1 MCI Care	16
1.1.2 Métier de MCI Care	16
1.1.3 Valeur ajoutée de MCI Care	17
1.1.4 MCI Care en chiffres	17
1.1.5 Le régime malien de sécurité sociale	17
1.1.5.1 Généralités	17
1.1.6 Maladie-maternité	21
1.1.7 Prestations familiales	24
1.1.8 Accidents du travail et maladies professionnelles	27
1.1.9 Invalidité, vieillesse et décès (survivants)	29
2 SOURCE ET ANALYSE DES DONNEES	32
2.1 Présentation du portefeuille	32
2.1.1 Présentation de la base de données	32
2.1.2 Purification de la base de données	32

2.2	Analyse de la Base d’Affiliés	33
2.2.1	Base d’affiliés	33
2.2.2	Répartition des affiliés par statut	33
2.3	Analyse de la base sinistre	34
2.3.1	Répartition de la base sinistre selon l’année de soin	34
2.3.2	Ratio de sinistralité	35
2.3.3	Consommation par année	36
2.3.4	Répartition par type de prestation	36
2.4	Analyse de l’inflation médicale	39
2.4.1	Coût moyen par personne	39
2.4.2	Coût moyen (mensuel) par personne	39
2.4.3	Estimation de la consommation des sinistres de 2019 et qui seront clôturé en 2020	41
2.4.4	Coût moyen par personne selon les types de prestations	42
2.4.5	Coût moyen par sinistre	44
2.4.6	Coût moyen mensuel par sinistre	44
2.4.7	Coût moyen de prestation par sinistre	47
2.5	Coût moyen par personne	48
2.5.1	Coût moyen par personne par prestation	48
2.5.2	Coût moyen par âge	49
2.6	Analyse par taux de remboursement	51
2.6.1	Répartition des affiliés selon le taux de remboursement	51
2.6.2	Répartition des affiliés par taux de remboursement et statuts	51
2.6.3	Analyse par taux de remboursement	52
2.6.4	Analyse par taux de remboursement(Hospitalisation)	54
2.6.5	Analyse par taux de remboursement (Pharmacie)	54
3	CADRE CONCEPTUEL ET THEORIQUE	57
3.1	Principe de tarification	58
3.2	Choix de modèle par sélection de variables	60
3.2.1	Pourquoi la sélection de variables	60
3.2.2	Critères de sélection de variables	60

3.3	Etude théorique selon l'approche GLM	62
3.3.1	Généralités	62
3.3.2	Rappel du modèle linéaire simple	63
3.3.3	Distributions de la famille exponentielle naturelle	63
3.3.4	Hypothèses et principe du Modèle Linéaire Généralisé	67
3.3.5	Estimation des paramètres	69
3.3.6	Choix des facteurs	70
3.3.7	Validation du modèle	72
4	TARIFICATION D'UN PRODUIT ASSURANCE SANTE	76
4.1	Segmentation des données en assurance santé	76
4.1.1	Sélection de variables tarifaires	76
4.1.2	Segmentation des variables tarifaires	77
4.2	Détermination de la prime pure	78
4.2.1	Modélisation de la fréquence des sinistres	78
4.2.1.1	Ajustement et choix du modèle	78
4.2.1.2	Analyse de Type III	80
4.2.2	Analyse des résultats	82
4.2.2.1	Estimation des paramètres du modèle établi	82
4.2.2.2	Validation du modèle	82
4.2.2.3	Prédiction de la fréquence	83
4.2.3	Modélisation de la charge des sinistres	84
4.2.3.1	Choix de la distribution	84
4.2.3.2	Le modèle tarifaire	85
4.2.4	Analyse des résultats	86
4.2.4.1	Estimation des paramètres du modèle établi	86
4.2.4.2	Analyse des résidus	87
4.2.4.3	Prédiction du coût moyen	88
	CONCLUSION	90
	Annexes	93

Table des figures

1.1	Les cotisations et les plafonds pour les salariés selon les branches	20
1.2	Les cotisations et les plafonds pour les indépendants selon les branches	21
1.3	Les plafonds trimestriels selon les classes	21
2.1	Répartition des affiliés de 2018 selon le statut	34
2.2	Répartition des affiliés de 2019 selon le statut	34
4.1	Sélection stepwise basée sur le F partiel de Fisher	77
4.2	Ajustement de la fréquence des sinistres par une loi de Poisson	79
4.3	Ajustement de la fréquence des sinistres par une loi Binomiale Négative	80
4.4	Analyse de type III des facteurs de sinistralité	81
4.5	Estimation des paramètres de la fréquence des sinistres	82
4.6	Test de déviance du modèle	83
4.7	Analyse de type III de la régression Gamma	86
4.8	Estimation des paramètres du coût moyen	86
4.9	Représentation des résidus de la déviance	88
4.10	Code d'importation de base de donnée	93
4.11	Code de détection des valeurs aberrantes	94
4.12	Code de Répartition de la base de sinistre 2019 selon l'année de soin	95
4.13	Code de calcul du nombre des personnes sinistrées	95
4.14	Code de répartition des sinistres selon les prestations	96
4.15	Code de la consommation mensuelle des sinistres	96

4.16	Code du calcul des nombres des personnes protégées mensuellement pour la base des actifs	97
4.17	Code du calcul des nombres des personnes protégées en janvier et février 2018 pour la base des retirés	97
4.18	Code de la répartition mensuelle des prestation consultation et pharmacie	98
4.19	Code de la répartition de la consommation selon les tranche d'âge	98
4.20	Code de la répartition de la consommation selon le taux de remboursement	99
4.21	Modélisation de la fréquence des sinistres : Modèle Binomiale négative	99
4.22	Modélisation du coût moyen des sinistres : Modèle Gamma	100

Liste des tableaux

2.1	Nombre de personne protégés pour les deux années	33
2.2	Répartition de la base de sinistre 2018 selon l'année de soin	35
2.3	Répartition de la base de sinistre 2019 selon l'année de soin	35
2.4	Ratio de sinistralité des deux exercices 2018 et 2019	35
2.5	Consommation des deux exercices 2018 et 2019	36
2.6	Répartition des sinistres 2018 selon la prestation	37
2.7	Répartition des sinistres 2019 selon la prestation	38
2.8	Coût moyen par personne pour les deux exercices	39
2.9	Coût moyen (mensuel) par personne pour exercice 2018	40
2.10	Coût moyen (mensuel) par personne pour exercice 2019	41
2.11	Estimation des sinistres de novembre 2019 qui ne sont pas déclarés encore	42
2.12	Estimation des sinistres de décembre 2019 qui ne sont pas déclarés encore	42
2.13	Coût moyen mensuel par personne (10mois) pour toutes les prestations	43
2.14	Coût moyen mensuel par personne (10mois) pour la prestation 'Pharmacie'	43
2.15	Coût moyen mensuel par personne (10mois) pour la prestation 'Consultation'	43
2.16	Coût moyen mensuel par personne (10mois) pour la prestation 'Biologie'	43
2.17	Coût moyen par sinistre pour les deux exercices	44
2.18	Coût moyen mensuel par sinistre pour l'exercice de 2018	45
2.19	Coût moyen mensuel par sinistre pour l'exercice de 2019	46
2.20	Comparaison entre la consommation mensuelle par sinistre entre les deux années 2018 et 2019	47

2.21	Coût moyen de chaque prestation par sinistre pour les deux années et leurs variations	48
2.22	Coût moyen par personne par prestation pour l'exercice de 2018	49
2.23	Coût moyen par personne par prestation pour l'exercice de 2019	49
2.24	Coût moyen par tranche d'âge pour l'exercice de 2018	50
2.25	Coût moyen par tranche d'âge pour l'exercice de 2019	50
2.26	Répartition des affiliés selon le taux de remboursement pour les deux exercices . . .	51
2.27	Répartition des affiliés par taux de remboursement et statuts pour l'exercice 2018 . .	52
2.28	Répartition des affiliés par taux de remboursement et statuts pour l'exercice 2019 . .	52
2.29	Comparaison entre le cout moyen par personne selon le taux de remboursement pour l'année 2018	53
2.30	Comparaison entre le cout moyen par personne selon le taux de remboursement pour l'année 2019	53
2.31	Comparaison entre le cout moyen par personne (Hospitalisation) selon le taux de remboursement pour l'année 2018	54
2.32	Comparaison entre le cout moyen par personne (Hospitalisation) selon le taux de remboursement pour l'année 2019	54
2.33	Comparaison entre le cout moyen par personne (Pharmacie) selon le taux de remboursement pour l'année 2018	55
2.34	Comparaison entre le cout moyen par personne (Pharmacie) selon le taux de remboursement pour l'année 2019	55
3.1	les foncions de liens classiques	68
3.2	Exemples de MLG correspond à des modèles statistiques classiques	69
4.1	Critère AIC et BIC de différents modèles	85

Liste des abréviations

CMSS	:	Caisse Malienne de Sécurité Sociale
ANAPE	:	Agence Nationale pour l’Emlpoi
ANAM	:	Agence National d’Assistance Médicale
INPS	:	Institut National de Prévoyance Sociale
AMO	:	Assurance Maladie Obligatoire
RAMED	:	Régime d’Assistance Médicale
FCFA	:	Franc de la Communauté Financière Africaine
GLM	:	Modèle Linéaire Généralisée
AIC	:	Critère d’information d’Akaike
BIC	:	Critère d’information Bayésien
TR	:	Taux de Remboursement

INTRODUCTION GENERALE

Le Maroc a réalisé des investissements directs de 37 milliards de DH dans les pays africains, entre 2013 et 2017, selon le ministre de l'Industrie, du Commerce, de l'Investissement et de l'Économie numérique, Moulay Hafid Elalamy. Il a souligné que l'Afrique est le premier bénéficiaire des investissements marocains, alors que Maroc est désormais parmi les premiers investisseurs africains de l'Union économique des États de l'Afrique de l'Ouest (UEMOA) et de la Communauté des États de l'Afrique centrale (CEMAC).

Cette forte croissance des investissements marocains en Afrique a été observée au cours de la dernière décennie à la suite de l'engagement fort du Maroc pour le développement du continent. Depuis l'année 2000, 50 visites royales ont été effectuées dans plus de 30 pays africains et plus de 1.000 accords ont été signés.

Dans ce cadre les entreprises marocaines ont commencé d'élargir leurs investissements dans l'Afrique, parmi ces entreprises, celles du secteur d'assurance qui ont commencé de chercher des nouveaux marchés d'assurance vu la grande progression de ce marché, mais cette mission n'est pas facile vu la grande concurrence au niveau de ce marché, et spécialement de la part des compagnies d'assurance du Sud d'Afrique qui ont la grande part du marché de l'Afrique.

Mci Care est une compagnie marocaine spécialisée dans l'assurance santé, elle a réussi d'intégrer quatre marchés africains, et elle cherche à augmenter ce nombre, en effet notre stage a été pour but d'étudier différents portefeuilles de différents pays africains visés comme des marchés potentiels par la compagnie. En effet, notre démarche commence par une étude globale sur le portefeuille donné, afin de conclure par une tarification, ensuite la compagnie fait une comparaison entre le tarif trouvé et les primes offertes par le marché.

Ce rapport résume les étapes suivies dans l'étude d'une portefeuille Malienne, en effet on a commencé par une présentation du secteur d'assurance santé en mali, parce que la première étape pour un investissement est d'étudier l'environnement potentiel où on peut l'effectuer.

Ensuite la deuxième partie est consacrée pour l'analyse de la base de données, d'une part d'assurer la cohérence des données, et d'autre part effectuer des analyses qui nous expliquent le comportement des personnes protégées, afin d'avoir des premières idées sur la consommation.

La troisième partie met l'accent sur le cadre conceptuel et théorique, dans lequel nous allons définir les différentes approches adoptées pour modéliser la prime pure de l'assurance maladie. Dans cette partie, on définit les étapes de la mise en place du modèle.

Enfin, la quatrième partie explicite le cadre pratique des éléments constitutifs de la tarification en assurance maladie de base. Il décrit les résultats des différents modèles et présente les limites et la possibilité d'amélioration de la méthode envisagée.

PRESENTATION DE L'ORGANISME D'ACCUEIL ET D'ENVIRONNEMENT POTENTIEL DU PROJET

1.1 Présentation générale de l'organisme d'accueil

1.1.1 MCI Care

Créée en 2013, MCI Care a pour objectif de proposer à des assurés un accès facilité à des soins de qualité, tout en assurant la maîtrise des dépenses de santé pour les assureurs :

-« Un accès facilité » via le tiers payant qui permet d'accéder aux soins sans avancer les frais (seul le ticket modérateur reste à la charge de l'assuré)

-« Des soins de qualité » à travers un réseau de prestataires de santé, soigneusement sélectionnés sur des critères de qualité et de couverture géographique

MCI Care a développé une expertise dans la mise en place et l'administration du tiers-payant santé. Nous avons été ainsi les premiers à proposer ce service innovant sur le marché marocain.

Présents dans quatre pays d'Afrique, MCI Care compte aujourd'hui 50 collaborateurs ayant à cœur d'assurer une qualité de service aux meilleurs standards du marché.

1.1.2 Métier de MCI Care

MCI Care est gestionnaire des prestations des contrats d'assurance santé pour le compte des garants (compagnies d'assurance, régimes d'assurance maladie obligatoire, mutuelles).

A ce titre, MCI Care assure à la fois un le traitement des dossiers de remboursement des assurés (en cas d'avance de frais par l'assuré) ainsi qu'un accès aux soins en tiers-payant auprès des prestataires de son réseau de soins conventionnés.

1.1.3 Valeur ajoutée de MCI Care

i. Auprès des garants

- Maîtriser les coûts des dépenses de santé.
- Donner de la visibilité et des leviers pour anticiper sur le risque.
- Organiser des comités de suivi périodiques sur l'évolution du régime.
- Mettre à disposition des reportings statistiques fiables, analysés et commentés.

ii. Auprès des assurés

- Faciliter l'accès aux soins et rembourser dans les meilleurs délais.
- Garantir des soins médicaux de qualité dans un réseau de soins dédié.

iii. Auprès des professionnels de santé

- Conventionner les prestataires de santé.
- Suivre l'activité et la qualité du réseau.
- Contrôler, liquider et régler les factures des prestataires de soins.
- Sensibiliser à la bonne pratique par des rencontres périodiques.
- Fixer des objectifs de gestion et des indicateurs de performance à faire respecter.

1.1.4 MCI Care en chiffres

- Une présence dans 4 pays (Maroc, Angola, Cameroun, Rwanda)
- 50 collaborateurs
- 70 000 personnes protégées
- Plus de 80 000 sinistres traités en 2016
- Un réseau de plus de 500 prestataires de soins dans nos pays de présence, ainsi qu'un réseau de 1500 prestataires à l'international

1.1.5 Le régime malien de sécurité sociale

1.1.5.1 Généralités

i. Structure

Le régime malien de sécurité sociale couvre les salariés des secteurs public et privé.

Il comprend les branches suivantes :

- Maladie-maternité.
- Accidents du travail et maladies professionnelles.
- Vieillesse, invalidité, décès-survivants.
- Prestations familiales.

Les travailleurs indépendants peuvent adhérer volontairement aux couvertures des assurances vieillesse, maladie et prestations familiales.

En vertu du Code du travail, les employeurs sont tenus, en cas d'absence d'un travailleur (jusqu'à 6 mois maximum) pour cause de maladie ou d'accident non professionnel, de lui verser :

-Pendant la 1^e année de présence : une indemnité égale au salaire, durant une période équivalente au préavis à effectuer en cas de licenciement ou démission,

-Au delà d'une année de présence : une indemnité égale à la moitié du salaire mensuel précédant l'arrêt de travail pendant la période restante.

En l'absence de conventions collectives, la durée du préavis est de :

- 8 jours pour le personnel rémunéré à la journée ou la semaine.
- 1 mois pour les travailleurs rémunérés mensuellement.
- 2 mois pour les agents de maîtrise et assimilés.
- 3 mois pour les cadres et le personnel de direction. Le contrat peut être rompu sans préavis en cas de faute lourde et sous réserve de l'appréciation de la juridiction compétente.

Conformément au Code du travail, une indemnité de licenciement doit être versée à tout employé sous contrat à durée indéterminée qui a au moins un an d'embauche. Le montant versé correspond à un pourcentage de la moyenne des rémunérations perçues au cours des 12 derniers mois qui ont précédé le licenciement :

- 20% pour chacune des 5 premières années de travail,
- 25% pour chaque année de la 6^e à la 10^e incluse,
- 30% pour chaque année au-delà de la 10^e.

Cette indemnité n'est pas due si le licenciement est motivé par une faute lourde du travailleur.

ii.Organisation administratif

Sous la tutelle du ministère de la Solidarité et de la Lutte contre la pauvreté, l'Institut National de Prévoyance Sociale (INPS) assure pour le secteur privé :

- La gestion des risques accident du travail - maladies professionnelles, vieillesse - invalidité - décès et le service des prestations familiales.
- Le recouvrement des cotisations.

L'INPS dispose de 9 directions régionales (Bamako, Gao, Kayes, Kidal, Koulikoro, Mopti, Sikasso, Ségou, Tombouctou). La loi n° 09-015 du 26 octobre 2009 a institué la mise en place d'un **régime d'Assurance Maladie obligatoire (AMO)** et d'un **Régime d'Assistance Médicale (RAMED)**.

La Caisse Nationale d'Assurance Maladie (CANAM), est chargée de la gestion du régime AMO. Certaines compétences (attestations de droits, paiement des prestataires de soins, recouvrement des cotisations) sont déléguées à :

-l'Institut National de Prévoyance Sociale (INPS) pour les travailleurs salariés et non-salariés, les pensionnés des régimes privés,

- Caisse Malienne de Sécurité Sociale (CMSS) pour les fonctionnaires civils et militaires, les députés et les pensionnés de la fonction publique.

L'Agence Nationale d'Assistance Médicale (ANAM), gère le Régime d'Assistance Médicale (RAMED), sous la tutelle du ministère de la Solidarité et de la Lutte contre la pauvreté.

La Caisse Malienne de Sécurité Sociale (CMSS), sous la tutelle du ministère de la Solidarité et de la Lutte contre la pauvreté, est chargée pour le compte des salariés des secteurs publics (administrations, militaires, parlementaires...), de la gestion de l'AMO et des régimes de pensions.

Il n'existe pas de régime couvrant la perte d'emploi mais le Mali a créé en 2001 l'Agence Nationale pour l'Emploi (ANPE). Sa mission est de :

-Contribuer à la mise en œuvre de la politique nationale de l'emploi.

-Procéder à la collecte, la centralisation, l'analyse et la diffusion des données relatives au marché de l'emploi et des offres d'emploi.

-Orienter les demandeurs d'emploi.

-Promouvoir l'auto-emploi.

-Concourir à la mise en œuvre des activités de formation professionnelle, de reconversion et d'insertion.

iii.Affiliation

Tout employeur est tenu de porter à la connaissance de l'INPS chaque embauche ou licenciement de personnel dans les 8 jours.

