INSTITUT DE FINANCEMENT DU DÉVELOPPEMENT DU MAGHREB ARABE

معهد تمويل التنمية للمغرب العربي

Mémoire de fin d'Etudes

Thème :

Tarification en Assurance Maladie Collective

Synthèse théorique avec application sur des données du GAT

Présenté et soutenu par :

Intibeh Bel Ibria

Encadré par :

Farouk Kriaa

Etudiant(e) parrainé(e) par :

GAT ASSURANCES

Table des matières

Liste des abréviations

- **APCI** : Affectation , prise en charge intégralement
- **AUTR** : Autre **(**Ascendant)
- **CNAM** : Caisse Nationale de l'assurance Maladie
- **CNRPS** : Caisse Nationale de Retraite et de Prévoyance Sociale
- **CNSS** : Caisse Nationale de Sécurité Sociale
- **CNJT** : Conjoint
- **CSP** : Catégorie socioprofessionnelle
- **CGA** :Comité Général des Assurances
- **ENFT** : Enfant
- **FTUSA** : Fédération Tunisienne des Sociétés des assurances
- **GLM** : Modèle Linéaire Généralisées
- **OMS** :Organisation Mondial de la santé
- **ODD** : Objectifs de développement durable
- **OCDE** : Organisation de coopération et de développement économique
- **PC** : Prime Commerciale
- **PIB** : Produit Intérieur Brut
- **RESP** : Responsable (assuré principale)

Liste des tableaux

Liste des figures

INTRODUCTION GENERALE

Dans la vie quotidienne, les individus sont exposés à des différents risques tels que la maladie, les accidents, le chômage, le décès. Ces risques impactent leur patrimoine. Ainsi, la maladie qui est un risque imprévisible, lorsqu'elle survient, elle a des résultats néfastes sur les patients et sur leur ménage, ce qui rend les familles plus vulnérables à la pauvreté. En fait, le mauvais état de santé affecte le bien-être des familles, et implique une diminution de leur capacité productive, et cela peut causer la perte de revenu et par la suite une forte probabilité d'exposition à la pauvreté et à la vulnérabilité surtout pour les indigents.

Au niveau microéconomique, un meilleur état de santé peut aider les individus et les ménages à sortir de la pauvreté, et accroitre leurs capacités productives. L'accès aux soins de santé est un facteur crucial et un résultat du développement humain.

Un proverbe français dit « **Toute personne qui est en bonne santé, est riche sans le savoir** ». De ce fait la consommation de soins médicaux ne cesse d'augmenter et les frais de santé occupent de nos jours une place prédominante dans le budget des ménages, du fait de l'amélioration des techniques médicales et le désengagement progressif de la Sécurité Sociale dans la part du remboursement de ces frais, et les organismes de complémentaire santé ont vocation à jouer un rôle prépondérant dans les remboursements des soins médicaux et plus généralement dans le fonctionnement du système de soins.

L'assurance est un élément indispensable lorsqu'on vit en société. L'être humain fait face à plusieurs situations défavorables qui peuvent arriver de manière imprévue et aléatoire. Les compagnies d'assurances sont parmi les meilleures alternatives pour se prémunir et gérer les multiples imprévus.

Souscrire à une assurance est plus qu'une nécessité au regard des avantages qu'elle présente. C'est une pratique de prévention qui permet à l'assuré d'anticiper sur les éventuelles situations qui pourraient subvenir. Il s'agit d'une prestation qui est offerte lorsqu'un événement inopiné ou inattendu survient dans la vie de l'assuré. Ce service d'ordre financier pour la plupart du temps peut

être octroyé à un individu, une association ou à une entreprise, en fonction du type d'activité exercée ou selon les besoins, on peut souscrire à tel ou tel type d'assurance.

Knight (1921) définie l'assurance maladie comme un mécanisme par lequel une personne se protège contre la perte financière causée par une maladie, un accident ou encore une invalidité. L'assurance maladie appartient, par définition, à la catégorie des assurances de personnes. Les assurances de personnes sont les assurances qui protègent les assurés contre les risques pouvant affecter directement l'intégrité de leur personne. Les principaux risques de la vie considérés comme portant directement atteinte à la personne sont : l'accident, la maladie, la vieillesse et la mort.

Elle fait partie aussi des assurances dites IARD (Incendie Accident et Risques Divers), également appelées « Assurances non-vie », car elles regroupent toutes les assurances qui n'appartiennent pas à la branche « Vie ». Ces assurances se distinguent par le mode de gestion des primes : en effet, pour les assurances « Non-Vie », les primes se gèrent par répartition, pour les assurances « Vie », les primes se gèrent par capitalisation.

Pourezza J (2007) a montré que l'assurance maladie influe positivement l'utilisation des services de santé. En effet, la possession d'une assurance maladie entraîne plus d'options de financement, ce qui a un impact sur le choix du prestataire de soins de santé effectué par les assurés.

Toutefois l'assurance maladie, comme les autres formes d'assurance, n'est pas à l'abri de problèmes. Ces principaux problèmes sont la sélection adverse, l'aléa moral et l'abus de l'assurance maladie.

Comme conçu, le marché de l'assurance santé est un marché très concurrentiel et certains assureurs ne souhaitent pas différencier leurs primes d'assurance santé et veuillent que tous les assurés paient la même prime. D'autre part, ils se doivent de proposer des tarifs compétitifs tout en couvrant leurs engagements dans le remboursement des frais de santé des assurés et leurs frais de fonctionnement.

Or au sein d'un portefeuille d'assurance hétérogène, les assurés ne sont pas tous égaux face aux risques, certains présentent un profil plus dangereux que d'autres. L'essence de la crédibilité consiste à calculer la pondération qui varie selon le niveau de risque, les caractéristiques des assurés et l'hétérogénéité des groupes. Pour cette raison, les entreprises d'assurances ont instauré une cellule d'actuariat qui permet de bien tarifier les primes d'assurances et de tenir compte de cette hétérogénéité au sein du même groupe.

Le risque que nous allons étudier dans ce mémoire est le risque santé. Plus précisément, les frais de santé dont les engagements sont à court terme. Dans ce type d'engagement, le principal risque auquel doit faire face l'assureur est celui de souscription qui correspond à la sous tarification du produit.

De ce fait, notre organisme de parrainage GAT ASSURANCES, nous a confié comme projet "la tarification de l'assurance maladie collective". En partant de cet objectif, la problématique de notre projet consiste à répondre aux questions suivantes :

- **Comment estimer la prime pure d'un contrat santé collectif ?**
- **Comment les caractéristiques des assurés affectent-elles le niveau de tarification ?**
- **Quelles sont les variables pertinentes pour une modélisation fiable du coût ?**

À partir de données relatives à un contrat d'assurance santé collectif commercialisé en Tunisie, nous proposons de mettre en œuvre la méthode des modèles linéaires généralisés. C'est une méthode stochastique qui permet, contrairement aux méthodes déterministes classiques, de déterminer un tarif adapté à chaque assurée.

Ce mémoire a pour but d'améliorer le processus de tarification de frais de santé collectifs à partir de la consommation des bénéficiaires du portefeuille. Plus précisément, le présent mémoire vise les objectifs suivants :

- Présenter une synthèse des connaissances théoriques relatives aux règles de fonctionnement de l'assurance maladie,
- Mener une analyse empirique permettant une éventuelle tarification prenant en compte les caractéristiques individuelles en matière de risque relative à un ensemble d'assurés d'un groupe à travers une analyse de cout de santé sur des données du GAT relativement à des années assez récentes.

Pour atteindre ces objectifs nous avons structuré notre travail en trois chapitres distincts. Le premier chapitre traite de la notion d'assurance maladie dans sa globalité. Le deuxième chapitre est consacré à présenter l'assurance santé privée et ses différentes fonctions ainsi que le système tunisien d'assurance santé et à analyser et traiter les données mises à notre disposition en faisant une étude descriptive globale et par critères tarifaires du portefeuille GAT ASSURANCES. Enfin, Le troisième chapitre est consacré à l'application empirique sur des données réelles relatives à 12 contrats collectifs contenant 15 433 assurés de la compagnie durant l'année 2018.Cette application

adopte la méthodologie modèles linéaires généralisés. Elle a pour objectif d'identifier les déterminants du coût des services et des actes de santé.

CHAPITRE PREMIER :

ASSURANCE MALADIE : LES CONCEPTS DE BASES

CHAPITRE PREMIER : ASSURANCE MALADIE : LES CONCEPTS DE BASE

Introduction

L'article 22 de la déclaration universelle des droits de l'Homme stipule que : « Toute personne, en tant que membre de la société a droit à la sécurité sociale, elle est fondée à obtenir la satisfaction des droits économique, sociaux et culturels indispensables à sa dignité et au libre développement de sa personnalité, grâce à l'effort national et à la coopération internationale, compte tenu de l'organisation et des ressources de chaque pays ».

La santé, comme composante du capital humain, est l'un des secteurs fondamentaux de développement et de croissance de toute économie et reste au centre de l'attention du public.

Ainsi, pour garantir à toutes et à tous l'égalité d'accès aux soins, les pouvoirs publics instaurent un mécanisme dit « couverture médicale ». Il a pour objectif de couvrir les risques et les frais de soins de santé inhérents à la maladie ou à l'accident. Toutefois, on constate une forte inégalité quant à l'accès aux soins de santé dans la plupart des pays surtout dans les pays en voie de développement. Dans ce chapitre, on va proposer une synthèse de la littérature relative à l'assurance maladie suivie d'une présentation synthétique des travaux empiriques réalisées sur ce thème en particulier on va étudie l'effet de l'assurance maladie sur la consommation de soins de santé.

Plus précisément, ce chapitre a pour objectifs, entre autres, de répondre à ces questions.

-Comment l'assurance maladie améliore-t-elle le niveau de dépense de soins de santé ?

-Quels sont les problèmes liés à la mise en place d'un système de sécurité équitable ?

À cet effet, ce chapitre est subdivisé en quatre sections. La première section illustre une analyse économique du marché de santé. La deuxième section présente l'effet de l'assurance maladie sur la consommation de soins de santé et on montre des exemples des études théoriques et empiriques antérieures qui ont traité l'effet de la couverture par un régime d'assurance maladie sur les dépenses en soins. La troisième section traite la tarification en assurance maladie alors que la dernière section expose des exemples des études empiriques antérieures effectuées sur la tarification en assurance ainsi une présentation des modèles adopté dans les études empiriques (GLM).

Section 1 : Analyse économique de la sante

1.1 Couverture sanitaire universelle

La couverture sanitaire universelle consiste à veiller à ce que l'ensemble de la population ait accès aux services préventifs, curatifs, palliatifs, de réadaptation et de promotion de la santé dont elle a besoin et à

ce que ces services soient de qualité suffisante pour être efficaces, sans que leur coût n'entraîne des difficultés financières pour les usagers.

Cette définition contient trois objectifs qui sont liés entre eux :

- L'accès équitable aux services de santé : tous ceux qui ont besoin des services de santé, quels que soient leurs moyens financiers, doivent pouvoir y accéder ;
- La qualité : les services de santé doivent être d'une qualité suffisante pour améliorer la santé de ceux qui en bénéficient ;
- La protection financière : le coût des soins ne doit pas exposer les usagers à des difficultés financières.

Pour des centaines de millions de gens, en particulier pour les plus vulnérables, la couverture universelle c'est l'espoir d'être en meilleure santé sans s'appauvrir.

La couverture universelle prend ses racines dans la constitution de l'**Organisation Mondiale de la Santé (OMS),** adoptée en 1948, qui fait de la santé l'un des droits fondamentaux de tout être humain, et dans la stratégie mondiale de la santé pour tous, lancée en 1979.

D'après Dr Tedros , Directeur général de l'OMS **:**

« Parvenir à une couverture sanitaire universelle en assurant la sécurité financière des patients est fondamental pour atteindre les objectifs sanitaires des objectifs de développement durable (ODD). La « santé pour tous » doit être le centre de gravité des efforts menés pour atteindre l'ensemble de ODD, car la bonne santé des individus est profitable à leur famille, leur communauté, et leur pays. Or, il reste encore beaucoup à faire. Environ 400 millions d'individus dans le monde, soit une personne sur 17, n'ont toujours pas accès aux services de santé ».

Le but de la couverture universelle en matière de santé est de faire en sorte que tous les individus aient accès aux services de santé sans encourir de difficultés financières. Pour cela, il faut :

- Un système de santé solide, efficace et bien géré ;
- Des soins à un coût abordable ;
- L'accès aux médicaments et technologies médicales ;
- Des personnels de santé en nombre suffisant, bien formés et motivés.

1.2 L'économie de la santé

L'économie de la santé est une discipline récente marquée par l'idée de maîtriser des dépenses en constante augmentation. Classiquement, l'approche économique comporte deux volets. Un premier volet macro-économique permet d'appréhender les grands axes d'équilibres financiers et le poids des dépenses de médicaments dans le PIB qui est en constante augmentation. Et l'approche micro-économique qui s'intéresse à la structure de la consommation médicale au sein des ménages.

On distingue deux types de facteurs (individuels et collectifs) qui expliquent la variabilité individuelle et la constante augmentation.

- Les facteurs individuels sont entre autres l'âge, le genre, le collège socioprofessionnel, le niveau de protection sociale etc.….
- Les facteurs collectifs peuvent être liés à la demande, le vieillissement démographique, la généralisation de la protection sociale, etc.…. Ils peuvent être également liés à l'offre : pression des industries de biotechnologie, modes de rétribution des professionnels.

1.2.1 Approche macro-économique

a) Dépenses de santé en proportion du PIB

La proportion de dépenses d'un pays en biens et services de santé par rapport aux dépenses totales dans l'économie peut varier dans le temps, en fonction des différences de croissance des dépenses de santé par rapport à la croissance économique globale.

Tout au long des années 90 et au début des années 2000, les dépenses de santé dans les pays de l'OCDE (Organisation de coopération et de développement économique) ont généralement augmenté plus rapidement que le reste de l'économie, ce qui a abouti à une hausse presque continue des dépenses de santé rapportées au PIB. Après une période d'instabilité pendant la crise économique, la proportion moyenne est demeurée relativement stable ces quelques dernières années, avec l'alignement de la croissance de ces dépenses sur celle de l'économie dans tous les pays de l'OCDE.

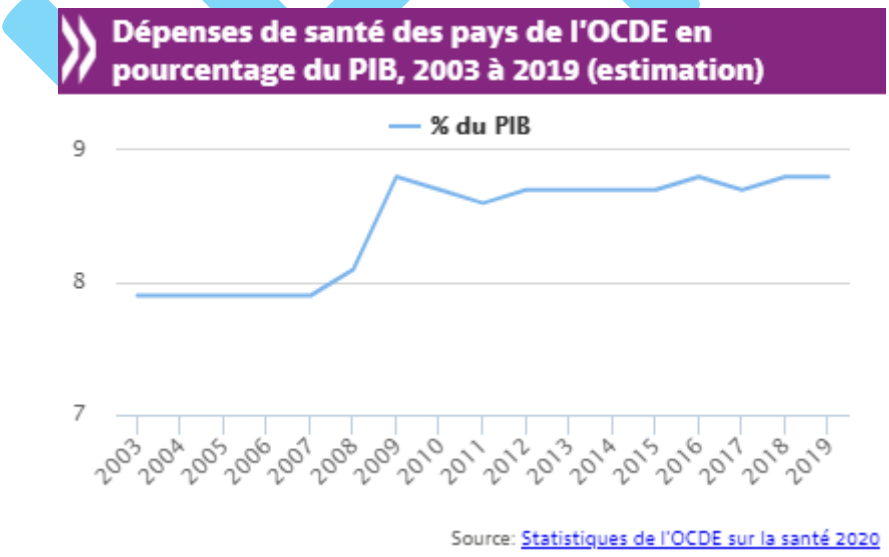


Figure 1:Croissance des dépenses de santé des pays de l'OCDE en pourcentage du PIB

En 2019, avant le début de la pandémie de coronavirus, on estime que les pays de l'OCDE (Organisation de coopération et de développement économique) ont dépensé, en moyenne, 8.8 % du PIB en soins de santé en 2019, chiffre plus ou moins stable depuis 2009. La croissance des dépenses de santé étant restée alignée avec la croissance économique globale depuis la dernière crise économique.

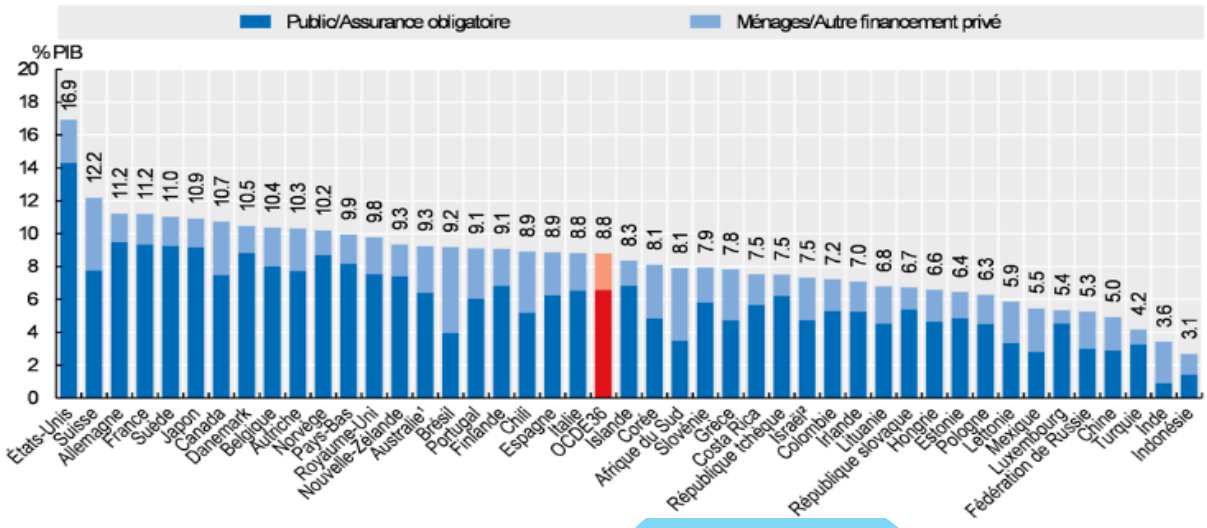


Figure 2 :Dépense de santé en pourcentage de PIB en 2018

Source : Statistiques de l'OCDE sur la santé 2019 ; Base de données de l'OMS sur les dépenses de santé mondiales

Ce graphique montre que les États-Unis enregistrent les dépenses les plus élevées en soins de santé, qui correspondent à 16.9 % du PIB, et devancent largement la Suisse, qui occupe la deuxième place avec 12.2 % du PIB. Vient ensuite un groupe de pays à revenu élevé, comprenant l'Allemagne, la France, le Japon et la Suède, qui ont consacré environ 11 % de leur PIB aux soins de santé. Un autre grand groupe de pays de l'OCDE, composé de nations européennes, ainsi que de l'Australie, de la Nouvelle-Zélande, du Chili et de la Corée, s'inscrit dans une fourchette de dépenses de santé comprise entre 8 et 10 % du PIB.

Bon nombre de pays d'Europe membres de l'OCDE, comme la Lituanie et la Pologne, ainsi que d'importants pays partenaires, ont consacré entre 6 et 8 % de leur PIB aux soins de santé. Enfin, quelques pays de l'OCDE dépensent moins de 6 % de leur PIB en soins de santé, dont le Mexique, la Lettonie, le Luxembourg et notamment la Turquie, dont les dépenses de santé ne dépassent pas 4.2 %. Les dépenses de santé de la Turquie en proportion du PIB sont comprises entre celles de la Chine et celles de l'Inde.

b) Le coût des systèmes de santé à travers le monde

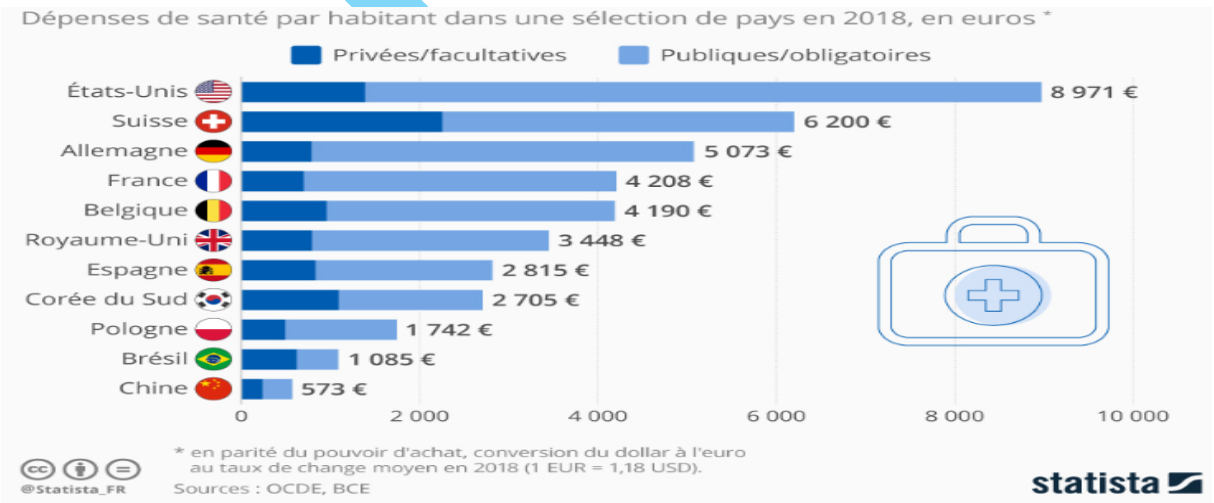


Figure 3 : dépenses de santé par personne dans une sélection de pays en 2018 (en euros).

L'organisation de coopération et de développement économique (OCDE) a publié dans son rapport annuel dressant le panorama de la santé à travers le monde. Un chapitre sur les dépenses de santé (publiques et privées) et révèlent les fortes disparités qui existent entre les pays.

Comme le montre le graphique, avec des dépenses par habitants qui atteignaient près de 9 000 euros en 2018, le système de santé américain est de loin le plus coûteux de la planète. Dans le détail, les dépenses privées s'élevaient à près de 1 400 euros et les dépenses publiques à environ 7 600 euros. La principale raison expliquant ces coûts exorbitants est le **prix des médicaments et des soins** .

En comparaison, les dépenses de santé sont plus de deux fois moins importantes en France où elles s'élevaient à environ 4 200 euros par habitant en 2018. La moyenne est similaire en Belgique même si la part des dépenses privées y est plus élevée qu'en France (27 % contre 20 %). Enfin, le Brésil fait partie des rares pays couverts par l'étude où les dépenses privées dépassaient les dépenses publiques (57 % du total).

1.2.2 Approche micro-économique : la consommation médicale

Un certain nombre de facteurs influent sur la consommation médicale. Il s'agit tout d'abord du nombre des malades dans un groupe donné et pendant un temps déterminé (la morbidité) qui explique la concentration logique des dépenses médicales sur une fraction réduite de la population.

Les travaux de recherche ont mis l'accent sur l'importance des caractéristiques individuelles et comportementales dans la détermination de l'état de santé de l'individu. La littérature économique classe les déterminants les plus souvent utilisés en quatre catégories : déterminants économiques, déterminants liés à la santé, déterminants liés à l'accès (l'offre), et déterminants démographiques.

Les variables liées à la démographie en particulier l'âge de l'individu, son sexe, son milieu de résidence, son statut sont systématiquement intégrés dans les études empiriques de l'offre de soins et même parfois avec des variables d'interaction.

a) L'âge

L'âge est parmi les variables démographiques les plus susceptibles d'influencer la décision d'un individu d'utiliser les services de soins de santé. En fait, plus la personne avance dans l'âge, plus ce que son état de santé se détériore. Ce qui l'incite à consommer plus de soins à l'âge avancée.

b) Le genre

Parmi les variables démographiques qui influencent la consommation de soins, on y trouve le genre. Les recherches empiriques effectuées sur les pays développés ont montré que les femmes sont plus sensibles à leurs besoins en matière de santé, car elles accordent plus d'attention à leur état de santé en particulier à leurs santés physiques. Grossman M. (1972) a montré que les femmes consomment plus de soins que les hommes.

c) La catégorie socioprofessionnelle (CSP)

La CSP joue également un rôle important sur la consommation médicale. De nombreuses études empiriques suggèrent que lorsque les personnes deviennent plus riches, elles exigeront des soins plus nombreux et de meilleure qualité et des soins plus coûteux.

d) Milieu de résidence

Le milieu de résidence est l'une des variables explicatives du recours aux fournisseurs de soins de santé. Les études réalisées par les chercheurs tels que Anderson, G, et Peter H.(2000) ont soulevé l'importance de l'environnement résidentiel dans le contexte d'utilisation des services de santé. Ils ont abouti à ce que les personnes appartenant à des zones urbaines et rurales présentent des perceptions différentes de l'utilisation des services de santé.

Section 2 -Assurance maladie : Concepts et objectifs

Lorsqu'une personne vit seul ou dans des groupes familiaux primitifs, chaque famille ou groupe de familles s'occupent autant que possible de ces proches. Quand la vie en communautaire devenait plus compliquée, les hommes ont reconnu la nécessité d'un système de protection qui leur permet de s'entraider en temps de crise et de maladie. Ainsi, les premiers régimes d'assurance ont été initiés et développés à partir de ce besoin.

L'assurance joue un rôle principal en économie de la santé. Aux Etats-Unis, plus de 80% des dépenses sont supportées par les assureurs.

2.1 Définitions

Knight, F. H. (1921) définie l'assurance maladie comme un mécanisme par lequel une personne « dite assurée » se protège contre la perte financière causée par une maladie, un accident ou encore une invalidité.

L'assurance maladie peut être définie encore comme un processus par lequel une personne se protège contre les pertes financières causées par un accident ou une invalidité afin d'améliorer l'utilisation des soins de santé et protéger les ménages contre l'appauvrissement des dépenses personnelles.

2.2 Objectifs de la couverture par une assurance maladie

L'assurance maladie offre aux assurés et leurs ayants droit, une couverture des risques et frais de soins de santé inhérents à la maladie ou l'accident et à la maternité. Par conséquent, elle donne aussi le droit à la prise en charge des frais de soins curatifs, préventifs et de réhabilitation médicalement requise par l'état de santé du bénéficiaire.

De sa part, Matthew J Eichner (1998), a montré que la couverture par une assurance maladie est devenue une obligation en raison de la nature imprévisible des dépenses de soins de santé. En fait, en vieillissant, les gens sont plus susceptibles de tomber malades. Ainsi, les individus ont une idée générale de leur besoin de services médicaux futurs. Toutefois, le montant exact qu'ils dépensent pour les soins de santé reste pour eux en grande partie incertain, (Asymétrie de l'information). Les dépenses de santé restent également beaucoup biaisées. Dans une telle situation, la possession d'une assurance maladie permet de protéger les individus contre l'éventuel dommage lié aux soins de santé.

2.3 Revue de littérature : Assurance maladie et demande de soins

2.3.1 Revue théorique

Plusieurs études ont été menées sur les différents aspects des services de santé et des domaines connexes. Ces études antérieures sur l'utilisation des services de soins de santé ont montré que l'assurance maladie a un impact positif sur l'utilisation des services de santé.

Dans ce sens, Arrow, K. J. (1963) a montré que la mise en place d'un système d'assurance incite les assurés à consommer d'avantage et plus que le taux de remboursement est élevé, cela peut pousser les gens à consommer plus de soins de santé. Pour limiter les comportements de surconsommation, il a préconisé une couverture incomplète. En d'autres termes, une partie de la charge de soins doit être supportée par l'assuré afin de modérer ses dépenses.

Rothschild et Stiglitz (1976) ont conclu qu'il y a une corrélation positive entre la couverture médicale et l'utilisation des soins de santé, en fait ils stipulent que les personnes les plus risquées sont ceux les plus disposées à souscrire une assurance maladie puisqu'elles savent que le montant qu'elles dépenseront pour les soins de santé sera supérieur à la prime qu'ils vont payer. Ce phénomène est appelé **la sélection adverse**.

Annear 2006 a soulignées que les familles qui se confrontent à un choc sanitaire à court terme peuvent subir la pauvreté à moyen et à long terme. Nous parlons des dépenses catastrophiques à ce niveau. En fait, les coûts de soins de santé élevées peuvent pousser les familles à puiser leurs économies ou à les inciter d'emprunter auprès des banques ou à vendre leurs avoirs en particulier dans les pays en développement. Cependant, les ménages qui ne peuvent pas facilement emprunter de l'argent auprès du marché bancaire, peuvent renoncer à des soins de grande valeur et de grande importance pour leurs états de santé et par la suite à chercher à des solutions alternatives (la médecine traditionnelle, par exemple).

Il convient de noter que la demande de soins de santé a certaines caractéristiques spécifiques et doit être prise en considération.

Nyman, (2008), dans son article a montré que la demande de l'assurance maladie est une demande dérivée de la demande de soin de santé. En effet, les services de l'assurance maladie ne sont pas demandés pour leurs propres caractéristiques, mais plutôt parce qu'on attend de ceux-ci un effet positif sur l'état financière de l'intéressé. La demande passe le plus souvent par un prescripteur (médecin, pharmacien) qui détermine quels soins sont nécessaires et les médicaments à consommer et avec quelle quantité.

2.3.2 Revue empirique

D'une manière générale, la qualité des soins est un argument présent dans la fonction d'utilité des patients mais son impact sur la demande et les dépenses peut varier considérablement selon

- Les pathologies : les malades chroniques disposant généralement d'informations pertinentes par exemple,
- La structure du marché : l'existence d'offreurs multiples permettant de comparer différentes stratégies de traitement,
- Les caractéristiques individuelles : les patients les plus éduqués étant aussi ceux qui ont un accès plus facile à l'information.

Au niveau empirique, il est difficile de mesurer l'impact de l'assurance maladie sur les dépenses de soins car l'état de santé des individus est difficilement mesurable et parfois inobservable. Il existe néanmoins plusieurs méthodes pour l'approcher. Le choix de l'approche conditionnera le modèle économétrique à préconiser.

Entre 1974 et 1982 une expérience aléatoire aux États-Unis nommée « Expérience RAND Health Insurance Experiment » réalisé par Joseph P (1993) a pour objectif d'examiner les effets de l'assurance maladie sur la santé des bénéficiaires. Cette expérience a été étudié près de 4000 personnes dans 2000 ménages. Certaines familles qui ont été assignées au hasard à un régime de soins gratuit, tandis que d'autres nécessitent des dépenses personnelle variables à leur charges (10%, 20%, 40%). L'étude a révélé que les personnes affectées à un régime de partage des coûts recherchaient moins de traitement que celles bénéficiant d'une assurance complète.

La majorité des études empiriques ont mis en évidence une relation significative entre l'assurance maladie et les dépenses de santé. Anderson, et Peter H (1978) et Grossman (1972) ont réalisé le premier travail sur les effets de l'assurance maladie dans les années 1900. Celles-ci ont montré que l'assurance maladie avait un effet considérable sur l'utilisation et les dépenses de santé.

Quelques années plus tard, Jowett al (2004) a montré que les personnes bénéficiant d'une assurance maladie sont plus susceptibles de demander des services de soins de santé. En outre, ils ont révélé que l'absence d'assurance maladie avait un impact significatif sur la décision de consultation individuelle et l'utilisation des services de soins de santé.

Dans le même ordre d'idée Ekman, B (2007) a montré que l'assurance maladie pouvait améliorer l'accès aux soins de santé et réduire les dépenses personnelles. Réellement, le comportement personnel est considéré comme un double processus décisionnel pour les besoins de soins de santé. La première décision est la possibilité de consulter un médecin et la deuxième détermine la qualité de soins à obtenir.

Toutes les études empiriques mentionnées dans cette section ont été basées sur des modèles économétriques à choix discrets tels que le modèle logit multinomial, la régression linéaire simple, le probit binomial, la logistique multinomiale emboitée, le probit multinomial indépendant, les équations simultanées ou encore les modèles censurés.

2.4 L'intérêt de l'assurance maladie

La gestion de la branche assurance maladie pose le plus souvent des problèmes aux assureurs. En effet, les résultats techniques de la branche maladie sont en général déficitaires. Il convient d'analyser cet intérêt tant du côté de l'assureur que de l'assuré.

2.4.1 L'assurance maladie : un produit d'appel pour l'assureur

Le produit d'appel se définit comme un produit qui attire la clientèle, en raison le plus souvent de son bas prix, tels que les produits de grande consommation dans les grandes surfaces de vente qui sont vendus pratiquement à perte, afin que le client assimile ces prix très bas à ceux de l'ensemble du magasin. C'est le cas de l'assurance maladie, qui en dépit du fait qu'il ne permet pas à l'assureur de réaliser des produits financiers conséquents et qu'il est en général en déficit, les assureurs en font un ticket d'entrée de l'assuré dans le portefeuille de la compagnie. En effet, celui-ci sera susceptible à confier ses affaires à la société ou à l'individu qui lui offre une bonne couverture. Les mauvais résultats enregistrés dans cette branche seront alors compensés par les autres affaires apportées. L'intérêt pour l'assureur de maintenir un tel de produit en portefeuille est donc d'attirer d'autres risques.

