

**Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuares**

Par : **Zélie NALIN**

Titre du mémoire : Quel est l'impact de la réforme des retraites de 2023 sur la tarification des contrats collectifs santé ?

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.

Membres présents du jury de la Signature : Entreprise :
filière :

Nom : Actuelia

Signature :



Directeur de mémoire en
entreprise

Membres présents du jury de
l'Institut des Actuares :

Signature :

Nom : Chloé Blanchard

Signature :



Invité :

Nom :

Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

Signature du responsable
entreprise :



Signature du candidat :



Résumé

Les risques liés à la santé représentent un enjeu majeur pour les assureurs et les mutuelles. Les montants à charge des organismes complémentaires sont influencés par divers facteurs externes tels que les lois, les réformes, les réglementations et l'état de santé général pouvant impliquer un changement de comportement des consommateurs. Ces évolutions exigent des assureurs une adaptation continue et une révision régulière de leur tarification. Ainsi, il est crucial d'anticiper ces changements en recourant à des outils de modélisation capables de prévoir les fluctuations des coûts.

La réforme des retraites de 2023, qui repousse l'âge légal de départ de 62 à 64 ans, aura nécessairement un impact sur les prestations de santé des contrats collectifs. En augmentant l'âge de la population active, les personnes concernées resteront plus longtemps affiliées à leurs contrats collectifs, ce qui pourrait entraîner une augmentation des soins.

Dans ce contexte, l'objectif de ce mémoire est **d'évaluer l'impact de la réforme des retraites de 2023 sur la tarification des contrats collectifs**. Il s'agit de quantifier cet impact depuis la mise en œuvre de la réforme en 2023 jusqu'à l'horizon 2030, en se concentrant sur le portefeuille d'une mutuelle standard gérant des contrats collectifs.

Pour atteindre cet objectif, une modélisation du coût moyen et de la fréquence des soins pour les personnes travaillant deux années supplémentaires, les 63 et 64 ans, est réalisée en utilisant des Modèles Linéaires Généralisés (GLM) sur les données de la mutuelle étudiée. Les primes pures, obtenues en multipliant le coût moyen par la fréquence pour chaque âge sont ensuite comparées à celles issues d'un outil de tarification marché afin d'en vérifier la fiabilité. Sur cette base, et en formulant des hypothèses sur le nombre de personnes concernées, l'impact annuel sur le portefeuille de la mutuelle est estimé et comparé aux prévisions d'articles assurantiels ayant étudié cette réforme.

Ce mémoire a ainsi pour but de fournir aux mutuelles, une estimation de l'impact de la réforme des retraites de 2023 sur la tarification de leurs contrats collectifs.

Mots clés : Retraite, Contrats collectifs, Santé, Âge de départ à la retraite, Décalage, Modèles Linéaires Généralisés, Impact, Coût moyen, Fréquence, Portefeuille, Primes pures, Tarification.

Abstract

Health risks represent a major challenge for insurers and mutual insurers. The amounts payable by supplementary insurers are influenced by various external factors, such as laws, reforms, regulations and the general state of health, which can lead to changes in consumer behaviour. These changes require insurers to continually adapt and review their pricing. Furthermore, it is crucial to anticipate these changes by using modelling tools capable of forecasting fluctuations in costs.

The pension reform of 2023, which raises the statutory retirement age from 62 to 64, is bound to have an impact on group contract healthcare benefits. By increasing the age of the working population, the people concerned will remain affiliated to their group contracts for longer, which could lead to an increase in the need for care.

In this context, the aim of this report is **to assess the impact of the 2023 pension reform on the pricing of group contracts**. The aim is to quantify this impact from the implementation of the reform in 2023 to 2030, focusing on the portfolio of a standard mutual managing group contracts.

To achieve this objective, the average cost and frequency of care for people working an extra two years, at ages 63 and 64, are modelled using Generalised Linear Models (GLM) on the data from the mutual insurer studied. The pure premiums, obtained by multiplying the average cost by the frequency for each age are then compared with those derived from a market pricing tool in order to check their reliability. On this basis, and by making assumptions about the number of people affected, the annual impact on the mutual's portfolio is estimated and compared with the forecasts of insurance companies that have studied this reform.

The purpose of this brief is to provide mutual insurers, with an estimate of the impact of the 2023 pension reform on the pricing of their group contracts.

Keywords : Retirement, Group contracts, Health, Retirement age, Shift, Generalised Linear Models, Impact, Average cost, Frequency, Portfolio, Pure premiums, Pricing.