Les membres non-salariés des professions libérales, artisanales, commerciales et industrielles ainsi que les travailleurs indépendants ont la possibilité d'adhérer volontairement à l'INPS. L'affiliation prend effet à compter du 1er jour du trimestre civil en cours à la date de réception de la demande.

iv.Financement

-Travailleurs salarié

Cotisations au 1 ^{er} janvier 2019			
Branches	Part patronale	Part salariale	Plafond
Prestations familiales Indemnités journalières de maternité	8 %	-	-
Assurance maladie obligatoire (AMO)	3,50 %	3,06 %	-
Accidents du travail, maladies professionnelles	de 1 à 4 % suivant les risques encourus	-	400 000 FCFA ¹
Vieillesse Invalidité, survivants	5,4 %	3,6 %	-
ANPE	1 %	-	
Total	entre 18,9 et 21,9 %	6,66 %	

FIGURE 1.1 – Les cotisations et les plafonds pour les salariés selon les branches

Les cotisations sont payées sur la totalité du salaire. Le montant du salaire pris en considération pour la base de calcul des cotisations ne peut, en aucun cas, être inférieur au montant du SMIG.

Le salaire minimum interprofessionnel garanti (SMIG) correspond depuis le 1er janvier 2016 à :

-40 000 FCFA par mois.

-230,77 FCFA par heure.

Le taux de la cotisation due au titre du régime de l'assurance maladie obligatoire (AMO) à la charge des pensionnés est fixé à 0,75%.

Le Régime d'Assistance Médicale (RAMED) est financé par les contributions des collectivités territoriales (15%) et par des subventions de l'Etat (85%).

Travailleurs indépendants

L'adhésion des travailleurs indépendants aux différents risques est volontaire.

Cotisations au 1 ^{er} janvier 2019		
Branches	Taux	Plafond trimestriel
Prestations familiales	8 %	5 classes de revenu : entre un minimum de 125 000 et un maximum de 1 000 000 FCFA
Assurance maladie obligatoire (AMO)	6,56 %	
Vieillesse Invalidité, survivants	9 %	

FIGURE 1.2 – Les cotisations et les plafonds pour les indépendants selon les branches

Les revenus trimestriels des travailleurs non-salariés sont classés en 5 catégories qui déterminent le montant des cotisations versées pour les prestations familiales, l'AMO et les pensions :

Classes	Revenus trimestriels	Plafonds trimestriels
1	Moins de 150 000 FCFA	125 000 FCFA
2	150 000 à 450 000 FCFA	400 000 FCFA
3	450 000 à 750 000 FCFA	675 000 FCFA
4	750 000 à 1 500 000 FCFA	900 000 FCFA
5	Plus de 1 500 000 FCFA	1 000 000 FCFA

FIGURE 1.3 – Les plafonds trimestriels selon les classes

Les cotisations sont versées trimestriellement.

1.1.6 Maladie-maternité

i. Assurance Maladie Obligatoire (AMO)

Bénéficiaire de ce régime :

- Les travailleurs salariés du secteur privé.
- Les travailleurs du secteur public.
- Les titulaires de pensions des secteurs public et privé.
- Les assurés volontaires déjà affiliés au régime d'assurance volontaire de l'INPS.

Par la suite, la couverture de l'AMO doit s'étendre aux :

-Travailleurs des secteurs agricole et informel.

-Etudiants.

Pour bénéficier de l'AMO, l'assuré doit avoir versé des cotisations pendant au moins 6 mois consécutifs.

Sont considérés comme bénéficiaires, les assurés ainsi que les membres de leur famille à charge :

-Conjoint marié légalement.

-Enfant de moins de 14 ans (ou 21 ans si scolarisé).

-Enfant, sans limite d'âge, atteint d'un handicap physique ou mental et dans l'impossibilité d'exercer une activité rémunérée.

-Ascendant direct dépendant économiquement de l'assuré.

Le droit aux prestations de l'AMO est maintenu pour l'assuré et ses ayants droit, pendant :

-6 mois maximum à compter de la date à laquelle les conditions ne sont plus remplies.

- 1 an aux ayants droit d'un assuré décédé qui ne bénéficient d'aucune pension de survivant.

Le panier de soins de l'AMO comprend les soins de santé inhérents à la maladie et à la maternité.

L'AMO garantit la prise en charge directe d'une partie des frais de santé par la caisse (70 ou 80%), l'autre partie restant à la charge de l'assuré sous forme de ticket modérateur (30 ou 20%).

L'assuré peut souscrire une assurance complémentaire pour couvrir les frais restant à sa charge.

Soins ambulatoires

La prise en charge des soins ambulatoires dans un établissement conventionné est de 70% pour tous les soins effectués hors hospitalisation :

-Les consultations générales ou spécialisées.

-Les soins préventifs et curatifs.

-Les soins bucco-dentaires.

-Les analyses de laboratoire.

-La radiologie et l'imagerie médicale.

-Les actes de médecine générale et chirurgicale.

Sauf autorisation préalable de la caisse, ne sont pris en charge :

-Qu'une seule consultation ou visite de praticien pour la même affection dans les 24

heures à moins que l'évolution du cas ne nécessite une assistance médicale urgente

-Qu'une seule consultation ou visite de praticien de la même discipline médicale par période de 7 jours, sauf évolution nécessitant une assistance médicale urgente.

La prise en charge des prestations bucco-dentaires et d'orthodontie est soumise à la présentation d'un devis détaillé. Concernant l'orthodontie, aucune prestation ne peut être prise en charge plus d'une fois.

Médicament

Les médicaments (inscrits dans une liste) sont pris en charge par l'AMO à 70% à condition qu'ils soient délivrés par une pharmacie conventionnée avec la CANAM.

Hospitalisation

Les frais d'hospitalisation sont pris en charge à 80% et comprennent :

- Les frais d'hôtellerie hospitalière à l'exclusion des dépenses à caractère personnel.
- Les frais correspondants aux soins et examens nécessaires lors de l'hospitalisation, aux actes médicaux, chirurgicaux.
- Les frais occasionnés par le transport du malade. Toute hospitalisation doit être déclarée à la CANAM par la structure dans les 48 heures.

Maternité

Les prestations de maternité de l'AMO comprennent les frais :

- De suivi de grossesse.
- Médicaux.
- Pharmaceutiques.
- D'hospitalisation relative à la grossesse, à l'accouchement et à ses suites.

La grossesse médicalement constatée d'une assurée ou conjointe d'un assuré doit être déclarée à la caisse au cours du 3^e mois.

Le droit de prise en charge des frais liés à la grossesse est ouvert à compter de la déclaration de grossesse jusqu'à la 8^e semaine après l'accouchement.

Les prestations en espèces liées à la maternité sont versées dans le cadre des prestations familiales.

ii. Régime d'assistance médicale (RAMED)

La loi n° 09-031 du 27 juillet 2009 a permis la mise en place d'un Régime d'Assistance Médicale (RAMED) afin d'assurer une couverture médicale aux personnes dépourvues de revenu.

Pour pouvoir bénéficier du régime et obtenir une carte RAMED, il faut :

- S'immatriculer auprès de l'ANAM.

-Etre reconnue comme "indigent" et disposer d'un certificat délivré par la mairie du lieu d'habitation.

-Ne pas être assujettie à l'AMO ni à aucun système d'assurance.

-Etre sans ressources.

Sont bénéficiaires de droit :

-Les pensionnaires d'orphelinats, les enfants abandonnés.

-Les personnes sans domicile fixe.

-Les pensionnaires d'établissements pénitentiaires.

-Les pensionnaires d'établissements de bienfaisance.

Sont considérés comme ayants droit à charge :

-Le conjoint.

-Les enfants âgés de moins de 14 ans.

-Les enfants âgés de 14 à 21 ans en cas de poursuite d'études.

-Les enfants handicapés, quel que soit l'âge, justifiant d'une impossibilité totale et définitive d'exercer une activité rémunérée.

La carte d'assuré du RAMED est valable pendant 3 ans et donne droit aux prestations médicales sur tout le territoire.

Le panier de soins du RAMED couvre :

-Les soins ambulatoires : consultations médicales, soins infirmiers, soins dentaires, imageries médicales, examens de laboratoire, petites chirurgies.

-Les hospitalisations : frais d'hôtellerie hospitalière, actes médicaux et chirurgicaux.

-Les produits pharmaceutiques inscrits sur une liste des médicaments admis.

-Les prestations de maternité : ensemble des frais médicaux, analyses, examens, hospitalisation liée à la grossesse, accouchement et ses suites.

Tous ces soins doivent être fournis dans un établissement public ou communautaire conventionné et sont pris en charge à 100%.

1.1.7 Prestations familiales

Le régime des prestations familiales comprend :

-La prime de premier établissement.

-Les allocations prénatales.

- Les allocations de maternité.
- Les indemnités journalières des femmes salariées.
- Le congé de naissance.
- Les allocations familiales.

i. Conditions générales

Les prestations familiales sont accordées :

- A tout travailleur salarié sous condition d'activité.
- A l'assuré volontaire.
- Au pensionné(e) veuve/veuf d'un bénéficiaire.

Pour bénéficier des prestations familiales, l'assuré doit :

- s'il est travailleur salarié :
 - justifier de 9 mois d'activités consécutives chez un ou plusieurs employeurs (1 mois étant validé par 18 jours ou 120 heures d'activité),
 - avoir une rémunération au moins égale au SMIG ;
- S'il est assuré volontaire : justifier d'au moins 2 trimestres d'affiliation consécutifs,
- Justifier d'une résidence au Mali (pour lui et les enfants bénéficiaires).

Les enfants doivent être déclarés à l'état civil et être à la charge effective de l'assuré pour ouvrir droit aux prestations.

La veuve d'un l'allocataire, lorsqu'elle assure la garde et l'entretien des enfants qui étaient à la charge de son époux, bénéficie des prestations familiales sans avoir à exercer une activité salariée.

ii. Prestations

• Prime de premier établissement

Seuls les travailleurs salariés ont droit à cette prestation.

Il s'agit d'une allocation qui permet au travailleur d'acquérir l'équipement nécessaire à son ménage.

Cette prime, versée une seule fois lors du premier mariage, est de 13 650 FCFA.

• Allocations prénatales

Des allocations prénatales sont versées à toute femme enceinte assurée ou ayant droit d'un assuré, ayant passé les examens médicaux obligatoires.

La déclaration de grossesse accompagnée d'un certificat médical doit être adressés à l'INPS dans les 3 premiers mois de la grossesse.

L'attribution des prestations est subordonnée à 3 examens médicaux :

- Avant la fin du 3^e mois.
- Vers le 6^e mois.
- Vers le 8^e mois.

Le montant des allocations prénatales est de 12 285 FCFA. Il est versé en 3 fractions :

- 2 730 FCFA.
- 5 460 FCFA.
- 4 095 FCFA.

• **Allocations de maternité**

Ces allocations sont versées à toute femme assurée ou ayant droit d'un assuré, qui a :

- Accouché sous contrôle médical.
- Donné naissance à un enfant viable.
- Soumis le nouveau-né aux examens médicaux obligatoires.

Le montant total est de 16 380 FCFA payé en 3 fractions pour chaque naissance :

- 1/2 à la naissance ou immédiatement après la demande (8 190 FCFA).
- 1/4 aux 6 mois de l'enfant (4 095 FCFA).
- 1/4 au 1 an de l'enfant (4 095 FCFA).

• **Indemnité journalière des femmes salariées**

Les assurées volontaires n'ouvrent pas de droit à cette prestation.

Cette indemnité est versée à toute assurée salariée enceinte qui suspend son activité professionnelle, réside au Mali et justifie de 9 mois consécutifs de cotisation sur la base de 18 jours ou 120 heures de travail par mois.

L'indemnité journalière est égale à 100% du dernier salaire sans plafond.

Elle est accordée pour la période réelle d'inactivité qui s'inscrit entre les 6 semaines qui précèdent et les 8 semaines qui suivent l'accouchement.

Le versement de l'indemnité journalière peut être prolongé pendant 3 semaines en cas de suites de couches pathologiques.

• **Congé de naissance**

A l'occasion de chaque naissance d'un enfant viable survenue dans son foyer, tout travailleur salarié a droit à un congé de 3 jours rémunéré par l'employeur, à 100% de son salaire journalier.

Ces jours peuvent être consécutifs ou non, après entente entre l'employeur et le bénéficiaire

mais devront être compris dans une période de 15 jours incluant la date de naissance.

- **Allocations familiales**

Elles sont attribuées pour chacun des enfants à charge d'un allocataire, âgé :

- De plus d'un an et de moins de 14 ans,

- Jusqu'à 18 ans pour les enfants placés en apprentissage.

- Jusqu'à 21 ans en cas de poursuite d'études.

- Jusqu'à 21 ans en cas de handicap rendant impossible l'exercice d'une activité professionnelle.

Le taux mensuel des allocations familiales est fixé, depuis juillet 2015, à :

- 3 500 FCFA par enfant.

- 4 000 FCFA par enfant souffrant d'un handicap.

1.1.8 Accidents du travail et maladies professionnelles

i. Déclaration, formalités, soins

Cette assurance ne concerne que les travailleurs salariés.

Est reconnu par l'assurance accident du travail - maladies professionnelles :

- L'accident survenu par le fait ou à l'occasion du travail.

- L'accident survenu sur le trajet domicile/entreprise.

- La maladie professionnelle inscrite sur le tableau des maladies professionnelles (liste de 44 maladies figurant en annexe dans le Code de la Prévoyance Sociale).

L'employeur est tenu de déclarer immédiatement ou au plus tard dans un délai de 48 heures tout accident du travail ou maladie professionnelle constaté(e) dans l'entreprise.

L'employeur doit :

- Faire assurer les soins de première urgence.

- Aviser le médecin chargé des services médicaux de l'entreprise ou le médecin le plus proche.

- Eventuellement diriger la victime sur la structure médicale la plus proche.

Si la victime n'a pas repris le travail dans les 3 jours, l'employeur doit faire établir un certificat médical.

Les frais, à l'exception de ceux qui concernent les soins de première urgence, sont à la charge de l'INPS (système du tiers payant). Il s'agit de :

-L'assistance médicale, chirurgicale et dentaire y compris les examens radiographiques.

-Les examens de laboratoire et les analyses.

-La fourniture des produits pharmaceutiques et accessoires. • L'hospitalisation.

-La fourniture, l'entretien et le renouvellement des appareils de prothèse et d'orthopédie nécessités par les lésions résultant de l'accident.

-La réadaptation fonctionnelle, la rééducation professionnelle et le reclassement de la victime.

ii.Prestations en espèces

• Incapacité temporaire

L'indemnité journalière est payée à la victime par l'INPS à partir du 1er jour qui suit l'arrêt du travail jusqu'à la guérison ou constatation de l'incapacité permanente.

Elle est égale à 100% du dernier salaire de l'assuré.

Le montant minimum de l'indemnité journalière est égal au SMIG journalier.

• Incapacité permanente

En cas d'incapacité permanente, la victime a droit à une rente égale au salaire annuel, multiplié par le taux d'incapacité préalablement réduit de moitié lorsque ce taux est compris entre 10 et 50% et augmenté de moitié lorsqu'il excède 50%.

Le plafond du salaire servant à calculer la rente est égal à 28 fois le SMIG.

Si la victime a besoin de l'assistance d'une tierce personne, la rente est majorée de 40%.

Le montant minimum de la rente, si le degré d'incapacité est égal ou supérieur à 10%, ne peut être inférieur à 1,3 fois le SMIG.

Le montant maximum de la rente est égal à 20 fois le SMIG.

Si le degré d'incapacité est inférieur à 10%, un versement unique est effectué.

Après un délai de 5 ans, la rente peut être convertie en capital en totalité ou en partie si :

-Le taux d'incapacité ne dépasse pas 20%, sur simple demande.

-Le taux d'incapacité est compris entre 20 et 50%. Dans ce cas, le rachat peut être opéré dans la limite du quart au plus du capital correspondant à la valeur de la rente.

L'assuré, victime d'un accident du travail et bénéficiaire d'une rente, doit se soumettre à des examens médicaux auprès d'un médecin agréé par l'INPS, durant les 2 premières années et ensuite chaque année.

• Décès

Au décès d'un assuré victime d'un accident du travail, le conjoint et les orphelins peuvent prétendre à un avantage versé sous forme de pension.

Le montant de la rente est égal à un pourcentage du salaire moyen annuel du défunt précédant le début de l'incapacité.

La rente due aux survivants s'élève à :

-30% du salaire annuel de la victime pour le conjoint, si le mariage a eu lieu avant le début de l'incapacité du défunt. En cas de remariage et s'il n'y a pas d'enfant à charge, la rente n'est plus versée, elle est remplacée par une allocation de remariage égale à 3 ans de rente.

-15% pour chacun des 2 premiers enfants, 10% pour chaque enfant suivant, 20% pour chaque enfant orphelin de père et de mère. Les orphelins bénéficiaires d'une rente doivent être âgés de moins de 14 ans (ou moins de 18 ans en cas d'apprentissage, 21 ans en cas d'études ou d'infirmité).

-10% pour chaque ascendant à charge jusqu'à concurrence de 30

Le total des rentes de survivants ne doit pas dépasser 85% du salaire moyen annuel du défunt, le cas échéant, les rentes seront réduites proportionnellement.

Le salaire servant de base de calcul ne peut être inférieur à 1,3 fois le SMIG.

Le plafond du salaire servant à calculer la rente est égal à 28 fois le SMIG.

Une allocation funéraire est versée pour aider les familles à faire face aux frais d'obsèques. Le montant du remboursement maximum est égal à 25% du SMIG annuel

1.1.9 Invalidité, vieillesse et décès (survivants)

i.Invalidité

• Conditions

Pour obtenir une pension d'invalidité, le travailleur doit :

- Avoir accompli au moins 8 ans d'assurance (10 ans pour l'assuré volontaire).
- Présenter une réduction de ses capacités physiques ou mentales d'origine non professionnelle d'au moins 66,66%.
- Ne plus pouvoir gagner plus d'1/3 de sa rémunération.

• Montant

La pension d'invalidité est calculée sur la base :

- Du salaire/revenu mensuel moyen des rémunérations soumises à cotisation au cours des 8 dernières années d'activité (10 pour les assurés volontaires) : RMM,
- Le taux d'annuité fixé par la loi à 2% : TA.
- Le nombre d'années d'activité ayant donné lieu à cotisation ou années ou durée d'assurance (DA).
- Le nombre d'années comprises entre 53 ans et l'âge effectif de l'invalidité, à raison d'un semestre par an ou durée d'assurance assimilée (DAA).

Le montant de la pension d'invalidité s'obtient en appliquant la formule suivante :

$$PP = RMM \times TA \times (DA + DAA)$$

Le montant de la pension ne peut excéder :

- 80% du salaire moyen mensuel de l'assuré salarié.
- 30% du revenu moyen trimestriel de l'assuré volontaire.

Le montant minimum sur lequel est calculée la pension d'invalidité est égal à 2 fois le SMIG.

La pension d'invalidité cesse d'être versée à partir de 53 ans (55 ans en cas d'assurance volontaire) et est remplacée par la pension de vieillesse d'un montant au moins égal à la pension d'invalidité.

ii. Vieillesse

• Pension

Pour prétendre à une **pension de vieillesse à taux plein**, il faut :

- Pour le salarié : être âgé de 58 ans et avoir accompli au moins 13 ans d'assurance.
- Pour l'assuré volontaire : être âgé de 60 ans et avoir accompli au moins 15 ans d'assurance.

La pension de retraite peut être liquidée de façon anticipée :

-**Sans coefficient d'anticipation**, pour raison de santé : en cas de reconnaissance d'inaptitude au travail, à partir de 53 ans (55 ans pour les assurés volontaires) et au moins 13 ans d'assurance (15 ans pour les assurés volontaires).