2.4.2 L'assurance maladie : une sécurité pour l'assuré.

Lorsque l'état de santé de l'assuré nécessite de consulter un médecin ou bien de suivre un traitement médical (après par exemple avoir contracté une maladie ou avoir subi un accident), la Sécurité Sociale au titre du régime obligatoire d'assurance maladie, prend en charge une partie de ces frais médicaux, mais non l'intégralité.

Certains soins, respecte les tarifs fixés par la Sécurité Sociale, parce qu'ils bénéficient d'un taux de remboursement relativement élevé, seront correctement pris en charge par l'assurance maladie obligatoire, alors que d'autres laisseront une part importante à la charge du patient. De ce fait, l'assurance maladie a été mise en place afin de permettre aux personnes de souscrire un contrat qui leur remboursera la part des frais de soins non prise en charge par la Sécurité Sociale. En cas d'absence d'assurance maladie, le malade déboursera la totalité de la somme des frais engagées et cela aurait eu des répercussions sur ses dépenses mensuelles. L'assurance maladie représente non seulement une façon de gérer les risques, mais bien plus encore, une sécurité pour celui qui en bénéficie.

2.5 Fraude en assurance maladie

La fraude est un acte intentionnel de la part d'un ou de plusieurs individus visant à obtenir un avantage injustifié ou illégal qui crée un préjudice réel direct et certain pour l'assurance maladie. Il s'agit également d'un acte de mauvaise foi destiné à tromper et à porter atteinte aux intérêts d'autrui.

En assurance Maladie, il existe deux types de fraudes en fonction de leur auteur : Fraude causée par les assurés et fraude causé par des professionnels de santé, (appelée encore fraude organisée). Elle peut prendre plusieurs formes :

- Une sur-tarification des actes médicaux effectués par les prestataires des soins de santé,
- Ajustez les frais en fonction du plafond de garantie en fonction des plafonds de garanties décrites aux tableaux de prestations,
- Factures multiples frauduleuses,
- Des remboursements pour des actes ou des frais médicaux de personnes non couvertes.

Le tableau suivant cite quelques risques de fraude avec leurs actes de maitrises

Tableau 1 : Risques de fraude avec quelques actes de maitrises

Risque (fraude)	Prévention et Maitrise
Fausse déclaration dans le questionnaire médical	Faire effectuer, en cas de doute, un examen médical
Fraude aux prestations de santé (Obtenir des remboursements frauduleux par la falsification d'ordonnances)	Faire contrôler par un médecin-conseil les dossiers de remboursement important

Cumul frauduleux de contrats de santé (Souscrire à plusieurs contrats de santé afin d'obtenir plusieurs remboursements)	-Développer le tiers payant pour limiter le risque de cumul frauduleux -Prévoir une clause rendant obligatoire la déclaration de souscription de contrats similaires et rappelant que le cumul des remboursements ne peut excéder le reste à charge de l'assuré
Fausse facturation par un professionnel de santé	Organiser des réseaux de professionnels de santé agréés en prévoyant le retrait d'agrément en cas de fraude
Majoration par le délégataire de gestion des prestations payées aux assurés	Vérifier le bulletin de soin envoyé par le délégataire avant paiement : rapprocher les montants indiqués avec les flux techniques

2.6 Assurance maladie : les principaux obstacles au développement rentable

Plusieurs intervenants peuvent être impliqués à la non rentabilité de l'assurances maladie, et cela constitue des obstacles pour sa non développement, parmi lesquels on peut citer par exemple :

- **De la part des assurés** : attitudes de fraude, anti sélection, faiblesse du pouvoir d'achat
- **De la part des prestataires de soins de santé** : inflation des prix appliqués aux assurés (tarifs différents selon qu'on est assuré ou pas), surfacturation et sur cotation des actes médicaux, équipements et procédures pas toujours favorables aux assurés (notamment dans les hôpitaux publics)
- **De la part des assureurs** : sous-tarification, non application des clauses contractuelles d'ajustement, approche de tarification plus comptable que réellement assurantielle (fréquence et coûts moyens), coûts d'acquisition et de gestion élevés, faible mutualisation.

Section 3 :la tarification en assurance maladie

3.1 Différentes approches de la tarification

La majorité des entreprises négocient avec les organismes assureurs des contrats d'assurance santé collectifs, auxquels les salariés et éventuellement leurs ayants-droits ont adhéré. Lors de la souscription du contrat, la compagnie d'assurance fixe un taux de prime annuel pour chaque entreprise ce qui explique la forte concurrence sur le marché en vue d'attirer le maximum des clients. En pratique, la prime demandée est proportionnelle à la masse salariale des assurés via ce taux de prime. De ce fait, les compagnies d'assurances doivent considérer des nouvelles méthodes de tarification rigoureuses permettant de proposer la meilleure prime.

3.1.1 Une première approche de tarification : Fréquence x Coût moyen

- Fréquence de consommation

En assurance santé, un sinistre se traduit par la consommation d'un ou plusieurs actes de soins. La fréquence est déterminée de façon déterministe en rapportant le nombre d'actes au nombre de bénéficiaires.

- Le coût moyen de consommation

Le coût moyen de consommation est la somme du coût moyen d'un acte de soin durant une année. Il peut être estimé par le quotient de la somme des consommations et du nombre d'actes dans l'année.

- Prime Pure

Sous les hypothèses que les coûts des sinistres sont indépendants, identiquement distribuées, et que les variables représentant le nombre des sinistres et les coûts des sinistres sont indépendantes, Denuit & Charpentier (2009), représente la prime pure annuelle comme l'espérance mathématique de la charge annuelle de sinistres et la définie par :

Prime pure = fréquence de consommation x coût de consommation.

3.1.2 Une seconde approche de tarification : Probabilité x Charges

En règle générale un acte de soins ne se pratique pas seul. Par exemple un acte de type consultation dentaire implique d'autres traitements donnant lieu à des prestations liées à la garantie couvrant le risque dentaire. Ceci signifie que l'hypothèse supposée dans un modèle fréquence coût classique de l'indépendance entre les variables aléatoires représentant les charges de sinistres est remise en cause.

- La probabilité de conso mmation durant une année

Il s'agit de la probabilité de consommer au moins une fois durant une année. Cette probabilité de consommer un poste médical est évaluée par un modèle de régression logistique avec comme variables explicatives le sexe et l'âge tel que la variable qualitative de modalités 1 et 0 (l'assuré consomme ou non pour un type d'actes donné).

- Charge de consommation durant une année

Une fois consommé, la charge de consommation est les frais réels engagés par un assuré.

- La prime pure

La prime pure de l'espérance mathématique de la charge sinistre

Prime pure = Probabilité de consommer x Charge de consommation

3.1.3 Choix de l'approche Fréquence x Coût moyen

Cette approche a pour hypothèses que les variables de nombre et de coût sont indépendantes, ce qui n'est pas vrai en assurance santé.

Pour démontrer cette dépendance : plus un assuré est couvert en régime haut de gamme, plus ses consommations seront élevées parce qu'il va plus chez le médecin comme il est plus couvert. Donc la fréquence et le coût moyen d'un sinistre ne sont pas indépendants.

En pratique, il n'est pas gênant que ces variables soient dépendantes puisque la consommation est l'estimation par la moyenne empirique des charges sinistres des différents assurés qui est égale au produit des moyennes des coûts par la moyenne empirique de nombre de sinistres.

Comme le coût moyen d'un sinistre est le rapport de la somme des consommations par le nombre d'acte et la fréquence est le rapport du nombre d'actes par le nombre de bénéficiaires.

Ces deux définitions ainsi posées permettent de dire que la consommation moyenne est le rapport entre la somme des consommations et le nombre de bénéficiaires. Le montant constitue est le tarif théorique des garanties, pour obtenir le tarif final, il faut y ajouter les frais. Ainsi, la contrainte de l'indépendance des variables coût moyen et fréquence est éliminée

3.2 Besoin de segmentation en classe homogène et apport de variables exogènes

Le risque supporté par un groupe dépend de certaines caractéristiques démographiques de ce groupe. Or dans un portefeuille, les assurées ne sont pas tous homogène face au risque, d'où la nécessité d'une segmentation.

Nous présentons ci-joint un exemple très simple qui permet de comprendre l'intérêt de la segmentation.

- **Exemple :**

-On suppose que la population G constituée de N personnes qui consomment un type d'acte par exemple « Consultation généraliste » avec une fréquence F et un coût moyen C.

-On considère ensuite que cette population est divisée en deux sous population G1 : les hommes de N1 personnes et G2 : les femmes de N2 personnes.

-On suppose également que la population G 1 (respectivement G2) consomme ce même type d'acte avec une fréquence F1 (respectivementF2) et un coût moyen C1 (respectivement C2)

Hypothèse : On considère deux sociétés d'assurance (A et B) qui veulent assurer ce risque. On suppose que pour ces assurées on a :

- F1 > F > F2
- C1> C > C2

Figure 4 société A qui ne pratique pas la segmentation

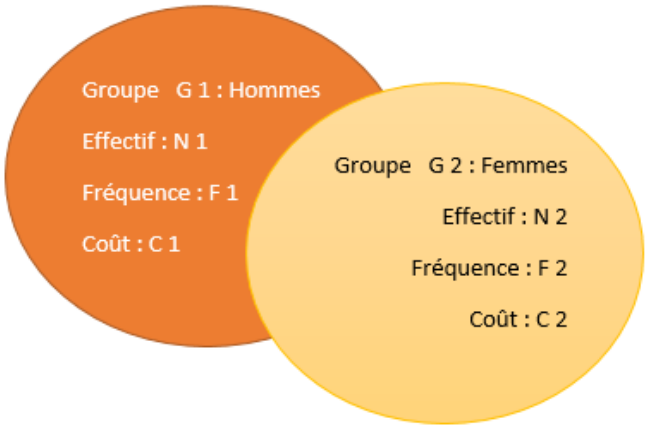


Figure 5 : société B qui pratique la ségmentation

	Choix	Prime demandé
Société A	Ne pas prendre en compte les sous populations.	Prime (G) = F * C
Société B	Existence de deux sous population homogène	Prime (G1) = F 1* C1
		Prime (G2) = F 2* C2

Que va-t-il se passer ?

Si A est seule sur le marché	**Si B entre sur le marché**
Elle propose le tarif moyen et son profit est nul	Les assurés vont se répartir entre les 2 sociétés à cause du tarif différencié. Les hommes vont plutôt aller chez A, et les femmes chez B.

Quel est le bilan pour chaque société ?

	Scénario	Recette -Charge	Bilan
Société A	A n'assure que les hommes avec le tarif moyen	N1*F*C - N1*F1*C1 < 0	La société A est en situation déficitaire.
Société B	B n'assure que les femmes avec le tarif adapté.	N2*F2*C2 - N2*F2*C2 = 0	La société B est en situation équilibrée

À travers cet exemple, on voit bien que la segmentation est un enjeu majeur dans le calcul de la prime pure. Il s'agit de prendre en compte les caractéristiques démographiques d'un groupe.

Généralement, les facteurs influant sur le risque santé sont l'âge, le secteur, le sexe, la catégorie socioprofessionnelle. Cette liste n'est pas exhaustive.

Section 4 : Les travaux empiriques portant sur la tarification

Les travaux empiriques réalisées sur la tarification des primes d'assurances peuvent se distinguer par leurs approches et les modèles actuarielles utilisées ainsi que les variables explicatives. Cependant ils présentent un cadre commun qui sert à la modélisation générale. En effet tous ces travaux s'intéressent à relier la variable à expliquer (Y) à des variables socioéconomique (X) qui se présentent sous la forme :

$$Y= f (X_1,.., X_n)$$

Les variables X_i désignent souvent l'âge, le sexe, ……

4.1 Le modèle linéaire gaussien : Un modèle peu adapté

Le modèle linéaire gaussien est un modèle de régression linéaire classique. C'est l'un des modèles les plus utilisés en statistique du fait de sa simplicité. Cependant, il n'est pas adapté au

contexte assurantiel puisque la variable à modéliser n'est pas nécessairement gaussienne. Cette section s'appuie sur le chapitre 9 de Denuit &Charpentier II (2009).

4.1.1 La formalisation du modèle

Le modèle linéaire classique consiste à établir une relation linéaire entre une variable Y à expliquer et un ensemble des une variables X explicatives. L'équation du modèle est de la forme :

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta j X_{ij} + \varepsilon_i \quad \text{avec} \quad \varepsilon_i \text{ indépendant et } N \sim (0, \sigma^2)$$

Elle peut se réécrire vectoriellement comme suit :

$$Y = X\beta + \varepsilon$$

Avec :

- p : Nombre de variables explicatives,
- n : Nombre d'observations,
- $Y = (Y_1,...,Y_n)^t$: Variables à expliquer supposée indépendantes et non identiquement distribuées
- $Xj = (x_{1j},...,x_{nj})^t$: j -ème variable explicative , j=0,1, . . . , p
- $\beta = (\beta 0 \dots \dots \beta p)^t$paramètres inconnus qui représente le lien existant entre Y et X, à estimer.
- X : la matrice représentant les observations relatives aux variables explicatives
- ε est une variable aléatoire qui représente l'écart entre la variable observée et la variable estimée.

Les hypothèses classiques sur les résidus :

- Espérance mathématique nulle : E(ε^) =0
- Homoscédasticité : var (ε^) =σ^2

4.1.2 Limites du modèle linéaire gaussien

Le modèle linéaire classique n'est souvent pas adapté aux problématiques d'assurance. Il présente par exemple les insuffisances suivantes :

- Dans le cas où la variable Y désigne le coût moyen elle doit être positive, et pour le cas de la fréquence, elle doit être un entier.
- La variable à modéliser n'est pas nécessairement gaussienne, on a souvent recours à d'autre lois continue comme loi gamma pour mesurer le coût de remboursement par exemple.
- Dans le cas où la variable à expliquer Y est qualitative (catégorie sociale, sexe, présence ou absence d'une maladie...) et possède un nombre fini de modalités $g_1,...,g_k$. Le problème consiste à expliquer l'appartenance d'un individu à un groupe à partir des p variables explicatives, on parlera d'analyse discriminante au lieu de régression.

Nous constatons que les modèles de régression classique ne convient pas, pour cette raison il faut utiliser les modèles linéaires généralisées.

4.2 Les modèles linéaires généralisées

4.2.1 Pourquoi un GLM ?

Longtemps, les actuaires se sont limités à utiliser le modèle linéaire gaussien lorsqu'il s'agissait de quantifier l'impact de variables explicatives sur un phénomène d'intérêt (fréquence ou coût des sinistres, ...). Vu que la complexité des problèmes statistiques qui se posent à

l'actuaire s'est considérablement accrue, il est crucial de se tourner vers des modèles tenant mieux compte de la réalité de l'assurance que ne le fait pas le modèle linéaire. Les modèles linéaires généralisés, introduits en statistique par Nelder & Wedderburn (1972), ont été introduits la première fois en assurance par des actuaires à la fin du 20 -ème siècle, permettent de s'affranchir de l'hypothèse de normalité, en traitant de manière unifiée des variables dont la distribution fait partie d'une famille de lois particulière : la famille exponentielle

4.2.2 Composante du GLM

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

- ❖ . Distribution de famille exponentielle
- ❖ Prédicteur linéaire
- ❖ Fonction lien

a) Distribution de famille exponentielle

Une variable Y a une loi faisant partie de la famille exponentielle si sa densité peut se mettre sous la forme :

$$f(y|\theta,\phi) = exp(\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)) \qquad y \in S \qquad (**)$$

- ✓ S est un sous-ensemble de N ou de R
- ✓ θ : entier naturel
- ✓ Ø est un paramètre de dispersion.
- ✓ La fonction b (respectivement c) est une fonction de θ (respectivement de Ø et y)
- ✓ La fonction b(.) est deux fois dérivable

Voici quelques exemples de lois usuelles dont la loi peut se mettre sous la forme (**).

Pour plus de détails voir Denuit & Charpentier II (2009) page 70

Tableau 2 : Les composantes de la famille exponentielle

Loi	**S**	**Ø**	θ	**b(θ)**	**c (y, Ø))**
Loi normale N (μ, σ^2)	R	σ^2	μ	$\frac{\theta^2}{2}$	$-\frac{1}{2}(\frac{y^2}{\sigma^2}+\ln(2\pi\sigma^2))$
Loi de Poisson P(λ)	N	1	Ln(λ)	λ	-ln (y!)
Loi binomiale B (n, p)	N	1	$\ln(\frac{p}{1-p})$	nln(1+exp(θ)	$\ln\binom{n}{y}$

b) Prédicteur linéaire

La composante systématique ηi, nommée prédicteur linéaire, correspond à une combinaison linéaire des variables explicatives.

Soit x_{ij} les observations de la variable explicative Xj , nous avons : $\eta_i = x_i^t\beta$

c) Fonction lien

La relation entre la composante aléatoire et le prédicteur linéaire est exprimée par la troisième composante appelée fonction de lien g, différentiable strictement monotone. Soit $\mu_i = E(Y_i)$, on pose : $g(\mu_i) = \eta_i = x_i^t\beta \Leftrightarrow \mu_i = g^{-1}(\eta_i) = g^{-1}(x_i^t\beta)$

Ainsi, l'espérance de Y correspond à une transformation du prédicteur linéaire $E[Y] = g^{-1}(\eta(x))$. Contrairement aux modèles linéaires simples, il s'agit ici de modéliser une transformation de l'espérance de la variable réponse. Le tableau ci-dessous nous renseigne sur les fonctions de lien classiques.

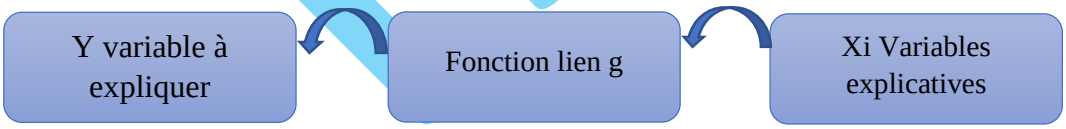
Tableau 3 : Fonctions lien associées à quelques lois de la famille exponentielle

Loi	**g(μ)**	**Fonction lien**
Loi normale $\mathcal{N}(\mu, \sigma^2)$	μ	Identité
Loi de Poisson $P(\lambda)$	$\ln(\mu)$	Log
Gamma	$\frac{1}{\mu}$	Inverse

On peut résumer un modèle GLM comme suit :

- ✓ Etape 1 : Estimer les paramètres $\beta = (\beta 0 \ldots\ldots \beta p)^t$.
- ✓ Etape 2 : Une fois cette estimation est réalisée, on disposera d'une estimation de $\eta(x)$
- ✓ Etape 3 : En choisissant la fonction lien approprié on obtient $E[Y] = g^{-1}(\eta(x))$

Ces composantes peuvent être schématisées de la manière suivante :



Y suit une loi exponentielle	g : fonction inversible	Combinaison linéaire des Xi
	$g(E(Y)) = g(\mu i) = \eta(x)$	$\eta = \beta_0 + \sum_{j=1}^{p} xj\beta j$

4.2.3 Estimation des paramètres par le maximum de vraisemblance

Principe

La vraisemblance du n-échantillon, notée L, est la densité de probabilité associée aux données observées. L'estimation des paramètres β est calculée en maximisant la log-vraisemblance du modèle linéaire généralisé. La fonction logarithme est une fonction

croissante, il est équivalent de maximiser le logarithme de la vraisemblance et la vraisemblance. Or, la maximisation d'une fonction consiste à déterminer la valeur de son paramètre qui annule sa dérivée tout en gardant sa dérivée seconde négative. Il sera alors plus aisé de dériver une somme plutôt qu'un produit.

$$L(\theta(\beta)\backslash y,\phi) = \Pi_{i=1}^{n} f(y_i\backslash\theta_i,\phi)$$

$$l = ln(L(\theta(\beta)\backslash y,\phi)) = \sum_{i=1}^{n} ln(f(y_i\backslash\theta_i,\phi))$$

Trouver les estimateurs du maximum de vraisemblance revient à trouver les paramètres ($\beta_0,\beta_1,...,\beta_p$) qui vérifient les équations $\frac{\partial l}{\partial \beta j} = 0$ j = 0, 1.....p

L'astuce est d'utiliser l'égalité suivante :

$$\frac{\partial l}{\partial \beta_j} = \frac{\partial l}{\partial \theta_j}\frac{\partial \theta_j}{\partial \mu_j}\frac{\partial \mu_j}{\partial \eta_j}\frac{\partial \eta_j}{\partial \beta_j}$$

Ainsi, les équations de vraisemblance s'écrivent :

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(Y_i)g'(\mu_i))}x_{ij}$$

Pour plus de détails on pourra se référer à Denuit & Charpentier II (2009).

Une fois l'estimateur du maximum de vraisemblance déterminé à l'aide d'une des procédures itératives telle que la méthode de Newton-Raphson (que nous ne développerons pas) il est possible d'étudier le modèle

4.3 Critères de choix du modèle

Afin de vérifier l'ajustement du modèle aux données utilisées nous présenterons :

- ❖ Les statistiques permettant d'apprécier l'adéquation du modèle aux données,
- ❖ L'analyse des résidus,
- ❖ Construction des intervalles de confiance des coefficients estimés.

4.3.1- Les statistiques permettant d'apprécier l'adéquation du modèle aux données

Dans le modèle linéaire classique (gaussien) on a :

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \qquad (***)$$

En d'autres termes, la variation totale est la somme de la variation résiduelle, et la variation expliquée. Le coefficient de détermination R^2 constitue une mesure de la qualité d'ajustement du modèle, et c'est le rapport de la variation expliquée et la variance totale. Plus il est proche de 1, meilleur est l'ajustement. L'égalité (***) n'est plus vérifiée dans le cadre d'un modèle linéaire généralisé, l'idée donc est de comparer l'écart entre un modèle

ajusté à un modèle idéale (où l'on aurait autant de variable explicative que d'observation) on appelle un modèle saturé. Il est caractérisé par $\hat{\mu} = y_i$.

Comparer ces deux modèles revient à comparer ces vraisemblances, cet écart se traduit par la notion de déviance.

a) Déviance

On introduit la statistique du rapport de vraisemblance suivante :

$$\Lambda = \frac{L}{\hat{L}}.$$

Le modèle est jugé "bon" si " Λ est proche de 1 ou encore si $\ln(\Lambda)$ est proche de 0.

On note D la déviance. Elle peut être définie comme l'écart, entre la vraisemblance du modèle estimé $L^{\wedge}$ et celle du modèle saturé L.

$$D = 2\ln\Lambda$$

❖ L'idéal serait donc d'avoir D = 0 mais ce n'est jamais le cas, le modèle saturé étant un idéal inatteignable en pratique.

❖ Le modèle est considéré comme mauvais au seuil α si :

$$D_{observé} > \chi^2_{n-p-1;1-\alpha}$$

Avec : $\chi^2_{n-p-1;1-\alpha}$ est le quantile d'ordre 1-α de la loi khi-deux à n-p-1 degrés de liberté.

b) Critère AIC et BIC

Plus on ajoute des variables explicatives au GLM, plus la déviance est petite. Cependant, elle ne prend pas en compte la complexité du modèle. C'est pour cela que l'AIC et le BIC sont deux critères qui seront plus adaptés dans le cadre d'une tarification. Ces deux critères permettent de comparer deux modèles non emboités et qui peuvent avoir des nombres de paramètres différents. Le choix se porte sur le modèle qui possède le plus petit AIC ou BIC.

$$AIC = -2Ln(L) + 2K$$

$$BIC = -2Ln(L) + K\,Ln(n)$$

Avec

- n : le nombre d'observation de l'échantillon,
- K : nombre de paramètre utilisé dans le modèle
- L : la fonction de vraisemblance du modèle.

4.3.2 L'analyse des résidus

L'analyse des résidus peut indiquer si le modèle peut être amélioré. Leur analyse permet de vérifier si l'erreur est aléatoire, et de repérer les valeurs aberrantes ou trop influentes.

Pour l'observation i le résidu observé $r_i = y_i - \hat{\mu}$ n'a qu'un intérêt très restreint. Deux types de résidus sont couramment utilisés dans le cadre des modèles linéaires généralisés : les résidus de Pearson et les résidus de déviance. Ils sont définis par :

❖ Les résidus de Pearson :

$$r_i^p = \frac{y_i - \hat{\mu}}{Var(\hat{\mu})}$$

❖ Les résidus de déviance

$$r_i^d = \text{signe } (yi - \hat{\mu}) \ \sqrt{} \ di$$

Avec d_i représente la contribution de l'observation i à la déviance $D = \sum di$ pour plus de détails on peut se référer à Denuit & Charpentier II (2009).

4.3.3 -Sélection de variables

Il est intéressant de déterminer la meilleure combinaison des variables $\boldsymbol{X_1,\ldots..,X_p}$ qui explique $\boldsymbol{Y}$. Or l'approche qui consiste à éliminer d'un seul coup les variables non significatives n'est pas bonne ; certaines variables peuvent être corrélées à d'autres, ce qui peut masquer leur réelle influence sur $\boldsymbol{Y}$.

On distingue trois approches :

- Approche en arrière,

- Approche en avant,

-Approche pas à pas.

a) <u>Approche en arrière (backward) :</u>

On part d'un modèle GLM avec toutes les variables explicatives X1,, Xp et on étudie leur significativité.

On retire la moins significative (donc celle qui a la plus grande p-valeur). Puis on refait un autre modèle avec les variables restantes et on retire de nouveau la moins significative. Autrement dit celle dont le retrait engendre une baisse significative de l'AIC. On itère ce processus jusqu'à n'avoir que des variables significatives.

b) <u>Approche en avant (forward)</u> :

Le modèle initial est celui ayant comme unique paramètre : β_0 (la référence). A chaque étape de la procédure, la variable la plus significative est rajoutée au modèle de l'étape précédente. La procédure d'ajout de variable est arrêtée lorsque toutes les variables restantes sont jugées comme pas assez significatives pour être intégrées dans le modèle autrement dit celle dont le retrait engendre une hausse significative de l'AIC.

c) Approche pas à pas (stepwise) :

Cette méthode est une combinaison des deux méthodes précédentes (forward et backward). Nous commençons par le même modèle retenu par la méthode forward, mais l'avantage réside dans la possibilité d'éliminer une variable dés lorsqu'elle est devenue moins significative suite à l'ajout des variables supplémentaires. Nous obtenons le meilleur modèle lorsque l'ajout ou la suppression d'une variable ne touche pas de manière remarquable le modèle**.**

4.4. Les travaux empiriques réalisés

4.4.1 Analyse de Matthieu Vautrin

Dans son application, intitulé « Élaboration d'une méthode de tarification avec indicateurs de risque pour des contrats complémentaires sante collectifs » publié dans le site officiel des mémoires des actuaires MATTHIEU VAUTRIN (2008), s'intéresse à la tarification du portefeuille assurance groupe maladie d'une compagnie d'assurance opérant en France avec des observations relatives à 25 contrats collectifs retenus pour l'étude, avec une moyenne d'environ 3 750 bénéficiaires par contrat. Le total se porte à près de 85 000. Il a choisi comme approche de tarification, l'approche fréquence * Coût moyen. Il a introduit des nouvelles variables tel que « IND » qui représente l'occurrence de sinistre. Il s'agit d'une variable binaire indiquant si l'assuré a déclaré ou non un sinistre durant l'année d'étude. Aussi la variable « DUR » qui représente la mesure de l'exposition au risque et la variable « NSIN » qui représente le nombre de sinistre.

Le premier tri parmi des variables tarifaires est effectuée avec un test d'indépendance de khi-deux sur la base de tables de contingence. Le croissement est effectué avec la variable « IND » au lieu de la variable « NSIN ». Une segmentation par région est appliquée au portefeuille étudié en plus de la segmentation induite par les variables tarifaires classiques telles que le sexe et l'âge des assurés. Dans ce mémoire VAUTRIN (2008) a exploité des informations sur le régime, le niveau de garantie (Un bon niveau de garantie augmente souvent l'aléa moral) et les modes d'adhésion au contrat. En effet, les contrats dont le mode d'adhésion est facultatif sont plus exposés au risque d'anti sélection que ceux à adhésion obligatoire. Ces deux dernières variables ont donc leur importance dans la modélisation des coûts lorsqu'elles sont disponibles et exploitables. Dans notre mémoire ces conditions ne sont pas remplies et nous n'incluons donc pas ces variables dans les modèles implémentés.

L'estimation des paramètres dans cette référence est effectuée sur le logicielle SAS. Les paramètres estimés permettent d'examiner la contribution de chacune des variables par rapport au modèle. L'ajustement des fréquences de l'acte examens de laboratoire est réalisé par une régression binomiale négative en raison de la sur dispersion constatée de certaines variables de comptage, et pour les coûts ils sont saisie par une régression gamma.

4.4.2 Analyse de NGUYEN

Une deuxième étude faite par NGUYEN (2013) dans son application intitulé « Construction de bases de tarification pour des contrats complémentaires santé collectifs » publié dans le site officiel

des mémoires des actuaires. L'objet du mémoire est donc de construire des bases de tarification dans le but d'estimer la prime pure d'un contrat complémentaire santé collectif par le modèle linéaire généralisé. Avant d'entamer la modélisation, il a commencé son étude par une analyse descriptive du portefeuille et il à insister sur la l'importance de constitution de la base de données. La phase d'extraction des données et leur nettoyage est très importante pour son étude.

Pour l'application du GLM, les variables « Coût » et « Fréquence » doivent suivre une des lois de la famille exponentielle. Pour l'analyse des coûts des sinistres, et en négligeant les sinistres graves, il a choisi comme modèles les modèles de régression Gamma. Comme, le modèle de Poisson est relativement contraignant, car il impose l'équidispersion des données il a utilisé le modèle Binomiale-Négative. Ce sont les lois pour lesquelles il a obtenu les meilleurs résultats sur les tests par le graphique et par l'indice de Kolmogorov-Smirnov. Pour l'ajustement du modèle, il a choisi la méthode Backward de sélection des variables explicatives.

Finalement NGUYEN (2013), a exposé un exemple détaillé de tarification d'un groupe d'assuré à partir du tableau des estimations trouvées,

Conclusion du premier chapitre

Le marché de la santé est caractérisé par l'omniprésence des phénomènes d'incertitude présente sous différentes formes. Elle débute chez le patient qui fait appel aux soins médicaux sur la base d'événements incertains mais réalisés : jambe cassée, infarctus...

Dans ce chapitre, on a appuyé notre partie théorique sur le rapport de l'OMS et une synthèse de littérature de l'assurance maladie basée sur des différents articles de recherches effectuées par plusieurs auteurs. Sur la base de données internationales établis en 2018 (pays de l'OCDE), on a montré le poids de dépenses de santé dans les ménages et la nécessité de l'assurance maladie pour leur sécurité financière. On a notamment illustré le rôle que joue l'assureur au sein de ce système et la nécessité de tarifier aussi correctement possible un contrat assurance maladie.

En fin du chapitre nous avons exposé, un petit aperçu sur les travaux empiriques réalisés par les chercheurs en présentant les différentes approches de tarification en assurance santé en insistant sur la nécessité d'adopter une méthodologie appropriée. Celles-ci sont centrés sur les modèles GLM relatifs aux distributions de la famille exponentielle.

Dans le prochain chapitre, nous allons consacrer notre effort à la présentation de la base de données du GAT. Cette base contient les observations relatives aux principales variables du modèle.

CHAPITRE DEUX :

GAT ASSURANCES : ETUDES STATISTIQUES PRELIMINAIRES

Chapitre Deux : GAT Assurances : Étude statistique préliminaire

Introduction :

L'assurance groupe maladie figure parmi les assurances des personnes qui a pour but de garantir la sécurité financière des ménages et de protéger les imprévus en matière de santé en Tunisie ou à

l'international. GAT ASSURANCES entant qu'une assurance pratiquant la branche santé, propose une panoplie de garanties telles que le remboursement des dépenses médicales, chirurgicales et pharmaceutiques, etc.…Le contrat d'assurance santé peut aussi comprendre des garanties optionnelles telles que : les prothèses dentaires, les soins optiques, les frais de maternité les appareils auditifs, etc...

D'après le rapport du Comité Général des Assurances (CGA) 2018 et le rapport de la Fédération tunisienne des sociétés d'assurances FTUSA on constate clairement que GAT Assurances figure parmi les plus puissant dans le marché d'assurance en Tunisie grâce à sa politique de souscription et sa bonne gestion de sinistre. Consciente de l'importance de l'optimisation de son processus d'assurance, GAT Assurances a opter pour un projet d'excellence opérationnelle qui vise l'optimisation de son processus d'assurance santé, plus particulièrement dans un projet de tarification des contrats santé collectifs. Et Comme toute étude actuarielle passe nécessairement par la fiabilisation des données de base, le traitement des données a été perçu comme un paramètre indispensable dans notre étude.