Note de Synthèse

Contexte et problématique

La réforme des retraites de 2023 est à l'origine de l'allongement de deux ans de l'âge légal de départ à la retraite passant de 62 à 64 ans. Ce décalage progressif a été mis en place le 1^{er} septembre 2023 et s'applique pour les personnes nées à partir du 01/09/1961.

Années	Génération	Âge légal + 3 mois par an	Durée de cotisation pour le taux plein Accélération du calendrier Touraine
2023	Entre le 01/09/1961 et 31/12/1961	62 ans et 3 mois	169 trimestres
2024	1962	62 ans et 6 mois	169 trimestres
2025	1963	62 ans et 9 mois	170 trimestres
2026	1964	63 ans	171 trimestres
2027	1965	63 ans et 3 mois	172 trimestres (43 ans)
2028	1966	63 ans et 6 mois	172 trimestres (43 ans)
2029	1967	63 ans et 9 mois	172 trimestres (43 ans)
2030	1968	64 ans	172 trimestres (43 ans)

TABLEAU 1 – Phase transitoire du décalage de 62 à 64 ans

Cette nouvelle mesure a prolongé la durée des carrières, ce qui pourrait augmenter l'exposition de l'assuré aux risques liés à la santé. En effet, cette réforme entraîne un vieillissement de la population active, obligeant les travailleurs à rester plus longtemps en activité avant de prendre leur retraite.

Prolonger la durée de travail de deux ans signifie que davantage de salariés âgés de 63 et 64 ans continueront de bénéficier des contrats de santé collectifs. Or, à mesure que les individus avancent en âge, leurs besoins en soins de santé augmentent, car ils sont plus susceptibles de développer des problèmes de santé liés au vieillissement. Ces contrats, qui couvrent les frais médicaux des employés, devront donc faire face à une hausse des demandes de remboursements, en raison de la fréquence accrue des consultations médicales, des hospitalisations et des traitements pour cette tranche d'âge.

Ce mémoire a pour objectif d'évaluer l'impact potentiel de la réforme des retraites de 2023 sur un portefeuille de contrats collectifs et de quantifier cet impact pour les années à venir.

Selon l'article de Garrouste C. et Perdrix E. [1], qui souligne que le comportement des actifs diffère de celui des retraités, il est essentiel de baser les simulations sur des données provenant de la population active. C'est pourquoi les modèles présentés par la suite sont construits à partir de données de personnes en activité.

Modélisation

Pour ce faire, une modélisation est réalisée afin d'estimer le coût moyen des soins de santé et la fréquence d'utilisation de ces soins par les personnes contraintes de travailler deux années supplémentaires en raison de la réforme des retraites, celles âgées de 63 et 64 ans. Des modèles linéaires généralisés ont été appliqués aux données fournies par la mutuelle étudiée, comprenant des bases de prestations pour trois garanties différentes ainsi que les informations sur les bénéficiaires de ces prestations. Il est important de noter que les montants de remboursement pris en compte se limitent à ceux versés par la mutuelle, excluant la part prise en charge par la sécurité sociale.

Schémas des étapes des modélisations

Les étapes de la modélisation du coût moyen sont présentées sur le schéma suivant :

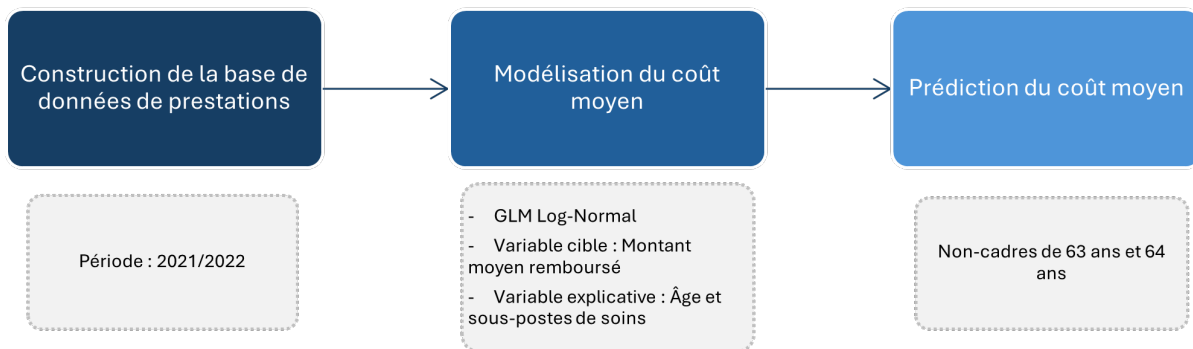


FIGURE 1 – Étapes pour la prédiction du coût moyen

Les étapes de la modélisation de la fréquence sont similaires à celles du coût moyen et sont présentées sur le schéma suivant :