-**Avec coefficient d'anticipation** : à partir de 53 ans (55 ans pour les assurés volontaires) et au moins 13 ans d'assurance (15 ans pour les assurés volontaires). Le montant de la pension est alors affecté d'un abattement de 5% par année d'anticipation.

La pension est égale à 26% du salaire/revenu moyen mensuel des 8 dernières années (10 dernières années pour les assurés volontaires) multiplié par la durée d'assurance majorée de 2% par période de 12 mois pendant 120 mois maximum (au-delà de 180 mois pour les assurés volontaires).

Le montant de la pension ne peut excéder :

- 80% du salaire moyen mensuel de l'assuré salarié,
- 30% du revenu moyen trimestriel de l'assuré volontaire.

Le montant minimum sur lequel est calculée la pension de vieillesse est égal à 2 fois le SMIG.

iii. Allocation vieillesse de solidarité

Le travailleur salarié âgé de 58 ans n'ayant pas accompli 13 ans d'assurance mais ayant accompli au moins 6 années pourra prétendre sous conditions de ressources, à une allocation de solidarité (60 ans pour l'assuré volontaire avec au moins 10 années d'assurance).

Cette allocation versée sous forme de rente mensuelle est égale à :

-52% du SMIG pour les travailleurs salariés,

-30% des revenus utilisés pour le calcul des cotisations pour les assurés volontaires.

iv. Pension de survivants

Au décès d'un assuré pensionné ou en activité qui aurait pu prétendre à pension, les ayants droit (conjoint et les orphelins) peuvent ouvrir droit à une pension de survivants.

Dans le cas où l'assuré ne bénéficiait pas d'une pension et ne remplissait pas les conditions d'ouverture de droits, une allocation de survivant peut être allouée sous forme de versement unique.

• Pension

Le conjoint ayant été marié au moins 2 ans avec le défunt peut prétendre à une pension de survivant à condition que l'assuré décédé ait bénéficié ou aurait pu bénéficier d'une pension de vieillesse ou d'une pension d'invalidité.

La pension versée aux survivants est égale à un pourcentage de la pension perçue par le défunt :

-50% pour le conjoint,

-10% pour chaque orphelin de moins de 14 ans (ou moins de 18 ans en cas d'apprentissage, 21 ans en cas d'études ou d'infirmité). Le total des pensions d'orphelin ne pourra pas dépasser 50% de la pension du défunt. Le montant de la pension d'orphelin ne peut être inférieur au montant des allocations familiales au titre du même enfant.

• Allocation

Lorsque l'assuré au moment du décès n'était pas titulaire d'une pension ou n'avait pas accompli au moins 13 ans d'assurance (15 ans pour les assurés volontaires), le conjoint peut prétendre à une allocation de survivant versée en une seule fois.

Cette allocation est égale à un mois de pension à laquelle aurait eu droit le défunt par semestres d'assurance accomplis.

SOURCE ET ANALYSE DES DONNEES

INTRODUCTION

Ce chapitre vise à présenter les données de base et les différents processus de détection et de correction des anomalies, en plus de l'analyse du portefeuille maladie. Nous présenterons dans ce chapitre quelques statistiques générales relatives à la répartition de la population sujette à l'étude au deux niveaux : composition du portefeuille et sinistralité. En effet, les études actuarielles demandent une pré connaissance de la composition du portefeuille étudié, ainsi que la distribution des individus composant la base de données.

2.1 Présentation du portefeuille

2.1.1 Présentation de la base de données

La base de données est malienne, elle est relative aux deux exercices 2018 et 2019. Elle se compose de deux fichiers.

Un fichier affilié regroupant toutes les informations relatives aux affiliés. à travers les variables suivantes : le matricule d'assuré, le matricule bénéficiaire, Le numéro de police, le numéro de la filiale, les dates relatives à la vie du contrat (date effet, date sortie), le taux de remboursement, la ville , la date de naissance, le statut et le genre.

Un fichier sinistre qui décrit la sinistralité maladie sur un horizon d'un an pour les deux années, via le nombre de sinistre et le montant de sinistre.

2.1.2 Purification de la base de données

On a commencé par l'importation de la base de données avec la fonction proc import, pour le choix du logiciel on a utilisé SAS car il plus efficace avec les bases de données de grandes tailles,

ensuite on a assuré que les données ont été bien importées, puis on a commencé l'analyse de chaque variable avec la fonction freq, en effet cette fonction nous permet de détecter les valeurs manquantes, les points aberrants et les erreurs de frappes.

Heureusement notre base de donnée ne contient pas beaucoup des erreurs, en effet la seule variable qui contient un nombre important des valeurs aberrantes est la date de naissance avec 306 valeurs (1/1/1900), mais ce nombre est très faible si on le compare avec le nombre des affiliés, comme on a remarqué 7 sinistres avec des montants de remboursement négatifs, on a les supprimés.

Après cette étape on peut dire que notre base de donnée est prête pour être traiter.

2.2 Analyse de la Base d'Affiliés

2.2.1 Base d'affiliés

Le nombre de personnes protégées est résumé dans le tableau suivant :

Nombre de personnes protégées	
2018	14 596
2019	19 847

TABLE 2.1 – Nombre de personne protégés pour les deux années

On remarque que le nombre de personnes protégées a augmenté par 36% entre 2018 et 2019.

2.2.2 Répartition des affiliés par statut

La répartition des affiliés par statut sera résumée dans les deux graphiques suivants :

-Année 2018

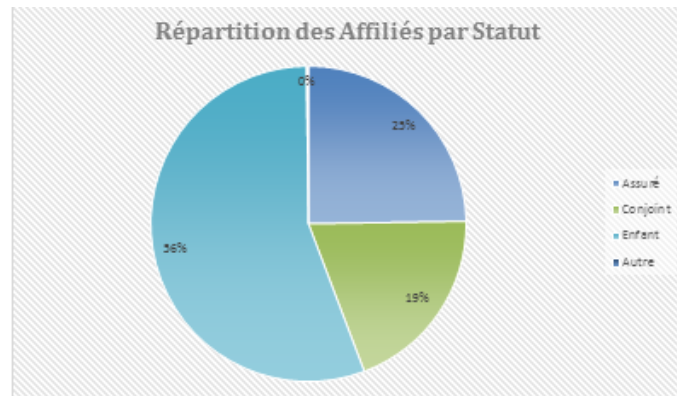


FIGURE 2.1 – Répartition des affiliés de 2018 selon le statut

-Année 2019

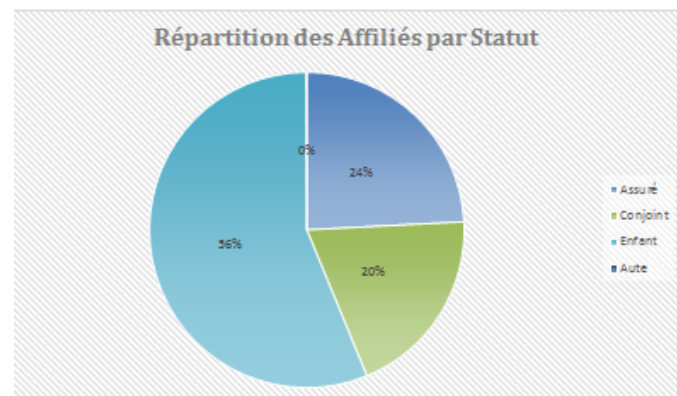


FIGURE 2.2 – Répartition des affiliés de 2019 selon le statut

Malgré le grand changement dans la population étudiée la répartition par statut reste la même entre les deux années 2018 et 2019. En effet, le statut E qui domine dans les deux années avec un pourcentage qui est proche de 56%, ensuite on trouve le statut A qui représente 25% du portefeuille, enfin on trouve les conjoints qui ont un pourcentage qui ne dépasse pas les 20%. Comme on remarque l'existence de deux autres statuts G et K avec des pourcentages très faibles.

2.3 Analyse de la base sinistre

2.3.1 Répartition de la base sinistre selon l'année de soin

-Année 2018

Année de soin	2016	2017	2018
Consommation	21.335	56.286.599	1.756.827.891
Pourcentage	00%	3,10%	96,89%

TABLE 2.2 – Répartition de la base de sinistre 2018 selon l'année de soin

-Année 2019

Année de soin	2017	2018	2019
Consommation	204.607	45.771.131	2.333.879.724
Pourcentage	00,1%	1,92%	98,07%

TABLE 2.3 – Répartition de la base de sinistre 2019 selon l'année de soin

On remarque que les deux bases de sinistres 2018 et 2019 contiennent en majorité des sinistres de l'année de la base de données. Comme elles contiennent des sinistres des deux années précédentes avec des pourcentages faibles. Donc on peut utiliser le pourcentage des sinistres de 2018 qui ont été traités en 2018 pour estimer les sinistres de 2019 et qui ceux seront traités en 2020.

2.3.2 Ratio de sinistralité

On résume le calcul du ratio de sinistralité dans le tableau suivant :

Année	Nombres de personnes protégées	Nombre de personnes sinistré	Ratio de sinistralité
2018	14 596	9898	67,81%
2019	19 847	13682	69,84%

TABLE 2.4 – Ratio de sinistralité des deux exercices 2018 et 2019

On voit que ce ratio a augmenté par 3% entre les deux années 2018 et 2019, comme on peut prendre cette augmentation en considération dans l'estimation du ratio de sinistralité de 2020.

2.3.3 Consommation par année

Année	Consommation
2018	1 802 599 022,00
2019	2 333 879 724,00

TABLE 2.5 – Consommation des deux exercices 2018 et 2019

La consommation a augmenté de 32,8% entre 2018 et 2019, sachant que la population a augmenté par 36%, ceci explique la grande augmentation de la consommation

2.3.4 Répartition par type de prestation

Nous résumons la répartition par type de prestation pour les deux années dans les deux tableaux suivants :

-Année 2018

Prestation	Consommation	Pourcentage
AUTRES EXAMENS	349 146 856	19,37%
AUXILIAIRE MEDICAUX	25 796 660	1,43%
BIOLOGIE	148 888 609	8,26%
CONSULTATION	192 496 054	10,68%
DENTAIRE	54 377 246	3,02%
HOSPITALISATION	75 123 037	4,17%
IMAGERIE ET EXAMENS SPECIALISES	64 883 218	3,60%
MATERNITE	30 703 417	1,70%
OPTIQUE	142 564 580	7,91%
PHARMACIE	716 277 483	39,74%
TRANSPORT MEDICALISE	2 341 862	0,13%
Total	1 802 599 022	100,00%

TABLE 2.6 – Répartition des sinistres 2018 selon la prestation

-Année 2019

Prestation	Consommation	Pourcentage
AUTRES EXAMENS	16 473 250	0,71%
AUXILIAIRE MEDICAUX	32 795 817	1,41%
BILAN DE SANTE	34 750	0,00%
BIOLOGIE	253 545 169	10,86%
CONSULTATION	288 094 618	12,34%
DENTAIRE	76 829 397	3,29%
HOSPITALISATION	112 067 786	4,80%
IMAGERIE ET EXAMENS SPECIALISES	99154786	4,25%
MATERNITE	41 560 869	1,78%
OPTIQUE	210 189 420	9,01%
PHARMACIE	1,201E+09	51,45%
TRANSPORT MEDICALISE	2 349 680	0,10%
Total	2 349 680	100,00%

TABLE 2.7 – Répartition des sinistres 2019 selon la prestation

Remarque

- Il est clair qu'il y' a plus de précision dans les prestations en 2019 que en 2018, parce que le pourcentage de la prestation (Autre examen) a démuni de 19,37% à 0,71%, comme on trouve une nouvelle prestation en 2019 (bilan de santé).

- Nous voyons une augmentation de 30% pour la prestation (pharmacie), pour les autres nous pouvons dire qu'elles sont restées stables.

- Nous pouvons déduire qu'on ne peut pas faire une tarification par prestation à cause du manque de précision dans les données de prestation de l'année 2018.

2.4 Analyse de l'inflation médicale

2.4.1 Coût moyen par personne

Année	Consommation	Nombre de personnes protégées	Coût moyen par Personne
2018	1 802 599 022,00	14 596	123 499,52
2019	2 333 879 724,00	19 847	117 593,57

TABLE 2.8 – Coût moyen par personne pour les deux exercices

Le coût moyen par personne protégé a diminué de 4,78% entre 2018 et 2019. Pour avoir des résultats plus précis sur la variation du coût moyen par personne, on peut comparer la consommation des mêmes personnes en 2018 et 2019, afin d'éliminer l'effet de la variation de la population étudiée.

2.4.2 Coût moyen (mensuel) par personne

Pour calculer le coût moyen par personne mensuel, on a déterminé le nombre de personnes protégées pour chaque mois, et ensuite on a calculé la consommation mensuelle, puis on a déduit ce coût, les résultats trouvés sont présentés dans les deux tableaux suivants :

2018	Consommation	Nombre de personnes protégées (ACTIF)	Nombre de personnes protégées (RETIRE)	total de personnes protégées	consommation par personne
Janvier	103 511 359	9 898	3 178	13 076	7 916,133298
Février	105 864 693	9 900	1 918	11 818	8 957,91953
Mars	119 793 571	9 988	1 860	11 848	10 110,86859
Avril	124 638 917	10 465	1 818	12 283	10 147,26997
Mai	130 876 823	10 475	1 380	11 855	11 039,79949
Juin	112 343 692	10 476	1 366	11 842	9 486,884986
Juillet	118 481 803	10 809	1 387	12 196	9 714,80838
Aout	200 247 512	10 826	1 170	11 996	16 692,85695
Septembre	135 550 853	10 979	1 008	11 987	11 308,15492
Octobre	168 347 361	11 172	988	12 160	13 844,35535
Novembre	204 614 572	11 268	951	12 219	16 745,60701
Décembre	278 327 566	11 336	908	12 244	22 731,75155

TABLE 2.9 – Coût moyen (mensuel) par personne pour exercice 2018

2018	Consommation	Nombre de personnes protégées (ACTIF)	Nombre de personnes protégées (RETIRE)	total de personnes protégées	consommation par personne
Janvier	125 754 449	11 122	1 835	12 957	9 705,52203
Février	133 373 465	11 371	1 875	13 246	10 068,9616
Mars	154 068 124	11 561	1 736	13 297	11 586,683
Avril	161 092 086	11 962	1 715	13 677	11 778,3203
Mai	150 871 142	12 047	1 649	13 696	11 015,7084
Juin	173 877 732	12 423	1 661	14 084	12 345,7634
Juillet	205 326 848	15 672	1 396	17 068	12 029,9302
Aout	191 374 828	15 950	1 323	17 273	11 079,4204
Septembre	236 716 077	16 260	1 243	17 503	13 524,3145
Octobre	286 661 990	16 417	1 239	17 656	16 235,9532
Novembre	287 055 529	16 497	979	17 476	16 425,6998
Décembre	227 707 454	17 749	966	18 715	12 167,1095

TABLE 2.10 – Coût moyen (mensuel) par personne pour exercice 2019

IL est clair que le coût moyen par personne se change d'un mois à un autre, ceci peut être expliqué par le changement climatique ou d'autres facteurs, donc on peut faire une tarification mensuelle pour les affiliés qui s'inscrivent pour des courtes périodes. Comme on observe que les coûts moyens des deux derniers mois de 2019 sont très faibles à cause des sinistres de ces deux mois ne sont pas encore clôturés, ainsi l'estimation de ces coûts sera l'objet de l'étape suivante de notre étude.

2.4.3 Estimation de la consommation des sinistres de 2019 et qui seront clôturés en 2020

On utilise l'augmentation du coût moyen mensuel, en effet on a remarqué que le coût moyen a augmenté par 9% durant les 10 premiers mois, donc on va supposer la même augmentation pour les deux derniers mois, donc l'estimation des sinistres de 2019 et qui seront clôturés en 2020 est comme suit :

Pour Novembre :

Coût moyen de novembre 2018	16 745,60701
Coût moyen de novembre 2019 déclaré	16 425,699
Estimation du coût moyen de novembre 2019	$16\,745,60701 \times 1,09 = 18\,252,7116$
Nombre de personne protégées de novembre 2019	17 476
Estimation des sinistres de novembre 2019 qui ne sont pas déclarés encore	$(18\,252,7116 - 16\,425,6998) \times 17\,476 = 319\,288,58,2$

TABLE 2.11 – Estimation des sinistres de novembre 2019 qui ne sont pas déclarés encore

Pour Décembre :

Coût moyen de décembre 2018	22 731,75155
Coût moyen de décembre 2019 déclaré	12 167,1095
Estimation du coût moyen de décembre 2019	$22\,731,75155 \times 1,09 = 24\,777,6092$
Nombre de personne protégées de décembre 2019	18 715
Estimation des sinistres de décembre 2019 qui ne sont pas déclarés encore	$(24\,777,6092 - 12\,167,1095) \times 18\,715 = 236\,005\,502$

TABLE 2.12 – Estimation des sinistres de décembre 2019 qui ne sont pas déclarés encore

Remarque

La somme de ces deux estimations est égale à 267 934 360 cette somme représente 11,48% de la consommation totale de l'année 2019.

2.4.4 Coût moyen par personne selon les types de prestations

D'abord on voit le coût moyen mensuel pour les dix premiers mois pour toutes les prestations, ensuite on le calcule pour chaque prestation.

TOUTES LES PRESTATIONS	
Coût moyen mensuel en 2018 (10 premiers mois)	10 921 ,9051
Coût moyen mensuel en 2019 (10 premiers mois)	11 937,0577
Augmentation	9%

TABLE 2.13 – Coût moyen mensuel par personne (10mois) pour toutes les prestations

PHARMACIE	
Coût moyen mensuel en 2018 (10 premiers mois)	4 808,34308
Coût moyen mensuel en 2019 (10 premiers mois)	6 026,43255
Augmentation	25%

TABLE 2.14 – Coût moyen mensuel par personne (10mois) pour la prestation ‘Pharmacie’

CONSULTATION	
Coût moyen mensuel en 2018 (10 premiers mois)	1 289,30421
Coût moyen mensuel en 2019 (10 premiers mois)	1 575,48745
Augmentation	22%

TABLE 2.15 – Coût moyen mensuel par personne (10mois) pour la prestation ‘Consultation’

BIOLOGIE	
Coût moyen mensuel en 2018 (10 premiers mois)	1 018,67151
Coût moyen mensuel en 2019 (10 premiers mois)	1 307,0976
Augmentation	28%

TABLE 2.16 – Coût moyen mensuel par personne (10mois) pour la prestation ‘Biologie’

Il est clair qu'il y a une augmentation importante pour le coût moyen mensuelle par prestation, ceci peut être expliqué par plusieurs facteurs la variation des prix, l'inflation et aussi l'augmentation dans la précision des prestations en 2019.

2.4.5 Coût moyen par sinistre

Dans les deux bases de données la variable qui définit un sinistre est NUMPREST, mais le problème est que cette variable considère chaque ligne comme un sinistre, donc on a créé une nouvelle variable qui définit le sinistre par la combinaison de la matriculeP et la date de soin, le tableau suivant résume les résultats trouvés :

	Consommation	Nombre de sinistre	Coût moyen par sinistre
2018	1 802 599 022,00	67 176	26 833,9738
2019	2 333 879 724,00	86 597	26 951,0459

TABLE 2.17 – Coût moyen par sinistre pour les deux exercices

On remarque que le coût moyen par sinistre a resté stable entre les deux années 2018 et 2019.