À ce titre l'objectif principal de ce chapitre consiste essentiellement à présenter les données de base et les différents processus de détection ainsi la correction des anomalies et la création de quelques variables, et par la suite l'analyse du notre portefeuille. Ce chapitre vise aussi à proposer quelques statistiques globales relatives à la répartition de la population étudiée, que ce soit au niveau de la consommation, ou bien, au niveau de la composition du portefeuille.

Ce chapitre est subdivisé en trois sections. La première section illustre une présentation du marché d'assurance santé en Tunisie et ses principaux intervenants. La deuxième section présentera notre compagnie de parrainage ainsi que quelques chiffres clés. La troisième section sera consacrée au traitement préliminaire des données vue d'identifier les interdépendances entre les

Section 1 : Le marché de l'assurance santé en Tunisie

1.1 La sécurité sociale en Tunisie

Le système de sécurité sociale en Tunisie est destiné à protéger les travailleurs et leurs familles contre les risques inhérents à la nature humaine, susceptibles d'affecter les conditions matérielles et morales de leurs existences tel que la vieillissement, l'incapacité, l'accouchement et la maladie. On distingue trois caisses de sécurité sociale CNSS, CNRPS et l'ETAT :

- **La Caisse Nationale de Retraite et de Prévoyance Sociale (CNRPS)** :

Créée en 1975 par la fusion de la Caisse Nationale de Retraite et de la Caisse de Prévoyance Sociale, et rassemble tous les actifs, les pensionnées et ses ayants droits du secteur public.

- **La Caisse Nationale de Sécurité Sociale (CNSS)** :

Créée par la loi du 14 décembre 1960, c'est l'organisme qui veille à assurer la couverture sociale des employés du secteur privé au profit des salariés non agricoles et la couverture s'est étendue aux affiliés des autres régimes.

1.2Le système d'assurance maladie

On peut regrouper le système d'assurance maladie en Tunisie en deux systèmes :

- **Le système de santé public**, qui est géré par la Caisse Nationale d'Assurance Maladie (CNAM) pour laquelle on va donner un aperçu dans la suite de cette partie. Ce système de protection sociale concerne toute la population du secteur privé et public. Par la suite, les employeurs du secteur privé sont tenus d'affilier leurs employés à la Caisse Nationale de Sécurité Sociale (CNSS) dans un délai d'un mois à compter de leur établissement. Vérifier que cela a bien été fait car l'immatriculation auprès de la Caisse Nationale d'Assurance Maladie est obligatoire pour bénéficier de la couverture maladie**.**

- **Le système de santé privé** qui est bien développé, aussi bien au niveau des infrastructures que de la capacité d'accueil et du personnel de santé. Le secteur privé propose entre autres des prestations de haute qualité de chirurgie esthétique, de thermalisme et de thalassothérapie. De même, les dentistes et opticiens ne se trouvent quasiment que dans le secteur privé. Certaines prestations toutefois dans le système privé sont prises en charge par l'assurance maladie. En effet les employeurs du secteur privé ainsi que certaines entreprises publiques contractent au profit de leurs employés des contrats assurances groupe pour couvrir le risque maladie, dans le cadre d'un contrat groupant plusieurs autres risques (invalidité, incapacité, décès). Ce régime s'est développé pour combler certaines insuffisances des régimes publics.

Tableau 4 : Les différences entre le système de santé public et le système de santé privé

Le système de santé public	**Le système de santé privé**
À but lucratif	Conforme aux objectifs commerciaux
Conçu pour couvrir un service de base	Conçu pour favoriser le choix, la flexibilité et l'efficacité
Par souci d'équité, une grande partie de la population est couverte.	Il est souvent réglementé et limité à un petit nombre.
Tarification non basée sur l'accessibilité financière.	Tarification basée sur l'expérience ou la mutualisation

1.3. Caisse Nationale d'Assurance Maladie (CNAM)

1.3.1. Fonctionnement de la CNAM

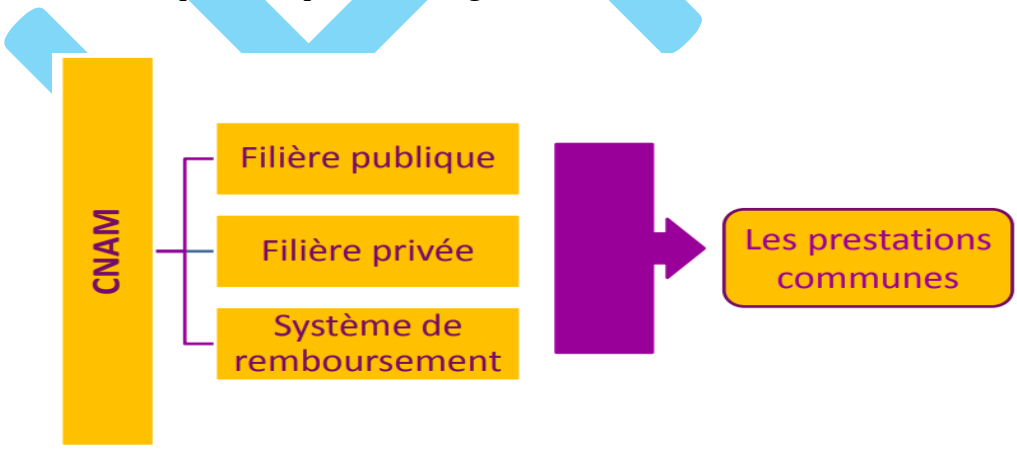
La Caisse Nationale d'Assurance-maladie (CNAM) est un régime d'assurance maladie tunisien. Il est mis en place en 2004 dans le cadre de la réforme visant à unifier les régimes d'assurance maladie et des prestations sanitaires auparavant assurées par la Caisse nationale de sécurité sociale (CNSS)et la Caisse nationale de retraite et de prévoyance sociale (CNRPS), mais aussi à élargir la couverture sanitaire aux prestataires privés de soins. Cette caisse a pour mission :

- La gestion des régimes d'assurance maladie,
- La gestion des régimes de réparation des dommages résultants des accidents du travail et des maladies professionnelles dans les secteurs public et privé,
- L'octroi des indemnités de maladie et de l'accouchement.

Une étude actuarielle et financière faite dans le cadre du projet de cette réforme à dégagée un taux de prime d'équilibre globale du nouveau système aux alentours de 6.75%[1] de la masse salariale repartie comme suit : 2,75% à la charge de l'employé et prélevés directement sur son salaire et 4% à la charge de l'employeur.

1.3.2. Les modes de couverture et prise en charge

Dans le cadre du nouveau régime d'assurance maladie, tout assuré social disposer la possibilité de choisir entre trois options de prise en charge.



Source : présentation GAT Figure 6 : Modalité de prise en charge des soins

a) Filière publique :

Le système de santé public permet l'accès aux différents soins prodigués par les structures publiques de santé aussi bien en ambulatoire qu'en hospitalisation. L'assuré n'aurait à payer qu'un ticket modérateur (C'est la différence entre la base de remboursement et le montant remboursé par

[1] Http ://www.cnam.nat.tn

l'assurance maladie obligatoire) plafonné à un montant annuel. Dépassant ce plafond, la prise en charge serait totalement supportée par la caisse. Ce principe de tiers payant permettrait une accessibilité garantie aux soins nécessaires dans le cadre d'une convention établit entre le ministre de la santé publique et les caisses de sécurité sociale.

b) Filière privée :

L'assuré social et ses ayants droit doivent obligatoirement passer par « le médecin de famille » avant de s'adresser à tout autre professionnel de santé, et ceux à l'exception de certaines spécialités limitatives tel que la pédiatrie, la médecine dentaire, l'ophtalmologie

c) Le système de remboursement :

Ce système permet aux assurés et à leurs ayants droits de choisir librement le prestataire de soins conventionné aussi bien en ambulatoire qu'en hospitalisation en lui réglant directement les frais occasionnés à cet effet. Le taux de remboursement varie selon qu'il s'agit d'une maladie lourde (APCI) ou maladies ordinaires.

1.4 Insuffisance de la CNAM et nécessité de l'assurance maladie

D'après la déclaration du ministre des Affaires sociales, le volume des dépenses de la Caisse Nationale d'Assurance Maladie (CNAM) dans le cadre de sa contribution à la couverture du coût des soins au profit des affiliés sociaux, a atteint 2028 millions de dinars au cours de l'année 2019, dont 43% des dépenses relatives aux soins, au profit des prestataires des services sanitaires privés, tandis que 57% de ces dépenses ont concerné les établissements sanitaires du secteur public.

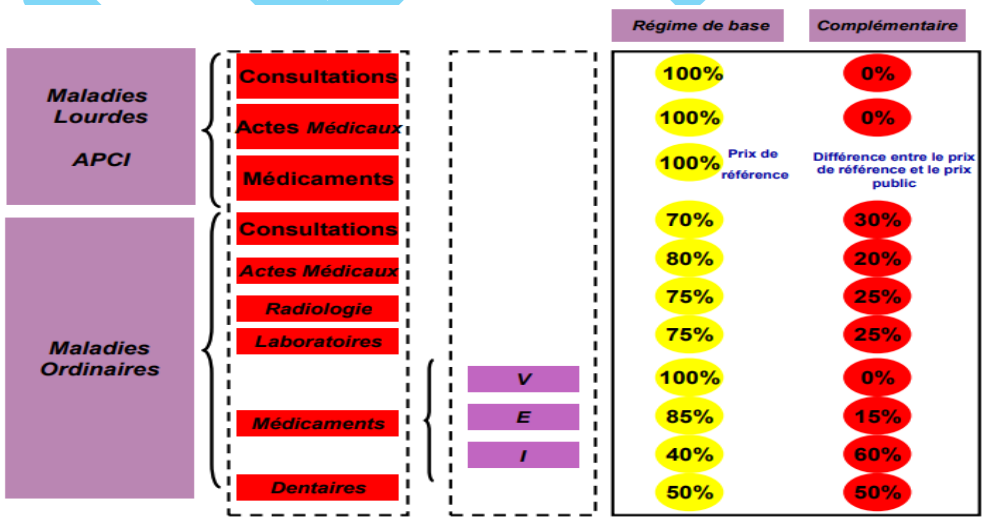


Figure 7 : taux de remboursement de la CNAM Source : présentation GAT

Malgré les améliorations successives apportées à ce système, les modalités de couverture sont très variables et complexes et la prise en charge est devenue au fil du temps trop partielle ce qui a généré, à la fois, l'augmentation des dépenses de santé mises à la charge des ménages. Pour combler ces lacunes et ses insuffisances, les employeurs ont fait recours à une assurance complémentaire

au profit de leurs salariés auprès des compagnie d'assurances sous formes des contrats d'assurances groupe.

1.5. Contrat synallagmatique : Prime contre Couverture

Le contrat d'assurance est désigné comme étant un contrat synallagmatique : c'est à dire une obligation pour les deux parties : moyennant le paiement d'une prime dont le montant est fixé a priori en début de période de couverture (C'est l'inversion du cycle de production, caractéristique du secteur de l'assurance), l'assureur s'engage à couvrir l'assuré pendant toute la période de couverture (disons une année).

Cette prime est censée refléter le risque associé au contrat. Pour chaque police d'assurance, la prime est fonction de variables dites de tarification (permettant de segmenter la population en fonction de son risque). Généralement, on considère des informations sur l'assuré, comme l'âge ou le sexe pour un particulier, ou le secteur d'activité et la taille mesurée souvent par le nombre de salariés pour une entreprise, etc.

La tarification consiste à déterminer la prime demandée au souscripteur, le prix que doit payer l'assuré pour bénéficier de la couverture du risque en cas de sa réalisation appelée Prime Commerciale (PC) elle est décomposition de la manière suivante :

Prime Commerciale = Prime Pure + Chargement de gestion + Chargement Commercial.

Avec : La Prime pure correspond au prix du risque. Le chargement de gestion correspond aux charges fixes. Le chargement Commercial représente l'ensemble des coûts relatifs au contrat (Agents, acquisition,)

En ce qui concerne les contrats collectifs, trois types de contrats sont proposés sur le marché :

- ✓ **Contrats à risque** : il s'agit des contrats conventionnels au sens de l'assurance et tous les risques et à la charge de la compagnie d'assurance ;
- ✓ **Les contrats à risque avec participation aux bénéfices** : il s'agit d'une variante des contrats à risque avec la possibilité de récupérer une partie de la prime par le souscripteur en cas où le contrat affiche un résultat bénéficiaire à la clôture de l'exercice. Le taux de la participation bénéficiaire est négocié lors de la souscription du contrat selon les taux de sinistralité antérieurs ;
- ✓ **Contrat de gestion de compte** : ce sont des contrats ou la compagnie d'assurance joue le rôle d'un simple prestataire de service, les risques inhérents à de tels contrats est pris en charge par le souscripteur. À la fin de l'exercice, si le contrat est déficitaire, le souscripteur rembourse le déficit, si le contrat profite, le souscripteur gagne du profit ;

Outre ces contrats complémentaires, le marché tunisien offre d'autres contrats d'assurance santé qui ne rentrent pas dans le cadre d'une complémentaire il s'agit des

- ✓ **Contrats d'assurance santé au premier dinar** (non complémentaire) : Certains assurés, pour une raison ou une autre, préfèrent le remboursement direct par leur assureur sans passer par la CNAM.

Dans ce cadre, ce type de contrat perd le qualificatif de "complémentaire" et tout simplement un contrat d'assurance santé sans aucune liaison avec le régime obligatoire d'assurance maladie.

Section 2 : GAT ASSURANCES et quelques chiffres clés

2.1. Présentation du GAT assurance

GAT assurances est une entreprise d'assurance et de réassurance opérant sur le marché depuis 45 ans, dotée d'un capital social de 45 millions de dinars tunisiens. Elle est classée parmi les premières compagnies grâce à ses valeurs humaines, ses engagements respectés, la richesse de ses produits et sa forte expertise. GAT assurances est une société anonyme à 100% tunisienne et à capitaux privés.

Activité : Toutes branches d'assurances et de réassurances

Création : le 18/07/1975

Forme juridique : SA de droit Tunisien

Chiffre d'affaire (2017) : 174.5 Millions de dinars en termes de primes acquises.

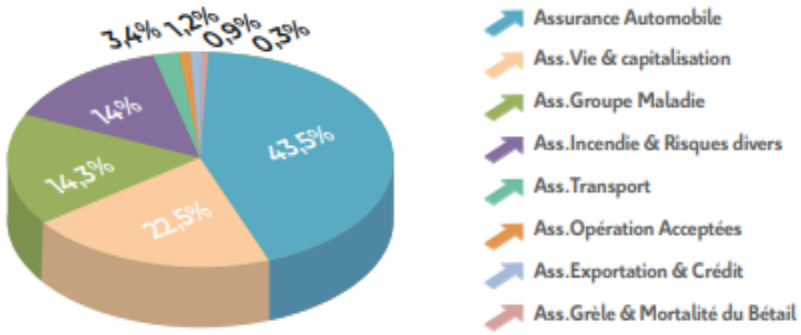
2.2. Historique de l'identité visuelle du GAT ASSURANCES

Figure 8 : évolution de l'identité visuelle de GAT assurances

2.3. Primes émises

D'après le rapport de CGA 2018, l'examen des données des cinq dernières années (2014-2018) fait ressortir une croissance régulière du chiffre d'affaires global du marché d'assurance à un taux annuel moyen de 9,8 %. Cependant, les primes émises en 2018 se sont accrues à un taux annuel de 7,9 % (contre 12,5 % en 2017) pour atteindre 2.252,4 MD (contre 2.087,9 MD l'année précédente).

L’assurance automobile reste toujours la locomotive du marché avec une part de 43,5 % des primes. La deuxième activité d’assurance réside dans la couverture des risques de la santé avec une part de marché de 14,3 % et un volume total des primes émises dépassant 322 MD



Sources : rapport FTUSA 2017

Figure 9: Répartition des primes émises totales par catégories et par branche d’assurance

Les primes émises par compagnie en assurance groupe maladie pour les trois entreprises sont retracées dans le tableau ci-après : (Unités DT)

Intitulé	2015	En %	2016	EN %	2017	EN %
1. STAR	75 520 995	31,72	81 698 913	30,79	85 264 961	29,28
2. MAGHREBIA	31 306 693	13,15	36 535 047	13,77	41 056 607	14,10
3. GAT	26 405 134	11,09	27 621 563	10,41	35 309 051	12,13

Sources : rapport FTUSA 2017

Figure 10:Les primes émises par entreprise en assurance groupe maladie

GAT assurances occupe la troisième place en termes de primes émises pour la branche assurance Groupe maladie avec un chiffre d’affaire réalisé en 2018 de 41.666 MD, soit une croissance annuelle de 17% par rapport à l’année 2017 (35 ,30) MD.

2.4. Sinistres payés

Au niveau des indemnisations, l’année 2018 a connu une hausse du rythme d’évolution de la valeur des sinistres réglés qui a avoisiné 19,9 % contre 3,6 % en 2017 en totalisant 1.262,8 MD contre 1.053,8 MD en 2017. Par ailleurs, les indemnisations payées au profit de l’assurance groupe maladie représente 22.5 % du total des indemnisations

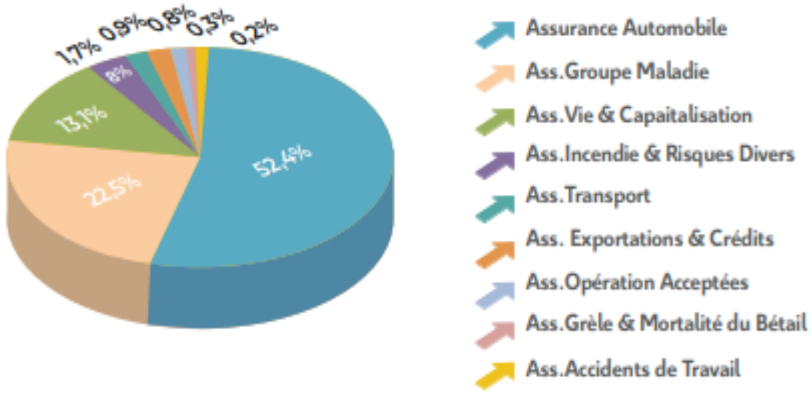


Figure 11 : Répartition des indemnisations réglées par catégories et par branche d'assurance

Sources : rapport FTUSA 2017

Les sinistres payés par les trois entreprises d'assurances durant les trois dernières années sont retracés dans le tableau suivant : (Unités DT)

Intitulé	2015	2016	En %	2017	En %	Evolution 17/16%
1- STAR	64 700 813	71 503 200	30,89	75 038 462	29,01	4,94
2- MAGHREBIA	27 306 811	29 040 089	12,55	32 358 278	12,51	11,43
3- GAT	22 011 516	24 734 159	10,69	28 568 111	11,05	15,50

Figure 12 Les sinistres payés par les entreprises d'assurances pour la branche Groupe maladie

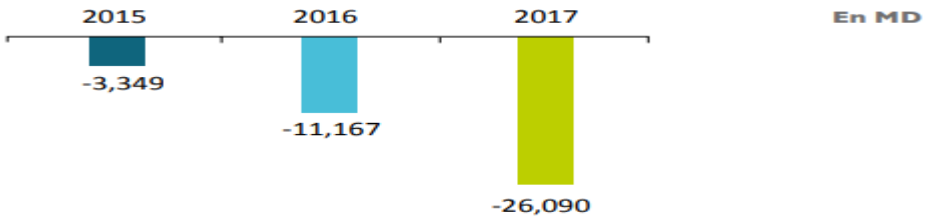
Sources : rapport FTUSA 2017

Le total des indemnisations pour GAT a marqué une augmentation pour atteindre 37.988 MD en 2018 soit une évolution de 50 % par rapport l'année 2016 et 33 % par rapport à l'année 2017 (28,568) MD. Elle occupe le troisième rang en termes d'indemnisation des sinistres.

2.5. Résultat technique

La branche Groupe Maladie connait une forte concentration des primes émises, cependant cette branche d'assurance enregistre un résultat déficitaire structurel.

Figure 13:Résultat technique de l'assurance Groupe maladie



Sources : rapport FTUSA 2017

GAT assurance n'est pas à l'abri de tout ça, malgré l'augmentation de son chiffre d'affaires, cette évolution est marquée par un résultat déficitaire structurel pour la branche maladie. En effet le résultat technique est négatif durant les trois années et il se creuse au cours du temps, pour atteindre -4.191 MD en 2018

Tableau 5 : Principaux indicateurs du GAT assurances pour la branche groupe maladie

	2016	**2017**	**2018**
Chiffre d'affaires	27,622	35,309	41,664
Sinistres réglés	24,734	28,568	37,988
Provisions techniques	3,505	7,243	8,212
Charges techniques	8,421	8,795	10,262
Résultat technique	-2,907	-2,438	-4,191
Primes cédées	0,307	0,539	1,786
Taux de cession	1,10%	1,50%	4,30%

Source : Rapports CGA 2018, 2017, 2016

La branche d'assurance groupe maladie a marqué un résultat technique négatifs durant les dernières années et cela dû peut-être à une mauvaise tarification qui ne tient pas compte de l'hétérogénéité de la population étudie. Pour cette raison, le comité général des entreprises d'assurance CGA a imposé aux entreprises d'assurances de créer une cellule d'actuariat comme une ligne de défense en effet, D'après **l'article 47 du code des assurances :**

« Les entreprises d'assurances et les entreprises de réassurance sont appelées à nommer un actuaire au sein de l'entreprise chargé de déterminer les provisions techniques à constituer, tarifer les primes d'assurances et s'assurer de leur adéquation avec les engagements effectifs de l'entreprise. »

2.6. Tarification au sein du GAT

L'approche de tarification en assurance santé suivie par le GA repose sur la formule suivante :

Prime pure= Taux de prime pure x Masse salariale

Il en découle que l'approche appliquée par les assureurs tunisiens n'est pas celle qu'on essaie de présenter dans cette étude. Cette approche de tarification ne prend pas en compte les critères liés à l'assuré tel que : le sexe, la catégorie socioprofessionnelle, la zone d'habitation, …, mais plutôt quelques informations générales sur le groupe à assurer à savoir :

-le nombre d'adhérent, nombre d'enfant, le nombre total du conjoint dans le contrat, et l'âge moyen.

-les informations liées au type du contrat à souscrire (contrat à gestion pour compte, contrat à risque avec participation au bénéfice)

-le choix des tableaux de prestation.

Le gestionnaire santé prend ces informations et essaye de trouver un taux d'équilibre à l'aide d'un simulateur sous Excel qui est spécifique au portefeuille de la compagnie.

Section 3 : Traitements préliminaires des données de l'études empirique

Consciente de l'importance de la tarification d'un produit d'assurance et de bien tenir compte de l'hétérogénéité de la population au sein d'un portefeuille, GAT assurances s'est donné pour objectif d'améliorer son système de tarification. À ce titre, la présente étude s'inscrit dans ce cadre-là. Pour ce faire, la compagnie a préparé des bases des données contenant les informations nécessaires pour toute analyse empirique relative au calcul des primes.

3.1. Présentation des données reçu

Les données utilisées pour cette étude ont été fournies par la direction santé. Ces données se décomposent en deux fichiers : le fichier des « bénéficiaires » concernant les informations sur la population assurée et le fichier des « Prestations » portant sur l'ensemble des prestations médicales et quelques tableaux de prestation relatifs à quelques entreprises permettant de préciser le niveau de garantit de chaque société.

3.1 .1 Le fichier « bénéficiaires »

La base de données « bénéficiaire 2018 » est composée de 153 216 lignes et 27 variables relatives aux informations de l'assuré. Parmi ces informations observées, nous proposons d'analyser les variables suivantes :

- Le numéro du contrat
- Le type du contrat, indiquant s'il s'agit d'une couverture collective ou individuelle
- État en 2019 : contrat résilié en 2019 ou en cours
- Nom du groupe
- Code famille : l'identifiant unique pour chaque famille
- Identifiant ayant droit : unique pour chaque personne bénéficiant de la couverture
- Nom et prénom du bénéficiaire
- La date de naissance du cotisant et de bénéficiaires
- Le lien de bénéficiaire dans le contrat (Adhérent principale, Conjoint, Enfant, AUTR)
- La date d'affiliation au contrat
- La date de résiliation

3.1.2 Le fichier « prestations

Cette base de données représente les montant de remboursement par date de soins et par bénéficiaire, les variables retenues sont les suivantes :

- Date de naissance
- Date de soin
- Montant dépensé
- Montant remboursé
- Libellé Acte
- Identité bénéficiaire : le numéro attaché à la personne adhérant au contrat,
- Qualité bénéficiaire (conjoint, enfant, ascendant, responsable, autre)
- Identité famille :
- Souscripteur
- Prénom bénéficiaire
- Nom bénéficiaire
- Code famille (référence sur SI)
- Collège
- Type de contrat
- Numéro de contrat

3.2. Statistiques descriptives et sélection des variables

Cette section décrit de façon synthétique les variables décrivant le portefeuille du GAT suivant plusieurs axes. Nous faisons l'analyse descriptive de ces variables en vue d'identifier leur lien avec le tarif à proposer

3.2.1 Analyse de la Base bénéficiaire

Il s'agit de l'ensemble des personnes ayant été présent pendant l'année durant toute la période 2018.

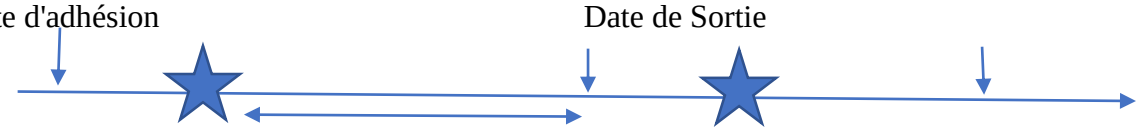
Pour chaque bénéficiaire, nous avons calculé l'exposition au risque durant l'année. Cette dernière mesure le temps durant lequel l'individu est exposé au risque. Il 'agit de la durée totale évacué par l'individu durant la période d'observation. Pour cela, nous avons considéré la date d'affiliation au contrat et la date de sorti de couverture.

La démarche suivie pour déterminer l'exposition au risque de chaque assuré de la branche maladie est comme suit : Pour chaque assuré i, l'exposition au risque est définie par la relation suivante :

$$Exposition\ i = \frac{\text{Nombre de jours d'assurance}}{365}$$

À partir de cette formule, nous remarquons que l'aléa réside dans le paramètre nombre de jours d'assurance pour chaque individu. Ainsi, on distingue deux cas possibles :

Cas 1 :



Nombre de jours

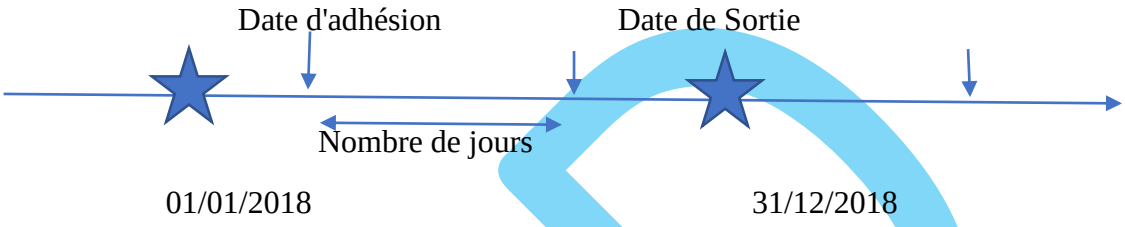
01/01/2018 31/12/2018

Ce cas concerne les adhérents qui s'assurent avant le début de l'exercice d'étude

Si la date de sortie est incluse dans l'exercice d'assurance, le nombre de jours d'assurance se présente comme suit : $Nombre\ de\ jours\ d'assurance_i = (Date\ de\ sortie - 01/01/2018)\ jours$

Dans le cas contraire, c'est-à-dire si la date de sortie n'est pas incluse dans l'exercice, le nombre de jours d'assurance est égale à 365, et l'exposition vaut 1.

Cas 2 :



Ce cas concerne les adhérents qui s'assurent au cours de l'exercice d'étude. Dans ce cas :

Si la date de sortie est incluse dans l'exercice d'assurance.

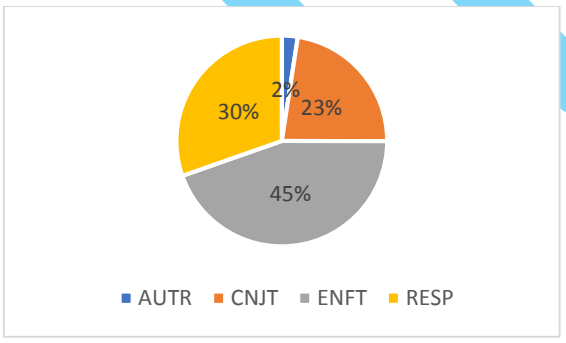
$$Nombre\ de\ jours\ d'assurance = (date\ de\ sortie - date\ adhésion)\ jours$$

Si non, c'est-à-dire si la date de sortie n'est pas incluse dans l'exercice,

$$Nombre\ de\ jours\ \ d'assurance = (\ 31/12/2018\text{-date d'adhésion})jours$$

a) Répartition de la population assurée selon le type de bénéficiaire :

Une décomposition du portefeuille des assurés par type bénéficiaire fournit les résultats suivants :



Bénéficiaires	Nombre de bénéficiaires
AUTR	3 738
CNJT	34 616
ENFT	68 335
RESP	46 526
Totale	153 215

Source : élaboré par nous même

Figure 14 : Composition du portefeuille par type de bénéficiaire

Tableau 6 : Nombre de bénéficiaires

Ce graphique montre que le nombre des ascendants est négligeable dans cette étude. Les poids des enfants et des conjoints présentent (68%) du portefeuille, avec une proportion de 23% pour le conjoint et 45% pour les enfants. Ainsi, nous en déduisons l'hypothèse que la majorité des assurés principaux adhèrent avec au moins 1 personne à charge.

b) Répartition de la population assurée par situation de famille et nombre d'enfants

En l'absence d'information sur la situation de famille dans les bases de données, la variable « Situation famille » est créée en comptant le nombre de bénéficiaires présents dans la variable « Code famille ». La variables « situation de famille » contient 4 modalités qui sont définies comme suit :

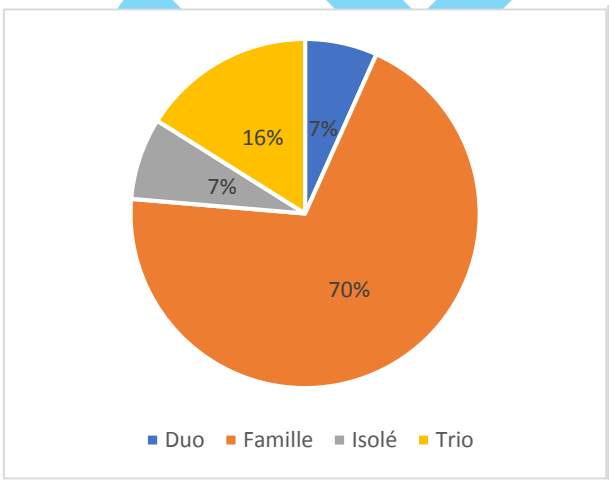
- « Isolé » : si 1 seul type de bénéficiaire est attaché au « Code famille ».
- « Duo » : si 2 types de bénéficiaires sont attachés au « Code famille ». Nous avons 3 possibilités :
  - Assuré principale, 1 conjoint ;
  - Assuré principale, 1 enfant ;
  - Assuré principale, 1 ascendant.

- « Trio » : si 3 types de bénéficiaires sont attachés au « Code famille ». Ainsi on a 3 possibilités :

  -Assuré principale, conjoint, enfant ;
  -Assuré principale avec 2 enfants ;
  -Assuré principale, conjoint, ascendant.

- « Famille » : si au moins 4 types de bénéficiaires sont attachés au « Code famille ». Nous avons 3 possibilités :

  -Assuré principale, conjoint avec au moins 2 enfants ;
  -Assuré principale avec au moins 3 enfants ;
  -Assuré principale, conjoint, ascendant, enfant.



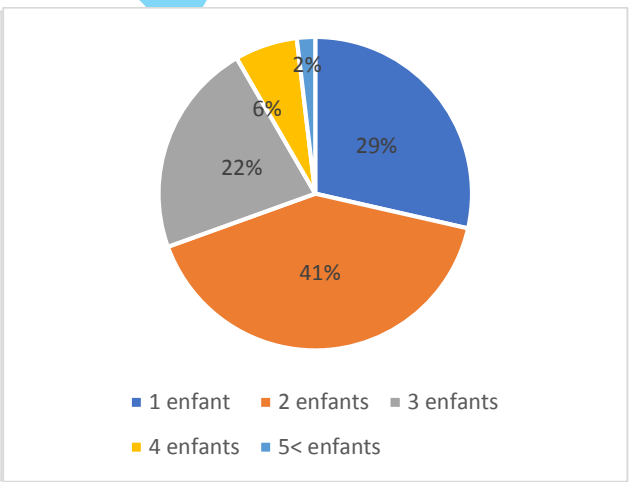


Figure 15 : Répartition par situation de famille

Figure 16 : Répartition par nombre d'enfants

Effectivement, seulement 7% des participants adhèrent en Isolé contre 93% aux structures où il y a au moins 2 personnes adhérentes.

En négligeant le nombre des ascendants, d'après le graphique, nous constatons qu'au minimum 86% des assurés (l'addition des poids du Trio 16% et de la Famille 70%) ayant des enfants se répartissent comme suit :

- Au minimum 29% des assurés ont au moins 1 enfant ;
- Au minimum 41% des assurés ont au moins 2 enfants ;
- Au minimum 22% des assurés ont au moins 3 enfants ;
- Au minimum 6% des assurés ont au moins 4 enfants ;
- Au minimum 2% des assurés ont au moins plus que 4 enfants.

c) Répartition de la population assurée selon la tranche d'âge :

Dans cette étude, nous ne sommes pas intéressés à l'âge du bénéficiaire au moment de l'étude, mais au moment de l'exécution d'un soin. L'âge est donc défini par la formule :

AGE = Année de la date de soins – Année de naissance

Ensuite, nous créons une variable tranche d'âge « Tranche âge » en regroupant l'âge en 14 tranches dont :

- 13 tranches de 5 ans pour l'âge allant de 0 jusqu'à 65 ans, telles que : [0-5[, [5-10[, [10-15[, [15-20[, [20-25[…., [60-65[

- 1 tranche pour l'âge supérieur ou égal à 6 ans « >=65 ».

Ce regroupement par tranche d'âge nous permet de fluidifier les analyses.

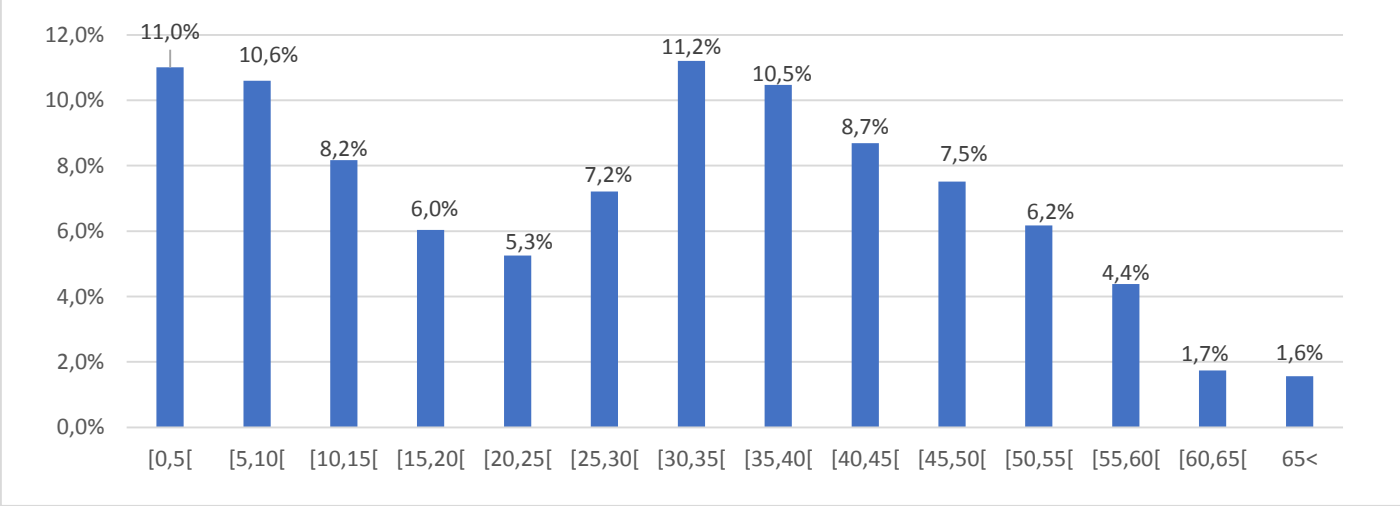


Figure 17 : Répartition selon tranche d'âge

On est dans le cadre d'un contrat « santé collective ». Ainsi, on constate que :

- Le pourcentage des personnes âgées supérieures à 50 ans (y compris des préretraités et retraités), de même que celles comprises entre 20 et 25 ans (les enfants à charge) sont moins importants que ceux dans les tranches d'âge entre 30 et 35 ans qui représentent 11,2 % du portefeuille et ceux de la tranche d'âge [0-5] avec un pourcentage de 11%.

À partir de la variable « Nom groupe » qui désigne le nom de l'entreprise souscripteur nous avons essayé de voir les caractéristiques liées à ce groupe en créant les variables

- « Secteur » qui prend deux modalités (Privé /public)

- « Secteur d'activité » qui prend 4 modalités : Finance, Commerce, Service, Industrie

- « Taille d'entreprise » en se basant sur le nombre d'employées. Elle prend 4 modalités :

<20, [20,100] ,[100,200], 200< .

d) Répartition du portefeuille par secteur public ou privé

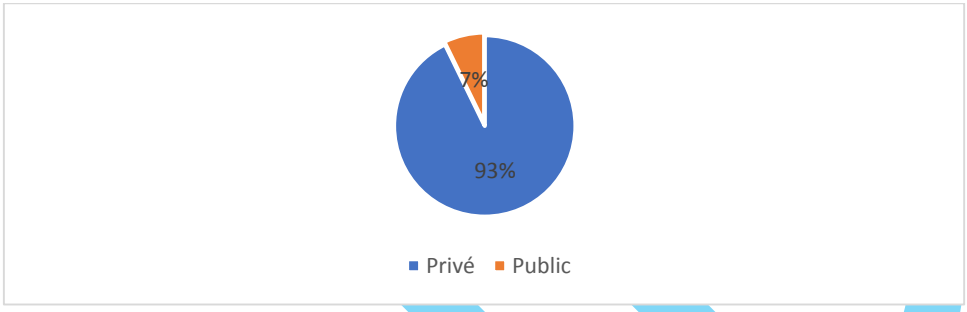


Figure 18:Répartition selon le secteur d'activité

Le portefeuille du GAT Assurances est composé de 153 216 adhérents répartis entre 401 entreprises réparties comme suit 7% des entreprises public et 93% entreprises privées.

e) Répartition du portefeuille selon la taille de l'entreprise

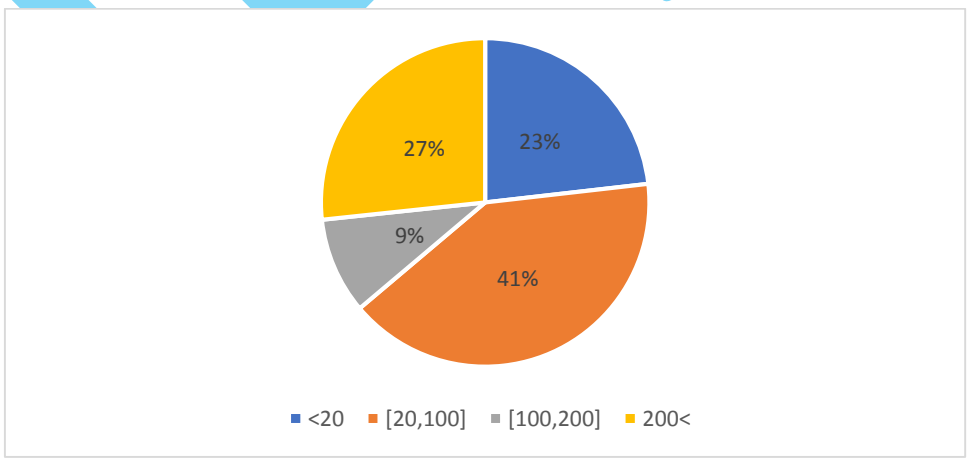


Figure 19 : Répartition selon la taille de l'entreprise

Nous distinguons que 41% des bénéficiaire du portefeuille sont des salariées des entreprises moyennes contenant entre 20 et 100 employées, et que 23% du portefeuille est composé par des microentreprises dont le nombre des salariées est inférieur à 20, et que 27 % du portefeuille est composé par des grands entreprises.

f) Répartition selon le secteur d'activité

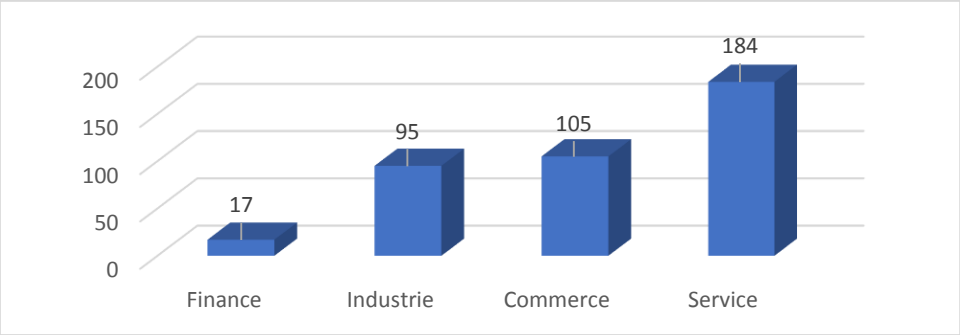


Figure 20 : Répartition des sociétés assurées selon le secteur d'activité

Pour les 401 entreprises composant notre portefeuille, on remarque que les plus grandes parts opèrent dans le secteur du service avec un nombre de 184 entreprises. Suivie par le secteur de commerce avec 105 entreprises et celle d'industrie avec 95 entreprises.

3.2.2 Analyse de la Base sinistre

La base prestation contient 656 675 lignes pour 85 931 bénéficiaires, une base riche en termes de nombre d'actes avec un montant total de prestation de 29 136 578,18

➢ **Création de la variable « Civilité Bénéficiaire » :**

Cette variable est créée en se basant sur les informations de 2 variables « CIV RESP » cette variable désigne la civilité de l'assuré principale (le salariés), et « Type bénéficiaire » créée de la manière suivante :

Conditions	**Civilité bénéficiaire**
Si (type bénéficiaire =RESP) alors	CIV (RESP) = {M, Mlle, Mme}
Si (type bénéficiaire =Conjoint) alors	Opposé (CIV(RESP)) = { M,Mme}
Si (type bénéficiaire = Enfant) alors	ENFT
Si (type bénéficiaire = AUTR) alors	AUTR

➢ **Création de la variable « SEXE Bénéficiaire » :**

La variable sexe est créée comme suit :

Dans un premier temps, nous avons défini une variable « CIV Bénéficiaire », elle peut nous renseigner sur le sexe du conjoint et l'assuré principale, il reste donc de trouver le sexe des ascendants et enfants. On a procédé comme suit : tout d'abord on a fait une extraction du tous les prénoms des ascendants et enfants présents sur la base et les séparer en prénoms masculins et féminins. La variable est créée de la manière suivante :

- Homme « M » si le prénom existe dans la liste des prénoms masculins

- Femme « F » si le prénom existe dans la liste des prénoms féminins.

Finalement la variable sexe prend 2 modalités : (F, M)

3.2.2.1 Analyse uni variée

a) La consommation VS la variable « Sexe »

Intéressons-nous maintenant à la répartition de consommation par sexe en négligeant les enfants

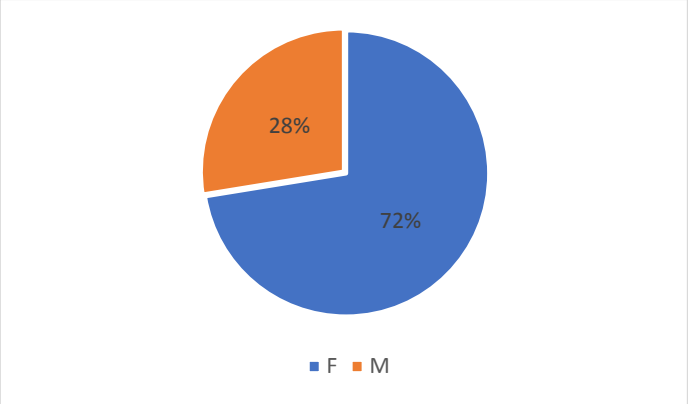


Figure 21 : Répartition de la consommation par sexe

Sur les 656 675 actes de consommation retenus pour cette l'étude, on dénombre 473 500 femmes et 183 175 hommes.

Tableau 7 : Répartition de la consommation par sexe

Sexe	Nombre d'actes	Montant remboursé	Coût moyen
F	473 500	20 303 201,2	42,8789887
M	183 175	8 833 377,01	48,2237042

Source : élaboré par nous même

On constate que les femmes consomment de manière générale plus de produits médicaux que les hommes avec un pourcentage de 72 % pour les femmes et 28 % pour les d'hommes.

Application du test de khi-deux

En se référant aux méthodes effectuées par Denuit & Charpentier II (2009), afin de sélectionner les variables du modèle, on opère une série de tests d'indépendance de Khi-deux. Pour cela, on choisit comme référence la variable" CONS", qui prenant la valeur 1 si l'individu étudié a consommé pendant l'exercice, et 0 s'il n'a pas consommé.

$$CONS \begin{cases} 1 & \text{Si } N > 0 \\ 0 & \text{Sinon} \end{cases}$$

Avec N est le nombre de sinistre durant une année.

Il est ainsi question d'examiner s'il existe une « corrélation » entre les différentes variables dont nous disposons et la variable CONS.

Pour ce test, on a préféré de retenir une variable indicatrice de consommation plutôt qu'une variable indiquant le nombre de sinistre d'un assuré pendant l'exercice car en croissant cette variable avec une autre, beaucoup d'effectifs seraient inférieurs à 5. Or, la condition d'application du test de Khi-deux exige que les effectifs doivent être supérieurs ou égaux à 5.

Soient les deux variables CONS et SEXE prenant chacune deux modalités. En croisant les fréquences de ces deux variables, nous obtenons le tableau de contingence suivant. (Les effectifs théoriques attendus sont notés entre parenthèses.)

Tableau 8 : tableau de contingence CONS et Sexe

CONS	Sexe		Total général
	Femme	**Homme**	
Pas de sinistre	65658 (63481)	21978 (24155)	87 636
Sinistre	473501 (475678)	183175 (180998)	656 676
Total général	539 159	205 153	744 312

Les hypothèses du test du χ^2 d'indépendance sont les suivantes :

— H0 : Les variables Sexe et CONS sont indépendantes

— H1 : Les variables Sexe et CONS ne sont pas indépendantes

En appliquant le test de Khi-2 sur le logiciel R sur ce tableau de contingence. On obtient la sortie suivante : > Pearson's Chi-squared test

X-squared = 306.98, df = 1, p-value < 2.2e-16

- **X-squared** : valeur classique renvoyée par un test de Khi2 permet de retrouver manuellement la p-value en s'aidant d'un tableau de χ^2
- **df :** degrés de liberté
- **p-value :** donne la probabilité de validation de H_0 –(la probabilité de ne voir aucun lien entre les critères). Plus p-value est petite, plus il y a un lien entre les critères (et donc pas d'indépendance).

Pour les variables Sexe et CONS, on a :

- la statistique de Khi-deux qui vaut 306.98, pour 1 degré de liberté.

- La p-value obtenue est très faible, inférieure à 2.2e-16, ce qui nous permet de rejeter l'hypothèse H_0 et penser qu'il y'a un lien entre la variable SEXE et la variable CONS

Conclusion :

Le SEXE est un critère important dans l'explication de la consommation

b) La consommation VS la variable « type bénéficiaire »

Le tableau suivant illustre la distribution des frais réels par type d'assuré

Tableau 9:Frais réels par type de bénéficiaire (DT)

Type d'Adhérent	**Frais réels**	**Pourcentage**
Ascendant (AUTR)	704 719,931	2,4%
Conjoint (CNJT)	7 731 276,97	26,5%
Enfant (ENFT)	8 260 362,28	28,3%
Adhérent principal (RESP)	12 485 154,2	42,8%

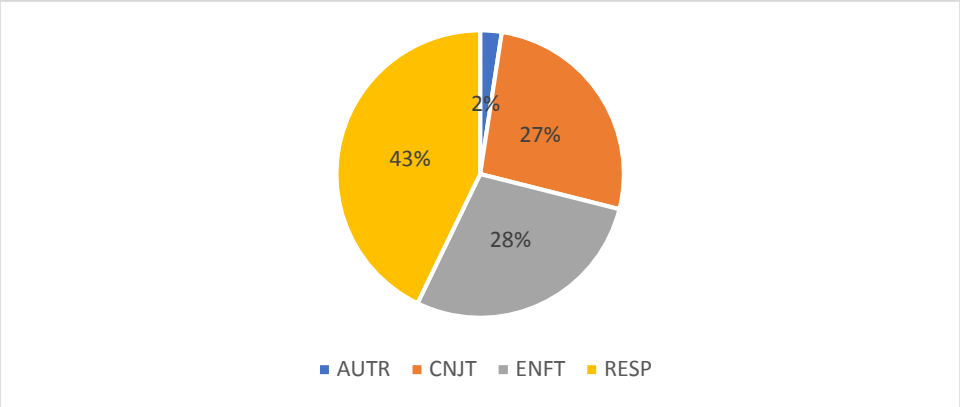


Figure 22 : Répartition du montant total remboursés par type de bénéficiaire

Selon le graphique, nous observons que le remboursement pour les assurés principaux (RESP) est plus important que ceux pour les autres types de bénéficiaires, en effet elle représente 42.8 % pour un montant de 12 485 154,2. Pour les conjoints et les enfants, le remboursement est assez équilibré, (repartie respectivement 27% et 28.3 % du portefeuille). Pour les ascendants bénéficiaires du contrat, la répartition demeure faible avec un pourcentage de 2.4%.

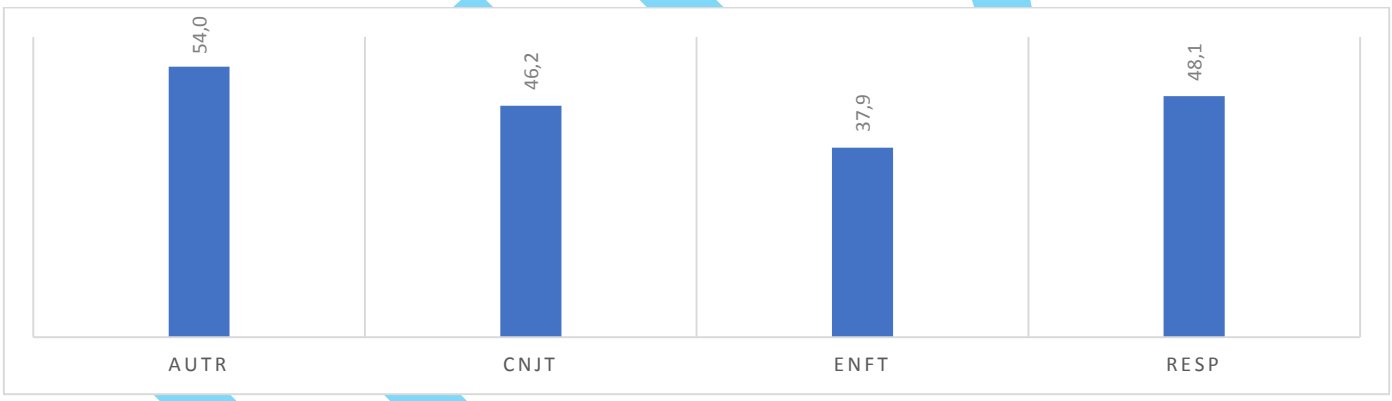


Figure 23 : Moyenne de consommation par type bénéficiaire

En moyenne de consommation, on remarque que les ascendants présentent une moyenne de consommation de 54 dt. Un montant très élevé par rapport aux autres bénéficiaires du contrat, et ceci est attendu puisque le nombre des ascendants sont très minimes par rapport aux assurés principales et conjoints qui domine le portefeuille.

	Type bénéficiaire				
CONS	**AUTR**	**CNJT**	**RESP**	**ENFT**	**Total général**
Pas de Sinistre	1557	22132	30941	33006	87636
Sinistre	13040	167136	217413	259087	656676
Total général	**14597**	**189268**	**248354**	**292093**	**744312**

Figure 24 : tableau de contingence CONS et type bénéficiaire

>Pearson's Chi-squared test

X-squared = 190.02, df = 3, p-value < 2.2e-16

Pour les variables type bénéficiaire et CONS la statistique de Khi-2 vaut 190.02, pour 3 degrés de liberté.

La p-value est très faible, < à 2.2e-16, ce qui permet de conclure que la qualité de bénéficiaire est un critère important dans l'explication de la sinistralité.

Conclusion : Le type de bénéficiaire est un critère important dans la détermination de la consommation.

c) La Consommation VS la variable « âge »

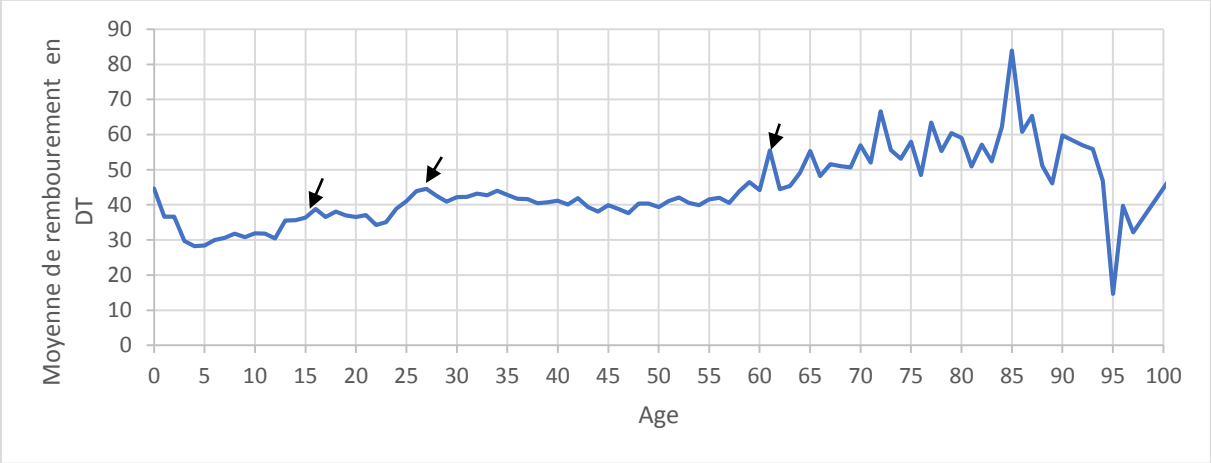


Figure 25 : Remboursements annuels moyens en fonction de l'âge

Ce graphique illustre que :

-Pour les très jeunes âges, une première forte consommation due principalement aux maladies infantiles. En effet, les enfants (très jeunes même) entre 0-10 ans représentent 20 % du portefeuille et consomment beaucoup en termes d'hospitalisation. Sachant que c'est la tranche d'âge où se développe le système immunitaire entraînant beaucoup de consultations et de vaccinations mais aussi parfois des problèmes à la naissance impliquant de longs et coûteux séjours à l'hôpital.

-Le pic du remboursement moyen à l'âge de 16 ans s'explique par des dépenses coûteuses en soins d'orthodontie et d'optique pour les adolescents. Et un autre pic pour les jeunes de tranche d'âge 20 et 39 ans dues principalement à la maternité et la stérilité, analyses et suivis gynécologiques ...

-Les dépenses de santé augmentent avec le vieillissement des assurés. En conséquence, les remboursements moyens des organismes assureurs suivent cette même tendance ce qui explique les dernières fortes croissances qui sont dû à l'augmentation des hospitalisations et de la consommation en pharmacie aux grands âges.

>Pearson's Chi-squared test

X-squared = 1378.8, df = 13, p-value < 2.2e-16

Pour les variables type bénéficière et CONS la statistique de Khi-2 vaut 1378.8, pour 13 degrés de liberté.

La p-value est très faible, < à 2.2e-16, ce qui permet de conclure que la variable âge est un critère important dans l'explication de la sinistralité.

Conclusion :

L'âge est un critère important dans l'explication de la consommation.

d) La consommation VS la variable « Collège professionnelle »

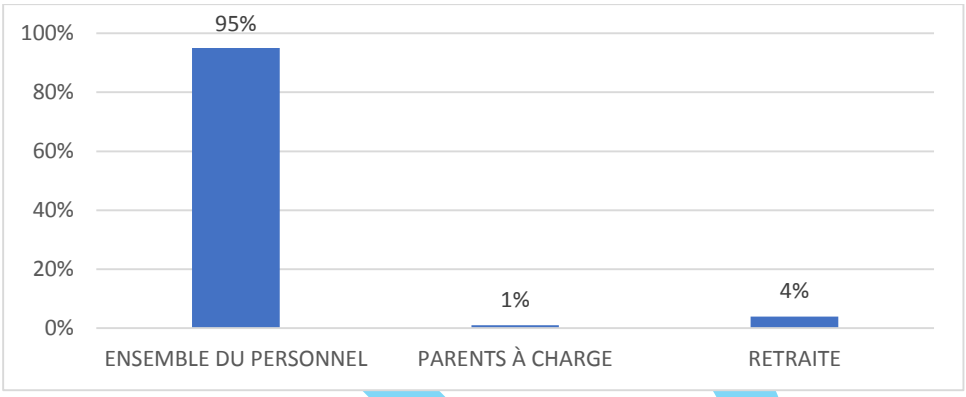


Figure 26 : Répartition par Collège professionnelle

D'après le graphique, le remboursement annuel pour les adhérents « Ensemble du personnel » est très élevé, elle représente 95 %. Le poids de remboursement pour les retraités et parents à charge n'est pas significatif, ils ne représentent que 4 % du montant totales de remboursement pour les retraités et 1% pour les parents à charges. Cela nous oriente sur plusieurs hypothèses telles que, la plupart des contrats santé sont souscrit sous la forme "Ensemble du personnel " et que les bénéficiaires des contrats sont généralement les enfants et le conjoint des salariées de l'entreprise.

>Pearson's Chi-squared test

X-squared = 94591 , df = 4, p-value < 2.2e-16

Pour les variables type bénéficière et CONS la statistique de Khi-2 vaut 94591, pour 4 degrés de liberté.

La p-value est très faible, < à 2.2e-16, ce qui permet de conclure que la variable collège socioprofessionnelle est un critère important dans l'explication de la sinistralité.

Conclusion :

Le collège socioprofessionnel est un critère important dans la variation de la consommation.

e) La consommation VS la variable « Secteur » :

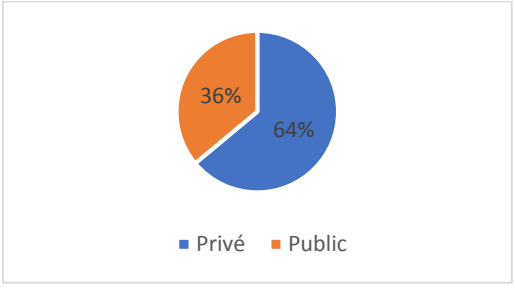


Figure 27 :Réparation par secteur Privé ou Public

Notre portefeuille est composé par 401 entreprises dont 7% sont des entreprises publiques et 93 % sont des entreprises privées. En faisant une projection sur leur niveau de consommation, cette figure montre que 36 % des montants remboursées sont affectées au secteur public pour un montant de 10 521 366 et les deux autres tiers (64%) sont affectées au secteur privé pour un montant de 18 660 147, ça fait un totale de 29 181 513 du montant remboursé pour l'année 2018.

>Pearson's Chi-squared test

X-squared = 70.096, df = 3, p-value = 4.071 e-15

Pour les variables type bénéficière et CONS la statistique de Khi-2 vaut 70.096, pour 3 degrés de liberté.

La p-value est très faible, =4.071 e-15, ce qui permet de conclure que la variable âge est un critère important dans l'explication de la sinistralité.

Conclusion :

La nature du secteur qu'il soit privé ou public est un critère important dans l'explication de la consommation.

f) La variable Taille d'entreprise VS consommation :

Taille entreprises	Montant de remboursement	Cout moyen	Nombre d'actes
[100,200]	4 784 808	40	120 964
[20,100]	5 908 121	48	122 413
<20	1 151 058	51	22 619
200<	17 292 591	44	390 679

Source : élaboré à partir de la base de données

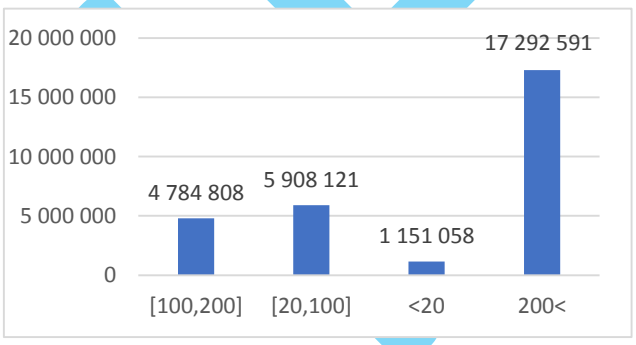


Figure 28 : Répartition du montant de remboursement

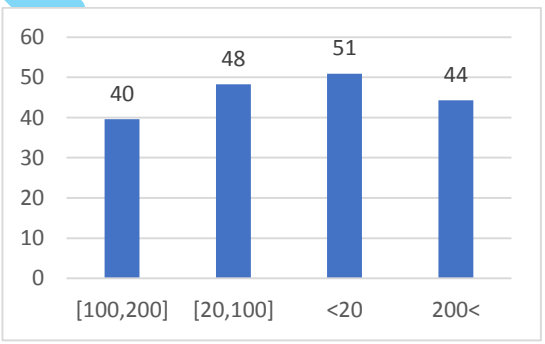


Figure 29 : répartition du cout moyen de remboursement

Rappelons que 27 % de notre portefeuille est composé par des salariées des entreprises de grand taille (200<employées) et qui présentent un nombre de consommation égale à 390 679 actes D'après cette figure, on voit clairement qu'elles présentent la part de remboursement annuelle la plus importante dans notre portefeuille par rapport aux PME. D'autre part, les microentreprises (<20), ont le coût moyen le plus important avec 51 dt

>Pearson's Chi-squared test

X-squared = 10373, df = 4, p-value < 2.2e-16

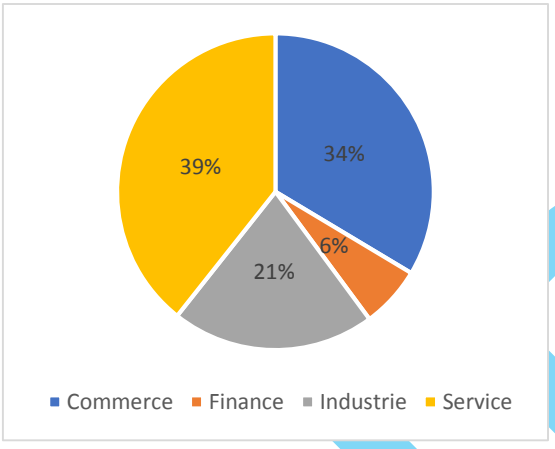
Pour les variables type bénéficière et CONS la statistique de Khi-2 vaut 10373, pour 4 degrés de liberté.

La p-value est très faible, < à 2.2e-16, ce qui permet de conclure que la variable taille entreprise est un critère important dans l'explication de la sinistralité.

Conclusion :

La variable taille d'entreprise est un critère important pour la détermination de la consommation.

g) La consommation VS la variable « secteur d'activité » :



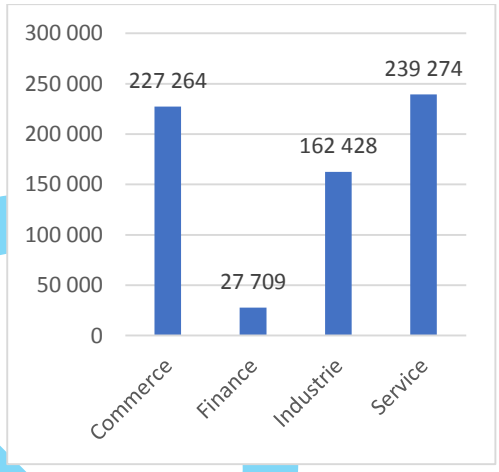


Figure 30 : Répartition de remboursement annuel par secteur d'activité Figure 31 : Répartition par nombre d'actes

Cette figure nous renseigne sur la différence du comportement de consommation selon la variable Secteur d'activité où nous constatons que les employés du secteur « Service » ont tendance à plus consommés (39% du montant remboursés) a l'opposé des employés du secteur « Financier » qui ont un faible nombre de consommation, 27709 actes expliquée peut-être par leur sous-représentation dans notre portefeuille (seulement 17 entreprises sur un total de 401 entreprises)

>Pearson's Chi-squared test

X-squared = 55.137, df = 1, p-value = 1.124 e-13

Pour les variables type bénéficière et CONS la statistique de Khi-2 vaut 70.096, pour 1 degré de liberté.