2.4.6 Coût moyen mensuel par sinistre

Pour calculer le coût moyen mensuel par sinistre, on détermine la consommation mensuelle puis on la divise par le nombre de sinistres mensuels, on résume les résultats obtenus dans les tableaux suivants :

2018	Consommation	Nombre de sinistres	consommation par personne
Janvier	103 511 359	4 740	7 916,133298
Février	105 864 693	4 531	8 957,91953
Mars	119 793 571	4 795	10 110,86859
Avril	124 638 917	4 822	10 147,26997
Mai	130 876 823	4 964	11 039,79949
Juin	112 343 692	5 188	9 486,884986
Juillet	118 481 803	5 681	9 714,80838
Aout	200 247 512	5 577	16 692,85695
Septembre	135 550 853	6 334	11 308,15492
Octobre	168 347 361	7 773	13 844,35535
Novembre	204 614 572	6 534	16 745,60701
Décembre	278 327 566	6 237	22 731,75155

TABLE 2.18 – Coût moyen mensuel par sinistre pour l'exercice de 2018

2019	Consommation	Nombre de sinistres	consommation par personne
Janvier	103 511 359	4 740	9 705,52203
Février	105 864 693	4 531	10 068,9616
Mars	119 793 571	4 795	11 586,683
Avril	124 638 917	4 822	11 778,3203
Mai	130 876 823	4 964	11 015,7084
Juin	112 343 692	5 188	12 345,7634
Juillet	118 481 803	5 681	12 029,9302
Aout	200 247 512	5 577	11 079,42045
Septembre	135 550 853	6 334	13 524,3145
Octobre	168 347 361	7 773	16 235,9532
Novembre	204 614 572	6 534	16 425,6998
Décembre	278 327 566	6 237	12 167,10955

TABLE 2.19 – Coût moyen mensuel par sinistre pour l'exercice de 2019

Maintenant on fait une comparaison entre la consommation mensuelle par sinistre entre les deux années 2018 et 2019 :

	Consommation par sinistre 2018	Consommation par sinistre 2019	Augmentation
Janvier	7 916,133298	9 705,52203	-3,75%
Février	8 957,91953	10 068,9616	-1,78%
Mars	10 110,868595	11 586,683	11,48%
Avril	10 147,26997	11 778,3203	11.65%
Mai	11 039,79949	11 015,7084	-3.03%
Juin	9 486,884986	12 345,7634	31.07%
Juillet	9 714,80838	12 029,9302	49,48%
Aout	16 692,85695	11 079,42045	-26,91%
Septembre	11 308,15492	13 524,3145	30.25%
Octobre	13 844,35535	16 235,9532	25,84%
Novembre	16 745,60701	16 425,6998	-8,56%
Décembre	22 731,75155	12 167,10955	-41,67%

TABLE 2.20 – Comparaison entre la consommation mensuelle par sinistre entre les deux années 2018 et 2019

On remarque une grande variation entre la consommation par sinistre des mois de 2018 et de 2019 malgré que la consommation par sinistre annuelle a resté constante.

2.4.7 Coût moyen de prestation par sinistre

Pour comparer le coût des prestations, on calcule le coût moyen de prestation pour chaque sinistre, les résultats trouvés sont résumés dans le tableau suivant :

	Coût moyen des prestations par sinistre 2019	Coût moyen des prestations par sinistre 2018	Augmentation
AUTRES EXAMENS	71 622,8261	1 587 031,16	-95,49%
AUXILIAIRE MEDICAUX	3 741,25222	3 791,39624	-1,32%
BIOLOGIE	13 919,5811	11 256,4156	23,66%
CONSULTATION	5 026,16267	4 289,98805	17,16%
DENTAIRE	51 015,5359	48 206,7784	5,83%
HOSPITALISATION	116 012,201	97 816,4544	18,60%
IMAGERIE ET EXAMENS SPECIALISES	17 583,7535	15 664,7074	12,25%
MATERNITE	127 879,597	117 188,615	9,12%
OPTIQUE	148 754,013	145 771,554	2,05%
PHARMACIE	17 328,5833	13 584,3856	27,56%
TRANSPORT MEDICALISE	195 806,667	146 366,375	33,78%

TABLE 2.21 – Coût moyen de chaque prestation par sinistre pour les deux années et leurs variations

Remarque

Il est clair que les prestations d'hospitalisation, dentaire, maternité et optique sont les plus coûteuses, pour le transport médicale les résultats trouvés ne sont pas significatifs car le nombre de cette prestation est très faible.

Comme on remarque une augmentation dans le coût moyen des prestations par sinistre avec des pourcentages importants, sauf le cas de la prestation (Autres Examen) qui a baissé par un pourcentage proche de 100%, ceci dit que en 2019 on a plus de précision dans les prestations, par contre en 2018, 20% de la consommation totale a été sous forme 'Autre Examen'.

2.5 Coût moyen par personne**2.5.1 Coût moyen par personne par prestation**

Année 2018

Statut	Consommation	Nombre de personnes protégées	Coût moyen par personne
A (assuré)	889 938 486	3 595	247 548,953
C (conjoint)	446 501 138	2 829	157 830,024
E (enfant)	444 124 346	8 007	55 467,0096
G	20 146 860	111	181 503,243
K	1 888 192	51	37 023,3725
Total	1 802 599 022	14 593	

TABLE 2.22 – Coût moyen par personne par prestation pour l'exercice de 2018

Année 2019

Statut	Consommation	Nombre de personnes protégées	Coût moyen par personne
A (assuré)	1,031E+09	4 705	219 095,95
C (conjoint)	615 498 962	3 822	161 041,068
E (enfant)	650 086 090	11 050	58 831,3204
G	35 673 653	186	191 793,833
K	1 774 575	84	21 125,8929
Total	2 333 879 724	19 847	

TABLE 2.23 – Coût moyen par personne par prestation pour l'exercice de 2019

On observe que la consommation par personne des enfants ne dépasse pas 26% de la consommation des Assurés et 36% de la consommation des conjoints pour les deux années.

Comme on remarque une diminution dans la consommation moyenne par personne pour les Assurés, par contre cette consommation augmenté pour les Enfant et les Conjoints, mais avec des pourcentages faibles.

2.5.2 Coût moyen par âge

Année 2018

Tranche d'Age	Coût moyen par personne	Nombre de personnes protégées
0-18	86 563,1314	9 836
19-30	116 123,918	3 351
31-40	154 537,773	3 262
41-50	166 694,674	2 045
51-60	188 375,509	1 100
61-65	195 872,518	170
Plus de 65	129 722,75	32

TABLE 2.24 – Coût moyen par tranche d'âge pour l'exercice de 2018

Année 2019

Tranche d'Age	Coût moyen par personne	Nombre de personnes protégées
0-18	94756,639	7059
19-30	112697,481	2399
31-40	164619,639	2695
41-50	167390,706	1502
51-60	205483,87	685
61-65	184466,08	112
Plus de 65	75040,2703	37

TABLE 2.25 – Coût moyen par tranche d'âge pour l'exercice de 2019

On remarque que les gens âgés moins de 18 ans consomment à peu près la moitié de la consommation des gens âgés plus de 30ans, comme il est clair que la consommation augmente avec l'âge, sauf pour la population âgée plus de 60 ans, car l'effectif de cette population est très faible, ceci explique les résultats trouvés.

2.6 Analyse par taux de remboursement

2.6.1 Répartition des affiliés selon le taux de remboursement

On analyse maintenant le coût moyen d'une personne selon le taux de remboursement auquel elle subit, le tableau suivant nous résume la répartition des affiliés selon le taux de remboursement.

TR	Nombre de personnes protégées en 2018	% en 2018	Nombre de personnes protégées en 2019	% en 2019
80%	5 425	37.17%	6 857	34.55%
90%	380	2.6%	622	3.13%
100%	8 477	58.08%	12 076	60.85%
Total	14 282	97.85%	19 356	98.53%

TABLE 2.26 – Répartition des affiliés selon le taux de remboursement pour les deux exercices

Remarque

- Le taux de remboursement dominant est 100% auquel environ 60% de la population protégée est soumise, comme on a d'autres taux de remboursements (30%, 70%) mais avec des pourcentages négligeables.

- Pour s'assurer de l'impact de taux de remboursement sur le comportement des assurés, il faut garantir que les populations soumises aux différents taux de remboursements sont similaires en termes de répartition enfant/adulte, seulement dans ce cas où la comparaison des coûts moyen est valide.

- Si ce n'est pas le cas, on ne peut pas déduire directement l'impact du taux de remboursement sur la consommation de deux personnes identiques.

- Ci-dessous, la répartition des assurés par taux de remboursement et par statut (enfant/adulte) pour les trois taux de remboursements les plus fréquents (80%, 90%, 100%)

2.6.2 Répartition des affiliés par taux de remboursement et statuts

Pour étudier l'effet du taux de remboursement sur le coût moyen par personne, on assure d'abord que la population étudiée a la même répartition selon le statut, car éliminer l'effet du statut.

En 2018			
TR	Enfant	Adulte	% des enfants
80%	2 769	2 656	51,04%
90%	210	170	55,26%
100%	4 845	3 632	57,15%

TABLE 2.27 – Répartition des affiliés par taux de remboursement et statuts pour l'exercice 2018

En 2019			
TR	Enfant	Adulte	% des enfants
80%	3575	3282	52.13%
90%	330	292	53.05%
100%	7036	5040	58,26%

TABLE 2.28 – Répartition des affiliés par taux de remboursement et statuts pour l'exercice 2019

D'après les tableaux ci-dessus, on peut déduire que les populations soumises aux différents taux de remboursement ont quasiment la même répartition (enfant/adulte). On peut considérer donc que les populations sont similaires et que le changement du coût moyen observé est dû au changement du taux de remboursement.

2.6.3 Analyse par taux de remboursement

Pour comprendre l'effet du taux de remboursement sur la consommation des affiliés, on fait comparaison entre le coût moyen par personne, pour les affiliés ayant un taux de remboursement de 100% avec les autres affiliés.

Année 2018

TR	Nombre de personnes protégées en 2018	CMP 2018	% de CMP TR 100%
80%	5 425	67 890,2216	46%
90%	380	33 663,02	23%
100%	8 477	145 831,793	

TABLE 2.29 – Comparaison entre le cout moyen par personne selon le taux de remboursement pour l'année 2018

Année 2019

TR	Nombre de personnes protégées en 2019	CMP 2019	% de CMP TR 100%
80%	6 857	60 851,0281	41%
90%	622	74 813,3553	50%
100%	12 076	149 401,968	

TABLE 2.30 – Comparaison entre le cout moyen par personne selon le taux de remboursement pour l'année 2019

En 2018 :

Les personnes ayant un TR de 80% consomment en moyenne environ 54% moins que celles ayant un remboursement total de 100%.

Les personnes ayant un TR de 90% consomment en moyenne environ 77% moins que celles ayant un remboursement total de 100%.

En 2019 :

Les personnes ayant un TR de 80% consomment en moyenne environ 59% moins que celles ayant un remboursement total de 100%.

Les personnes ayant un TR de 90% consomment en moyenne environ 50% moins que celles ayant un remboursement total de 100%.

2.6.4 Analyse par taux de remboursement(Hospitalisation)

Pour bien comprendre l'effet du taux de remboursement sur la consommation des personnes protégées, on prend en considération aussi la prestation.

Année 2018

TR	Nombre de personnes protégées en 2018	CMP 2018	% de CMP TR 100%
80%	5 425	4 395,17493	127%
90%	380	0	0%
100%	8 477	3 900,33786	

TABLE 2.31 – Comparaison entre le cout moyen par personne (Hospitalisation) selon le taux de remboursement pour l'année 2018

Année 2019

TR	Nombre de personnes protégées en 2019	CMP 2019	% de CMP TR 100%
80%	6 857	6 355,29634	108%
90%	622	3 128,30707	53%
100%	12 076	5 870,94319	

TABLE 2.32 – Comparaison entre le cout moyen par personne (Hospitalisation) selon le taux de remboursement pour l'année 2019

Remarque

On remarque que le taux de remboursement n'a pas un effet sur la consommation pour la prestation d'hospitalisation.

2.6.5 Analyse par taux de remboursement (Pharmacie)

Année 2018

TR	Nombre de personnes protégées en 2018	CMP 2018	% de CMP TR 100%
80%	5 425	22 085,6308	36%
90%	380	14 590,7158	24%
100%	8 477	60 574,1509	

TABLE 2.33 – Comparaison entre le cout moyen par personne (Pharmacie) selon le taux de remboursement pour l'année 2018

Année 2019

TR	Nombre de personnes protégées en 2019	CMP 2019	% de CMP TR 100%
80%	6 857	25 241,3694	31%
100%	12 076	79 697,1191	

TABLE 2.34 – Comparaison entre le cout moyen par personne (Pharmacie) selon le taux de remboursement pour l'année 2019

En 2018 :

Les personnes ayant un TR de 80% consomment en moyenne pour la prestation pharmacie environ 64% moins que celles ayant un remboursement total de 100%.

Les personnes ayant un TR de 90% consomment en moyenne environ 76% plus que celles ayant un remboursement total de 100%.

En 2019 :

Les personnes ayant un TR de 80% consomment en moyenne environ 69% moins que celles ayant un remboursement total de 100%.

Remarque

Il faut prendre en considération que les résultats de 2019 sont plus significatifs car, comme on a déjà vu, on a plus de précision dans les prestations de 2019 qu'en 2018.

CONCLUSION

Dans ce chapitre, nous avons présenté les données de base et les différents processus de détection des anomalies, ainsi que l'analyse du portefeuille maladie. Nous avons exposé aussi quelques statistiques globales sur la répartition de la population étudiée au deux niveaux : sinistralité

et composition du portefeuille. Le chapitre suivant sera consacré au cadre théorique, dans lequel nous allons définir les différentes approches adoptées pour modéliser la prime pure de la branche assurance maladie. En effet, l'élaboration du tarif maladie implique un processus de modélisation approprié s'adaptant avec la complexité de la branche étudiée

CADRE CONCEPTUEL ET THEORIQUE

INTRODUCTION

Un cycle de production inversé caractérise l'activité d'assurance : en contrepartie d'une prime d'un montant connu à la souscription du contrat, un risque de montant inconnu couvert par un engagement de l'assureur quel que soit la date de réalisation. La tarification proposée par l'assurance à pour objectif est d'évaluer la prime nécessaire pour couvrir les engagements et les frais de fonctionnement de l'assureur. Le tarif de l'assurance est nommée prime commerciale, qui est en fonction de la prime pure et des frais de fonctionnement.

Les engagements de l'assureur vis-à-vis les assurés sont couverts par une partie de la prime commerciale, c'est la prime pure. Cette dernière est un paramètre qui représente également le coût futur des risques, du coup, l'assureur doit calculer la prime pure correspondante à chaque groupe d'assurés. Ce qui lui amène à segmenter son portefeuille afin de répartir ses assurés en groupes homogènes. En effet , dans un marché fluide, si deux compagnies identiques ont les mêmes offres, les mêmes frais ainsi la même distribution, la compagnie la moins segmentée court le risque d'anti sélection.

L'actuaire cherche à définir des classes de risques homogènes, autrement dit ayant le même coup du risque (prime pure). En se basant de deux grandes classes de variables : les variables exogènes, et les variables endogènes. Les variables exogènes représentent les informations relatives au risque à l'exclusion des données relatives aux réalisations du risque. Et les informations permettant la réalisation du risque sont représentés par les variables endogènes.

En assurance maladie, les critères des assurés (âge, sexe, taux remboursement...) représentent les variables exogènes. À l'opposé les variables endogènes sont liés à la réalisation du risque assuré : le nombre de sinistres par garantie, le coût des sinistres...

Le choix de la méthode de segmentation et des variables tarifaires se base sur des objectifs commerciaux et aussi sur des méthodes statistiques. Dans ce chapitre, nous allons commencer par sélectionner les variables de tarification en utilisant un algorithme de sélection par optimisation. Par la suite, nous allons segmenter les assurés.

Ce présent chapitre vise à proposer une méthode de tarification pour les assurés de branche

maladie. Plusieurs facteurs de tarification sont intégrés dans cette méthode, en introduisant les modèles linéaires généralisés comme étant une approche statistique qui consiste à modéliser la prime pure de l'assurance maladie.

3.1 Principe de tarification

L'élaboration d'un tarif en assurance santé se fait classiquement par une approche de type fréquence x coût, afin de déterminer les primes pures. À partir des primes pures nous déduisons les primes commerciales payées par les assurés.

Tout d'abord, on doit étudier le montant global de primes de base à amasser afin de payer les sinistres potentiels. Par la suite, On choisit les caractéristiques des assurés liés à la survenance du risque santé pour proposer une prime pure adaptée à chacun.

On appelle aussi ces caractéristiques des facteurs de tarification, puisque les primes sont directement déterminées à partir de ces éléments. Par conséquent, l'effet des variables explicatives sur le niveau du risque est estimé par des modèles de régression de type GLM.

Cette estimation n'est fiable que si certaines hypothèses fortes sont vérifiées. Si on considère un échantillon de deux séquences de variables aléatoires : $(X_i)_{1 \leq i \leq n}$ et $(N_i)_{1 \leq i \leq A}$, alors ces hypothèses sont :

- L'indépendance et l'identique distribution des variables aléatoires X_i représentatives des « Coûts de consommation annuels » d'un assuré i .
- L'indépendance entre les variables aléatoires X_i et les variables aléatoires N_i représentatives des « Fréquences de consommation annuelles » relativement à un assuré i .

Soit A le nombre total d'assurés exposé au risque santé. Alors, le nombre de consommations total en biens et services de santé est :

$$N = \sum_{i=1}^A N_i$$

Et le coût total de toutes les consommations est :

$$S = \sum_{i=1}^N X_i$$

On s'intéresse à l'espérance de la distribution :

$$E(S) = E\left(\sum_{i=1}^N X_i\right) = \sum_{k=1}^{\infty} Pr(N = k) \times E\left(\sum_{i=1}^k X_i\right) = \sum_{k=1}^{\infty} Pr(N = k) \times k \times E(X_1) = E(N) \times E(X_1)$$

Ainsi, l'idée directrice de notre approche est de considérer le coût moyen de toutes les consommations en santé égal au produit du nombre moyen de consommations et de leur coût moyen.

Par la suite, cela nous mène à percevoir la prime pure annuelle π comme étant un coût moyen de consommations annuelles d'un individu multiplié par la fréquence annuelle moyenne

$$\pi = E(S) = E(N) \times E(X_1)$$

Regardons maintenant, la variance de la variable aléatoire S définie par :

$Var(S) = Var(E(S|N)) + E(Var(S|N))$ Sachant que :

$$E(S|N) = E\left(\sum_{i=1}^N X_i | N\right) = N \times E(X_1)$$

Et que :

$$\begin{aligned} Var(S|N) &= Var\left(\sum_{i=1}^N X_i | N\right) \\ &= Var\left(\sum_{i=1}^N X_i\right) \quad \text{par indépendance de } N \text{ et } X_i \\ &= \sum_{i=1}^N Var(X_i) \quad \text{par indépendance des } X_i \\ &= N \times Var(X_1) \end{aligned}$$

Ce qui nous donne :

$$Var(E(S|N)) = Var(N \times E(X_1)) = E^2(X_1) \times Var(N)$$

Et :

$$E(Var(S|N)) = E(N \times Var(X_1)) = E(N) \times Var(X_1)$$

En conclusion :

$$\text{Var}(S) = E^2(X_1) \times \text{Var}(N) + E(N) \times \text{Var}(X_1)$$

On remarque que la variance du coût total des consommations dépend de deux composantes : l'une liée à l'incertitude de la fréquence annuelle de consommation et l'autre due à l'incertitude sur le coût de consommation.