La p-value est très faible, égale à 1.124 e-13, ce qui permet de conclure que la variable âge est un critère important dans l'explication de la sinistralité.

Conclusion : La variable taille d'entreprise est un critère important pour expliquer la consommation

h) Répartition de remboursement par famille d'actes

La consommation médicale n'est pas homogène pour les types d'actes, les dépenses sont reparties en 6 catégories d'actes qui seront divisés en sous catégories d'actes comme le montre le tableau suivant :

Tableau 10 : Différents catégorie d'actes

Catégorie d'actes	Actes

Hospitalisation	Accompagnement, Accouchement, Anesthésiste, Chambre individuelle, Chirurgie, Frais chirurgicaux, Poche de sang, Médecin réanimateur,
Optique	Lentille, Monture, Vari lux, Verres, Verre progressifs
Dentaire	Orthodontie, Orthopédie, Prothèses dentaire, Prothèses dentaires, Soins dentaires, Appareillage
Pharmacie	Pharmacie Chronique, Pharmacie ordinaire, Pharmacie relative à la stérilité
Soins courants	Acte de prélèvement biologique, Analyses, Consultations, Échographie, Scanner, IRM, Visites médicales ….
Autres	Actes spéciaux, Circoncision, Couveuse, Endoscopie, Interruption volontaire de grossesse, Injections, Massage et rééducation, Réanimation lourde, Péridurale…

Nous avons reparti les prestations en famille d'actes de la manière suivante

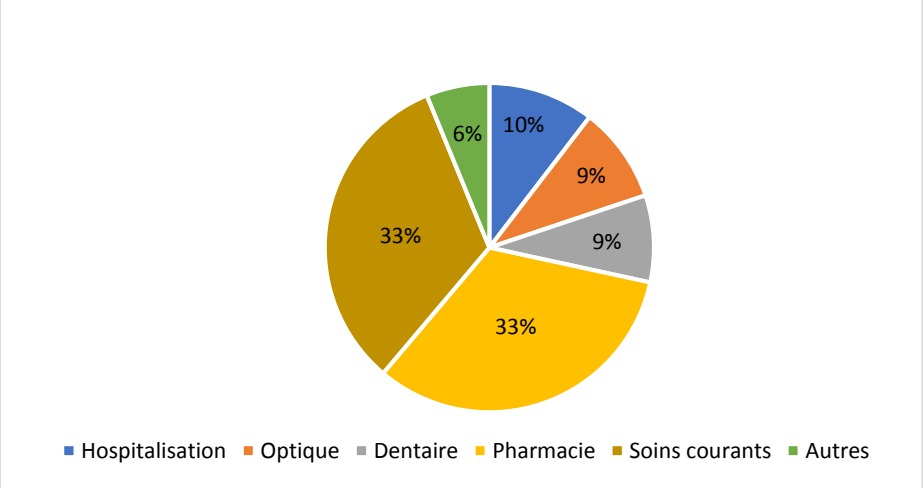


Figure 32 : Répartition par famille d'actes

Cette figure nous montre que les garanties « Pharmacie » et « Soins courants » coûtent plus chères pour l'assureur que les autres garanties. Elles présentent plus que 66% des remboursements au sein du notre portefeuille. En revanche, les garanties « Hospitalisation », « Optique » et « Dentaire » représentent respectivement 10%,9 % et 9% des remboursements annuels.

i) La consommation VS variable « Exposition » :

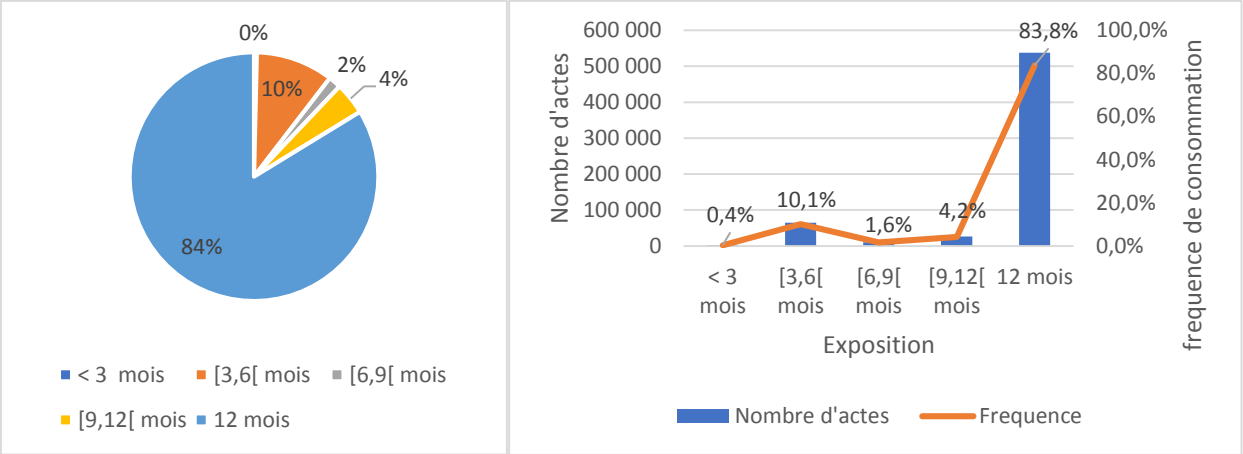


Figure 33 : remboursement selon la durée l'exposition

L'analyse de cette variable nous indique que 84 % de nos assurés ont une durée d'exposition d'une année, soit 12 mois, c'est à dire qu'ils ont été présents dans le portefeuille pendant toute la période d'étude. Seulement 16 % d'assurés ont adhéré pour une durée inférieure à 12 mois. La proportion importante de la population ayant adhérée toute la période d'observation laisse imaginer une bonne gestion du portefeuille. Les 16 % d'assurés présents sur moins de 6 mois s'expliquent soit par une sortie du contrat lors de la période d'observation ou à l'inverse d'une entrée sur la fin de la période.

3.2.2.2 Analyse Bi variée

a) Consommation VS « SEXE » et « Catégorie d'actes »

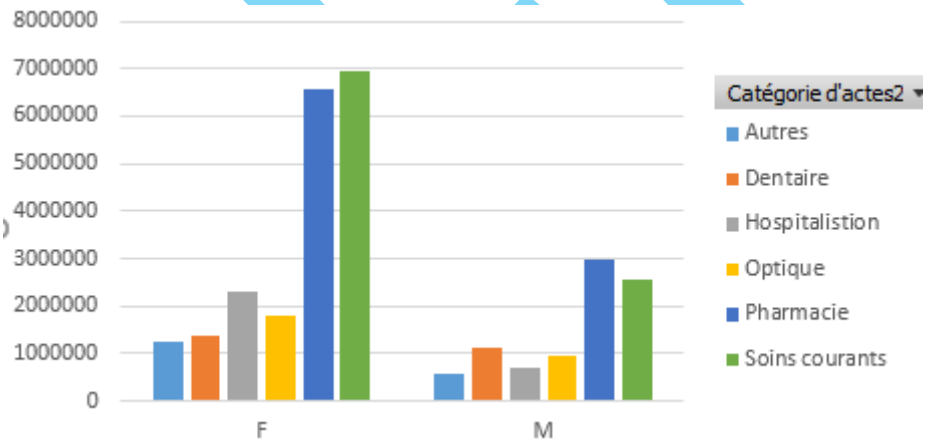


Figure 34:Répartition des remboursement annuels suivant le sexe et la catégorie d'actes

Suivant la figure, nous constatons que la plus forte différence de dépenses s'observe bien évidemment chez les femmes pour toutes les catégories d'actes vu qu'il y a des actes adressés uniquement aux femmes tels que l'accouchement, ce qui explique l'importante variation pour les soins courants, l'hospitalisation et pharmacie. En effet, cet acte engendre des analyses et visites des médecins spécialistes et médicaments. Toutefois, et de façon générale, pour les postes dentaires on remarque une légère variation par rapport au sexe masculin. L'impact de la variable sexe sur le coût de sinistres sera certainement constaté lors de l'étape de modélisation.

b) Consommation VS « Qualité bénéficiaire » et « Catégorie d'actes »

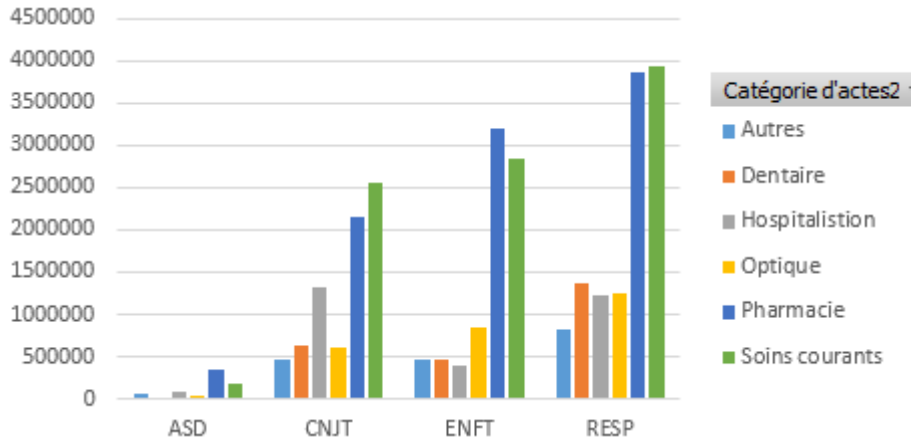


Figure 35:Répartition des coûts par statut bénéficiaire et par catégorie d'actes

En outre, si l'on compare la consommation des assurés par rapport à leur statuts (qualité bénéficiaire), nous observons une très forte consommation de la part des assurés principaux (RESP) des actes pharmacie et soins courants, qui dominent avec un énorme écart par rapport aux autres catégories de bénéficiaires. Les dépenses de santé chez les enfants sont souvent inférieures, mais pas très éloignées des assurés principales (RESP) et conjoints. Logiquement, les enfants ont une tendance à attraper plus de virus ou de maladies, ce qui implique un important nombre de consultations chez le pédiatre ou d'hospitalisations.

c) Age et Catégorie d'actes VS consommation moyenne

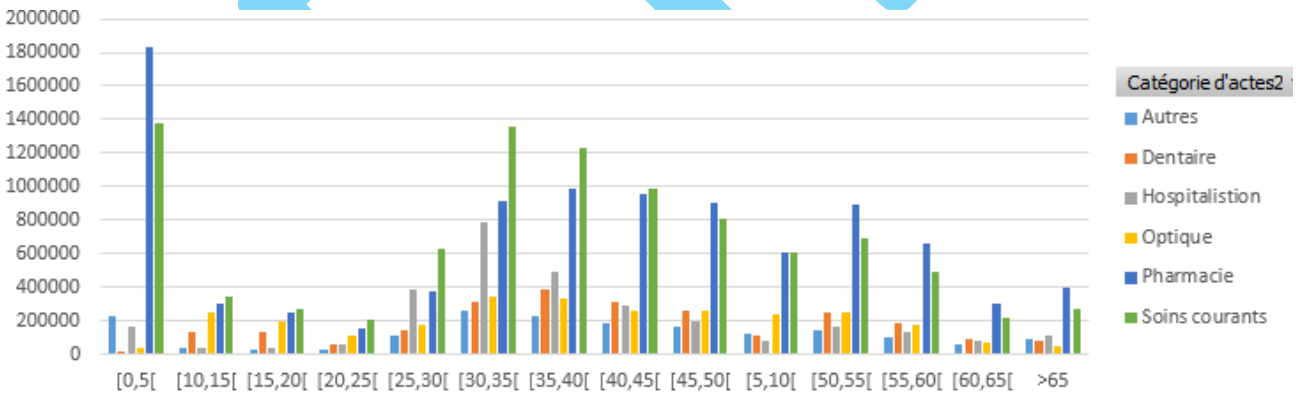


Figure 36: Répartition par âge et par catégorie d'acte

L'impact de l'âge est très important par rapport aux consommations des frais de santé. En effet, et en liaison avec notre conclusion des points précédents, les enfants (très jeunes même) entre 0-5 ans consomment beaucoup en termes de pharmacie et soins courants. Sachant que c'est la tranche d'âge où se développe le système immunitaire entraînant beaucoup de consultations et de vaccinations mais aussi parfois des problèmes à la naissance impliquant de longs et coûteux séjours à l'hôpital. Finalement, l'hospitalisation telle qu'on pourrait l'imaginer engendre des dépenses qui s'effectuent entre 25 et 35 ans et la tranche des 30-35 ans étant la plus forte à cause de la maternité.

d) SEXE et Secteur d'activité VS consommation :

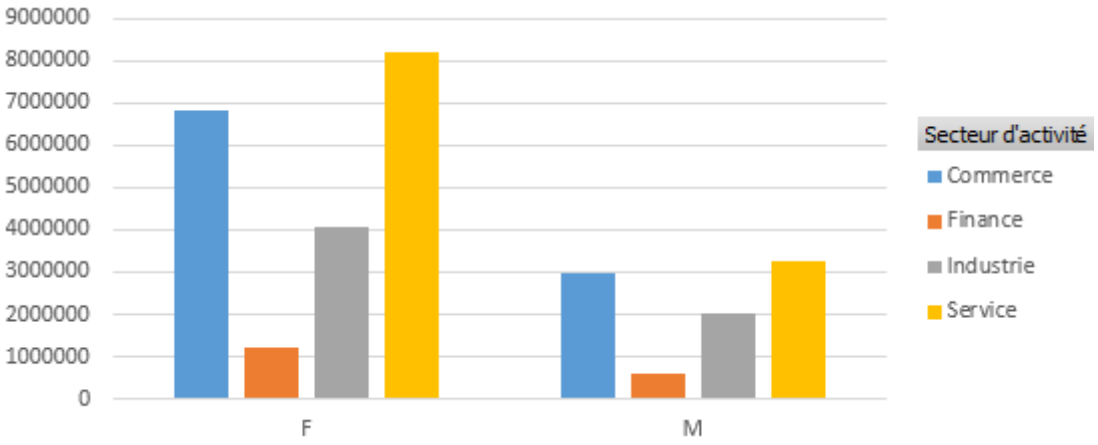


Figure 37 : Sexe et secteur d'activité vs consommation annuelle

L'analyse de croisement entre les deux variables sexe et secteur d'activité, montre clairement que la part de consommation des bénéficiaires de sexe féminin est plus élevée que celui de sexe masculin dans tous les secteurs d'activité, et que cette différence est très flagrante surtout pour le secteur d'activité service.

Conclusion : La consommation dépend simultanément des variables sexe et secteur d'activité.

e) Secteur d'activité et Catégorie d'entreprise VS consommation :

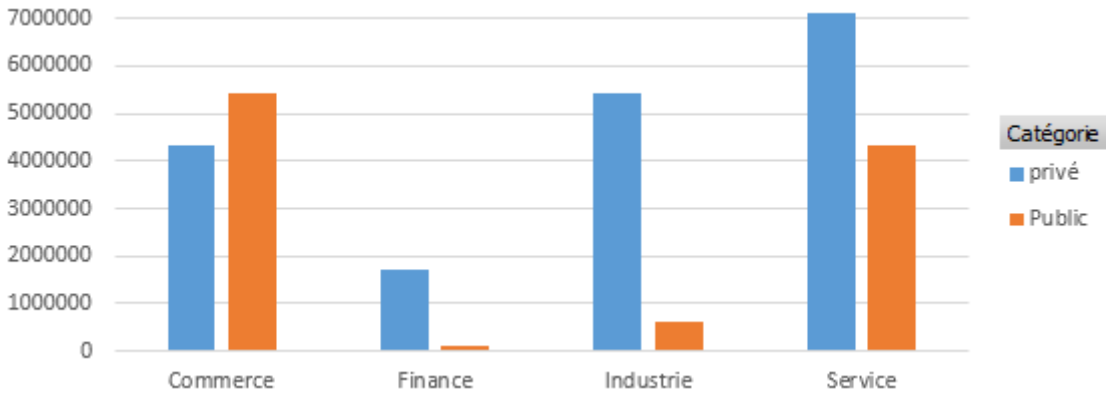


Figure 38:Secteur d'activité et catégorie d'entreprise vs consommation

Ce graphique monte que quel que soit le secteur d'activité de l'entreprise, les sociétés privées consomment plus que les sociétés publiques, et que la différence de consommation est flagrante surtout pour le secteur d'industrie et Finance.

Conclusion : Les variables Secteur d'activité et Catégorie d'entreprise interviennent simultanément pour expliquer la consommation.

Conclusion :

Dans ce chapitre, on a décrit d'une manière synthétique le système Tunisien d'assurance santé et ses différents acteurs, tout en évoquant l'insuffisance de la CNAM en matière de remboursement, ce qui a engendré l'augmentation de dépenses de santé mis à la charge des ménages, d'où le recours à l'assurance maladie pour pallier ces insuffisances. On a notamment illustré le rôle que joue

l’assureur au sein de ce système et les différents types de contrats offerts par l’assurance santé privée.

Dans la deuxième partie du chapitre on a présenté la compagnie de parrainage et son positionnement dans le marché de la santé à travers d’une analyse statistique du marché d’assurance groupe maladie et à travers quelques chiffres clés.

À la fin du chapitre on a entamé une analyse descriptive du portefeuille de GAT à partir d’une analyse uni variée et une analyse bi variée.

Or dans la pratique et lors de souscription d’un contrat assurance maladie groupe collectif, toutes les variables interviennent simultanément, par la suite, une analyse binaire ne suffit pas, il faut une analyse multivariée. D’où la nécessité d’un instrument statistique qui permet de relier la variable à expliquer qui est le coût et la fréquence à diverses variables explicatives, et qui permet de mesurer son impacte. À ce propos, nous avons prouvé que pour tells analyses, les modèles linéaires GLM sont les instruments les plus adaptées pour cette étude.

CHAPITRE TROIS :

TARIFICATION D'UN CONTRAT ASSURANCE MALADIE (APPLICATION EMPIRIQUE : CAS DU GAT)

Chapitre trois : Tarification d'un contrat assurance maladie ((Application empirique : cas du GAT)

Introduction

L'activité d'assurance est caractérisée par un cycle de production inversé : en contrepartie d'une prime d'un montant connu à la souscription du contrat, l'assureur s'engage à couvrir un risque de montant inconnu dont il ignore la date de réalisation. La tarification de l'offre d'assurance consiste à évaluer la prime nécessaire pour couvrir les engagements et les frais de fonctionnement de l'assureur. Le prix de l'assurance est appelé prime commerciale, qui est déterminé en fonction de la prime pure et des frais de fonctionnement.

La prime pure est la part de la prime commerciale qui couvre les engagements de l'assureur vis à vis les assurés. Ce paramètre représente également le coût futur des risques et, par conséquent, l'assureur est invité à calculer la prime pure correspondante à chaque groupe d'assurés. C'est la raison pour laquelle, l'assureur est amené 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 même frais et la même distribution, la compagnie la moins segmentée court le risque d'anti sélection.

La tarification constitue le cœur métier de l'actuariat, et dans un marché concurrentiel, les compagnies d'assurances doivent considérer des nouvelles méthodes de tarification rigoureuses qui prend en compte les caractéristiques individuelles d'une personne ou son besoin en matière de niveau de garanti. Or dans un groupe les individus ne sont pas tous égaux face aux risques, certaines personnes sont exposées plus aux risques que des autres. Le sens de crédibilité consiste à bien mesurer cette variation et proposer le tarif le bien approprier à chaque groupe d'assuré en tenant compte de ses caractéristiques.

La méthode de tarification en assurance Non vie la plus répandues actuellement par un grand nombre d'assureurs est la méthode de tarification « Fréquence × Coût » développée avec les modèles linéaires généralisés (GLM).

Nous étudions dans ce chapitre, la tarification de l'acte pharmacie ordinaire Notamment, on se penche sur les modélisations des variables à expliquer qui sont le coût des sinistres remboursés par « Assureur » et la fréquence annuelle, ainsi que les paramétrages nécessaires à la prise en compte des variables explicatives.

À cet effet, la première section illustre le champ sur lequel on va faire notre étude et le choix de l'acte à modéliser, la deuxième section sera conservée à la modélisation de la fréquence en faisant appel aux modèles linéaires généralisés comme étant une approche justifiée pour élaborer le tarif de la branche assurance maladie. La troisième section est réservée à la modélisation des coûts moyen.

Section 1 : La tarification d'un acte classique : La pharmacie ordinaire

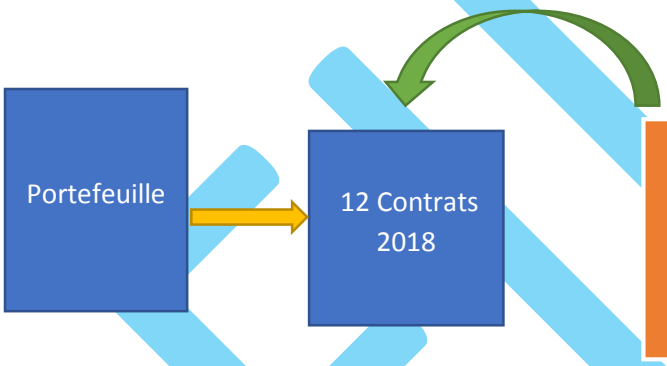
Vue que cette garantie est souscrite par environ 31,14% des assurés du portefeuille du GAT on l'a choisi pour estimer sa prime pure.

Tous les contrats n'ont pas été conservés pour cette étude. Au final, ce sont seulement 12 contrats collectifs qui sont retenus pour l'étude, avec 15 433 bénéficiaires, avec une moyenne de 1 285 bénéficiaires par contrat, et un montant total remboursé par GAT de 1 850 927 DT, une moyenne de 37,6 Dt par bénéficiaires.

Sur la base de ces contrats groupes nous obtenons au final un modèle optimal, noté GLM_op Les coefficients du modèle sont ensuite appliqués sur la base restante. Si le modèle ajuste bien les données restantes, nous pouvons conclure que ce modèle a une bonne consistance.

Après la validation du modèle, nous allons l'appliquer sur la totalité du portefeuille. Enfin, nous obtenons les coefficients qui sont les coefficients définitifs de ce modèle.

1.1-Extraction d'un échantillon



Dessin	Opération
	Extraction de données
	Figer le modèle et l'appliquer er sur une autre base ciblée

Le principe consiste à faire une extraction de sous base qui ne contient que les contrats choisis. Ensuite nous appliquons le modèle sur cette sous base et analysons chaque variable. Si ce modèle s'ajuste bien par rapport à la sous base, nous pouvons alors conclure qu'il est stable.

La validation consiste à reprendre l'estimation sur un nombre important d'individu, par exemple 15 000 et garder comme témoin la population restante 433 personnes.

Suite à cela, il s'agit de :

- Conclure pour la population témoin les coûts estimés à travers le modèle GLM.
- Comparer les coûts estimés au coût réel garder en témoin.

La validation sera retenue si l'écart entre les deux vecteurs co ût estimés et cout réel est faible.

1.2. Traitement des sinistres graves

Avant de commencer la modélisation, les études de tarification consistent à distinguer les sinistres graves du reste des sinistres. Cela permet ainsi d'éviter quelques gros sinistres influencent trop le calcul des coefficients et les indicateurs renvoyés par le modèle.

En pratique, le gestionnaire doit segmenter les sinistres après leurs évaluations en deux types. À savoir les sinistres "typiques" (classe majoritaire) et les sinistres "atypiques" et cela en se basant sur un seuil de gravité bien précis. Pour appliquer cette division, on utilise la technique de « nuage de point » pour détecter les valeurs aberrantes.

Cette technique sera appliquée sur les charges des sinistres pour toute l'année 2018 en vue de détecter les sinistres graves

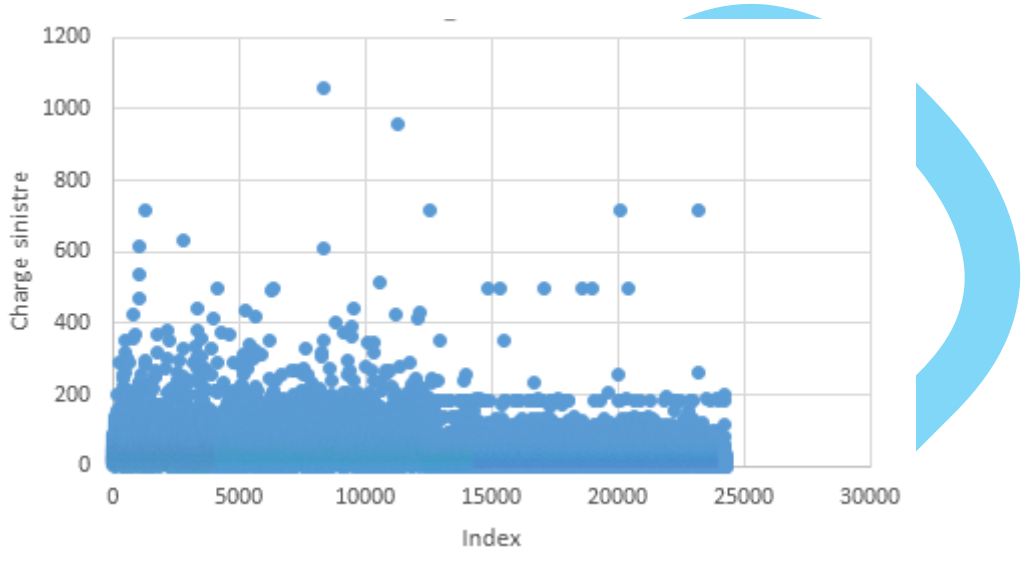


Figure 39 : Nuage de point des montants remboursées :

Nous remarquons d'après le nuage de point qu'au-dessus de 400 DT la dispersion de la charge des sinistres augmente fortement.

Nous considérons alors 400 DT comme seuil de gravité et par la suite les sinistres dépassant ce seuil sont classés "atypiques" pour ne pas biaiser les résultats u Modèles Linéaire Généralisé.

Nous passons d'une base qui contient 49 274 lignes à une base qui contient 49 212 lignes pour l'acte pharmacie ordinaire.

Section 2 : Modélisation de la fréquence de consommation

D'après la littérature, Denuit & Charpentier II (2009), nous disposons de deux lois statistiques discrètes qui peuvent éventuellement modéliser la fréquence de consommation en assurance Non-vie : la loi de Poisson et la loi Binomiale-Négative. Cette même référence indique que les évènements de comptage sont généralement modélisés par une loi de Poisson qui suppose une équidispersion (variance=moyenne) des données. Cependant, cette caractéristique n'est pas forcément compatible avec les données de fréquences de consommation en assurance santé. Nous serons ainsi amenés à utiliser une loi négative binomiale. Un test de " Kolmogorov-Smirnov" permettra de vérifier notre hypothèse.

La modélisation de la fréquence est effectuée avec les Modèles Linéaires Généralisés. On commence tout d'abord à traiter l'adéquation des lois empiriques avec les lois théoriques, et cela se fait en se basant sur trois critères, La démarche retenue est la suivante :

- Critères Espérance-Variance
- Critère graphique
- Critères des tests non paramétriques

Une fois que l'on a choisi la loi et on a estimé ses paramètres, on applique finalement la modélisation retenue par les Modèles Linéaires Généralisés.

2.1 Choix entre les deux lois

2.1.1 Critères Espérance-Variance

On désigne par N_i une variable de comptage dont on cherche la loi, qui désigne le nombre de consommation déclaré par l'assuré i durant l'année 2018, i = {1, 2,..., n} .On choisira :

- La loi de Poisson si E(N) = Var(N)
- La loi Binomiale Négative si E(N) < Var(N).

Le tableau ci-dessous contient les fréquences empiriques pour la variable comptant le nombre de consommation pour l'acte « pharmacie ordinaire »

Tableau 11 : Distribution du nombre de consommations du poste « pharmacie ordinaire »

Nombre de consommation N	Effectif	Fréquence
0	0	0
1	5171	0,33506123
2	3204	0,20760708
3	2080	0,13477613
4	1464	0,09486166
5	1086	0,07036869
6	705	0,04568133
7	499	0,03233331
8	397	0,0257241
9	266	0,01723579
10	179	0,01159852
11	118	0,00764595
12	84	0,00544288
13	58	0,00375818
14	35	0,00226787
15	26	0,0016847
16	15	0,00097194
17	12	0,00077755
18	10	0,00064796
19	6	0,00038878

20	6	0,00038878
21	2	0,00012959
22	2	0,00012959
23	4	0,00025918
25	1	6,4796E-05
26	1	6,4796E-05
28	2	0,00012959
Total	15433	1

Nous remarquons que :

- La distribution de la fréquence qui ne comporte pas des observations est nulle et cela est cohérent avec la nature des postes médicaux à forte consommation comme l'acte pharmacie.

-La fréquence de consommation diminue avec l'augmentation de la valeur de l'observation pour atteindre des valeurs presque nulles ou nulle pour les observations maximales.

Tableau 12 : Statistiques descriptives du nombre de consommation (pharmacie ordinaire)

Moyenne	Variance	Écart-type	Valeur max de N (Sur tout l'échantillon
3.192186	7.466432	2,73	28

La moyenne de nombre de consommation pour l'acte pharmacie ordinaire est de l'ordre de deux actes avec un écart-type d'environ 2.73. Il s'agit ici d'une variable décrivant des valeurs très dispersées, puisque l'écart-type est important et le nombre maximum d'acte par personne sur la période observée atteint le nombre de 28 actes.

D'après ce critère, on devrait choisir la loi Binomiale Négative pour estimer notre fréquence de consommation. (E(N) < Var(N))

2.1.2 Critère graphique

Afin de juger l'ajustement de ces distributions aux modèles classiques de comptage, nous allons d'abord estimer les paramètres de ces modèles par maximum de vraisemblance.

Ensuite nous visualisons le modèle obtenu et le comparons graphiquement aux données empiriques

a) Estimation de paramètre de la loi de poisson

On teste dans un premier temps l'ajustement à la loi de Poisson, loi la plus classiquement utilisée pour modéliser des fréquences de sinistre.

- La fonction "fitdistr" sous R nous permet d'obtenir le paramètre $\lambda = 3.19218558$.

- La fonction "goodfit" du logiciel , fournit un bon indicateur visuel de l'écart entre la loi théorique et la loi empirique pour des variables de comptage.
- La fonction "plot ()" de , nous permet d'obtenir une vision graphique, elle permet d'afficher sur le même graphique les valeurs observées et les valeurs prévues par la distribution de la loi avec laquelle on teste l'adéquation. Ainsi on obtient le graphe suivant :

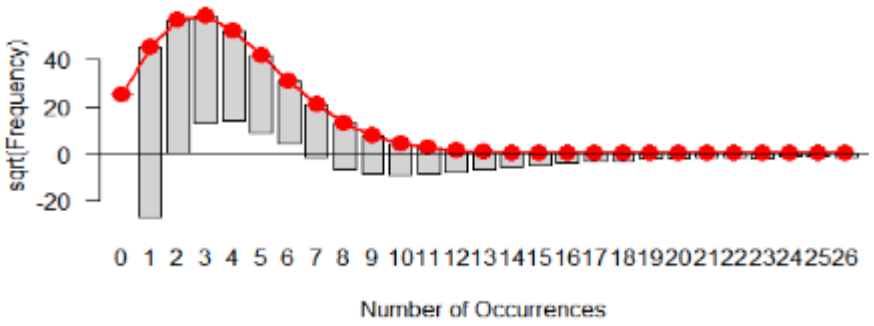


Figure 40 : Ajustement des fréquences de pharmacie ordinaire par une loi de Poisson

Interprétation :

On interprète le graphique de la manière suivante : les points en rouge 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 un mauvais ajustement des observations par la loi théorique. On remarque ainsi une :

-surestimation de la probabilité d'avoir un nombre de consommation compris entre 3 et 6.

-sous-estimation considérablement de la probabilité d'avoir 1 seule acte de consommation et les probabilités d'avoir un nombre de consommation compris entre 7 et 26.