3.2 Choix de modèle par sélection de variables

3.2.1 Pourquoi la sélection de variables

La sélection de variables est une étape primordiale de la modélisation. Dans les études réelles, nous sommes face à des bases de données qui contiennent un grand nombre d'assurés. Ce sont autant de variables explicatives potentielles. Certaines d'entre elles sont redondantes, d'autres n'ont aucun rapport avec la variable dépendante. La méthode statistique doit nous donner des indications sur le sous-ensemble des bonnes variables explicatives à inclure dans le modèle. Dans l'idéal, elles devraient être fortement liées avec la variable dépendante.

Plus un modèle contient de variables explicatives, plus il est précis mais moins il est robuste. À l'inverse, moins un modèle a de variables explicatives, plus il est robuste mais moins il est précis. En effet, l'ajout d'une nouvelle variable explicative apporte des informations supplémentaires sur la variable à expliquer mais impose de fait une contrainte supplémentaire au modèle.

Prenons un exemple concret. Si on décrit un individu en donnant sa taille, son poids, ses caractères physiques, sa coupe de cheveux et ses vêtements, on aura une description très précise de cet individu. On comprendra immédiatement de qui il s'agit. Par contre, si la personne change de vêtements ou de coupe de cheveux par exemple, on n'est plus sûr que ce nouvel individu soit le même que l'individu initial. Le modèle n'est donc pas robuste. Ne considérer que la taille, le poids et les différents caractères physiques comme la couleur de la peau, des yeux, aurait certainement suffi dans les deux cas à identifier l'individu.

Il s'agira donc, dans notre étude, de déterminer la combinaison de variables explicatives qui permettra d'obtenir le meilleur compromis entre précision et robustesse.

3.2.2 Critères de sélection de variables

Une variable quantitative Y dite à expliquer (ou réponse) est mise en relation avec p variables explicatives X_1, \dots, X_p (ou indépendantes). Les données sont supposées provenir de l'observation d'un échantillon statistique de taille n .

Cependant, la détermination des variables explicatives nécessite un processus de sélection de variables approprié. Ils existent plusieurs méthodes qui ont pour objectif la sélection des variables explicatives. Ces méthodes nécessitent un cadre théorique précis. Elles reposent principalement sur la régression. Dans la suite nous allons présenter quelques-unes de ces méthodes.

La régression ascendante et descendante

La méthode ascendante ou forward consiste à l'introduction au fur et à mesure des variables en démarrant au début par un modèle n'ayant qu'une constante. Si on possède un modèle avec potentiellement k variables, on réalise k régressions en changeant à chaque fois la variable explicative introduite. On retiendra la variable explicative la plus significative au sens du test de Student à chaque itération. On réitère ce processus jusqu'au moment où aucune nouvelle variable explicative introduite n'est significative au test de Student.

La méthode descendante ou backward consiste à considérer un modèle complet puis on retire au fur et à mesure les variables explicatives non significatives au test de Student sur le même principe que la méthode en avant.

L'ordre d'introduction peut avoir de l'importance notamment si on introduit des variables explicatives corrélées du fait de l'instabilité des coefficients de régression. Sur ce point il faut noter que l'élimination successive de deux variables par la procédure descendante n'équivaut pas à la recherche du couple de variables à éliminer maximisant la log-vraisemblance sur tous les couples éliminés possibles. Il en sera de même par similitude pour la procédure ascendante. Souvent l'exécution des deux procédures donne des résultats incompatibles entre eux.

Ces méthodes sont basées sur le calcul des sommes des carrés et peuvent donner des estimations non optimales en fonction des cas. La dichotomisation de variables qualitatives permet leurs introductions sous réserve de retirer une modalité afin d'assurer l'indépendance ou la non corrélation dans le modèle vis à vis de ces variables.

La régression stepwise

La méthode stepwise corrige les principales lacunes de la méthode en avant. Elle introduit de nouvelles variables en essayant d'augmenter le plus possible le coefficient de corrélation multiple R^2 . A chaque étape, elle contrôle la pertinence de l'introduction des variables antérieures. Une variable significative précédemment ne le sera pas obligatoirement à l'étape suivante, notamment en raison de corrélation. La méthode retire alors les variables les moins significatives d'entre elles. Le processus s'arrête lorsque plus aucune variable ne peut être retirée ou introduite.

Cette procédure semble meilleure que les deux précédentes. Néanmoins la multicollinéarité peut perturber le modèle en fonction de l'ordre d'introduction des variables corrélées. Cette méthode ne se soucie pas des interrelations entre les variables non encore sélectionnées et elle introduit une nouvelle variable à la fois. Par conséquent des variables importantes peuvent être "oubliées" pour la sélection.

Cette méthode est basée sur le calcul des sommes des carrés et peut donner des estimations non optimales en fonction des cas. La dichotomisation de variables qualitatives permet leurs introductions sous réserve de retirer une modalité afin d'assurer l'indépendance ou la non corrélation dans le modèle vis à vis de ces variables.

Généralement, nous fixons un risque plus exigeant pour la sélection (ex. 5%, nous ne faisons entrer la meilleure variable que si elle est significative à 5%) que pour la suppression (ex. 10%, nous supprimons la variable la moins pertinente si elle est non significative à 10%).

La recherche exhaustive

On considère tous les modèles de régression possibles. Pour k variables explicatives potentielles il y a $\sum_{r=0}^k \binom{k}{r} = 2^k$ régressions à réaliser. Il faut souligner que tous les modèles comprennent un terme constant. Ce procédé n'a de sens qu'en croisant les différents modèles pour valider le plus adéquat. La procédure "Leaps bonds" de Furnival Wilson (pour toutes natures de données) propose un équivalent retenant les régressions optimisées à l'aide du R^2 , R^2 ajusté ou C_p de Mallows pour toutes les tailles de modèle. Par conséquent, elle ne fait pas intervenir de critères externes pour influencer sur les critères en entrée ou sortie mais nécessite beaucoup de calculs.

Toutes les méthodes n'aboutissent pas forcément à la même solution de sélection. La lecture des résultats peut permettre d'identifier ou de mesurer la corrélation entre variables ou d'un ensemble de variables. Les principaux éléments de lecture dans cette optique sont :

-La régression avec la lecture du "Facteur d'inflation de la variance". Cette approche permet d'indiquer la présence d'un problème de corrélation entre plusieurs variables. $VIF_j = \frac{1}{1-R_j^2}$ où R_j^2 est le coefficient de détermination du modèle. Si VIF est supérieur à 100 on considère qu'il y a un risque de colinéarité. Si les variables sont corrélées, les variables explicatives expliquent ou contiennent une part de la variable à expliquer.

-La règle de Klein montre que si le coefficient de corrélation partielle est plus élevé que le R^2 il y a alors multicolinéarité.

3.3 Etude théorique selon l'approche GLM

3.3.1 Généralités

le modèle linéaire généralisé (MLG) souvent connu sous les initiales anglaises GLM est une généralisation souple de la régression linéaire. Le GLM généralise la régression linéaire en permettant au modèle linéaire d'être relié à la variable réponse via une fonction lien et en autorisant l'amplitude de la variance de chaque mesure d'être une fonction de sa valeur prévue.

Les modèles linéaires généralisés ont été formulés par John Nelder et Robert Wedderburn comme un moyen d'unifier les autres modèles statistiques y compris la régression linéaire, la régression logistique et la régression de Poisson. Ils proposent une méthode itérative dénommée méthode des moindres carrés repondérés itérativement pour l'estimation du maximum de vraisemblance des paramètres du modèle. L'estimation du maximum de vraisemblance reste populaire et est la méthode par défaut dans de nombreux logiciels de calculs statistiques. D'autres approches incluant les statistiques bayésiennes et la méthode des moindres carrés convenant aux réponses à variance stabilisées, ont été développées.

Aujourd'hui les modèles linéaires généralisés sont couramment utilisés par les professionnels de l'assurance pour tarifier les branches personnelles.

3.3.2 Rappel du modèle linéaire simple

Le modèle linéaire simple consiste à établir une relation du type linéaire entre une variable à expliquer (aléatoire) et des variables explicatives (déterministes).

L'équation du modèle est donc de la forme :

$$Y = \alpha_0 + \sum_{i=1}^n \alpha_i X_i + \varepsilon$$

où :

-Y est la variable à expliquer.

$-\varepsilon$ est une variable aléatoire de loi normale de moyenne nulle et de variance σ^2 .

$-(X_1, \dots, X_n)$ sont les variables explicatives déterministes.

$-(\alpha_1, \dots, \alpha_n)$ sont les paramètres du modèle à estimer.

Ce modèle suppose donc que la variable à expliquer suit une loi normale de moyenne $(\alpha_0 + \sum_{i=1}^n \alpha_i X_i)$ et de variance σ^2 .

Les conditions sur ε est difficile à satisfaire et demande des tests de normalité. Un des avantages du Modèle Linéaire Généralisé (MLG) est de supprimer cette contrainte.

3.3.3 Distributions de la famille exponentielle naturelle

Les distributions de la famille exponentielle naturelle sont indispensables pour la mise en place d'un MLG. C'est pourquoi, nous allons présenter ici les principales distributions qui sont généralement utilisées.

Forme générale

Soit Y une variable aléatoire et y une observation de Y.

La loi de probabilité de Y appartient à la famille exponentielle naturelle si et seulement si elle peut se mettre sous la forme :

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

où :

-a est une fonction non nulle définie sur \mathbb{R} .

-b est une fonction définie sur \mathbb{R} , deux fois dérivable.

-c est une fonction définie sur \mathbb{R}^2 .

$-\theta$ paramètre canonique ou paramètre de la moyenne.

$-\phi$ paramètre de dispersion.

L'espérance de Y s'écrit alors $\mu = E(Y) = b'(\theta)$

On introduit la fonction de lien canonique g telle que : $\theta = g(\mu)$

La variance de Y s'écrit : $\text{Var}(Y) = b''(\theta) \times a(\phi)$

La fonction variance se définit par : $b''(\theta) = V(\mu)$

Loi normale

Soit Y une variable aléatoire suivant une loi normale d'espérance μ et de variance σ^2 . Y est à valeurs réelles.

Sa fonction de densité est :

$$f_{\mu, \sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

qui peut s'écrire sous la forme :

$$f_{\mu, \sigma}(y) = \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2 + \ln(2\pi\sigma^2)}{2}\right\}$$

ainsi la loi gaussienne appartient à la famille exponentielle naturelle avec :

$$-\theta = \mu$$

$$-\phi = \sigma^2$$

$$-a(\phi) = \phi$$

$$-b(\theta) = \frac{\theta^2}{2}$$

$$-c(y, \phi) = -\frac{1}{2} \left[\frac{y^2}{\phi} + \ln(2\pi\phi) \right]$$

Loi de Poisson

Soit Y une variable aléatoire suivant une loi de Poisson de paramètre λ . Y est à valeurs discrètes.

Sa fonction de densité est de la forme :

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

qui peut être mis sous la forme :

$$P(Y=y) = \exp(y \ln \lambda - \lambda - \ln(y!))$$

ainsi la loi de Poisson appartient à la famille exponentielle naturelle avec :

$$-\theta = \ln(\lambda)$$

$$-\phi = 1$$

$$-a(\phi) = 1$$

$$-b(\theta) = \exp(\theta)$$

$$-c(y, \phi) = -\ln(y!)$$

Loi binomiale négative

Soit Y une variable aléatoire suivant une loi binomiale négative de paramètres r et p (r étant un entier strictement positif, p un réel compris entre 0 et 1 exclus).

L'expérience consiste en une série de tirages indépendants, donnant un "succès" avec une probabilité p (constante durant toute l'expérience) et un "échec" avec une probabilité complémentaire. Y représente le nombre de tirages nécessaires pour obtenir r succès. Y est à valeurs entières.

Sa fonction de densité est :

$$P(Y=y) = \binom{y+r-1}{y} p^r (1-p)^y$$

qui peut s'écrire sous la forme :

$$P(Y=y) = \exp(y \ln(1-p) + r \ln p + \ln \frac{y^{r-1}}{\Gamma(r)})$$

où

$$\Gamma(x) = \int_0^{\infty} e^{-u} u^{x-1} du$$

ainsi la loi binomiale négative appartient à la famille exponentielle naturelle avec :

$$-\theta = \ln(1-p)$$

$$-\phi = 1$$

$$-a(\phi) = 1$$

$$-b(\theta) = -r \ln(1 - e^{\theta})$$

$$-c(y, \phi) = \ln \frac{y^{r-1}}{\Gamma(r)}$$

Loi Gamma

Soit Y une variable aléatoire suivant une loi Gamma de paramètre r et α (tous deux strictement positifs).

La densité s'écrit :

$$f_{r,\alpha}(y) = \frac{\alpha^r}{\Gamma(r)} y^{r-1} \exp(-\alpha y)$$

où

$$\Gamma(x) = \int_0^{\infty} e^{-u} u^{x-1} du$$

La fonction de densité peut également être mise sous la forme :

$$f_{r,\alpha}(y) = \exp(-r \ln \alpha - \alpha y + (r-1) \ln y - \ln \Gamma(\alpha))$$

ainsi la loi Gamma appartient à la famille exponentielle naturelle avec :

$$-\theta = -\alpha$$

$$-\phi = 1$$

$$-a(\phi) = 1$$

$$-b(\theta) = -r \ln(-\theta)$$

$$-c(y, \phi) = (r-1) \ln y - \ln \Gamma(\alpha)$$

Cas particulier de la loi log-normale

Soit Y une variable aléatoire suivant une loi log-normale de paramètres μ et σ^2 .
Y est à valeurs réelles.
Sa fonction de densité est :

$$f_{\mu, \sigma}(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\right\}$$

Cette loi n'est pas une exponentielle naturelle. On ne peut donc pas appliquer un MLG sur la variable Y directement. L'astuce consiste à poser $X = \ln(Y)$. X suit alors une loi normale de paramètres μ et σ^2 .

3.3.4 Hypothèses et principe du Modèle Linéaire Généralisé

Hypothèses

Le modèle linéaire généralisé part du même principe que celui du modèle linéaire simple. La différence est qu'au lieu de modéliser la variable à expliquer directement, c'est une fonction de l'espérance de cette variable (appelée fonction lien) qui est modélisée.

Une variable aléatoire Y relève du modèle linéaire généralisé si la loi de Y sachant $\{X_1 = x_1, \dots, X_n = x_n\}$ est telle que :

-Il existe une fonction lien g strictement monotone de \mathbb{R} dans \mathbb{R} et des coefficients $(\alpha_0, \dots, \alpha_n)$ tels que :

$$g(E[Y]) = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$$

-La loi de probabilité de Y doit appartenir à la famille exponentielle naturelle.

Les paramètres $(\alpha_0, \dots, \alpha_n)$ sont les coefficients de régression et la quantité $g(E[Y])$ est le prédicteur linéaire.

L'estimation des paramètres de ces lois se fait par maximum de vraisemblance et le choix de la distribution pour le modèle linéaire généralisé se fait en observant les données.

Fonctions de lien

Les fonctions de lien classiques sont résumés dans le tableau suivant :

Fonction identité	$g : z \rightarrow z$
Fonction logarithme	$g : z \rightarrow \ln(z)$
Fonction inverse	$g : z \rightarrow \frac{1}{z}$
Fonction logit	$g : z \rightarrow \ln\left(\frac{z}{1-z}\right)$
Fonction probit	$g : z \rightarrow \Phi(z)$

TABLE 3.1 – les fonctions de liens classiques

où Φ est la fonction de répartition de la loi normale $N(0,1)$

Il y a deux cas où le choix n'est pas laissé à l'opérateur pour la fonction de lien. Ce sont les cas du modèle additif ou du modèle multiplicatif. Dans le cas d'un modèle additif, il faudra opter pour une fonction de lien identité et pour un modèle multiplicatif pour une fonction de lien logarithme.

En effet, si la fonction de lien est l'identité, on a :

$$E(Y) = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$$

Ou encore :

$$E(Y) = \sum_{i=0}^n \beta_i$$

Avec :

$$-\beta_0 = \alpha_0$$

$$-\beta_i = \alpha_i X_i$$

Il s'agit bien d'un modèle additif.

Si la fonction de lien est le logarithme, on a :

$$E(Y) = \exp\left(\alpha_0 + \sum_{i=1}^n \alpha_i X_i\right)$$

Ou encore :

$$E(Y) = \prod_{i=0}^n \beta_i$$

Avec :

$$-\beta_0 = \exp(\alpha_0)$$

$$-\beta_i = \exp(\alpha_i X_i)$$

Il s'agit bien d'un modèle multiplicatif.

En choisissant le couple densité de la variable à expliquer/fonction de lien adéquat, le MLG correspond à des modèles statistiques classiques. Ainsi, par exemple :

Type de régression	Loi des observations	Fonction de lien
Régression linéaire classique	Normal	Identité
Régression logistique	Bernoulli	Logit

TABLE 3.2 – Exemples de MLG correspond à des modèles statistiques classiques

Mis à part ces cas particuliers où le choix de la fonction de lien ne se pose pas, le statisticien est libre de sélectionner la fonction de lien qu'il jugera pertinente.

3.3.5 Estimation des paramètres

L'estimation des paramètres du MLG se fait par maximum de vraisemblance dont nous rappelons ici la méthode.

Rappelons tout d'abord que la vraisemblance est, par définition, un produit de fonctions de densité. Pour en déterminer le maximum, il suffit de déterminer la valeur du paramètre de la fonction qui l'annule tout en gardant la dérivée seconde négative. Pour des raisons pratiques, on préfère dériver le logarithme de la vraisemblance. On dérive ainsi une somme plutôt qu'un produit. De plus, comme la fonction logarithme est strictement croissante, maximiser le logarithme de la fonction équivaut à maximiser la fonction.

Concrètement, prenons l'exemple d'une loi exponentielle de paramètre θ et de densité f . Soit Y une variable aléatoire suivant cette loi et (y_1, \dots, y_p) p observations de cette variable.

Rappelons que notre modèle est de la forme :

$$g(\mu) = \alpha_0 + \sum_{i=0}^n \alpha_i X_i$$

$$\text{avec } \mu = E(Y)$$

Nous avons de plus :

$$\theta = g_c(\mu)$$

On en déduit :

$$\theta = g_c(g^{-1}(\alpha_0 + \sum_{i=0}^n \alpha_i X_i))$$

Si on note la fonction de vraisemblance :

$$L(y_1, \dots, y_p, \theta) = \prod_{i=1}^p f(y_i, \theta)$$

Cette fonction de vraisemblance s'écrit également :

$$L(y_1, \dots, y_p, \alpha_0, \dots, \alpha_n) = \prod_{i=1}^p f(y_i, g_c(g^{-1}(\alpha_0 + \sum_{i=0}^n \alpha_i X_i)))$$

L'équation à résoudre sera :

$$\frac{\partial \ln(L(y_1, \dots, y_p, \alpha_0, \dots, \alpha_n))}{\partial \alpha} = \sum_{i=1}^p \frac{\partial \ln f(y_i, \alpha_0, \dots, \alpha_n)}{\partial \alpha} = 0$$

On obtient ainsi $\hat{\alpha} = (\hat{\alpha}_0, \dots, \hat{\alpha}_n)$ les estimations des paramètres du modèle.

Ces estimations nous seront données par la suite par le logiciel SAS et la procédure GENMOD.

3.3.6 Choix des facteurs

Dans cette section nous décrivons le processus de choix des facteurs intervenants dans la modélisation. Nous étudions préalablement les distributions des dérivées partielles de la fonction de log-vraisemblance, des estimateurs du maximum de vraisemblance et de la logvraisemblance afin de déterminer les propriétés de la déviance. Des tests statistiques sur la déviance nous permettent alors de sélectionner les facteurs pertinents.

Généralités

Si $\hat{\theta}$ est un estimateur consistant du paramètre θ et $var(\hat{\theta})$ variance de l'estimateur, alors pour un grand échantillon de données, la statistique

$$\frac{\hat{\theta} - \theta}{\sqrt{var(\hat{\theta})}}$$

suit une loi $N(0,1)$ et son carré suit une loi du χ^2 à un degré de liberté. La généralisation de ce résultat à n paramètres donne :

$$(\hat{\theta} - \theta)^T V^{-1} (\hat{\theta} - \theta) \sim \chi_n^2$$

où $\hat{\theta}$ est un estimateur convergent de θ , V la matrice de variance covariance de $\hat{\theta}$. $\hat{\theta}$ est asymptotiquement un estimateur sans biais de θ et V est une matrice inversible.

Inférence sur les paramètres

Faire de l'inférence statistique signifie émettre des conclusions concernant une population à partir de résultats obtenus sur un échantillon de cette population. Nous avons vu comment ajuster un GLM à des données provenant d'un échantillon. Nous allons maintenant chercher à établir des conclusions sur la population cible à l'étude à partir des estimations calculées pour les coefficients.

En plus d'estimer le vecteur α de paramètres, on calcule une matrice de variance-covariance asymptotique pour le vecteur des estimations obtenues.

Cette matrice est $\hat{\sigma}^2(\hat{\alpha}) = var(\hat{\alpha}) = I^{-1}(\hat{\alpha})$ où I est la matrice d'information observée dans l'algorithme de Newton-Raphson.

Les erreurs-types des paramètres sont la racine carrée des éléments sur la diagonale de cette matrice.

Ainsi, $\hat{\sigma}(\hat{\alpha}_j)$ est la racine carrée de l'élément correspondant à α_j de la matrice $var(\hat{\alpha})$.

On peut se servir de ces erreurs-types pour construire des intervalles de confiance de Wald de niveau $(1-\beta)\%$ pour tous les paramètres du vecteur comme suit :

$$\alpha_j \in [\hat{\alpha}_j - z_{\beta/2} \hat{\sigma}(\hat{\alpha}_j), \hat{\alpha}_j + z_{\beta/2} \hat{\sigma}(\hat{\alpha}_j)]$$

Où $j = 1, \dots, n$

Test de Wald sur un paramètre

Tout d'abord, un paramètre j nul signifie que la variable explicative dont il est le coefficient, x_j , n'a pas de lien avec la variable réponse N . Il est donc d'intérêt de tester si la valeur d'un paramètre est nulle dans la population étudiée. Nous pouvons faire ce type d'évaluation grâce au test de Wald pour confronter les hypothèses suivantes :

$H_0 : \alpha_j = 0$ (il n'y a pas de lien entre x_j et N) Vs $H_1 : \alpha_j \neq 0$ (il y a un lien entre x_j et N)

Nous utilisons le test de Wald, basé sur la distribution asymptotique de $\hat{\alpha}_j$. La statistique du test est : $Z = \frac{\hat{\alpha}_j}{\hat{\sigma}(\hat{\alpha}_j)}$.

Sous H_0 , la statistique du test suit approximativement une loi normale $N(0,1)$. Il est à noter que le carré de la statistique du test de Wald suit approximativement une loi de khi-deux à un degré de liberté.

Soit $Z_w^2 = \left(\frac{\hat{\alpha}_j}{\hat{\sigma}(\hat{\alpha}_j)}\right)^2 \sim \chi_{(1;1-\beta)}^2$ où β est le seuil de signification du test.

En général, nous prenons $\beta=0.05$, c'est-à-dire si la p-value observée est inférieure de ce seuil, nous rejetons l'hypothèse nulle du test. Dans ce cas, et seulement dans ce cas, la variable explicative x_j est significative (il y a un lien entre x_j et la variable réponse).

Prédiction de la variable réponse

Une valeur prédite par le modèle est définie par :

$$\hat{\lambda}_j = g^{-1}(\hat{\alpha}_0 + \hat{\alpha}_1 x_{j,1} + \dots + \hat{\alpha}_n x_{j,n})$$

Donc, pour une observation n_i associée au vecteur de variables explicatives x_i , sa valeur prédite par le modèle ayant le logarithme de fonction de lien est :

$$\hat{n}_i = \exp(\hat{\alpha}_0 + \hat{\alpha}_1 x_{i,1} + \dots + \hat{\alpha}_n x_{i,n})$$

Ainsi, nous prédisons la variable réponse N par son espérance, qui est elle fonction des valeurs prises par les variables explicatives.

Nous pouvons aussi prédire la valeur moyenne de N pour les valeurs de x non observées dans l'échantillon étudié. D'où l'importance des modèles linéaires généralisés dans ce type de traitement.

3.3.7 Validation du modèle

L'étape de la validation du modèle est qualifiée de l'une des parties les plus importantes des modèles linéaires généralisés. En effet, il est indispensable de s'assurer que le modèle considéré s'ajuste bien aux données avant de pratiquer l'inférence statistique. A cet effet, nous utilisons la déviance comme statistique de test qui permet de valider un modèle linéaire généralisé.

La déviance

La déviance D est l'écart en terme de log-vraisemblance entre le modèle saturé d'ajustement maximum et le modèle considéré.

$$D = 2 \sum_{i=1}^m (l_{\text{saturé}}(n_i) - l(\hat{\lambda}_i))$$

Si le modèle était parfait (modèle saturé), $\hat{\lambda}_i$ serait égale à n_i ($E(N_i | x_i) = \lambda_i = n_i$)

Un modèle présente un bon ajustement aux données (Goodness-of-fit), si les valeurs prédites par le modèle sont similaires aux valeurs observées. Le modèle saturé présente donc un ajustement parfait puisque ses valeurs prédites sont toutes égales aux valeurs observées.

Un tel modèle n'est pas désirable puisqu'il comporte trop de paramètres, il est trop complexe. De plus, trop coller aux données n'est pas toujours une bonne chose puisque celles-ci proviennent

d'un échantillon aléatoire et non de la population complète. On désire un modèle parcimonieux qui présente des relations théoriques interprétables.

La valeur de la déviance obtenue en se basant sur le modèle considéré, permet de tirer la qualité de l'adéquation de ce modèle aux données de l'étude. En effet, si la déviance est petite, le modèle considéré sera adéquat. Dans le cas contraire, c'est-à-dire si la déviance est grande, alors le modèle considéré est loin du modèle saturé et donc il n'est pas très adéquat.

*** La statistique du test**

La statistique de déviance permet de tirer le modèle qui présente le bon ajustement aux données de base. Pour cela, nous confrontons les deux hypothèses suivantes :

H_0 : Le modèle considéré à n paramètres est adéquat

Vs

H_1 : Le modèle considéré à n paramètres n'est pas adéquat

La statistique du test est notée D, et sa formule est la suivante :

$$D = 2 \sum_{i=1}^m (n_i \log(\frac{n_i}{\lambda_i}) - (n_i - \hat{\lambda}_i)) \sim \chi_{m-n}^2$$

Le modèle est adéquat si la valeur de D s'approche de m-n. Il est à préciser que m représente le nombre d'observation et n le nombre de paramètres à estimés.

Pratiquement, si $D_{observé} > \chi_{m-n, 1-\beta}^2$ alors H_0 est rejetée. Autrement dit, le modèle considéré n'est pas adéquat.

Il existe d'autres statistiques de déviance, c'est le cas de la statistique de Pearson qui permet de mesurer l'adéquation du modèle considéré aux données de l'étude. Cette statistique se base sur les mêmes hypothèses précédemment citées, mais la différence réside dans la formule de la statistique du test :

$$Q^P = \sum_{i=1}^m (\frac{n_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}}) \sim \chi_{m-n}^2$$

Le modèle est adéquat si la valeur de Q^P est proche de m-n.

Analyse des résidus

Considérons un modèle Normal dont les variables de réponse Y_i sont modélisées par :

$$Y_i = \mu_i + e_i$$

Où les termes d'erreur e_i sont supposés indépendants et identiquement distribués $N(0, \sigma^2)$ et les espérance μ_i sont des fonctions de combinaisons linéaires des paramètres du vecteur α .

Pour ce modèle $\frac{y_i - \mu_i}{\sigma} \sim N(0, 1)$

Le résidu associé à Y_i est défini par $(y_i - \hat{\mu}_i)$ où $\hat{\mu}_i$ est déterminé par le modèle. Il est calculé à partir de l'estimation du maximum de vraisemblance. Le résidu standardisé est défini comme $r_i = \frac{y_i - \hat{\mu}_i}{\hat{\sigma}}$ où $\hat{\sigma}$ est un estimateur de σ . Par conséquent, les résidus standardisés suivent approximativement la loi $N(0,1)$.

Pour les autres modèles linéaires généralisés, les résidus sont définis par analogie avec le cas Normal $r_i = \frac{y_i - \hat{\mu}_i}{s_i}$ où s_i est la valeur estimée de l'écart type de la valeur modélisée $\hat{\mu}_i$.

Les résidus standardisés peuvent être comparés à la distribution Normale pour évaluer la qualité du choix de la distribution dans le modèle.

Choix entre différents modèles

Plusieurs modèles concurrents sont en compétition, le but est de choisir le plus adéquat aux données de l'étude. Pour cela, et lorsque les modèles sont emboîtés, il est possible d'utiliser le test de déviance entre modèles ou bien des critères de choix de modèles (AIC et BIC).

H_0 : le modèle simple à n_1 paramètres est adéquat

Vs

H_1 : le modèle simple à n_2 paramètres est adéquat ($n_2 > n_1$)

Pour comparer deux modèles emboîtés, nous allons utiliser l'écart de déviance qui mesure la différence entre les deux modèles en question : $\Delta D = D_{simple} - D_{grand}$.

Si ΔD est grande, le fait de passer d'un modèle simple à un modèle plus grand a donc apporté un écart de déviance significatif et donc le modèle général est acceptable.

Si l'écart est faible, le modèle simple et celui plus général sont voisins et par parcimonie le modèle simple est conservé. Sous l'hypothèse nulle, la statistique du test d'hypothèse (H_0 contre H_1) précédemment citée ΔD suit une loi de khi-deux de $n_2 - n_1$ degré de liberté.

Pratiquement, on rejette H_0 si $\Delta D_{obs} > \chi_{n_2 - n_1; 1 - \beta}^2$. Dans ce cas, le modèle à n_2 est le plus adéquat aux données de l'étude.

Il est à préciser qu'un mauvais ajustement d'un modèle peut s'expliquer de différentes façons :

- Des variables explicatives importantes ont été oubliées lors de l'écriture de la composante systématique du modèle ;
- La distribution théorique qui permet d'ajuster les observations est mal choisie ;
- La fonction de lien est inappropriée.

Finalement, il reste à noter que pour le choix entre différents modèles non emboîtés, nous comparons les valeurs d'un indice d'ajustement tel le critère d'information d'Akaike (AIC) ou le critère d'information bayésien (BIC). En effet, le modèle qui présente le plus faible critère est sélectionné.

CONCLUSION

Comme nous l'avons exposé, ce chapitre a introduit le cadre théorique et conceptuel qui servira à mettre en place le tarif de l'assurance maladie. En effet, ce chapitre a défini les modèles linéaires généralisés comme une approche justifiée pour élaborer le tarif de l'assurance maladie.

Le chapitre suivant sera consacré à la mise en pratique des approches théoriques adoptées. En effet, nous allons commencer par sélectionner les variables tarifaires. Ensuite, nous allons segmenter la population assurée en groupes homogènes, en nous basant sur les descripteurs de la sinistralité que nous allons déterminer par la procédure de sélection stepwise.

Finalement, nous allons modéliser la prime pure de l'assurance maladie, en nous basant sur la sinistralité individuelle des assurés. Le tarif sera ainsi en fonction du profil de risque de chaque assuré de la branche maladie de base.

TARIFICATION D'UN PRODUIT ASSURANCE SANTE

INTRODUCTION

Après que nous avons introduit les modèles linéaires généralisés et que nous avons expliqué la structure de cette approche statistique, nous allons consacrer ce chapitre à la modélisation de la prime pure de la branche assurance santé. Tout d'abord nous allons sélectionner les variables dites de tarification, puis nous allons segmenter le portefeuille étudié en se basant sur les variables tarifaires sélectionnées. Cette segmentation de la population va nous permet de proposer des tarifs différents pour les assurés de la branche santé selon la structure de la population en matière de la sinistralité. En vue de cela, nous allons modéliser séparément le coût moyen et le nombre de sinistres afin d'isoler les effets des facteurs sur la fréquence et la sévérité. Une meilleure compréhension de l'influence des facteurs sur le risque est obtenue par cette méthodologie .

Des différents facteurs impactent la fréquence de sinistres sans pour autant être significatifs quant à l'explication de la sévérité des sinistres. Cette distinction donne de plusieurs possibilités dans le processus de tarification.

4.1 Segmentation des données en assurance santé

4.1.1 Sélection de variables tarifaires

Afin de déterminer les variables potentiels d'affecter la sinistralité de l'assurance santé, nous allons appliquer le processus de sélection stepwise aux observations (durant les deux derniers exercices d'assurance 2018 – 2019).

Les variables susceptibles d'expliquer la sinistralité de l'assurance santé sont :

-Taux de remboursement ;

-Age de l'assuré ;

-Sexe de l'assuré.

Après avoir appliqué le processus de sélection stepwise par le logiciel SAS, plus précisément la requête selection=stepwise de la procédure phreg, nous avons obtenu les résultats suivants :

Summary of Stepwise Selection							
Step	Effect		DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
	Entered	Removed					
1	age		1	1	553.8275		<.0001
2	taux		1	2	387.7616		<.0001
3	sex		1	3	64.6219		<.0001

FIGURE 4.1 – Sélection stepwise basée sur le F partiel de Fisher

La figure montre que toutes les variables proposées pour expliquer la sinistralité des assurés de l'assurance santé sont significatives au seuil 5%. Et par conséquent, les variables que nous allons utiliser pour modéliser la fréquence et la sévérité de l'assurance santé sont présentées comme suit :

-Taux de remboursement ;

-Age de l'assuré ;

-Sexe de l'assuré.

4.1.2 Segmentation des variables tarifaires

Après avoir sélectionné les variables tarifaires, l'étape qui se suit est une étape primordiale qui consiste à segmenter la population des assurés en groupes homogènes. En nous basant sur les variables précédemment sélectionnées et l'étude faite dans le chapitre II, nous allons déterminer les classes d'assurés pour chaque variable tarifaire comme suit :

*Taux de remboursement

Concernant la variable "taux de remboursement", nous avons choisi 3 segments : le moins risqué, le plus risqué et l'intermédiaire de ces deux derniers segments :

- Taux de remboursement 100% ;

-Taux de remboursement 90% ;

-Taux de remboursement 80% ;

***Age de l'assuré**

Pour la variable "age de l'assuré", nous avons choisis sept groupes d'âge qui se représentent comme suit :

- Groupe d'âge [0,18] pour les assurés ayant un âge inférieure ou égale à 18ans ;
- Groupe d'âge [18,30[pour les assurés ayant un âge supérieur à 18ans et strictement inférieur à 30ans ;
- Groupe d'âge [30,40[pour les assurés ayant un âge supérieur à 30ans et strictement inférieur à 40ans ;
- Groupe d'âge [40,50[pour les assurés ayant un âge supérieur à 40ans et strictement inférieur à 50ans ;
- Groupe d'âge [50,60[pour les assurés ayant un âge supérieur à 50ans et strictement inférieur à 60ans ;
- Groupe d'âge [60,65[pour les assurés ayant un âge supérieur à 60ans et strictement inférieur à 65ans ;
- Groupe d'âge [65,[pour les assurés ayant un âge supérieur à 65ans.

***Sexe de l'assuré**

Pour la variable "sexe de l'assuré", nous avons gardé la distinction naturelle du genre qui est une distinction entre hommes et femmes.

4.2 Détermination de la prime pure**4.2.1 Modélisation de la fréquence des sinistres****4.2.1.1 Ajustement et choix du modèle**

Maintenant que nous avons sélectionné les variables explicatives de la fréquence et on a les segmenté, nous allons choisir la distribution la plus convenable pour modéliser la fréquence. La littérature sur la question suggère de tester pour les fréquences les lois théoriques suivant :

- La loi de Poisson ;
- La loi binomiale négative ;
- La loi binomiale ;

A l'aide du logiciel R, nous avons effectué un ajustement graphique de la fréquence de

consommation par les lois cités précédemment, les résultats sont présentés ci-dessous.

***Loi de Poisson**

Soit N une variable aléatoire qui suit une loi de Poisson de paramètre λ , sa fonction de probabilité s'écrit :

$$P(N = n) = \exp(-\lambda) \frac{\lambda^n}{n!} \text{ où } n \in \mathbb{N}$$

$$\text{Avec } E(N) = \lambda \text{ et } \text{Var}(N) = \lambda$$

L'ajustement par la loi de Poisson est le suivant :

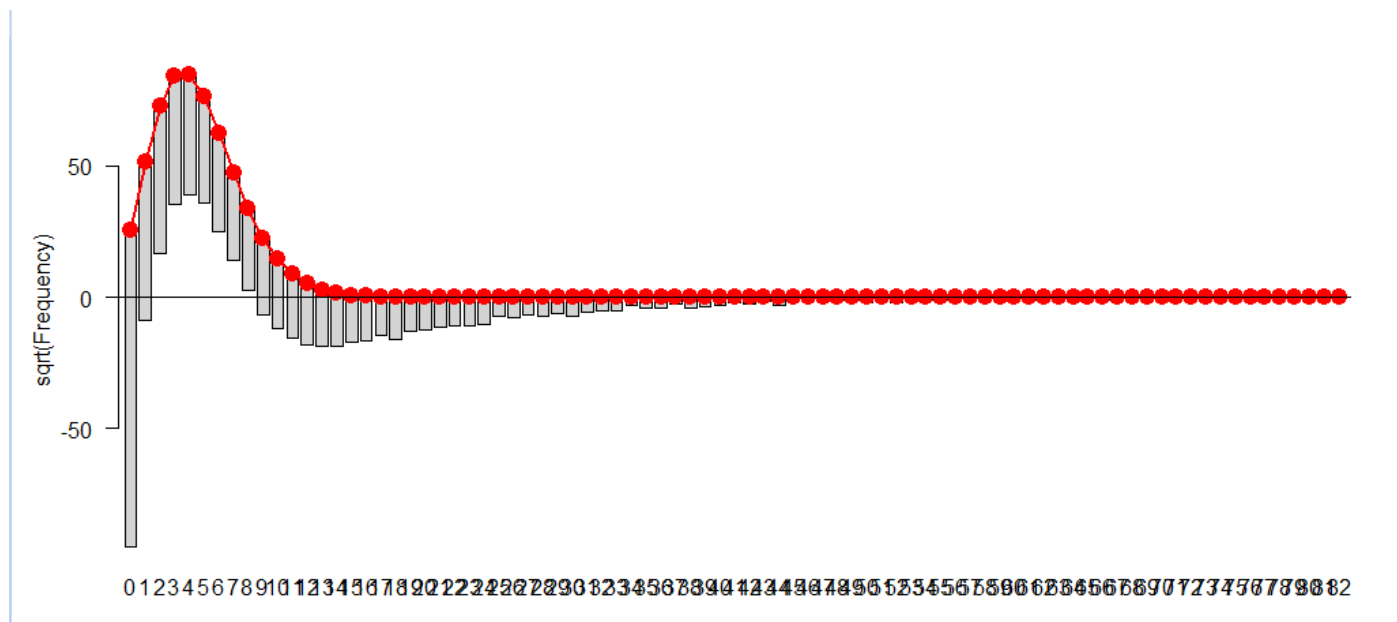


FIGURE 4.2 – Ajustement de la fréquence des sinistres par une loi de Poisson

Les points rouges représentent la loi théorique et les histogrammes représentent les fréquences observées, qui sont collés par le sommet à la loi théorique. Tout écart de la base d'un histogramme avec l'axe des abscisses indique donc un mauvais ajustement des observations par la loi théorique.