On constate par la suite que l'ajustement par la loi poisson n'est pas adapté et s'ajuste mal aux données à cause des écarts observés et la présence de plusieurs mauvais ajustements graphiques et cela quel que soit le nombre d'occurrences.

b) Estimation des paramètres de la loi binomiale négative

L'estimation des paramètres de la loi Binomiale-Négative par la méthode de maximum de vraisemblance sur nous fournit les résultats suivants :

Size	mu
3.41464274	3.19215095
(0.07614197)	(0.02000487)

La fonction "plot () "de nous permet d'obtenir le graphe suivant :

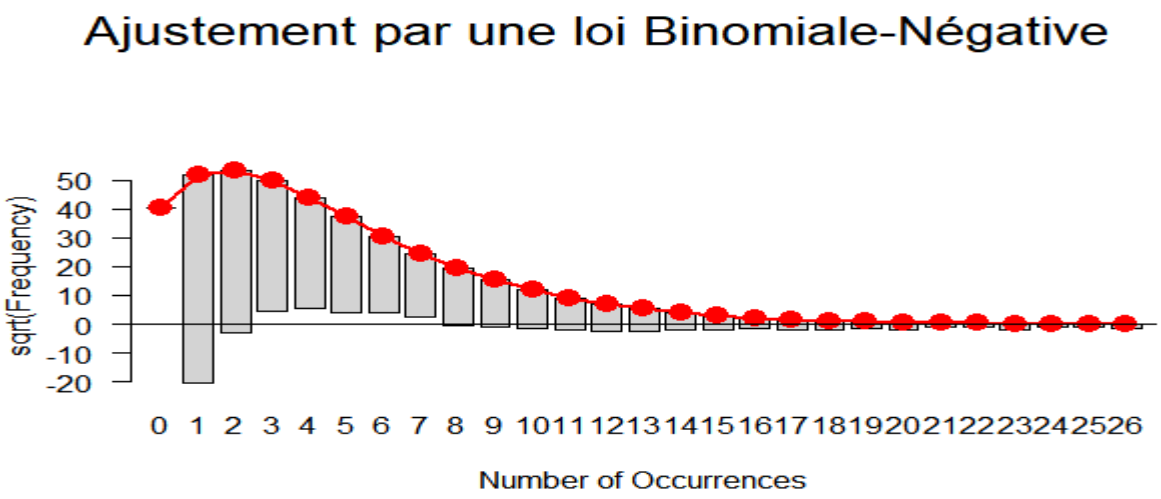


Figure 41 : Ajustement des fréquences de pharmacie ordinaire par une loi binomiale négative

Interprétation :

On remarque que :

-Les modalités de fréquences, 8 et 9 sont correctement estimées.

-Une légère surestimation des modalités 3 ,4,5 et 7 du nombre de consommations.

-Il existe néanmoins des sous-ajustements des fréquences 1,2 , et celles supérieures à 10 est marquée par de légères sous-estimations.

Ce modèle est beaucoup plus satisfaisant que le modèle de Poisson et cela sera confirmée par les résultats du test de Kolmogorov-Smirnov.

1.1.3 Critères des tests non paramétriques (test Kolmogorov-Smirnov) :

Le test de Kolmogorov-Smirnov est un test utilisé pour déterminer si un échantillon suit bien une loi donnée connue par sa fonction de répartition continue.

C'est un test d'ajustement non paramétrique car il vise à vérifier si les données observées sont compatibles avec un modèle théorique donné. Son principe est simple, on mesure l'écart maximum qui existe entre la fonction de densité cumulée observée F et la fonction de répartition théorique F_0 en posant 2 hypothèses : Les hypothèses du test sont les suivantes :

-H0 : $F = F_0$ Les données sont bien ajustées avec la loi théorique.

- H1 : $F \neq F_0$ les données ne sont pas ajustées avec la loi théorique.

Le test réalisé sur avec la fonction « chisq.test » affiche :

Tableau 13 : tableau récapitulatif du test de Kolmogorov- Smirnov

	Loi de Poisson	**Binomiale négative**
D	D = 0.6923	7.2308
P-value	7.465e-05	< 2.2e-16

Les résultats des tests non paramétriques de Kolmogorov-Smirnov appuient le résultat graphique. Nous obtenons une p-value très faible (p-value < 0.05) pour le modèle de Poisson. Cela signifie le rejet de l'hypothèse qui stipule que les données sont ajustables à une loi de Poisson, et c'est avec une confiance de plus de 95 % que nous choisissons la loi Binomiale négative pour modéliser la fréquence de consommation.

2.2. Estimation des coefficients de la régression Généralisée

Après la validation du choix de la loi suivie par notre fréquence de sinistre, on passe à l'estimation des coefficients attribués à chaque modalité de chaque variable.

La fréquence de consommation est modélisée par des variables explicatives telles que : le sexe, la tranche d'âge, le collège, le type de bénéficiaire la durée d'exposition…. Une variable contient plusieurs modalités. Par exemple, le variable « sexe » contient 2 modalités telles que femme et homme. Pour chaque variable, nous allons prendre une modalité de référence.

La fréquence est exprimée sous la forme :

$$\begin{bmatrix} H & \beta_{s1} \\ F & \beta_{s2} \end{bmatrix} + \begin{bmatrix} Resp & \beta_{t1} \\ Conj & \beta_{t2} \\ Enft & \beta_{t3} \\ ASD & \beta_{t4} \end{bmatrix} + \begin{bmatrix} [0.5[& \beta_{A1} \\ \vdots & \vdots \\ > 65 & \beta_{A16} \end{bmatrix} + \begin{bmatrix} EP & \beta_{c1} \\ R & \beta_{c2} \\ PA & \beta_{c3} \end{bmatrix} + \begin{bmatrix} Privé & \beta_C \\ Public & \beta_C \end{bmatrix}$$

Sexe type bénéficiaire Age Collège Secteur

À partir de ce modèle, on construit une matrice des variables explicatives X de la taille n × (p -1) avec :

- n est le nombre d'individu
- p−1 est le nombre de p variables explicatives et un « coefficient constant » dont les valeurs sont égales à 1.

Le modèle présenté ci-dessus ne peut être estimés sous cette forme en raison de la multi linéarité entre les variables qualitatives. Pour cette raison, nous estimons le modèle avec constante tous en retenant une modalité de référence pour chaque variable.

Le coefficient constant ou « Intercept » nous permet d'exprimer le coût et la fréquence à partir des modalités de référence.

On peut choisir la modalité de référence sous R avec la fonction "relevel(x, ref, ...)". En cas d'absence de choix, les modalités seront choisies arbitrairement par R.

En ajoutant cet élément, le prédicteur linéaire que nous retenons pour l'étude est quelque peu différent de ce que l'on a présenté dans la partie précédente. Il devient :

$$\eta = X\beta = \beta_0 + \sum_{j=1}^{p} xj\beta j$$

Avec $X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nn} \end{bmatrix}$ et $\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}$

Les valeurs de la matrice explicative X ne prennent que la valeur 0 ou 1 selon les valeurs des variables explicatives correspondantes. Autrement dit :

- Pour un individu possédant les mêmes modalités des modalités de référence, le vecteur servant à le caractériser est égal à 0.
- Pour un individu possédant une modalité différente de la modalité de référence le vecteur servant à le caractériser est égal à 1.

Prenons l'exemple d'un individu possédant les modalités suivantes :

-Sexe : Homme (modalité normale)

- Age : [0,5[ans **(modalité de référence)**

- Collège : Ensemble du personnel (modalité normale)

- Bénéficiaire : AUTR (**modalité de référence)**

Le vecteur explicatif est modélisé par :

Sexe Collège

x= (1 , 1 , 0 , 1 , 0)

Constante Age Bénéficiaire

Le vecteur des paramètres est modélisé par :

Sexe Collège

β = (β_0 , β_1 , β_2 , β_3 , β_4)

Constante Age Bénéficiaire

Le prédicteur linéaire sera :

$$\eta = X\beta = \beta_0 + \beta_1 + \beta_3$$

Le coefficient β_j est un coefficient d'aggravation de consommation, par la suite on peut supposer deux cas :

—Si $\beta_j > 0$ il y'a une surconsommation par rapport à l'individu de référence.

— Si $\beta_j < 0$ cela indiquera un facteur améliorant la sinistralité par rapport à l'individu de référence.

2.3. Estimation de la Fréquence

Dans cette partie du rapport, nous allons estimer les paramètres du modèle par la méthode de maximum de vraisemblance. L'estimation est réalisée sous le logiciel à l'aide de la fonction « glm.nb». Les sorties du logiciel sont les suivants à l'aide de la fonction « summary» :

Dans notre cas, Un Individu de référence est choisi comme étant celui qui représente les caractéristiques suivantes :

- **Sexe** : Femme,
- **Collège** : ensemble du personnel
- **Tranche d'âge** : [0,5[,
- **Bénéficiaire** : AUTR,
- **Secteur** :Public,
- **Taille entreprise** : 200<
- **Plafond acte** :]400 ;500],
- **Plafond annuel** :]2000, 3000]
- **Situation famille** : Famille

```
 Call:
glm.nb(formula = Nombredesinistre ~ plfondacte + plafond.annuel +
    Sexe + bénéficiaire + Tranche.age + Situation.famille + Secteur +
    Collège + Taille.entreprise + offset(log(Exposition)), data = phor,
    init.theta = 4.137330431, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0288  -0.8821  -0.3400   0.4234   4.5083

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)                0.99359    0.11408   8.709  < 2e-16 ***
plfondacte<400             0.15105    0.05089   2.968 0.002999 **
plfondacte]500;600]        0.25415    0.05752   4.418 9.95e-06 ***
plfondacte]600;700]       -1.67008    1.11917  -1.492 0.135633
plfondacte]700;800]        0.52695    0.06115   8.617  < 2e-16 ***
plafond.annuel=<2000       0.12184    0.05107   2.386 0.017049 *
plafond.annuel]3000;4000]  0.15881    0.06141   2.586 0.009711 **
plafond.annuel]4000;5000]  1.93805    1.11840   1.733 0.083116 .
SexeM                     -0.10594    0.01330  -7.966 1.64e-15 ***
bénéficiaireCNJT           0.06891    0.04676   1.474 0.140551
bénéficiaireENFT           0.09221    0.06324   1.458 0.144773
bénéficiaireRESP           0.21392    0.04635   4.615 3.93e-06 ***
Tranche.age[5,10[         -0.45023    0.02569 -17.526  < 2e-16 ***
Tranche.age[10,15[        -0.62227    0.02931 -21.229  < 2e-16 ***
Tranche.age[15,20[        -0.60932    0.03274 -18.611  < 2e-16 ***
Tranche.age[20,25[        -0.58050    0.04339 -13.377  < 2e-16 ***
Tranche.age[25,30[        -0.27330    0.05880  -4.648 3.35e-06 ***
Tranche.age[30,35[        -0.30145    0.05298  -5.690 1.27e-08 ***
Tranche.age[35,40[        -0.33830    0.05108  -6.623 3.51e-11 ***
Tranche.age[40,45[        -0.27365    0.04969  -5.507 3.64e-08 ***
Tranche.age[45,50[        -0.24559    0.04949  -4.963 6.95e-07 ***
Tranche.age[50,55[        -0.24447    0.05200  -4.701 2.59e-06 ***
Tranche.age[55,60[        -0.18821    0.05422  -3.471 0.000518 ***
Tranche.age[60,65[        -0.18399    0.08322  -2.211 0.027045 *
Tranche.age>=65           -0.06541    0.07737  -0.845 0.397847
Situation.familleDuo       0.05861    0.03238   1.810 0.070305 .
Situation.familleIsolé     0.01809    0.04385   0.413 0.679932
Situation.familleTrio      0.08945    0.02023   4.422 9.78e-06 ***
SecteurPublic              0.25329    0.05125   4.942 7.71e-07 ***
CollègeACTIFS             -1.32087    1.11829  -1.181 0.237540
CollègeRETRAITE           -1.35132    1.11686  -1.210 0.226306
```

```
Taille.entreprise<20      -1.65497    1.13915  -1.453 0.146278
Taille.entreprise[20,100]  0.01534    0.05098   0.301 0.763456
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(4.1373) family taken to be 1)

    Null deviance: 15930  on 15432  degrees of freedom
Residual deviance: 14207  on 15400  degrees of freedom
AIC: 65555

Number of Fisher Scoring iterations: 1


              Theta:  4.137
          Std. Err.:  0.102

 2 x log-likelihood:  -65487.351
```

2.3.1 Interprétation des résultats :

- "Std. Error" : L’écart type estimé de chaque paramètre est faible, ce qui implique une bonne précision et un intervalle de confiance étroit.

- Pr (>|z|) : La significativité du modèle est déterminée en lisant la p-value et en la comparant au niveau de significativité souhaité. Plus la p-value d’une variable est petit, plus cette variable a un pouvoir explicatif important. Les étoiles accompagnées de la p-value indiquent le degré de significativité.

Certaines modalités sont dites de « référence » et leurs paramètres sont incorporés dans l’intercept. L’effet des autres modalités est interprété par rapport à ces modalités choisies arbitrairement par .

2.3.2 Adéquation du modèle

Tout d’abord, il est indispensable de s’assurer que le modèle considéré s’ajuste bien aux données. À cet effet, nous utilisons la déviance comme statistique de test qui permet de valider un modèle linéaire généralisé. Elle permet ainsi de tirer le modèle qui présente le bon ajustement aux données de base. Pour cela, nous confrontons les deux hypothèses suivantes :

- **H0** : Le modèle considéré à p paramètres est adéquat
- **H1** : le modèle considéré à p paramètres n’est pas adéquat.

Dans notre cas la déviance égale 14207 elle est inférieure au nombre de degrés de liberté qui est égale à 15 400 . Le modèle retenu est donc jugé de bonne qualité. Il utilise la loi de Binomiale négative.

Nous constatons sur les sorties de que certaines modalités des variables ont des p-values non convenables avec le niveau de signifiance de 5%, et que le coefficient de certaines modalités sont très proches deux à deux. Il est nécessaire de faire un regroupement.

Il existe plusieurs techniques de regroupement, soit grouper une modalité dont la valeur estimée approximative ou ayant le même signe (soit positive, soit négative), soit regrouper une modalité ayant la p value non significative avec une modalité ayant la p-value significative. L’objectif est d’obtenir un modèle dont les p-values des modalités sont toutes inférieures à 5%. Dans ce cas, le regroupement est effectué comme suit :

-tranche âge : fusionner « [10,15[» et « [15,20[» par « [10,20[»

« [30,35[» et « [35,40[» par « [30,40[»

« [55,60[» et « [60,65[» par « [55,65[»

-Collège : fusionner « Actifs et « ensemble du personnel » par « ensemble du personnel ».

2.3.3 Procédure de sélection des variables

Nous partons d'un modèle saturé, c'est à dire, contenant toutes les variables. Nous avons utilisé la méthode descendante "backward " pour faire la sélection des variables pertinentes.

Nous procédons à un enlèvement de variable, les unes après les autres. L'idée est de **retirer à chaque étape**, la variable dont **l'élimination maximise le critère choisi**. L'opération s'achève lorsqu'il n'existe plus de variables dont l'absence fait augmenter le critère.

Dans notre cas, le critère choisi est l'AIC. Ainsi, pour sélectionner les variables, nous utilisons la fonction «stepAIC» de On a les sorties suivante

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

```
Step:  AIC=65550.04
Nombredesinistre ~ plfondacte + plafond.annuel + Sexe + bénéficiaire +
    Tranche.age + Situation.famille + Secteur + offset(log(Exposition))

                    Df   AIC
<none>                  65550
- Situation.famille  3 65565
- Secteur            1 65587
- Sexe               1 65612
- bénéficiaire       3 65621
- plfondacte         4 65691
- plafond.annuel     3 65861
- Tranche.age       13 66246
```

Figure 21: Procédure de sélection des variables du modèle

Nous remarquons une hausse de l'AIC. Pour la modélisation de la fréquence de l'acte pharmacie ordinaire on va prendre seulement 7 variables explicatives (plafond acte, plafond annuel, Sexe, bénéficiaire, Tranche. Age, Situation famille, Secteur)

Nous affichons alors les coefficients finaux du modèle :

```
summary(Frequence.)
```

```
Call:
glm.nb(formula = Nombredesinistre ~ plfondacte + plafond.annuel +
    Sexe + bénéficiaire + Tranche.age + Situation.famille + Secteur +
    offset(log(Exposition)), data = phor1, init.theta = 4.122153703,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0194  -0.8815  -0.3404   0.4213   4.4675

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              1.62431    0.05821  27.903  < 2e-16 ***
plafondacte]400;500]    -0.14703    0.03468  -4.240 2.24e-05 ***
plafondacte]500;600]     0.10291    0.03248   3.168 0.001532 **
```

```
plafondacte]600;700]        -0.48082    0.05433  -8.849  < 2e-16 ***
plafondacte]700;800]         0.39084    0.04839   8.077 6.63e-16 ***
plafond.annuel=<2000         0.10761    0.02014   5.342 9.17e-08 ***
plafond.annuel]3000;4000]    0.14639    0.03966   3.691 0.000223 ***
plafond.annuel]4000;5000]    0.61174    0.03551  17.228  < 2e-16 ***
SexeM                       -0.10605    0.01330  -7.972 1.57e-15 ***
bénéficiaireAUTR            -0.20196    0.04369  -4.622 3.79e-06 ***
bénéficiaireCNJT            -0.14441    0.01810  -7.978 1.48e-15 ***
bénéficiaireENFT            -0.12329    0.04544  -2.714 0.006656 **
Tranche.age[5,10[           -0.45088    0.02570 -17.542  < 2e-16 ***
Tranche.age[10,20[          -0.61817    0.02521 -24.519  < 2e-16 ***
Tranche.age[20,25[          -0.57831    0.04345 -13.311  < 2e-16 ***
Tranche.age[25,30[          -0.27580    0.05884  -4.687 2.77e-06 ***
Tranche.age[30,40[          -0.32699    0.04963  -6.589 4.42e-11 ***
Tranche.age[40,45[          -0.27674    0.04971  -5.567 2.59e-08 ***
Tranche.age[45,50[          -0.24900    0.04950  -5.030 4.89e-07 ***
Tranche.age[50,55[          -0.24857    0.05194  -4.786 1.70e-06 ***
Tranche.age[55,65[          -0.19163    0.05352  -3.580 0.000343 ***
Tranche.age>=65             -0.09820    0.07103  -1.382 0.166839
Situation.familleIsolé       0.02126    0.04374   0.486 0.626911
Situation.familleDuo         0.06122    0.03237   1.891 0.058600 .
Situation.familleTrio        0.08982    0.02017   4.454 8.42e-06 ***
SecteurPrivé                -0.25070    0.03918  -6.399 1.57e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(4.1222) family taken to be 1)

    Null deviance: 15902  on 15421  degrees of freedom
Residual deviance: 14197  on 15396  degrees of freedom
AIC: 65507

Number of Fisher Scoring iterations: 1
```

2.3.4 Amélioration du modèle

Une fois le modèle construit, nous souhaitons d’abord tester si les variables ont un effet significatif.

Considérons l’hypothèse linéaire générale : $H_0 : L\,\beta = 0$

Avec L est un vecteur dont la taille est le nombre de modalités de la variable que nous voulons tester moins un, et β est le vecteur des coefficients du modèle. Cette hypothèse nulle indique que le vecteur des coefficients d’une variable est égal à 0. Autrement dit que cette variable n’a pas d’influence pour la variable réponse. Cette hypothèse peut être testée à l’aide d’un test du rapport des vraisemblances.

Pour ce faire, nous utilisons la fonction « anova » dans le logiciel . L’objectif de cette fonction est de savoir si une variable numérique a des valeurs significativement différentes selon plusieurs catégories. Nous présentons le test du rapport de vraisemblance dans la suite.

La statistique du test est : $$S = -2\log\frac{\text{la vraisemblance du modèle sans la variable testée}}{\text{la vraisemblance du modèle avec la variable testée}}$$

Soient les hypothèses suivantes :

- **H_0 :** La variable testée n’a pas d’influence dans le modèle. Les coefficients β_j pour toutes les modalités de cette variable sont nulles.
- **H_1 :** La variable testée a une influence dans le modèle.

Sous H_0, S suit approximativement une loi du Khi-Deux à r degrés de liberté, r est le nombre des modalités de la variable testée moins un.

La p-valeur associée utilise la loi du Chi-deux : si $*$, l'influence de X sur Y est significative, si $**$, elle est très significative et si $***$, elle est hautement significative.

```
> anova(Frequence., test = "Chisq")
Analysis of Deviance Table

Model: Negative Binomial(4.1222), link: log

Response: Nombredesinistre

Terms added sequentially (first to last)


                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                15421      15902
plfondacte         4   296.22      15417      15606 < 2.2e-16 ***
plafond.annuel     3   336.81      15414      15269 < 2.2e-16 ***
Sexe               1    17.44      15413      15252 2.967e-05 ***
bénéficiaire       3   247.05      15410      15005 < 2.2e-16 ***
Tranche.age       10   749.18      15400      14256 < 2.2e-16 ***
Situation.famille  3    17.88      15397      14238 0.0004646 ***
Secteur            1    40.87      15396      14197 1.623e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Les résultats sont satisfaisants, car les p-values sont toutes très faibles. Nous pouvons conclure que les 7 variables influent toutes sur le modèle que nous avons construit.

2.3.5 Résidus

Les tests concernant les coefficients du modèle et les statistiques de l'adéquation du modèle indiquent globalement comment le modèle s'ajuste aux données. Ces statistiques sont complétées par une analyse précise qui compare les valeurs observées et les valeurs estimées, appelée résidus. Les résidus indiquent les distances entre les valeurs estimées et observées, observation par observation. Nous calculons les résidus de la déviance pour notre base de données, ils sont représentés dans le graphique suivant :

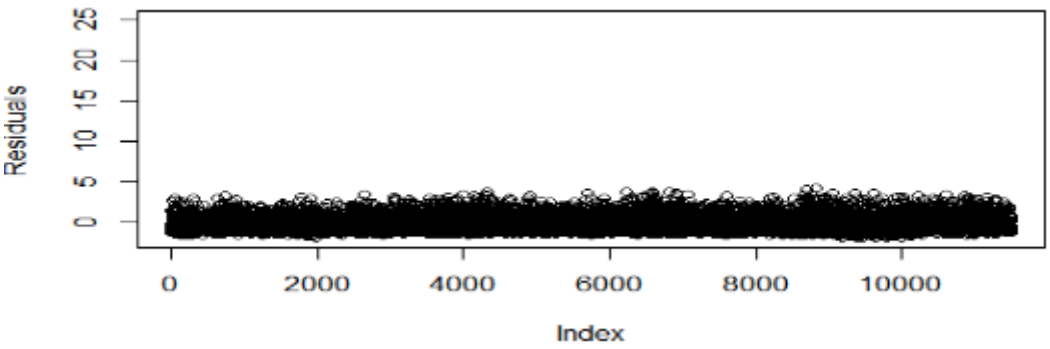


Figure 42 : Résidus du modèle GLM de la fréquence

Nous constatons que les résidus observés se situent globalement autour de l'axe des abscisses et sont répartis dans une bande horizontale régulière autour de 0, ce qui vérifie l'hypothèse d'espérance nulle.

L'hypothèse de variance constante des erreurs semble vérifiée, puisque les résidus sont répartis de façon symétrique. Au final, il n'y a pas des points qui sont très éloignés de l'abscisse. Le modèle est donc acceptable.

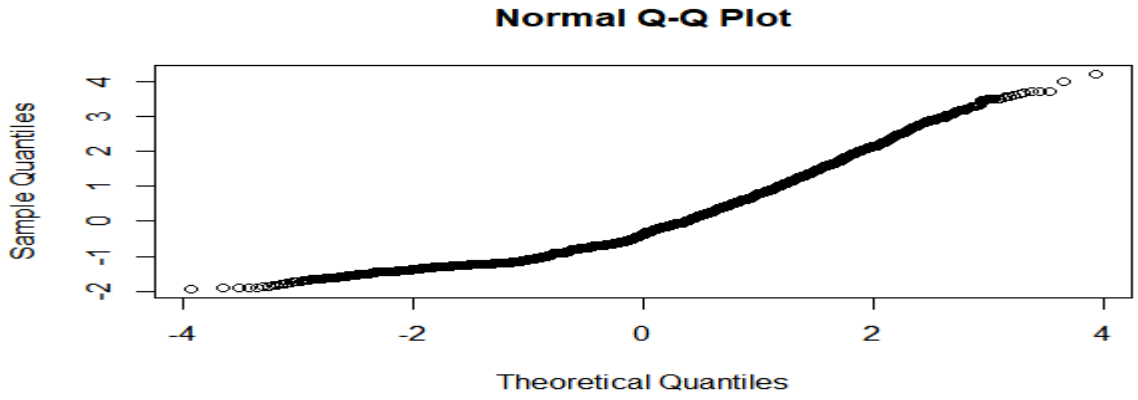


Figure 43 : Graphique Q-Q plot de la fréquence

Le graphique Q-Q plot (quantile-quantile plot) est un graphique "nuage de points" qui vise à confronter les quantiles de la distribution empirique et les quantiles d'une distribution théorique normale, de moyenne et d'écart type estimés sur les valeurs observées. On dit que la distribution est compatible avec la loi normale, lorsque les points forment une droite.

- Si les points sont alignés sur la première bissectrice c'est que la distribution suit probablement une loi de distribution gaussienne normalisé.

- Si les points sont alignés sur une autre droite, c'est que la distribution observée suit une loi normale d'espérance b et d'écart type a.

La commande « qqnorm » de trace la QQplot qui compare la loi d'un vecteur avec la loi N (0,1)

Dans notre cas le nuage de point est proche de la bissectrice on peut conclure que la loi normale s'ajuste bien aux valeurs observées.

2.3.6 Interprétation des coefficients de la régression

Pour vérifier la cohérence des différents coefficients, nous regarderons le sens des estimateurs par exemple :

a) Variable « Sexe »

Pour le Sexe, nous avons choisis femme comme référence. La valeur du paramètre associé pour le sexe masculin est de (-0,10605), ceci indique que toute choses égales, la variable homme améliore de 10% (=1-exp (-0,10605),)) le niveau de sinistralité de la fréquence par rapport à la femme. Ce point est bien vérifié dans le graphique suivant.

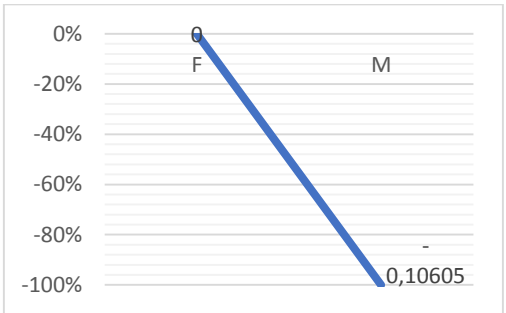


Figure 44 : Coefficients appliqués sur la fréquence par sexe

Figure 45:Nombre de consommation d'acte par sexe

Le nombre d'acte attendus pour le sexe masculin est de 25287*(1-10%) =2275

b) Variable « Tranche âge »

Nous prenons comme autre exemple la variable tranche d'âge, les estimations pour les tranches d'âge représentent des assurés de la tranche d'âge [5,10[à >65 ans ont des valeurs négatives. La référence choisie est la tranche d'âge [0,5[ans. Les estimations négatives correspondent à l'amélioration de la sinistralité par rapport à la modalité de référence. Cet aspect est bien illustré dans les deux graphiques suivants :

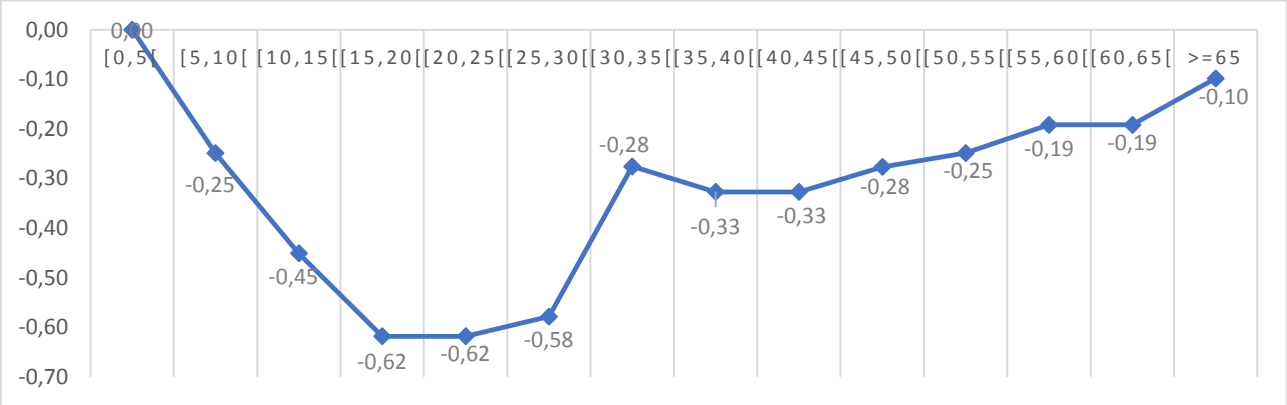


Figure 46 : coefficients appliqués sur la fréquence par tranche d'âge

Cette courbe représente les coefficients appliqués à la fréquence de consommation moyenne par tranche d'âge. Il s'agit d'une courbe décroissante jusqu'à la tanche d'âge [15,20[, une légère bosse à partir de la tranche [25.30[pour atteindre son pic à la tranche d'âge [30,35[. En effet, la fréquence de consommation augmente pour cette tranche d'âge vu la maternité, aux examens et suivis gynécologique chez les femmes d'où le recours aux médicaments. Ces résultats sont approuvés par l'histogramme de consommation de l'acte pharmacie ordinaire par tranche d'âge. En effet, toute choses égales, la tranche d'âge [10.15[améliore la sinistralité de l'ordre de 36 % (1- exp (-0,45)).

L'histogramme suivant suit la même allure de courbe de coefficient

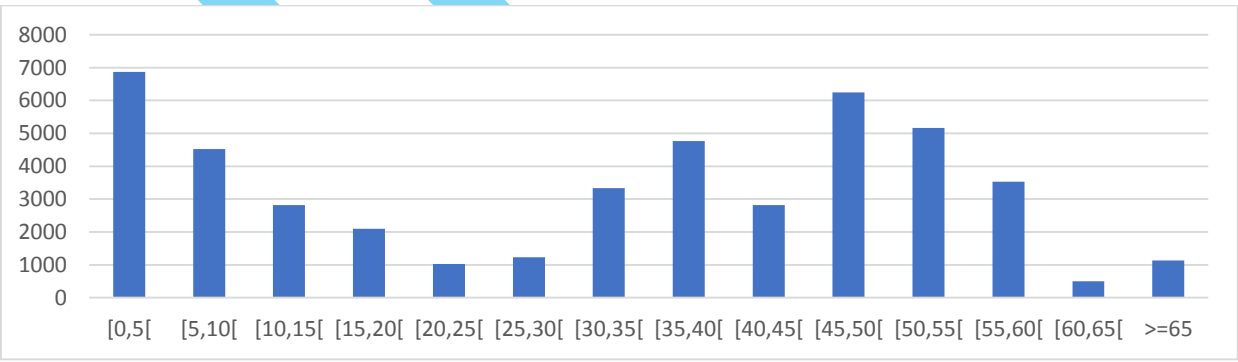


Figure 47 : La somme de nombre d'acte par tranche d'âge

c) Variable « type de bénéficiaire »

Pour la variable type de bénéficiaire, on a choisi la modalité de référence celle (RESP) (assuré principale). Les coefficients appliqués aux autres modalités sont marqués par des signes négatifs, d'où une courbe de coefficients décroissante. La modalité enfant améliore la sinistralité de 1- exp (0.13), soit 12% par rapport à l'assuré principal (RESP).

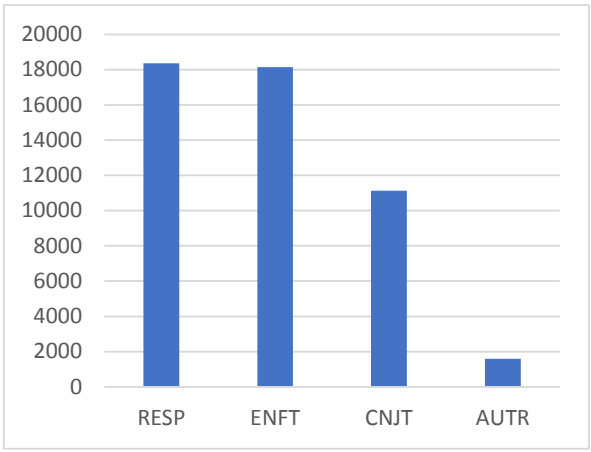


Figure 48 : nombre d'actes par bénéficiaire

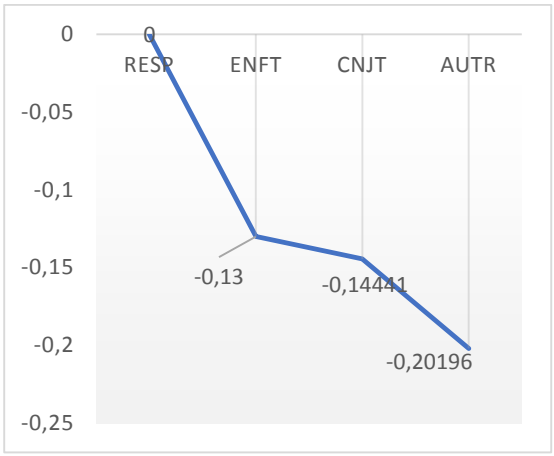


Figure 49 : Coefficients appliqués par bénéficiaire

d) Variable « Plafond acte »

Intéressons-nous maintenant à la variable « plafond acte ». Cette variable concerne le niveau du garantie accordé au bénéficiaire, on a choisi comme référence la modalité « <400 » ,

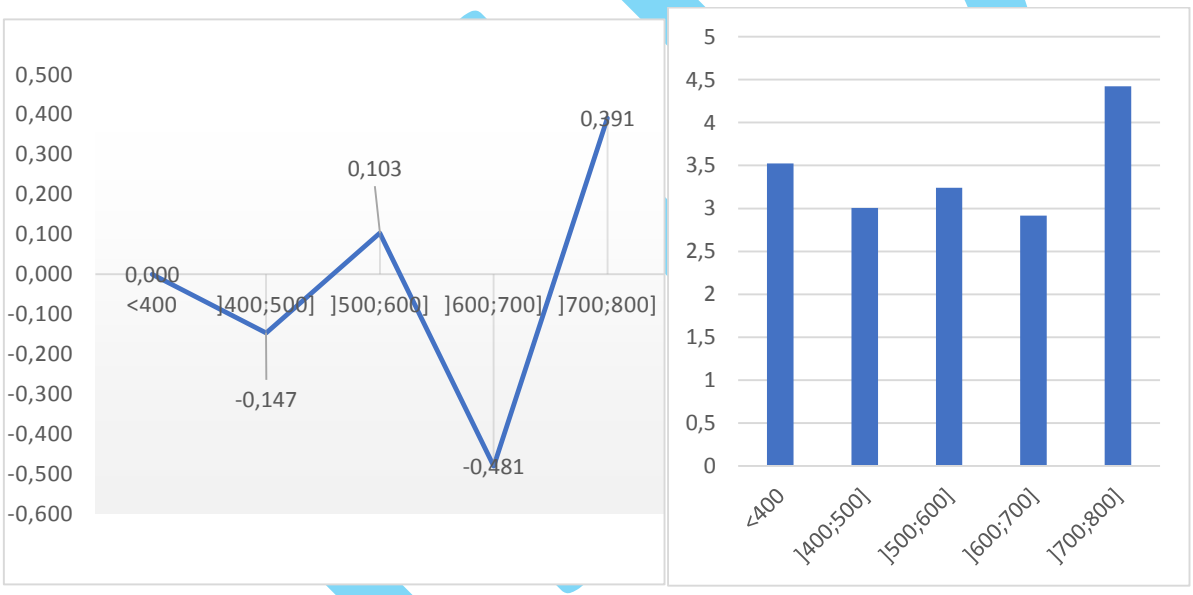


Figure 50 : coefficients appliqués par plafond d'actes

La courbe de coefficient est marquée par des signes positifs et des signes négatifs. Or on sait que la fréquence de consommation est très sensible (voir proportionnel) au niveau de la garantie, néanmoins, la fréquence, quant à elle, est parfois aléatoire. En effet, en plus des variables explicatives présentées dans cette étude, la fréquence peut être influencée par des éléments tels que le comportement, la psychologie de l'assuré.

Nous obtenons donc le modèle suivant pour la variable Fréquence :

$$g(E[Y]) = \beta_0 + \beta_1 \times plafond\ acte + \beta_2 \times plafond\ acte + \beta_3 \times sexe + \beta_4 \times tranche\ age + \beta_5 \times situation\ famille + \beta_6 \times secteur + \beta_7 \times taille\ \ entreprise$$

$$\ln(y) = \beta_0 + \beta_1 1_{plafond\ acte} + \beta_2 1_{plafond\ acte} + \beta_3 1_{sexe} + \beta_4 1_{tranche\ age} + \beta_5 1_{situation\ famille} + \beta_6 1_{secteur} + \beta_7 1_{taille\ d'entreprise}$$

$$y = \exp(\beta_0 + \beta_1 1_{plafond\ acte} + \beta_2 1_{plafond\ acte} + \beta_3 1_{sexe} + \beta_4 1_{tranche\ age} + \beta_5 1_{situation\ famille} + +\beta_6 1_{secteur} + \beta_7 1_{taille\ d'entreprise})$$

e) Interprétation des résultats

Pour expliquer simplement ces résultats, nous prenons un exemple pour déterminer la fréquence de consommation de l'acte pharmacie ordinaire d'un individu possédant les caractéristiques suivantes :

-**Sexe** : Homme,

-**Bénéficiaire** : Conjoint,

-**Age** : 31

-**Situation famille**= TRIO
-**Secteur** =Privé

-**Taille d'entreprise** = 20<

-**Plafond acte** :]400 ;500]

Plafond annuel :<2000

La fréquence de consommation s'écrit sous la spécification suivante :

Fréquence $= \exp\left(\beta_{référence} + \beta_M^{Sexe} + \beta_{Conjoint}^{Bénéficiaire} + \beta_{[30,35[}^{Age} + \beta_{TRIO}^{Situation\ famille} + \beta_{Privé}^{Secteur} + \beta_{]400;500]}^{plafond\ acte} + \beta_{<2000}^{Plafond\ annuelle}\right)$

Les résultats de l'estimation présenté dans la sortie de R permet d'écrire pour l'individu spécifié ci-dessus

= exp (1,62431+ (-0,1061)+ (-0,1444) +(-0,327) +0,08982+ (-0,2507) +-0,14703+0,10761)

=exp (0,84656)

=2,331612294

Ainsi, le nombre d'actes attendu d'un assuré qui possède les caractéristiques suivantes est de l'ordre de 2 actes :

Si on change le sexe de l'assuré tous en gardant les autres caractéristiques, alors on aura 3 actes durant une année.

f) Comparaison avec les données brutes

Tableau 14 : tableau comparatif de nombre de sinistres réel et nombre de sinistres estimés par le modèle

Assuré n :	Nombre de sinistres	Nombre de sinistres estimés par le modèle
1	8	7,04
2	3	2,75
3	6	4,48

4	5	4,71
5	4	3,20
6	4	4,09
7	6	4,48
8	3	2,99
9	2	3,15
10	3	3,07
11	5	4,71
12	3	3,04
13	6	4,48
14	5	4,17
15	6	5,59
16	5	4,71
17	6	5,97
18	5	4,71
19	1	2,27
20	6	5,97
21	4	3,15
22	3	2,17
23	3	2,73
24	8	7,04
25	4	3,14
26	3	2,73
27	4	3,52
28	4	4,00
29	3	2,97
30	6	4,48
31	8	7,04
32	5	4,71
33	6	5,97

D'après le tableau ci-dessus, on remarque que le modèle s'ajoute bien aux données. En effet l'écart entre les valeurs estimés par le modèle et les valeurs réels est pratiquement faible. Ainsi on peut déduire que le modèle appliqué pour calculer la fréquence est bon.

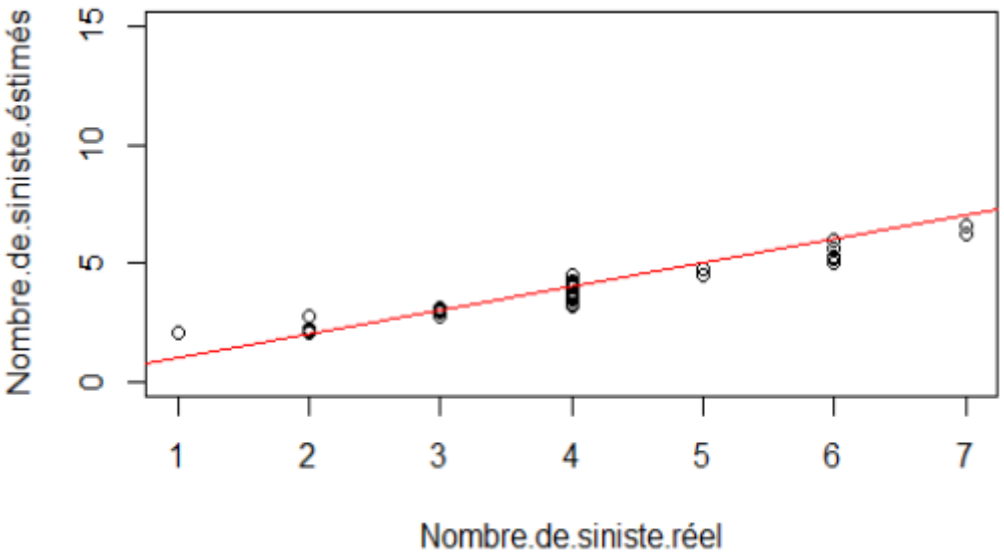


Figure 51 : Nombre de sinistre réel en fonction de nombre de sinistre estimés

Cette figure appuie notre conclusion précédente, en effet le nuage de points sont repartis d'une façon symétrique et se situent globalement autour la première bissectrice.

Conclusion : On adopte ce modèle pour calculer la fréquence de consommation de l'acte pharmacie ordinaire

Section 3 : Modélisation du coût de consommation

3.1. Quel type de montant modéliser ?

Deux méthodes différentes permettent d'obtenir une tarification : une où l'on modélise le remboursement Assureur, et un autre où l'on modélise les montants de frais réels (dépenses totales pour un acte).

3.1.1. Modélisation des frais réels

a) Les avantages :

-Permettre de connaître et de suivre l'évolution du coût réel d'un sinistre,

- Faciliter de l'actualisation des coûts en cas de changement législatif ou contractuel,

- Une seule modélisation suffit, au lieu de plusieurs. En effet, les divers scénarios de remboursement de Sécurité sociale n'impactent pas les montants de frais réels.

b) Les inconvénients :

- Il faut passer une étape intermédiaire pour obtenir le coût en charge par l'Assureur.
- Risque de surestimation du coût de remboursement Assureur en l'absence d'information.
-Le frais réel comprend le remboursement de la Sécurité Sociale qui varie d'un bénéficiaire à un

autre, le remboursement Assureur, le restant à charge de l'assuré n'est pas souvent connu dans la tarification d'un nouveau contrat.

- La variation des taux de remboursement de la Sécurité Sociale peut engendrer une surestimation ou une sous-estimation du coût de remboursement Assureur.

3.1.2 Modélisation des remboursements « Assureur »

a) Les avantages :

- Obtention directe des coûts des remboursements « Assureur » sans passer par les étapes intermédiaires
- Adaptation au contexte de souscription actuelle où la majorité des contrats collectifs sont souscrits sous la forme « en complément du remboursement de la Sécurité Sociale ».

b) Les inconvénients :

- Difficulté de l'actualisation du coût en cas de changement législatif
-Un changement de remboursement de la Sécurité sociale entraîne également un changement des coûts.

Choix final du montant à modéliser

Pour notre étude et pour ne pas tomber dans le cas d'une surestimation du coût nous choisissons pour la modélisation du coût de consommation la variable remboursement assureur.

3.2. Choix de la loi

D'après le référentiel de (Denuit & Charpentier II (2009), les lois les plus courantes utilisées pour la modélisation du coût de consommation sont la loi gamma et la loi log-normale De même que dans le paragraphe précédent, on estime donc les paramètres de ces lois grâce à la méthode de l'estimateur du maximum de vraisemblance, et on compare ensuite la loi théorique à la loi empirique par une représentation graphique.

3.2.1- Estimation des paramètres des lois

Avant toute analyse et modélisation de la charge de sinistre, il est indispensable de travailler sur les montants des sinistres supérieurs à zéro, autrement, les résultats seront certainement amenés à l'erreur.

- **Log-Normale**

Une variable est dite Log-Normale si son logarithme suit une loi Normale. Elle présente l'avantage d'être positive, donc adaptée à la modélisation de coût. La méthode utilisée pour estimer les paramètres est la méthode du maximum de vraisemblance.

La densité de la loi Log-Normale est de la forme suivante :

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

Cette loi vérifie que :

$$E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \qquad VAR(X) = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2}$$

La fonction « fitdist « sur R permet d'estimer les paramètres des lois

fitdist(coutmoyen, "lnorm")

Fitting of the distribution ' lnorm ' by maximum likelihood

Parameters:

	estimate	Std. Error
meanlog	3.252723	0.006747897
sdlog	0.837990	0.004771453

➢ **Loi Gamma**

La densité de la loi Gamma est de la forme :

$$f(y) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y) \quad \text{Pour y >0}$$

L'espérance et la variance sont calculées via les formules ci-dessous

$$E(y) = \frac{\alpha}{\beta} \qquad VAR(y) = \frac{\alpha}{\beta^2}$$

fitdist(coutmoyen,"gamma")

Fitting of the distribution ' gamma ' by maximum likelihood

Parameters:

	estimate	Std. Error
shape	1.69418126	0.017704900
rate	0.04741337	0.000575421

3.2.2 Représentations graphiques

Le graphique ci-dessous représente l'ajustement de la densité du coût de l'acte pharmacie ordinaire et par la loi Gamma et la loi Log-Normal.

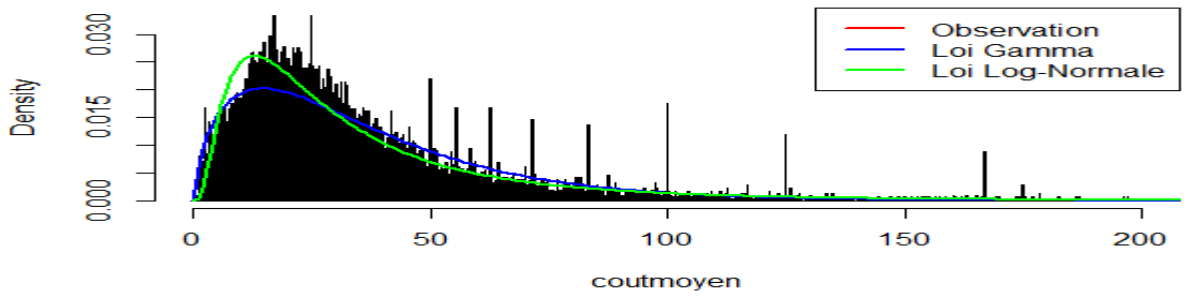


Figure 52 : Ajustement des coûts moyen de l'acte pharmacie ordinaire par les deux lois

Dans ce graphique, la courbe rouge représente les observations des couts moyen, la courbe en bleu est construite par les simulations de la loi Gamma, et la courbe verte est l'ajustement de la loi log Normale. Selon le point de vu graphique, il semble que la loi log Normale adapte mieux les observations. Nous aimerons faire plus de tests pour comparer les deux hypothèses.

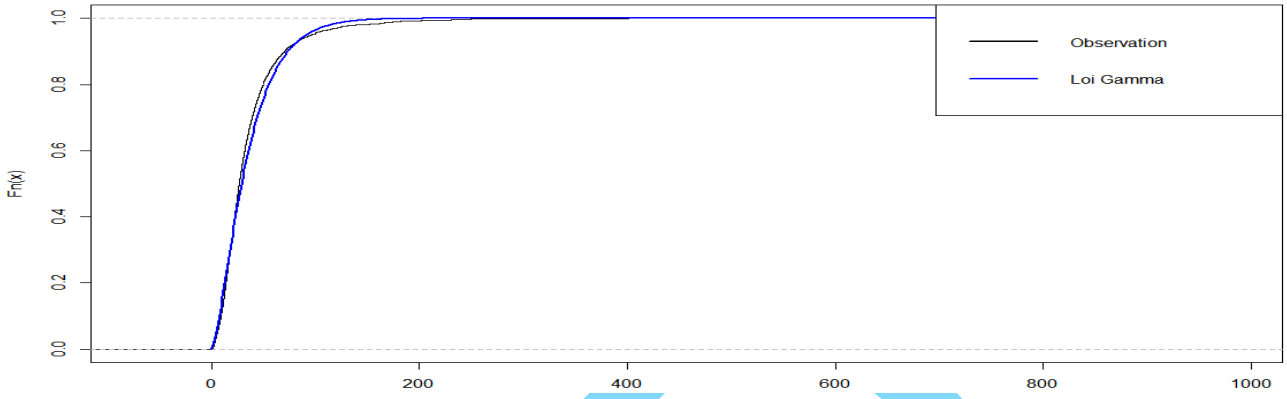


Figure 53 : Ajustement des coûts moyen de l'acte pharmacie ordinaire par une loi Gamma, (Fonction de répartition)

Graphiquement, la loi Gamma s'ajuste bien aux observations. On note néanmoins certaines écarts. En effet la loi Gamma surestime légèrement les frais à partir de 40dt et sous-estime légèrement les frais de coût moyen (entre 40 et 100 dt).

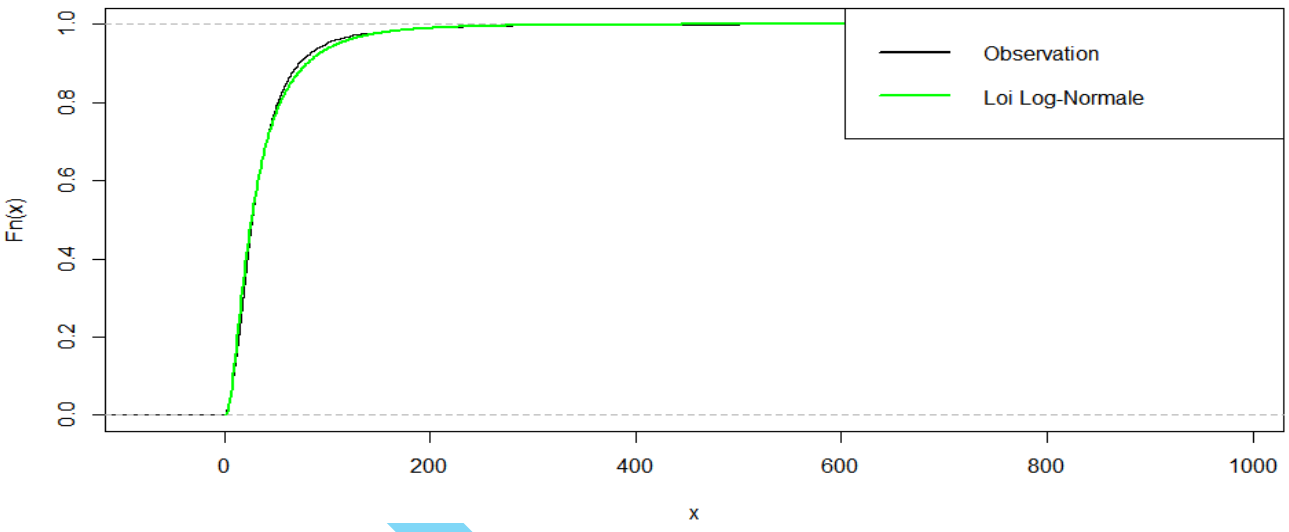


Figure 54 : Ajustement des coûts de l'acte pharmacie ordinaire par une loi Log-Normale, (Fonction de répartition)

La modélisation par une loi Log-Normale atténue quelque peu les écarts observés pour les faibles frais. On se rend compte que l'ajustement est meilleur pour les frais de plus de 150 dt qu'il ne l'était pour la loi Gamma.

3.2.3. Test de Kolmogorov-Smirnov

Globalement, nous constatons que la loi Log-Normal ajuste mieux les observations que la loi Gamma. Les résultats du test de Kolmogorov-Smirnov nous permettent de confirmer notre constatation :

Avec les paramètres estimées $\beta = 1.69418126$ et $\alpha = 0.04741337$, les résultats Kolmogorov-Smirnov p our la loi Gamma sont les suivants :

Δ = 0.054773 p-value = < 2.2e-16

Cela nous amène à rejeter l'hypothèse que les observations suivent une loi Gamma. En respectant le critère de p-value, nous choisissons la loi Log-Normale pour modéliser les coûts de remboursement moyenne.

3.3. Estimation des coûts moyen

Après avoir validé le choix de la loi suivie par les coûts moyens, on va estimer les coefficients attribués à chaque modalité de chaque variable.

La sélection des variables se faisant par la suite, on commence par intégrer l'ensemble des variables dans le modèle, à savoir : Le sexe, le bénéficiaire, la tranche d'âge, la situation familiale, le collège, Secteur (public ou privé), le secteur d'activité, la taille de l'entreprise, plafond acte, plafond annuel.

Pour chacune des variables explicatives, nous allons prendre une modalité de référence. Cette modalité de référence représente l'individu le plus représenté dans le portefeuille. L'ensemble de ces modalités de références constitue notre intercepte, et donc la consommation d'un individu quelconque s'interprètera ainsi comme une surconsommation ou une sous consommation par rapport à l'individu de référence.

Les modalités de l'individu de référence (Intercept) sont les suivantes :

- **Sexe** : Femme,
- **Collège** : ensemble du personnelle
- **Tranche d'âge** : [0,5[,
- **Bénéficiaire** : RESP,
- **Secteur** : Public,
- **Taille entreprise** : 200<
- **Plafond acte** : <400
- **Plafond annuel**]3000;4000]
- **Situation famille** : Famille

L'estimation est réalisée à l'aide de la fonction « glm » du logiciel affiche :

summary(logcout)

```
Call:
glm(formula = Logcoutmoyen ~ plfondacte + plafond.annuel + Sexe +
    bénéficiaire + Tranche.age + Situation.famille + Collège +
    Taille.entreprise + Secteur + offset(log(Exposition)), family = gaussian(l
ink = "identity"),
    data = phor1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3671  -0.2188   0.0032   0.2192   3.6717

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                2.93910    0.042194  40.239  < 2e-16 ***
plfondacte]500;600]       -0.015211   0.017254  -0.882 0.378011
plfondacte]400;500]       -0.173185   0.026724  -6.481 9.42e-11 ***
plfondacte]700;800]       -0.119777   0.038985  -3.072 0.002127 **
plfondacte]600;700]       -0.014347   0.030599  -0.469 0.639164
plafond.annuel]2000;3000] -0.042678   0.032477  -1.314 0.188827
plafond.annuel]4000;5000]  0.006674   0.038130   0.175 0.861051
```

```
plafond.annuel=<2000       0.076630   0.018294   4.189 2.82e-05 ***
SexeM                     -0.004375   0.007079  -0.618 0.536555
bénéficiaireAUTR           0.026676   0.024560   1.086 0.277435
bénéficiaireCNJT          -0.020116   0.009830  -2.046 0.040730 *
bénéficiaireENFT          -0.048289   0.024241  -1.992 0.046387 *
Tranche.age[10,20[        -0.136144   0.013114 -10.381  < 2e-16 ***
Tranche.age[20,25[        -0.092416   0.021842  -4.231 2.34e-05 ***
Tranche.age[25,30[        -0.109395   0.031615  -3.460 0.000541 ***
Tranche.age[30,40[        -0.174252   0.026593  -6.553 5.83e-11 ***
Tranche.age[40,45[        -0.165060   0.026663  -6.191 6.14e-10 ***
Tranche.age[45,50[        -0.169534   0.026578  -6.379 1.84e-10 ***
Tranche.age[5,10[         -0.127892   0.013631  -9.383  < 2e-16 ***
Tranche.age[50,55[        -0.170888   0.027951  -6.114 9.96e-10 ***
Tranche.age[55,65[        -0.158051   0.029003  -5.449 5.13e-08 ***
Tranche.age>=65           -0.144696   0.042012  -3.444 0.000574 ***
Situation.familleDuo       0.043957   0.017852   2.462 0.013815 *
Situation.familleIsolé     0.023438   0.024040   0.975 0.329606
Situation.familleTrio      0.020273   0.011121   1.823 0.068323 .
CollègeRETRAITE            0.016547   0.044586   0.371 0.710559
Taille.entreprise[20,100]  0.124543   0.026776   4.651 3.33e-06 ***
Taille.entreprise<20       0.282783   0.115327   2.452 0.014217 *
SecteurPrivé              -0.162804   0.026893  -6.054 1.45e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.1605918)

    Null deviance: 2534.5  on 15421  degrees of freedom
Residual deviance: 2472.0  on 15393  degrees of freedom
AIC: 15592
```

Call:
glm(formula = Logcoutmoyen ~ plfondacte + plafond.annuel + Sexe + bénéficiaire + Tranche.age + Situation.famille + Collège + Taille.entreprise + Secteur + offset(log(Exposition)), family = gaussian(link = "identity"), data = phor1)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-5.6850	-0.4399	0.0400	0.4921	3.2330

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.93910	0.08588	34.222	< 2e-16 ***
plafondacte]700;800]	0.03766	0.07935	0.475	0.635077
plafondacte]400;500]	-0.17376	0.05439	-3.194	0.001404 **
plafondacte]500;600]	-0.05834	0.03512	-1.661	0.096680 .
plafondacte]600;700]	-0.09151	0.06228	-1.469	0.141759
plafond.annuel=<2000	0.25161	0.03724	6.757	1.46e-11 ***
plafond.annuel]2000;3000]	0.24978	0.06610	3.779	0.000158 ***
plafond.annuel]4000;5000]	0.34896	0.07761	4.496	6.97e-06 ***
SexeM	0.01106	0.01441	0.768	0.442786
bénéficiaireAUTR	0.28281	0.04999	5.657	1.57e-08 ***
bénéficiaireCNJT	-0.10849	0.02001	-5.422	5.97e-08 ***
bénéficiaireENFT	0.26026	0.04934	5.275	1.35e-07 ***
Tranche.age[10,20[	-0.32264	0.02669	-12.087	< 2e-16 ***
Tranche.age[20,25[	-0.01675	0.04446	-0.377	0.706303
Tranche.age[25,30[	0.26091	0.06435	4.055	5.05e-05 ***
Tranche.age[30,40[	0.26300	0.05413	4.859	1.19e-06 ***
Tranche.age[40,45[	0.38453	0.05427	7.086	1.45e-12 ***
Tranche.age[45,50[	0.36473	0.05410	6.742	1.62e-11 ***
Tranche.age[5,10[	-0.30781	0.02774	-11.094	< 2e-16 ***
Tranche.age[50,55[	0.42774	0.05689	7.519	5.84e-14 ***
Tranche.age[55,65[	0.45997	0.05903	7.792	7.03e-15 ***

Tranche.age>=65	0.52852	0.08551	6.181	6.55e-10 ***
Situation.familleDuo	0.12630	0.03634	3.476	0.000511 ***
Situation.familleIsolé	0.14095	0.04893	2.881	0.003975 **
Situation.familleTrio	0.07339	0.02264	3.242	0.001188 **
CollègeRETRAITE	0.36431	0.09075	4.014	5.99e-05 ***
Taille.entreprise[20,100]	0.05236	0.05450	0.961	0.336680
Taille.entreprise<20	-0.07158	0.23474	-0.305	0.760407
SecteurPrivé	-0.10782	0.05474	-1.970	0.048898 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.6653312)

Null deviance: 11175 on 15421 degrees of freedom
Residual deviance: 10241 on 15393 degrees of freedom
AIC: 37513

Comme la déviance égale 10241 qui est inférieure au nombre de degrés de liberté qui est égale à 15393 on peut juger que le modèle est de bonne qualité.

3.3.1. Procédure de sélection des variables

De la même façon que pour les fréquences des sinistres, on va utiliser la procédure descendante « Backward » comme méthodes de sélection des variables pour éliminer les variables les moins significatives.

```
Step:  AIC=37508.5
Logcoutmoyen ~ plfondacte + plafond.annuel + bénéficiaire + Tranche.age +
    Situation.famille + Collège + Secteur + offset(log(Exposition))

                   Df Deviance   AIC
<none>                   10243 37508
- Secteur           1    10245 37510
- Collège           1    10253 37523
- Situation.famille 3    10260 37528
- plfondacte        4    10261 37528
- plafond.annuel    3    10288 37570
- bénéficiaire      3    10322 37622
- Tranche.age      10    10510 37886
```

Figure 55: Procédure de sélection des variables du modèle

Le modèle complet n'est pas le bon choix. La méthode « backward » a fait ressortir un nouveau modèle dans lequel toutes les variables sont acceptées sauf les variables « sexe » et taille entreprise

Call:
glm(formula = Logcoutmoyen ~ plfondacte + plafond.annuel + bénéficiaire + Tranche.age + Situation.famille + Collège + Secteur + offset(log(Exposition)), family = gaussian(link = "identity"), data = phor1)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-5.6624	-0.4404	0.0396	0.4943	3.2005

Coefficients:

```
                Estimate Std. Error t value Pr(>|t|)
```

(Intercept)	2.91252	0.07591	38.369	< 2e-16 ***
plafondacte]700;800]	0.09355	0.05412	1.729	0.083904 .
plafondacte]400;500]	-0.13622	0.03759	-3.624	0.000291 ***
plafondacte]500;600]	-0.05402	0.03491	-1.547	0.121792
plafondacte]600;700]	-0.07536	0.06032	-1.249	0.211571
plafond.annuel=<2000	0.25101	0.03723	6.743	1.61e-11 ***
plafond.annuel]2000;3000]	0.29752	0.04288	6.938	4.14e-12 ***
plafond.annuel]4000;5000]	0.40220	0.05152	7.806	6.29e-15 ***
bénéficiaireAUTR	0.27560	0.04942	5.577	2.49e-08 ***
bénéficiaireCNJT	-0.11382	0.01870	-6.087	1.18e-09 ***
bénéficiaireENFT	0.25697	0.04920	5.223	1.79e-07 ***
Tranche.age[10,20[	-0.32307	0.02668	-12.111	< 2e-16 ***
Tranche.age[20,25[	-0.01820	0.04437	-0.410	0.681746
Tranche.age[25,30[	0.25745	0.06422	4.009	6.13e-05 ***
Tranche.age[30,40[	0.25923	0.05402	4.799	1.61e-06 ***
Tranche.age[40,45[	0.38217	0.05422	7.049	1.88e-12 ***
Tranche.age[45,50[	0.36279	0.05407	6.710	2.01e-11 ***
Tranche.age[5,10[	-0.30779	0.02774	-11.095	< 2e-16 ***
Tranche.age[50,55[	0.42685	0.05688	7.504	6.51e-14 ***
Tranche.age[55,65[	0.46007	0.05903	7.794	6.89e-15 ***
Tranche.age>=65	0.52926	0.08548	6.192	6.10e-10 ***
Situation.familleDuo	0.12662	0.03631	3.487	0.000490 ***
Situation.familleIsolé	0.14146	0.04890	2.893	0.003820 **
Situation.familleTrio	0.07423	0.02261	3.284	0.001027 **
CollègeRETRAITE	0.36498	0.09063	4.027	5.68e-05 ***
SecteurPrivé	-0.07513	0.04222	-1.780	0.075133 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

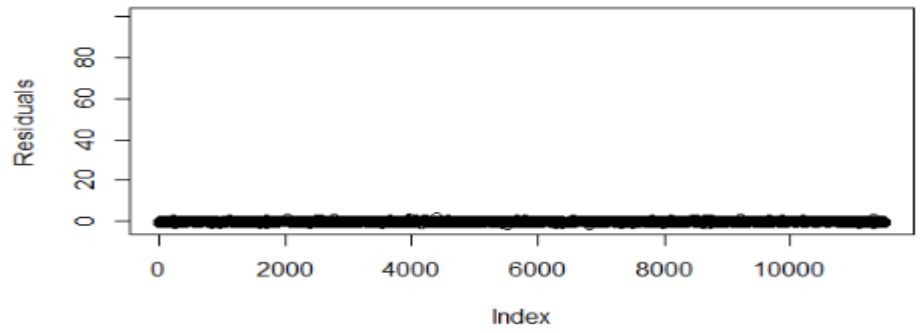
(Dispersion parameter for gaussian family taken to be 0.6652788)

Null deviance: 11175 on 15421 degrees of freedom
Residual deviance: 10243 on 15396 degrees of freedom
AIC: 37508

Number of Fisher Scoring iterations: 2

Nous remarquons une amélioration dans la qualité du modèle puisque la valeur de l'AIC du modèle à baissé, elle passe de 37513 à 37508. Nous obtenons donc un modèle acceptable où la majorité des paramètres sont significatifs. Selon ces résultats le montant de remboursement d'une consommation de l'acte pharmacie ordinaire est égal à exp (2,91252), soit 8,403116 dinars pour l'individu de référence

3. 3.2Analyse des Résidus



D'après cette figure on constate que les résidus observés se situent autour de l'axe des abscisses et qu'ils vérifient l'hypothèse d'espérance nulle. On remarque aussi qu'ils sont repartis d'une façon symétrique ce qui vérifie l'hypothèse de variance constance.

Le modèle est donc acceptable.

3.3.3 Exemple de tarification

Les modèles du coût et de la fréquence établis ci-dessus nous permettent d'obtenir les montants du coût et de la fréquence annuelle de la garantie « pharmacie ordinaire » présentés dans le tableau suivant :

		Coût moyen		Fréquence		Prime
Variable	Modalités	Valeur estimé	EXP(valeur estimé)	Valeur estimé	Exp(valeur estimé)	
Intercept		2,9125	18,4027	1,62431	5,074916	93,3924
Plafond acte	<400	0	1	0	1	1
	]400;500]	-0,1362	0,87265	-0,147	0,863268	0,753331
	]500;600]	0,05402	1,05551	0,10291	1,108392	1,169914
	]600;700]	-0,0754	0,92741	-0,4808	0,618276	0,573395
	]700;800]	0,09355	1,09807	0,39084	1,478222	1,623185
Plafond. Annuel	<2000	0,25101	1,28532	0,10761	1,113613	1,431353
	]2000;3000]	0,29752	1,34652	0	1	1,346515
	]3000;4000]	0	1	0,14639	1,157648	1,157648
	]4000;5000]	0,4022	1,49511	0,61174	1,843637	2,75644
Sexe	F	0	1	0	1	1
	M	0	1	-0,1061	0,89938	0,89938
Bénéficiaire	CNJT	-0,1138	0,89242	-0,1444	0,865533	0,772418
	ENFT	0,25697	1,29301	-0,1233	0,884007	1,143027
	RESP	0	1	0	1	1
	AUTR	0,2756	1,31732	-0,20196	0,817128	1,076419
Tranche d'âge	[0,5[	0	1	0	1	1
	[5,10[	-0,3078	0,73507	-0,4509	0,637067	0,468289
	[10,15[	-0,3231	0,72392	-0,6182	0,53893	0,390144
	[15,20[	-0,3231	0,72392	-0,6182	0,53893	0,390144
	[20,25[	-0,0182	0,98196	-0,5783	0,560845	0,55073
	[25,30[	0,25745	1,29363	-0,2758	0,758965	0,981817
	[30,35[	0,25923	1,29593	-0,327	0,721091	0,934485
	[35,40[	0,25923	1,29593	-0,327	0,721091	0,934485
	[40,45[	0,38217	1,46546	-0,2767	0,758252	1,111188
	[45,50[	0,36279	1,43733	-0,249	0,77958	1,120517
	[50,55[	0,42685	1,53242	-0,2486	0,779915	1,19516
	[55,60[	0,46007	1,58418	-0,1916	0,825612	1,307922
	[60,65[	0,46007	1,58418	-0,1916	0,825612	1,307922
	>65	0,52926	1,69768	-0,0982	0,906468	1,538888
Situation famille	Isolé	0,14146	1,15195	0,02126	1,021488	1,176707
	Duo	0,12662	1,13499	0,06122	1,063133	1,20664

	Trio	0,07423	1,07705	0,08982	1,093977	1,178273
	Famille	0	1	0	1	1
Collège	Ensemble du Pesonnel	0	1	0	1	1
	ACTIFS	0s	1	0	1	1
	RETRAITE	0,36498	1,44049	0	1	1,440485
Secteur	Privé	-0,07513	0,92762	-0,2507	0,778256	0,721928
	Public	0	1	0	1	1

➢ **Exemple de tarification**

-Sexe : Homme,

-Bénéficiaire : Conjoint,

-Age : 31

-Situation famille= TRIO
-Secteur =Privé

-Collège = ensemble du personnel

-Plafond acte :]400 ;500]

Plafond annuel :<2000

✓ La fréquence de consommation est donc estimée par :

Fréquence $=\beta_{référence} * \beta_{]400\ ;500]}^{plafond\ acte} * \beta_{<2000}^{Plafond\ annuelle} * \beta_{M}^{Sexe} * \beta_{Conjoint}^{Bénéficiaire} * \beta_{[30,35[}^{Age} * \beta_{TRIO}^{Situation\ famille} * \beta_{Privé}^{Secteur} * \beta_{ensemble\ du\ persnnel}^{collège}$

=5.07*0.86*1.113*0.899*0.865*0.788*1.093*1*0.778

=**2,33**

✓ Le coût de consommation est donc estimé par :

Coût $=\alpha_{référence} * \alpha_{]400\ ;500]}^{plafond\ acte} * \alpha_{<2000}^{Plafond\ annuelle} * \alpha_{M}^{Sexe} * \alpha_{Conjoint}^{Bénéficiaire} * \alpha_{[30,35[}^{Age} * \alpha_{TRIO}^{Situation\ famille} * \alpha_{Privé}^{Secteur} * \alpha_{ensemble\ du\ persnnel}^{collège}$

=18.40*0.872*1.285*1*0.892*1.2951*1.077*1*0.927

=**23,87 dt**

✓ **La prime pure**

Prime pure = PP =Coût * Fréquence =2.33*23,87= 55 ,60 dt

3.3.4 Comparaison avec les données brutes

D'après le tableau ci-dessous, on remarque que le modèle construit pour estimer le coût moyen de l'acte pharmacie ordinaire s'ajuste bien aux données fournit par la base de données. En effet l'écart entre les valeurs estimés par le modèle et les valeurs réels est très faible. En vue de ces résultats on peut juger que le modèle appliqué pour calculer le coût moyen est bon.

Tableau 15 : tableau comparatif des coût réels et estimés

Assuré n :	Coût moyen réel	Coût moyen estimé par le modèle
1	19,93	19,62
2	24,82	23,87
3	28,40	29,67
4	29,99	29,67
5	34,23	32,58
6	30,05	29,18
7	27,07	26,21
8	27,57	26,69
9	24,73	23,87
10	26,79	26,48
11	31,95	31,63
12	27,32	26,99
13	33,68	32,74
14	18,53	16,87
15	21,40	23,31
16	30,33	30,01
17	34,88	33,94
18	34,47	33,62
19	23,35	22,46
20	43,38	43,08
21	33,69	32,76
22	29,30	28,43
23	24,21	23,35
24	33,32	31,63
25	31,65	32,70
26	29,97	29,67
27	20,51	19,62
28	21,42	20,54

29	21,90	20,24
30	26,69	27,96
31	28,82	27,96
32	26,76	28,02
33	25,63	26,69
34	37,19	35,54

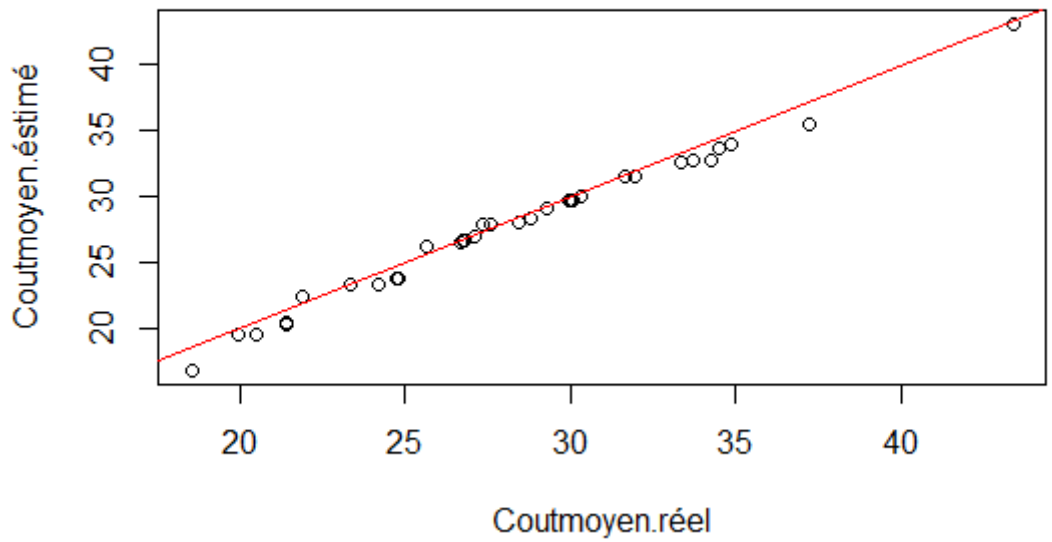


Figure 56 : coût moyen estimé en fonction du coût moyen réel

A partir de ce Graphique, on remarque que le nuage de points ainsi construite s'aligne sur la première bissectrice. A cet effet on peut déduire que l'écart entre les valeurs estimés et les valeurs réel est très réduit.

Conclusion : On peut adopter ce modèle pour estimer le coût moyen de consommation de l'acte pharmacie ordinaire

Conclusion du chapitre trois

Dans ce chapitre dont l'objectif central consiste à mener une application empirique relative à la détermination de la prime on a choisi quelques contrats groupe relatifs à un nombre d'entreprises qui ont contracté avec le GAT durant l'année 2018 pour bénéficier de ses produits en matière de santé. Ce choix est basé en fait sur les tableaux de prestation qui sont à notre disposition. On a choisi de tarifier l'acte pharmacie ordinaire en utilisant les modèles GLM sur le logiciel R avec la loi

Binomiale Négative pour la fréquence et la loi log Normale pour le coût moyen. L'utilisation de la fonction de lien "log" permet de prendre seulement les valeurs positives qui correspondent effectivement à la nature de la variable réponse. Les coefficients estimés permettent d'indiquer si cette modalité influence significativement (avec un signe positif ou négatif) les modalités des références.

Nous avons validé les modèles avec le test du rapport de vraisemblance concernant les coefficients du modèle et la déviance, la validation est complétée également par une étude des résidus. À la fin du chapitre on a essayé d'illustré les résultats à travers un exemple de faire un exemple simplifié pour tarifier l'acte pharmacie ordinaire.

Nous rappelons aussi que, la prime calculée est une prime pure qui est d'ordre technique. Il revient à la compagnie de déterminer les chargements nécessaires en fonction de ses frais de gestion et de ses relations avec ses intermédiaires pour avoir une prime commerciale.

Conclusion Générale

Il est bien connu que la Caisse Nationale d'Assurance Maladie (CNAM) en Tunisie rencontre des difficultés énormes, dans le financement des coûts des engagements en soins de santé pris envers ses adhérents. De ce fait, les organismes complémentaire santé tiennent un rôle prépondérant dans les remboursements des soins médicaux. Le marché des assurances santé aujourd'hui est extrêmement concurrentiel, ce qui rend la tarification des primes un enjeu pour tous les assureurs qui sont appelés à identifier leurs consommateurs et segmenter leurs portefeuilles en classes de risque homogène afin de pouvoir identifier les facteurs expliquant la sinistralité. Le choix de la méthode de tarification est dans ce cas, une étape primordiale. Dans ce projet de mémoire on s'est basé sur le calcul des deux grandeurs « fréquence » et « coût moyen » des sinistres. Le but de ce mémoire est de présenter l'apport des modèles dans la tarification des contrats collectifs de Frais de Santé.

Pour ce faire, nous avons présenté dans un premier temps, les concepts de base de l'assurance maladie et le coût de dépenses pour les ménages ainsi que son poids par rapport au PIB de chaque pays à partir d'une approche macroéconomique et microéconomique. Cela montre l'intérêt de l'assurance maladie qui désigne une sorte de protection financières. Dans la première partie on a mis aussi un focus sur les obstacles qui freinent le développement de l'assurance maladie tel que la fraude et on a présenté quelques actes de maitrise et de prévention.

Dans le deuxième chapitre, nous avons présenté le marché d'assurances tunisiens et ses principaux intervenants. Nous avons aussi traité les données à notre disposition en faisant des études statistiques afin de mieux comprendre le portefeuille du GAT ASSURANCES et le risque qu'on souhaite traiter.

Le présent projet de fin d'études propose une méthodologie de tarification intégrant la sinistralité des assurés comme facteur déterminant de la prime. Une analyse préliminaire de la sinistralité du portefeuille justifie l'utilisation d'une segmentation, afin de séparer les assurés. Pour se faire on a présenté une approche de tarification en se basant sur le principe « Coût*Fréquence » où le coût et la fréquence sont estimés par le Modèle Linéaire Généralisé (GLM). Pour l'application du GLM. Les variables « Coût » et « Fréquence » doivent suivre une des lois de la famille exponentielle. Les lois retenues sont la loi Gamma pour le coût et la loi Binomiale-Négative pour la fréquence . Ce sont les lois pour lesquelles nous avons obtenu les meilleurs résultats sur les tests par le graphique et par l'indice de Kolmogorov-Smirnov.

Les modèles fournissent les primes pour un assuré de référence ayant des caractéristiques précisées ; ainsi que des coefficients correcteurs sous forme de pourcentage pour les assurés autres que l'assuré de référence. Concrètement, la prime d'un assuré est déterminée en appliquant les coefficients de manière multiplicative sur la prime de référence. Ces coefficients nous permettent également d'effectuer des analyses sur la sensibilité des primes de chaque famille d'actes par rapport aux variables explicatives.

Les résultats obtenus sont acceptables. Néanmoins, ils peuvent être optimisés par les études supplémentaires telles que :

- une étude sur la fréquence avec une période plus affinée (semestrielle, trimestrielle…) que la fréquence annuelle.

- une étude plus affinée sur le collège contractuel "ensemble du personnel" en décomposant plus précisément par collège "cadre", "employé" et "ouvrier"…

- la l'ajout de quelques variables explicatives tel que par exemple : la pratique d'une activité sportive, ou même la consommation du tabac peuvent perfectionner cette modélisation

De telles études nécessitent plus d'informations et plus de temps pour pouvoir obtenir des résultats plus consistants.

BIBLIOGRAPHIE

Articles et ouvrages

- Anderson, G, and Peter H. (2000). Population Aging : A Comparaison Among Industrialized Contries. Health Affairs, 191-203
- Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. The American Economic Review, Vol LIII, No. 5, 941-973.
- -Annear PL, Wilkinson D, Men RC and Van Pelt M. (2006). Increasing access to health services for the poor: Health financing and equity in Cambodia. . Working Paper. Phnom Penh, Cambodia
- [Anh Tuan Nguyen] Nguyen .A . Conception des méthodes d'évaluation en assurance expatriée
- [Denuit & Charpentier I] Charpentier A. & Denuit M. : Mathématiques de l'assurance non vie, tome 1, Tarification et Provisionnement, décembre 2009
- [Denuit & Charpentier II] Charpentier A. & Denuit M. : Mathématiques de l'assurance non vie, tome 2, Tarification et Provisionnement, décembre 2009
- [E & T] Allain E& Brenac.T, Modéles Lnéaires Généralisés appliqué à l'étude de nombre d'accident sur des sites routiers
- Ekman,B (2007) the impact of the helth insurance of outpatient utilization and national household surveydata .Health Reaserch Policy and Systems

- Grossman M. (1972). On the concept of health capital and the demand for health. Journal of Political Economy 80 , 223-225
- Joseph P. (1993). Newhouse and the Insurance Experiment Group. Free for All? Lessons from the RAND Health Experiment, Cambridge, Mass.: Harvard University Press
- Jowett M,Deolalikar .A , Martinsson.P (2004) health insurance and treatment seeking behavior :evidence for effects on a ccess to care and health outcomes .Medical Care Reaserch and review , vol .57 , n 3 298-318
- Knight, F. H. (1921). Risk, Uncertainty and Profit. New York: Harper and Row, 1965
- Matthew J Eichner (1998). The Demand for Medical Care: What People Pay Does Matter.
- [MATTHIEU Vautrin] Vautrin. M, Élaboration d'une méthode de tarification avec indicateurs de risque pour des contrats complémentaires santé collectifs 2008/2009.
- Nyman, J. (2008). Health Insurance theory: the case of the missing welfare gain. European Journal of Health economics, 9, 369-380.
- NGUYEN (2013) NGUYEN Ngoc Trung Phuong Construction de bases de tarification pour des contrats complémentaires santé collectifs par le Modèle Linéaire Généralisé
- [Ohlsson& Johansson] Ohlsson E& Johansson B, Non-Life Insurance Pricing with Generalized Linear Models, 7 décembre 2009.
- Othman EL JAMYLY (juin 2015) Tarification en assurance maladie de base
- [P & JA] Mccullagh .P & Nelder JA, Generalized Linear Models
- Pourezza J,S.A (2007). Effet of supplemental enzyme on nutrientdigestibility and performance of broiler chiks on diets containing triticale.Int J Poult Sci6 115-117
- Rothschild, M. & Stiglitz, J.E. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economies of Imperfect Information. Quarterly Journal of Economies, 90,, 629-650
- Zukevas, S. (2014). Health Care Demand, Empirical Determinants of. Dans A. Culyer (Éd.), Encyclopedia of Health Economics (pp. 343- 354). Oxford: Elsevier

WEB bibliographie

- https://www.oecd.org/fr/els/systemes-sante/base-donnees-sante.htm
- https://fr.statista.com/infographie/8662/depenses-de-sante-par-habitant-dans-le-monde/
- SITE OMS https://www.who.int/universal_health_coverage/fr/
- https://lapresse.tn/40464/les-depenses-de-la-cnam-relatives-a-la-couverture-du-cout-des-soins-ont-atteint-2028-millions-de-dinars-en_2019/#:~:text=Le%20volume%20des%20d%C3%A9penses%20de,des%20Affaires%20sociales%2C%20Mohamed%20Trabelsi.
- https://www.oecd.org/fr/els/systemes-sante/base-donnees-sante.htm
- https://fr.statista.com/infographie/8662/depenses-de-sante-par-habitant-dans-le-monde/
- http://www.cga.gov.tn/fileadmin/contenus/pdf/Rapport_CGA_FR-ANG_-_2018_final.pdf
- http://ftusanet.org/wp-content/uploads/2015/10/Rapport-FTUSA-DEFINITIF.pdf

ANNEXES

TABLEAU DE PRESTATIONS

PRESTATIONS GARANTIES	MONTANT
CONSULTATIONS & VISITES: 100% des frais engagés avec un maximum pour	
FRAIS PHARMACEUTIQUES : max/an/prestataire = **DT**	% des frais eng.
ACTES DE PRATIQUE MEDICALE COURANTE ET AUXILIAIRES MEDICAUX : AM= PC=	
ANALYSES : B=	
ELECTRORADIOLOGIE : R =	
TRAITEMENTS SPECIAUX : KE=	
ORTHOPEDIE- PROTHESE (non dentaires) : max/an/prestataire = **DT**	100% des frais eng.
FRAIS CHIRURGICAUX : Y compris accessoire, FSO, anesthesiste ... Kc=	
HOSPITALISATION :	00/J
OPTIQUE: **MONTURE : 100%** des frais eng. avec possibilité de renouvellement tous les deux ans avec un max par monture = **VERRES : 100%** des frais eng. avec un max par an et par prestataire= Les renouvellements des verres ne sont remboursés que si l'assuré présente une modification dans l'acuité visuelle. **OU LENTILLES: 100%** des frais eng. avec un max par an et par prestataire= Les renouvellements des lentillesne sont remboursés que si l'assuré présente une modification dans l'acuité visuelle. En cas de présentation de lunette et lentilles, seul les lunette sont remboursable	
SOINS ET PROTHESES DENTAIRES : max/an/prestataire = **DT** D=	
ODF pour les enfants - 16 ans max/an/prestataire = **DT**	100% des frais eng.
ACCOUCHEMENT :	
TRANSPORT DU MALADE : (sur prescription médicale & en cas d'hospitalisation): 100% des frais	
FRAIS FUNERAIRES: Indemnité forfaitaire	
CIRCONCISION : Indemnité forfaitaire	

MAXIMUM DES REMBOURSEMENTS POUR L'ENSEMBLE DES PRESTATIONS PAR AN ET PAR PRESTATAIRE	

Annexe 1

```
# Chargement des bibliothèques nécessaires à l'ajustement
library(stats4);library(MASS); library(grid)
```

```
 library(vcd)

 Nombresinstre=c(5171,3204, 2080, 1464, 1086,705, 499,397,266,179, 118, 84, 58
, 35,26, 15,12, 10,6, 6,2,2, 4,1,1,2)
x=rep(1:26,Nombresinstre)


#Calcul de la moyenne empirique et de la variance empirique de l'échantillon
> mean(x)
[3.192186

>
> var(x)
7.466432
# Estimation des paramètres d'ajustement à la loi de Poisson

 fitdistr(x,"poisson")
     lambda
  3.19218558
 (0.01438199)
# Estimation des paramètres d'ajustement à la loi de Binomiale-Négative

 fitdistr(x,"negative binomial")
      size          mu
  3.10400286   3.17319775
 (0.07603252) (0.02362547)
 pois=goodfit(x,type="poisson",method="ML")
 pois


Observed and fitted values for nbinomial distribution
with parameters estimated by `ML'

 count observed       fitted pearson residual
     0        0 1.291946e+03     -35.94364381
     1     4000 2.027204e+03      43.81607422
     2     2379 2.102840e+03       6.02223380
     3     1538 1.808535e+03      -6.36151606
     4     1071 1.395124e+03      -8.67770886
     5      774 1.002022e+03      -7.20340218
     6      462 6.841583e+02      -8.49345231
     7      373 4.498027e+02      -3.62130888
     8      244 2.871816e+02      -2.54812577
     9      176 1.791121e+02      -0.23253466
    10      130 1.095936e+02       1.94928281
    11       96 6.599749e+01       3.69312350
    12       63 3.921206e+01       3.79880411
    13       58 2.303030e+01       7.28688924
    14       46 1.339172e+01       8.91065451
    15       23 7.719232e+00       5.49994215
    16       20 4.415296e+00       7.41683495
    17       10 2.508225e+00       4.73043978
    18        9 1.416145e+00       6.37288730
    19       10 7.951517e-01      10.32266065
    20        6 4.442445e-01       8.33550870
    21        1 2.470706e-01       1.51475954
    22        0 1.368418e-01      -0.36992132
    23        3 7.550314e-02      10.64311283
    24        1 4.151375e-02       4.70424460
    25        1 2.275184e-02       6.47883011
    26        2 1.243204e-02      17.82587005
    27        0 6.774271e-03      -0.08230596
    28        0 3.681798e-03      -0.06067782
    29        1 1.996224e-03      15.11888739
 plot(nb,main="Ajustement par une loi Binomiale-Négative ")
```

Annexe (2)

```
> # test kologorov-smirnovpour loi poisson


ks.test( Nombresinstre,rpois(18,lambda = 2.10581506 ))

        Two-sample Kolmogorov-Smirnov test

data:  Nombresinstre and rpois(29, lambda = 3.17317561)
D = 0.63333, p-value = 1.457e-05
alternative hypothesis: two-sided

> # test kologorov-smirnovpour loi binomiale négative

ks.test(Nombresinstre, rnbinom,size = 1.16700263,mu=  2.10581771 )

        One-sample Kolmogorov-Smirnov test

data:  Nombresinstre
D = 7.2333, p-value < 2.2e-16
alternative hypothesis: two-sided


library(corrplot)

M<-cor(cbind(cout,sexe,bénéficiaire,tranche.age,situation.famille,collège,Sect
eur.d.activité,secteur,Taille.entreprise,plafond.acte,plafond.annuel,Taux.de.r
emboursement))
 corrplot(M, method="circle")
```

Fréquence de consommation

```
phor <- read.csv2("G:/intibeh base nette/ phor.csv")
  View(phor)
 summary(phor)
library(MASS)

# détermination des variables explicatives


Secteur<-phor$secteur
Taille.entreprise<-phor$Taille.entreprise
Secteur.d.activité<-phor$Secteur.d.activité
Tranche.age<-phor$tranche.age
plfondacte<-phor$plafond.acte
taux <-phor$Taux.de.remboursement
bénéficiaire<-phor$Qualité.bénéficiaire
plafond.annuel<-phor$plafond.annuel
Sexe<-phor$Sexe
Situation.famille<-phor$Situation.famille
Collège<-phor$Collège
Logcoutmoyen<-phor$Log.Cout.moyen.
Cout.moyen<-phor$Cout.moyen
Nombredesinistre<-phor$Nombre.de.Sinistre


# détermination de l'individu de référence
# détermination de l'individu de référence
phor$secteur <- relevel(phor$secteur, ref="Public")
phor$tranche.age<- relevel(phor$tranche.age, ref="[0,5[")
phor$Sexe <- relevel(phor$Sexe, ref="F")
phor$plafond.annuel <- relevel(phor$plafond.annuel, ref="]2000;3000]")
phor$Taille.entreprise<-relevel(phor$Taille.entreprise, ref="200>")
phor$Collège <- relevel(phor$Collège, ref="ENSEMBLE DU PERSONNEL")
```

```
phor$Qualité.bénéficiaire <-relevel(phor$Qualité.bénéficiaire, ref="RESP")
phor$Situation.famille <-relevel(phor$Situation.famille, ref="Famille")
phor$plafond.acte <- relevel(phor$plafond.acte, ref="<400")

summary(phor)


# Estimation des paramètres


Frequence. <-glm.nb (formula= Nombredesinistre ~ plfondacte +plafond.annuel +
Sexe + bénéficiaire  +Tranche.age + Situation.famille + Secteur+ Collège +Tail
le.entreprise+offset(log(Exposition)), data =  phor)

summary(Frequence.)
```

Regroupement

```
Secteur<-phor1$secteur
Taille.entreprise<-phor1$Taille.entreprise
Secteur.d.activité<-phor1$Secteur.d.activité
Tranche.age<-phor1$tranche.age
plfondacte<-phor1$plafond.acte
Taux.de.remboursement<-phor1$Taux.de.remboursement
bénéficiaire<-phor1$Qualité.bénéficiaire
plafond.annuel<-phor1$plafond.annuel
Sexe<-phor1$Sexe
Situation.famille<-phor1$Situation.famille
Collège<-phor1$Collège
Logcoutmoyen<-phor1$Log.Cout.moyen.
Cout.moyen<-phor1$Cout.moyen
Nombredesinistre<-phor1$Nombre.de.Sinistre


# détermination de l'individu de référence
phor1$secteur <- relevel(phor1$secteur, ref="Public")
phor1$tranche.age<- relevel(phor1$tranche.age, ref="[0,5[")
phor1$Sexe <- relevel(phor1$Sexe, ref="F")
phor1$plafond.annuel <- relevel(phor1$plafond.annuel, ref="]2000;3000]"
)
phor1$Taille.entreprise<-relevel(phor1$Taille.entreprise, ref="200<")
phor1$Collège <- relevel(phor1$Collège, ref="ENSEMBLE DU PERSONNEL")
phor1$Qualité.bénéficiaire <-relevel(phor1$Qualité.bénéficiaire, ref="R
ESP")
phor1$Situation.famille <-relevel(phor1$Situation.famille, ref="Famille
")
phor1$plafond.acte <- relevel(phor1$plafond.acte, ref="<400")

summary(phor1)

# Estimation des paramètres

Frequence. <-glm.nb (formula= Nombredesinistre ~ plfondacte + plafond.a
nnuel + Sexe + bénéficiaire + Tranche.age + Situation.famille + Secteur
+ offset(log(Exposition)), data =  phor1)

summary(Frequence.)
```

Les résidus

```
e = residuals(Frequence.)
summary(residuals(Frequence.))


plot(residuals(Frequence.), xlab="Index", ylab = "Residuals ",ylim=c(-2,25),ma
in="Graphique des résidusdu modèles GLM de la frequence ")

qqnorm(residuals(Frequence.))
qqnorm(scale(e))
```

Degré de significativité

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

anova(Frequence., test = "Chisq")
```

COUT moyen

Estimation de paramètres d'ajustement aux lois

```
library(fitdistrplus)
coutmoyen<-phor1$Cout.moyen
summary(coutmoyen)
fitdist(coutmoyen,"gamma")
fitdist(coutmoyen, "lnorm")
```

 # Représentation graphique

```
hist(coutmoyen, pch=2000, breaks=2000, prob=TRUE, main="" ,xlim =c(0,200),col
= 'red' )
fitgamma<-fitdist(coutmoyen,"gamma")
x<-seq(0,300,1)
y<-dgamma(x,fitgamma$estimate[1],fitgamma$estimate[2] )
lines(x,y,lwd=2,col='blue')

fitlnorm<- fitdist(coutmoyen, "lnorm")
y2<-dlnorm(x,fitlnorm$estimate[1],fitlnorm$estimate[2] )
lines(x,y2,lwd=2,col='green')

legend<-c("Observation","Loi Gamma","Loi Log-Normale" )
legend("topright",leg=legend,lty=1,lwd=2,col=c("red","blue","green"))
```

Fonction de répartition

```
# Représentation graphique loi gamma


x<-seq(0,3000,1)
 Fn<-ecdf(coutmoyen)
 plot(Fn , main ="")
 y<-pgamma(x,fitgamma$estimate[1],fitgamma$estimate[2])
 lines(x,y,lwd=2,col='blue')
 legend<-c("Observation","Loi Gamma")
 legend("topright",leg=legend,lty=1,lwd=2,col=c("black","blue")

# Représentation graphique loi log normal


x<-seq(0,3000,1)
 Fn<-ecdf(coutmoyen)
 plot(Fn , main ="")
 y1<-plnorm(x,fitlnorm$estimate[1],fitlnorm$estimate[2])
 lines(x,y1,lwd=2,col='green')
 legend<-c("Observation","Loi Log-Normale" )
 legend("topright",leg=legend,lty=1,lwd=2,col=c("black","green"))
```

Test kolmogorv

```
# test kologorov-smirnov pour loi gamma

  ks.test(coutmoyen, pgamma,shape= 1.5387695  ,rate = 0.0407228  )



# test kologorov-smirnov pour loi lognormal


ks.test(coutmoyen, plnorm,meanlog=3.2729534   ,sdlog=0.8695704  )
```

Estimation des paramètres

```
# détermination des variables explicatives


Secteur<-phor1$secteur
Taille.entreprise<-phor1$Taille.entreprise
Secteur.d.activité<-phor1$Secteur.d.activité
Tranche.age<-phor1$tranche.age
plfondacte<-phor1$plafond.acte
Taux.de.remboursement<-phor1$Taux.de.remboursement
bénéficiaire<-phor1$Qualité.bénéficiaire
plafond.annuel<-phor1$plafond.annuel
Sexe<-phor1$Sexe
Situation.famille<-phor1$Situation.famille
Collège<-phor1$Collège
Logcoutmoyen<-phor1$Log.Cout.moyen.
Cout.moyen<-phor1$Cout.moyen
Nombredesinistre<-phor1$Nombre.de.Sinistre


# détermination de l'individu de référence
phor1$secteur <- relevel(phor1$secteur, ref="Public")
phor1$tranche.age<- relevel(phor1$tranche.age, ref="[0,5[")
phor1$Sexe <-relevel(phor1$Sexe,ref="F")
phor1$plafond.annuel <- relevel(phor1$plafond.annuel, ref="]3000;4000]")
phor1$Taille.entreprise<-relevel(phor1$Taille.entreprise, ref="200<")
phor1$Collège <- relevel(phor1$Collège, ref="ENSEMBLE DU PERSONNEL")
phor1$Qualité.bénéficiaire <-relevel(phor1$Qualité.bénéficiaire, ref="RESP")
phor1$Situation.famille <-relevel(phor1$Situation.famille, ref="Famille")
phor1$plafond.acte <-relevel(phor1$plafond.acte, ref="<400")
```

```
summary(phor1)
logcout <-glm(formula = Logcoutmoyen ~ plfondacte + plafond.annuel + Sexe +
                  bénéficiaire + Tranche.age + Situation.famille + Collège +Ta
ille.entreprise + Secteur + offset(log(Exposition)), family = gaussian(link =
"identity"), data = phor1)
summary(logcout)
```

regroupement 2

```
# détermination des variables explicatives


Secteur<-phor2$secteur
Taille.entreprise<-phor2$Taille.entreprise
Secteur.d.activité<-phor2$Secteur.d.activité
Tranche.age<-phor2$tranche.age
plfondacte<-phor2$plafond.acte
Taux.de.remboursement<-phor2$Taux.de.remboursement
bénéficiaire<-phor2$Qualité.bénéficiaire
plafond.annuel<-phor2$plafond.annuel
Sexe<-phor2$Sexe
Situation.famille<-phor2$Situation.famille
Collège<-phor2$Collège
Logcoutmoyen<-phor2$Log.Cout.moyen.
Cout.moyen<-phor2$Cout.moyen
Nombredesinistre<-phor2$Nombre.de.Sinistre


# détermination de l'individu de référence
phor2$secteur <- relevel(phor2$secteur, ref="Public")
phor2$tranche.age<- relevel(phor2$tranche.age, ref="[0,5[")
phor2$Sexe <-relevel(phor2$Sexe,ref="F")
phor2$plafond.annuel <- relevel(phor2$plafond.annuel, ref="=<2000")
phor2$Taille.entreprise<-relevel(phor2$Taille.entreprise, ref="200>")
phor2$Collège <- relevel(phor2$Collège, ref="ENSEMBLE DU PERSONNEL")
phor2$Qualité.bénéficiaire <-relevel(phor2$Qualité.bénéficiaire, ref="RESP")
phor2$Situation.famille <-relevel(phor2$Situation.famille, ref="Famille")
phor2$plafond.acte <-relevel(phor2$plafond.acte, ref="]400;500]")

summary(phor2)
```

Validation du modèle

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

anova(logcout ,test = "Chisq")
```

Test du rapport de vraisemblance
Nous réduisons le modèle :
m=step(logcout)
Test du rapport de vraisemblance : si les variables sont significatives ?
anova(m,test="LRT")
Corrélation entre les modalités :
corr=summary(m, cor = TRUE)


IFID