Nous remarquons que les écarts sont importantes, donc la loi de Poisson n'ajuste pas bien la fréquence des sinistres.

***Loi Binomiale Négative**

Soit N une variable aléatoire qui suit une loi binomiale négative de paramètre p et r, sa fonction de probabilité s'écrit :

$$P(N = k) = \binom{k+r-1}{k} p^r (1-p)^k \text{ où } k \in \mathbb{N}$$

$$\text{Avec } E(N) = \frac{r(1-p)}{p} \text{ et } \text{Var}(N) = \frac{r(1-p)}{p^2}$$

L'ajustement par la loi Binomiale Négative est le suivant :

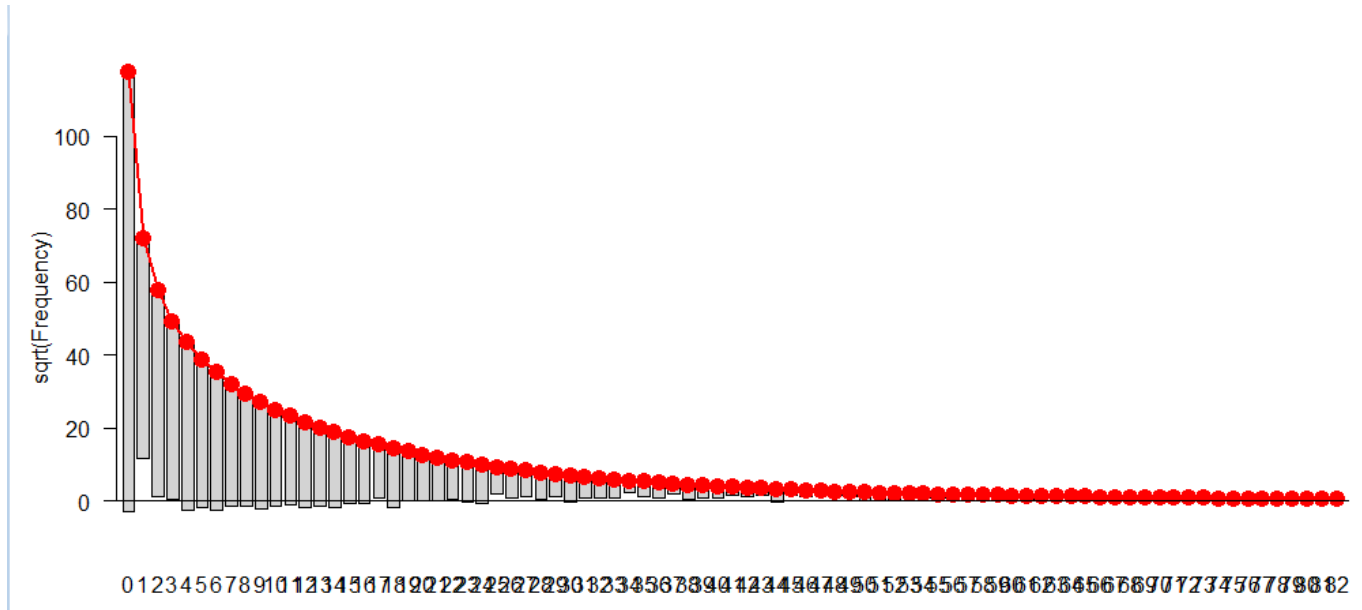


FIGURE 4.3 – Ajustement de la fréquence des sinistres par une loi Binomiale Négative

Nous voyons que les écarts sont faibles par rapport au loi de Poisson, donc la loi binomiale négative ajuste mieux la fréquence de sinistres. A partir de ces résultats, nous allons tentés de modéliser la fréquence des sinistres par une loi Binomiale-Négative.

Pour ajuster la fréquence, l'hypothèse de distribution sélectionnée est la loi Binomiale Négative.

$$\text{Or } \text{Fréquence} = \frac{\text{Nombre de sinistres}}{\text{exposition}}$$

Le Modèle Linéaire Généralisé s'écrit comme suit :

$$\ln\left(\frac{E[\text{Nombre de sinistres}]}{\text{exposition}}\right) = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p$$

ou encore :

$$\ln(E[\text{Nombre de sinistres}]) = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_p x_p + \ln(\text{exposition})$$

Avec $\ln(\text{exposition})$ en variable offset

Nombre de sinistres \sim Binomiale négative

4.2.1.2 Analyse de Type III

Afin d'étudier la significativité de l'impact d'une variable introduite dans le modèle, on utilise une analyse de type III. Elle peut être effectuée suivant le test de Wald ou du rapport de vraisemblance.

Dans notre cas, nous avons effectué cette analyse grâce au logiciel SAS avec le critère du rapport de vraisemblance, en testant la significativité des variables explicatives une à une. Ce test peut être aussi effectué par le critère de l'AIC (le meilleur modèle étant celui qui possède la valeur de l'AIC la plus faible).

***Le principe du test**

Le test de rapport de vraisemblance est un test d'hypothèse qui compare l'adéquation de l'ajustement de deux modèles afin de déterminer celui qui offre le meilleur ajustement pour les données échantillons : un modèle non contraint dont tous les paramètres sont libres et son modèle contraint correspondant avec moins de paramètres pour tester l'hypothèse nulle. Comme son nom l'indique, le test s'appuie sur le rapport de vraisemblance.

***La statistique du test**

On teste les hypothèses suivantes :

H_0 : La variable X introduite n'est pas influente dans le modèle
Vs

H_1 : La variable X introduite est influente dans le modèle

Sous H_0 $\lambda = -2\ln(R)$ suit asymptotiquement χ_n^2 , où ;

$$R = \frac{\text{Vraisemblance du modèle sans } X}{\text{Vraisemblance du modèle avec } X}$$

Où : n = nombre de paramètres du modèle complet - nombre de paramètres du modèle réduit

***Les résultats du test**

Les résultats du test donnés par le logiciel SAS se représentent comme suit :

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
GENRE	1	57.36	<.0001
TR	4	1906.59	<.0001
grp_age	6	1017.18	<.0001

FIGURE 4.4 – Analyse de type III des facteurs de sinistralité

La figure ci-dessus montre que la p-value de chaque variable tarifaire introduite dans le modèle est largement inférieure à 5% (seuil de significativité). On en déduit donc que toutes les variables choisies ont un effet significatif sur la fréquence de consommation de la branche assurance santé. On voit aussi que la variable "taux de remboursement" est la plus significative au sens de l'analyse de type III.

4.2.2 Analyse des résultats

4.2.2.1 Estimation des paramètres du modèle établi

Dans cette partie, nous allons estimer les paramètres du modèle par la méthode de maximum de vraisemblance. Il est à signaler que la qualité de l'estimation des paramètres, repose sur la convergence de l'algorithme de Newton Raphson, comme étant la méthode numérique qui permet de prédire les valeurs des paramètres introduits dans le modèle. La figure suivante, présente l'estimation de tous les paramètres du modèle adopté pour modéliser la fréquence des sinistres, donné par le logiciel SAS grâce à la procédure proc genmod :

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	0.2618	0.1116	0.0430	0.4806	5.50	0.0190
GENRE	F	1	0.1217	0.0161	0.0902	0.1532	57.36	<.0001
GENRE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
TR	100%	1	0.4848	0.0454	0.3958	0.5737	114.19	<.0001
TR	80%	1	-0.1489	0.0464	-0.2399	-0.0579	10.29	0.0013
TR	90%	0	0.0000	0.0000	0.0000	0.0000	.	.
grp_age	[0,18]	1	0.7014	0.1026	0.5002	0.9026	46.69	<.0001
grp_age]18,30	1	1.0874	0.1038	0.8839	1.2909	109.66	<.0001
grp_age]30,40	1	1.2453	0.1037	1.0420	1.4486	144.14	<.0001
grp_age]40,50	1	1.1912	0.1050	0.9854	1.3970	128.67	<.0001
grp_age]50,60	1	1.2670	0.1080	1.0554	1.4787	137.68	<.0001
grp_age]60,65	1	0.9405	0.1356	0.6747	1.2063	48.10	<.0001
grp_age]65,]	0	0.0000	0.0000	0.0000	0.0000	.	.
Dispersion		1	1.8363	0.0189	1.7997	1.8736		

FIGURE 4.5 – Estimation des paramètres de la fréquence des sinistres

La figure montre que toutes les variables retenues par le modèle pour modéliser la fréquence de la consommation des assurés sont significatives (la p-value de chaque paramètre est largement inférieur au seuil de significativité 5%).

4.2.2.2 Validation du modèle

Tout d'abord, il faut s'assurer que le modèle considéré s'ajuste bien au donné. Afin de vérifier cela, nous allons utiliser la déviance citée précédemment comme statistique du test suivant :

H_0 : Le modèle considéré à p paramètres est adéquat

Vs

H_1 : Le modèle considéré à p paramètres n'est pas adéquat

Les résultats du test sont présentés ci-dessus, données par le logiciel SAS grâce au procédure proc genmod.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	36E3	37558.8469	1.0404
Scaled Deviance	36E3	37558.8469	1.0404
Pearson Chi-Square	36E3	53731.4178	1.4884
Scaled Pearson X2	36E3	53731.4178	1.4884
Log Likelihood		143992.7023	
Full Log Likelihood		-83326.1141	
AIC (smaller is better)		166678.2282	
AICC (smaller is better)		166678.2383	
BIC (smaller is better)		166788.6552	

FIGURE 4.6 – Test de déviance du modèle

La figure montre que la valeur de la déviance est $D = 37558.8469$, or, on a 36917 observations, et 14 paramètres à estimer, d'où D est proche de $m-n$ (avec m est le nombre d'observations et n le nombre d paramètres à estimer). Alors, le modèle mis en place pour modéliser la fréquence des sinistres des assurés est adéquat.

4.2.2.3 Prédiction de la fréquence

Pour une observation n_i associée au vecteur de variables explicatives x_i , sa valeur prédite par un modèle ayant le logarithme comme fonction de lien est :

$$\hat{n}_i = \exp(\alpha_0 + \alpha_1 x_{i,1} + \dots + \alpha_n x_{i,n})$$

Cependant, les résultats obtenus précédemment nous permettent de prédire la fréquence de consommation du poste assurance santé pour chaque assuré i . Or, la fréquence obtenue change de manière multiplicative en passant d'une classe à une autre classe. Ainsi, un coefficient d'une telle classe inférieur à 0, correspond à une réduction de la fréquence de consommation. Par contre, un coefficient supérieur à 0 indique qu'il s'agit d'une majoration de la fréquence des sinistres.

Prenons par exemple le cas d'un assuré homme, qui a un taux de remboursement égale à 80% et appartient au groupe d'âge]60,65], sa fréquence de consommation est calculé à partir de la figure 4.5 comme suit :

$$\text{fréquence}_{\text{assuré } i} = \exp(0.2618 - 0.1489 + 0.9405) = 2.87$$

D'où l'assuré i a une fréquence de consommation égale à 2.87.

4.2.3 Modélisation de la charge des sinistres

4.2.3.1 Choix de la distribution

De la même façon que pour les fréquences de sinistres, nous cherchons à modéliser le coût moyen des sinistres par des lois usuelles. Classiquement, nous utilisons :

-La loi Log_Normale;

-La loi Gamma;

Le paragraphe suivant présente le critère qui va nous aider à choisir la distribution la plus convenable.

-Critère d'information d'Akaike ;

Le critère d'information d'Akaike, (en anglais Akaike information criterion ou AIC) est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973.

Le critère d'information d'Akaike s'écrit comme suit :

$$AIC = 2k - 2\ln(L)$$

où k est le nombre de paramètres à estimer du modèle et L est le maximum de la fonction de vraisemblance du modèle.

Si l'on considère un ensemble de modèles candidats, le modèle choisi est celui qui aura la plus faible valeur d'AIC. Ce critère repose donc sur un compromis entre la qualité de l'ajustement et la complexité du modèle, en pénalisant les modèles ayant un grand nombre de paramètres, ce qui limite les effets de sur-ajustement (augmenter le nombre de paramètre améliore nécessairement la qualité de l'ajustement).

Lorsque tous les modèles candidats sont des modèles linéaires généralisés (ou GLM), une approche possible pour sélectionner le modèle est de choisir celui qui maximise la vraisemblance L . Celle-ci peut être facilement estimée et le modèle ayant la plus grande vraisemblance sera celui qui a le plus petit AIC.

-Comparaison au BIC ;

Il existe de nombreux critères d'informations inspirés du critère d'Akaike. Parmi ceux-ci, le critère d'information bayésien est l'un des plus populaires. Il se définit comme suit :

$$BIC = -2\ln(L) + \ln(n)k$$

avec n le nombre d'observations dans l'échantillon étudié et k le nombre de paramètres.

L'AIC pénalise le nombre de paramètres moins fortement que le BIC. Burnham et Anderson (2002 et 2004) proposent une comparaison de l'AIC au BIC. Les auteurs montrent que l'AIC peut

être construit dans le même contexte bayésien que le BIC, mais avec des hypothèses différentes. Burnham et Anderson défendent aussi l'idée que l'AIC possède certains avantages théoriques sur le BIC : d'abord parce que l'AIC est dérivé des principes de la théorie de l'information, au contraire du BIC, ensuite parce que les hypothèses sous-tendant la dérivation du BIC sont discutables, et enfin parce que certaines de leurs simulations suggèrent que l'AIC serait en pratique plus performant que le BIC.

Le tableau suivant présente les valeurs de AIC et BIC données par le logiciel SAS de différents modèles :

Loi théorique	Log_Normale	Gamma
Critère	Value	Value
AIC (smaller is better)	614281.8199	489886.4976
BIC (smaller is better)	614385.8066	489990.4731

TABLE 4.1 – Critère AIC et BIC de différents modèles

Le tableau montre que le AIC et le BIC les plus faibles sont celles de la loi Gamma. Par conséquent, nous allons modéliser la charge des sinistres par la loi Gamma.

Pour ajuster le coût moyen, l'hypothèse de distribution sélectionnée est la loi Gamma.

$$\text{Or } \text{Coût_moyen} = \frac{\text{Montant des sinistres}}{\text{Nombre de sinistres}}$$

Le Modèle Linéaire Généralisé s'écrit alors :

$$\ln(E[\text{Coût_moyen}]) = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_q x_q$$

Avec Coût_moyen \sim Gamma

4.2.3.2 Le modèle tarifaire

Pour la modélisation de la charge des sinistres, nous avons retenu trois facteurs qui sont : l'âge de l'assuré, le taux de remboursement et le sexe de l'assuré.

Comme précédemment, nous allons effectuer une analyse de type III afin de tester la significativité des variables introduites dans le modèle.

Les résultats obtenues grâce au logiciel SAS se présentent comme suit :

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
GENRE	1	448.06	<.0001
taux	4	106.23	<.0001
grp_age	6	722.20	<.0001

FIGURE 4.7 – Analyse de type III de la régression Gamma

Nous voyons que toutes les variables introduites dans le modèle sont significatives (les p-value sont largement inférieur au seuil de significativité 5%).

A noter que la variable la plus significative est la variable "groupe d'âge".

4.2.4 Analyse des résultats

4.2.4.1 Estimation des paramètres du modèle établi

Dans cette partie, nous allons estimer les paramètres du modèle par la méthode de maximum de vraisemblance. La figure suivante présente l'estimation de tous les paramètres du modèle adopté pour modéliser le coût moyen des sinistres, donné par le logiciel SAS.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	10.2998	0.1016	10.1007	10.4989	10282.4	<.0001
GENRE	F	1	-0.2463	0.0114	-0.2686	-0.2240	468.60	<.0001
GENRE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
taux	0.8	1	-0.0687	0.0128	-0.0937	-0.0437	29.03	<.0001
taux	0.9	1	-0.0674	0.0326	-0.1312	-0.0036	4.29	0.0384
taux	1	0	0.0000	0.0000	0.0000	0.0000	.	.
grp_age	[0,18]	1	-0.0860	0.1017	-0.2853	0.1132	0.72	0.3974
grp_age]18,30	1	-0.0257	0.1023	-0.2262	0.1748	0.06	0.8014
grp_age]30,40	1	0.1384	0.1022	-0.0619	0.3388	1.83	0.1757
grp_age]40,50	1	0.2993	0.1029	0.0976	0.5010	8.46	0.0036
grp_age]50,60	1	0.3132	0.1045	0.1083	0.5180	8.98	0.0027
grp_age]60,65	1	0.5289	0.1234	0.2871	0.7707	18.37	<.0001
grp_age]65,]	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	1.4752	0.0128	1.4503	1.5006		

FIGURE 4.8 – Estimation des paramètres du coût moyen

La figure montre que la plupart des paramètres du modèle considérée sont significatives au seuil de 5% (la p-value est inférieur à 5%). À l'exception de certains paramètres de "groupe d'âge" que nous allons les traiter dans le paragraphe suivant.

-Le traitement des coefficients non significatifs

On rappelle que par cohérence et par soucis de lisibilité des tables tarifaires, les mêmes variables et les mêmes modalités ont été retenues pour la fréquence et le coût moyen. Il peut ainsi arriver que certains coefficients ne soient pas significatifs pour la fréquence des sinistres ou pour le coût moyen.

C'est l'exemple de certains paramètres du "groupe d'âge" dans la modélisation du coût moyen.

Dans ce cas, on pourra être amenés à exclure une variable de l'analyse en conservant des coefficients à 100% pour chacune des modalités de la variable, dans le but d'obtenir des bases de tarification facilement compréhensibles et homogènes.

Dans notre cas, même si certains coefficients sont non significatives, nous allons garder tous les variables.

4.2.4.2 Analyse des résidus

L'analyse des résidus permet une analyse plus poussée que les statistiques vues dans le paragraphe précédent. Elle permet en effet de comprendre d'où proviennent les éventuels écarts entre les valeurs prédites et les données en détectant les observations particulières.

En effet l'existence de certaines valeurs aux caractéristiques très atypiques peut biaiser fortement les coefficients calculés dans le modèle. Il conviendra donc éventuellement d'ôter ces valeurs atypiques et de relancer le modèle afin d'obtenir des résultats plus stables.

Deux types de résidus sont classiquement utilisés pour les modèles linéaires généralisés : les résidus de Pearson et les résidus de la déviance. C'est cette dernière approche que nous utilisons dans les calculs.

Pour que nous obtenions un modèle valide, il faut que les résidus soient assez proches de 0 et soient répartis de manière assez uniforme autour de l'axe des abscisses.

La figure suivante présente le graphique des résidus :

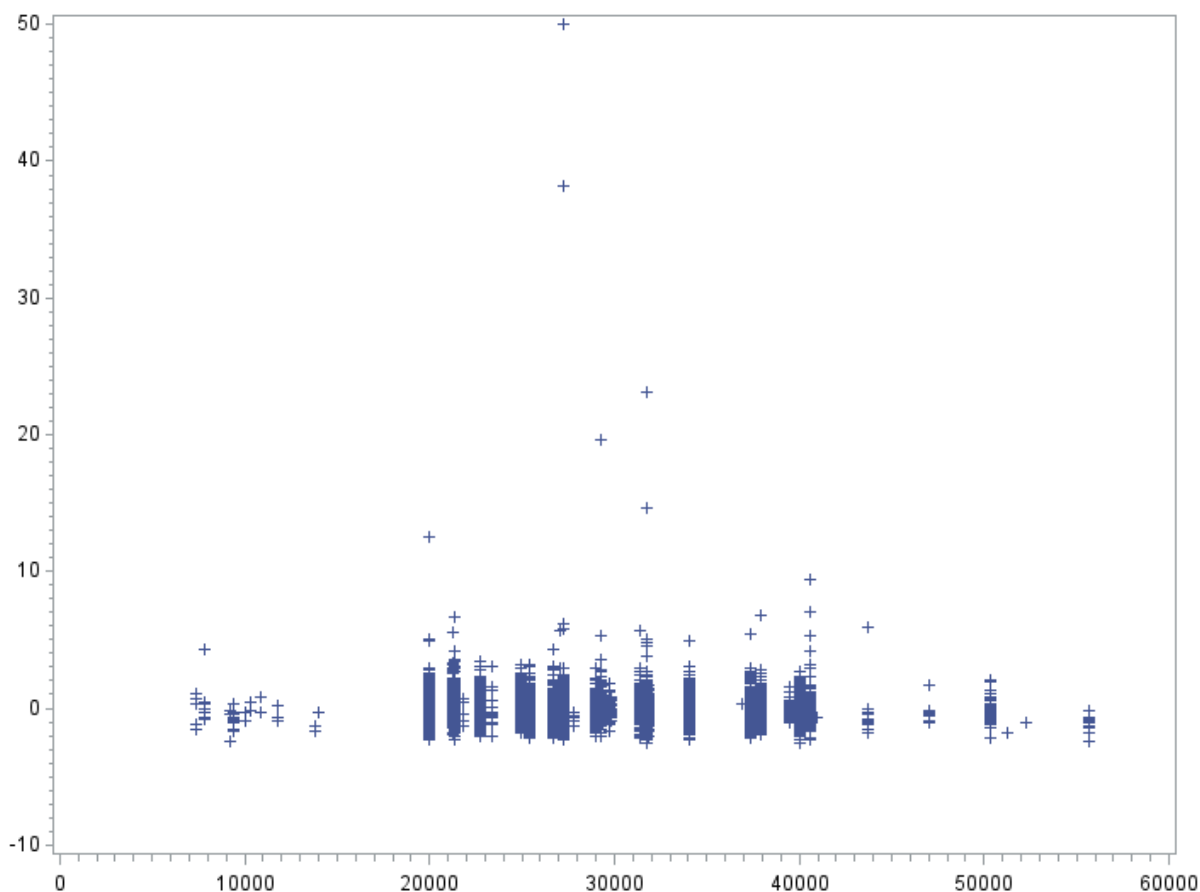


FIGURE 4.9 – Représentation des résidus de la déviance

Les résidus sont correctement répartis sur l'axe des abscisses et sont d'une manière générale relativement proches de 0. On n'observe pas de valeurs étonnamment éloignées de 0 pour les résidus de la déviance.

Aucune structure particulière ne se dégage de ce graphique, mais on constate des valeurs assez élevées pour certains résidus ne correspondant pourtant pas à des observations aberrantes, ce qui peut traduire une certaine justesse du modèle.

Nous concluons alors que le modèle adopté pour modéliser la charge des sinistres est adéquat.

4.2.4.3 Prédiction du coût moyen

Les résultats obtenus précédemment nous permettent de prédire le coût moyen de consommation du poste assurance santé pour chaque assuré j . Or, les frais obtenus changent de manière multiplicative en passant d'une classe à une autre classe. Ainsi, un coefficient d'une telle classe inférieur à 0, correspond à une réduction de frais de consommation. Par contre, un coefficient supérieur à 0 indique qu'il s'agit d'une majoration de frais de consommation.

Prenons par exemple le cas d'un assuré femme, qui a un taux de remboursement égale à 80% et appartient au groupe d'âge $[30,40]$, ses frais de consommation sont calculés à partir de la figure 4.8 comme suit :

$$\text{coût_moyen}_{\text{assuré } j} = \exp(10.2998 - 0.2463 - 1.0687 + 0.1384) = 9165.48 \text{XAF}$$

D'où l'assuré j a un coût moyen de consommation égale à 9165.48 XAF.

CONCLUSION

Dans ce chapitre, nous avons mis en place des différentes approches théoriques adoptées pour modéliser la prime de l'assurance santé. Nous avons utilisé dans cette modélisation les techniques des modèles linéaires généralisés, comme une approche justifiée pour combiner le tarif de la branche assurance santé.

Pourtant, pour réussir à avoir une segmentation assez précise de la population assurée, la sélection des variables tarifaires était une étape nécessaires . En effet, nous avons constater que la construction des classes de risque homogènes, se base sur le pouvoir explicatif des discripteurs de la sinistralité.

D'un autre côté, pour estimer la prime de l'assurance santé, l'approche fréquence - sévérité était la méthodologie adaptée. Pour cela, nous avons modéliser le coût moyen et le nombre de sinistres séparément dans le but d'isoler les effets des facteurs sur la charge et la fréquence des sinistres.

CONCLUSION

Au terme de notre étude, notre objectif principal a été la détermination de La prime qu'un individu, doit verser à la compagnie pour avoir une couverture maladie individuelle de base, ce travail a été fait pour deux bases de données de deux pays différents, ensuite la compagnie décide par rapport au marché le plus intéressant ou le plus rentable pour l'investissement.

Ainsi, l'approche adoptée tout au long de ce travail est basée sur une modélisation mathématique actuarielle. Ces méthodes se sont avérées très efficaces, dans la mesure où elles sont arrivées à se mettre au service des stratégies commerciales et ont pu aussi donner des tarifs en phase avec le marché.

Par ailleurs, il s'est avéré dans la partie préliminaire que le comportement de la consommation médicale change en fonction de la nature des bénéficiaires et le taux de couverture. Les bénéficiaires ayant un taux de remboursement complet consomment plus par rapport aux autres, nous avons remarqué aussi que les femmes ont un nombre moyen de sinistres plus que les hommes, et les enfants enregistrent la consommation la plus faible. L'estimation des coefficients des modèles a été basé sur les modèles linéaires généralisés GLM. L'application de ce dernier a nécessité le passage par les étapes suivantes :

- La détermination des facteurs de risque, autrement dit, les variables tarifaires possibles.
- La segmentation de chaque variable tarifaire.
- Le choix de la distribution de la variable cible.

Ensuite, chaque modèle retenu a fait l'objet d'une validation par des tests statistiques d'adéquation et de validation, qui ont affirmé la bonne qualité de toutes les grandeurs estimées.

Enfin, le présent projet n'est qu'un précurseur pour améliorer les pratiques de gestion de risque et les méthodes de tarification appliquées en assurance maladie.

BIBLIOGRAPHIE

- « Assurance Non Vie II : tarification et provisionnement » Professeur MARRI Fouad; INSEA.
- « Cours GLM INSEA ».
- « LANGAGE SAS » Axelle Chauvet-Peyrard; ENSAE.
- « Assurance non-vie : modélisation, simulation. (Assurance Audit Actuariat) » PARTRAT, C., BESSON, JL. Paris : Economica.
- « An introduction to Generalized Linear Models. Third edition. » DOBSON, A. J., BARNETT, A. G.
- « Choix du modèle. Université de Strasbourg » Frédéric Bertrand.

Webographie

- « Site officiel de MCI CARE : » <https://www.mccicare.com/>.
- « <https://support.sas.com/en/knowledge-base.html> »
- « <https://www.actuarialab.net/> »
- « <http://www.ressources-actuarielles.net/> »
- « https://www.cleiss.fr/docs/regimes/regime_mali.html ».

Annexes

Annexe 1

```
libname ma 'C:\Users\ASUS\Desktop\pfe sas';  
  
proc import datafile='C:\Users\ASUS\Desktop\pfe sas\DETAILS DES BENEFICIAIRES 2018.xls'  
out=ma.ben18  
replace;  
getnames=yes;  
run;  
  
proc import datafile='C:\Users\ASUS\Desktop\pfe sas\DETAILS DES BENEFICIAIRES 2018.xls'  
out=ma.benr18  
replace;  
getnames=yes;  
sheet=Retirés;  
run;  
  
proc import datafile='C:\Users\ASUS\Desktop\pfe sas\DETAILS DES BENEFICIAIRES 2019.xls'  
out=ma.ben19  
replace;  
getnames=yes;  
run;  
  
proc import datafile='C:\Users\ASUS\Desktop\pfe sas\DETAILS DES BENEFICIAIRES 2019.xls'  
out=ma.benr19  
replace;  
getnames=yes;  
sheet=Retirés;  
run;
```

FIGURE 4.10 – Code d'importation de base de donnée

Annexe 2

```
proc freq data=m.benr19; tables MATBENEF ;run;
proc freq data=m.benr19; tables MATASSURE ;run;

proc freq data=m.bena19; tables COLLEGE ;run;
proc freq data=m.benr19; tables CODCOL ;run;
proc freq data=m.benr19; tables STATUT ;run;
proc freq data=m.benr19; tables datefeet ;run;
proc freq data=m.benr19; tables datenais ;run;

proc freq data=m.benr19; tables DATESORTIE ;run;
proc freq data=m.benr19; tables genre ;run;

proc freq data=m.bena19; tables MATBENEF ;run;
proc freq data=m.bena19; tables MATASSURE ;run;

proc freq data=m.bena19; tables COLLEGE ;run;
proc freq data=m.bena19; tables CODCOL ;run;
proc freq data=m.bena19; tables STATUT ;run;
proc freq data=m.bena19; tables datefeet ;run;
proc freq data=m.bena19; tables datenais ;run;

proc freq data=m.bena19; tables DATESORTIE ;run;
proc freq data=m.bena19; tables genre ;run;

proc sql;
create table m.benr18w as
select matbenef,datefeet,datesortie from m.benr18
where datefeet > datesortie;
quit;

proc print data=m.benr18w;run;
```

FIGURE 4.11 – Code de détection des valeurs aberrantes

Annexe 3

```

        DATA M.sinANNES0IN;
    SET M.SIN;
    IF DATEDES0IN >"31dec2018"d and datedesoIn < "1jan2020"d then datedesoIn="1jan20
    IF DATEDES0IN >"31dec2017"d and datedesoIn < "1jan2019"d then datedesoIn="1jan20
    IF DATEDES0IN >"31dec2016"d and datedesoIn < "1jan2018"d then datedesoIn="1jan
    IF DATEDES0IN >"31dec2015"d and datedesoIn < "1jan2017"d then datedesoIn="
run;

proc sql;
create table m.sinREPDATES0IN as
select datedesoIn, SUM(rembours_) from M.sinANNES0IN
where annes=2018
group by datedesoIn;
quit;

proc print data= m.sinREPDATES0IN ;run;

```

FIGURE 4.12 – Code de Répartition de la base de sinistre 2019 selon l'année de soin

Annexe 4

```

proc sql;
create table m.sinPP19 as
select distinct(MATRICULEP) from m.SIN
where annes=2019;
quit;

proc sql;
create table m.sinPP18 as
select distinct(MATRICULEP) from m.SIN
where annes=2018;
quit;

```

FIGURE 4.13 – Code de calcul du nombre des personnes sinistrées

Annexe 5

```

proc sql;
create table m.sinconsPrest18 as
select prestations, sum(rembours_) from m.sin
where datedesoin >"31dec2017"d and datedesoin <"1jan2019"d
group by prestations;
quit;

proc print data= m.sinconsPrest18;run;

proc sql;
create table m.sinconsPrest19 as
select prestations, sum(rembours_) from m.sin
where datedesoin >"31dec2018"d and datedesoin <"1jan2020"d
group by prestations;
quit;

proc print data= m.sinconsPrest19;run;

```

FIGURE 4.14 – Code de répartition des sinistres selon les prestations

Annexe 6

```

DATA M.SIN18mois;
SET M.SIN;
IF DateDeSoin < "1jan2018"d then delete;
else if DateDeSoin < "1feb2018"d then datedesoin="1jan2018"d;
else if DateDeSoin < "1mar2018"d then datedesoin="1feb2018"d;
else if DateDeSoin < "1apr2018"d then datedesoin="1mar2018"d;
else if DateDeSoin < "1may2018"d then datedesoin="1apr2018"d;
else if DateDeSoin < "1jun2018"d then datedesoin="1may2018"d;
else if DateDeSoin < "1jul2018"d then datedesoin="1jun2018"d;
else if DateDeSoin < "1aug2018"d then datedesoin="1jul2018"d;
else if DateDeSoin < "1sep2018"d then datedesoin="1aug2018"d;
else if DateDeSoin < "1oct2018"d then datedesoin="1sep2018"d;
else if DateDeSoin < "1nov2018"d then datedesoin="1oct2018"d;

else if DateDeSoin < "1dec2018"d then datedesoin="1nov2018"d;
else if DateDeSoin < "1jan2019"d then datedesoin="1dec2018"d;
else if datedesoin > "31dec2018"d then delete; run;

proc sql;
create table m.sin18moiscon as
select datedesoin, sum(rembours_) from M.SIN18mois
group by datedesoin;
quit;

```

FIGURE 4.15 – Code de la consommation mensuelle des sinistres

Annexe 7

```

DATA M.ben18EFFECTact;
SET M.bena18;
if      Datefeet < "1feb2018"d then Datefeet="1jan2018"d;
else if Datefeet < "1mar2018"d then Datefeet="1feb2018"d;
else if Datefeet < "1apr2018"d then Datefeet="1mar2018"d;
else if Datefeet < "1may2018"d then Datefeet="1apr2018"d;
else if Datefeet < "1jun2018"d then Datefeet="1may2018"d;
else if Datefeet < "1jul2018"d then Datefeet="1jun2018"d;
else if Datefeet < "1aug2018"d then Datefeet="1jul2018"d;
else if Datefeet < "1sep2018"d then Datefeet="1aug2018"d;
else if Datefeet < "1oct2018"d then Datefeet="1sep2018"d;
else if Datefeet < "1nov2018"d then Datefeet="1oct2018"d;

else if Datefeet < "1dec2018"d then Datefeet="1nov2018"d;
else if Datefeet < "1jan2019"d then Datefeet="1dec2018"d;

run;

proc freq data= M.ben18EFFECTact ; tables datefeet;run;

```

FIGURE 4.16 – Code du calcul des nombres des personnes protégées mensuellement pour la base des actifs

Annexe 8

```

DATA M.ben18EFFECTret2;
SET M.benr18;

WHERE Datefeet < "1mar2018"d and datesortie >"31jan2018"d;

RUN;

DATA M.ben18EFFECTret3;
SET M.benr18;

WHERE Datefeet < "1APR2018"d and datesortie >"28FEB2018"d;

RUN;

```

FIGURE 4.17 – Code du calcul des nombres des personnes protégées en janvier et février 2018 pour la base des retirés

Annexe 9

```

proc sql;
create table m.sin19moiscon as
select datedesoin, sum(rembours_) from m.SINMOIS19
WHERE PRESTATIONS="PHARMACIE"
group by datedesoin;
quit;

proc sql;
create table m.sin18moisconPHAR as
select datedesoin, sum(rembours_) from M.SIN18mois
WHERE PRESTATIONS="CONSULTATION"
group by datedesoin;
quit;

```

FIGURE 4.18 – Code de la répartition mensuelle des prestation consultation et pharmacie

Annexe 10

```

data m.sintrancheage19;
set m.sin;
IF Datanais < "1jan1905"d or Datanais > "31dec 2019"d then datenais="1jan1905";
else if Datanais > "31dec2001"d then Datanais="18dec2018"d;
else if Datanais > "31dec1989"d then Datanais="30dec2030"d;
else if Datanais > "31dec1979"d then Datanais="4dec2040"d;
else if Datanais > "31dec1969"d then Datanais="5dec2050"d;
else if Datanais > "31dec1959"d then Datanais="2dec2060"d;
else if Datanais > "31dec1954"d then datenais="30dec2065"d;
else Datanais="31dec2065"d ; run ;

proc sql ;
create table m.Sintrancheage18M as
select datenais, sum(rembours_) from m.sintrancheage18
where datedesoin >"31dec2017"d and datedesoin <"1jan2019"d
group by datenais;
quit;

proc print data= m.sintestdatnai18 ; run;

```

FIGURE 4.19 – Code de la répartition de la consommation selon les tranche d'âge

Annexe 11

```

                                DATA M.SINTR;
SET M.SIN;
IF lbc_college= "100% AF" THEN lbc_college ="100%";
if lbc_college="100% AF/EU" then lbc_college="100%";
IF lbc_college= "100% ME" THEN lbc_college ="100%";
if lbc_college="100% SNG" then lbc_college="100%";
IF lbc_college= "80% AF" THEN lbc_college ="80%";
if lbc_college="80% AF/EU" then lbc_college="80%";
IF lbc_college= "80% AFO" THEN lbc_college ="80%";
if lbc_college="80% ME" then lbc_college="80%";
IF lbc_college= "80% SNG" THEN lbc_college ="80%";
if lbc_college="80%ME" then lbc_college="80%";
IF lbc_college= "90 % ME" THEN lbc_college ="90%";
if lbc_college="90% AF" then lbc_college="90%";
IF lbc_college= "90% AF-EU" THEN lbc_college ="90%";
if lbc_college="90% AFMO" then lbc_college="90%";
IF lbc_college= "90% M E" THEN lbc_college ="90%";
if lbc_college="90% M.E" then lbc_college="90%";

if lbc_college="90% M.E." then lbc_college="90%";
IF lbc_college= "90% ME" THEN lbc_college ="90%";
if lbc_college="90% SNG" then lbc_college="90%";
IF lbc_college= "90%ME" THEN lbc_college ="90%";
run;

proc sql ;
create table m.sintrcon as select
lbc_college,sum(rembours_) from m.sintr
where datedesoin >"31dec2017"d and datedesoin <"1jan2019"d
group by lbc_college;
quit;

```

FIGURE 4.20 – Code de la répartition de la consommation selon le taux de remboursement

Annexe 12

```

proc genmod data=bib.genmod1;
class GENRE taux grp_age ;
model CM=GENRE taux grp_age / dist=gamma
link=log
type3;

run;

```

FIGURE 4.21 – Modélisation de la fréquence des sinistres : Modèle Binomiale négative

Annexe 13

```
proc genmod data=bib.genmod1;  
class GENRE TR grp_age;  
model nmb_sin=GENRE TR grp_age  
/ dist=NB link=log offset=log_exp type3;  
make 'modelfit' out=negbinmodel;  
run;
```

FIGURE 4.22 – Modélisation du coût moyen des sinistres : Modèle Gamma