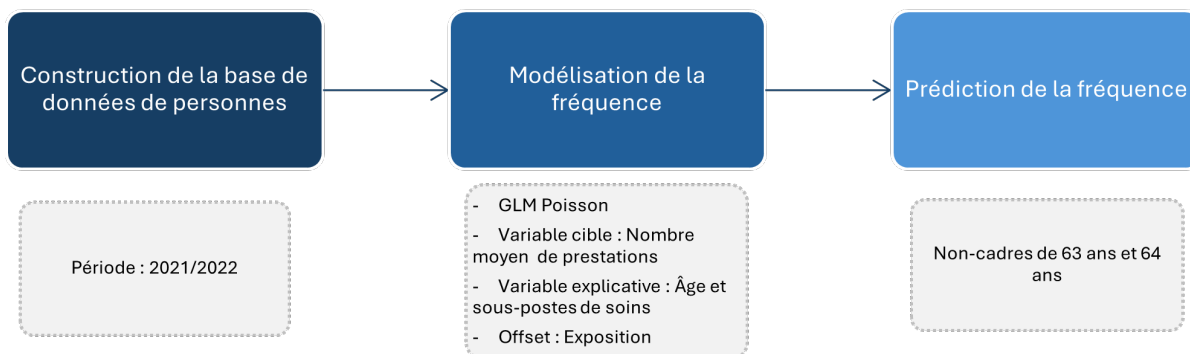


FIGURE 2 – Étapes pour la prédiction de la fréquence

Descriptif des étapes des modélisations

Dans un premier temps, le traitement des bases de données relatives aux prestations et aux personnes associées est effectué afin d’obtenir des données propres pour la réalisation des modèles c’est à dire de supprimer les lignes ayant des données manquantes par exemple. L’étude se concentre sur les prestations de santé et les personnes concernées pour les années 2021 et 2022. Les modèles de régression généralisée (GLM) ont pour objectif de prédire le coût moyen et la fréquence. Ces deux types de modèles sont élaborés indépendamment pour chaque grand poste de soins (optique, dentaire, soins de ville, etc.).

Des indicateurs de performance, tels que l’AIC et le pseudo- R^2 , sont utilisés pour sélectionner le type de modèle. En particulier, pour le modèle de coût moyen, le modèle Log-Normal est comparé au modèle Gamma avec une fonction de lien « identité ». Selon ces indicateurs, le modèle Log-Normal se révèle mieux adapté à la modélisation des coûts moyens. En outre, le modèle Poisson est utilisé pour modéliser la fréquence.

Les mêmes indicateurs permettent également d’évaluer la pertinence de l’âge dans les GLM de coût moyen et de fréquence. Une comparaison est effectuée entre les modèles avec et sans l’âge comme variable explicative. Les résultats montrent que l’inclusion de l’âge améliore la performance des modèles selon ces métriques.

Ensuite, une analyse des résidus est réalisée pour examiner le comportement des résidus studentisés et détecter la présence éventuelle de valeurs aberrantes. Pour les résidus studentisés, l’objectif est que 95% d’entre eux se situent dans l’intervalle de -2 à 2.

Postes	Proportions de résidus entre -2 et 2	
	Coût moyen	Fréquence
Optique	94,55%	96,03%
Dentaire	96,15%	98,10%
Soins de ville	95,99%	70,84%
Pharmacie	96,28%	68,77%
Divers	93,42%	97,62%
Hospitalisation	95,75%	84,21%
Aides auditives	97,35%	99,88%

TABLEAU 2 – Proportions de résidus studentisés pour les modèles de coût moyen et de fréquence pour chaque poste de soins

La plupart des modèles présentent une proportion de résidus se situant entre les seuils de -2 et 2 d'environ 95%. Cependant, les modèles pour les fréquences des soins de ville, de la pharmacie et de l'hospitalisation ne respectent pas cette norme. Cela n'est toutefois pas surprenant, étant donné qu'il s'agit de modèles de fréquence ; leur utilisation pour les prédictions reste donc appropriée.

Enfin, cette analyse utilise les modèles pour prédire le coût moyen et la fréquence pour les personnes âgées de 63 et 64 ans. Les prédictions sont faites pour différents sous-postes de soins tels que les verres simples, les prothèses dentaires, les consultations de médecine générale, etc.

Résultats

Les prédictions obtenues sont ensuite utilisées pour calculer les primes pures de ces individus pour chaque sous-poste de soins, en multipliant le coût moyen par la fréquence pour les deux âges concernés. En additionnant les primes pures de tous les sous-postes de soins pour les deux âges, la prime annuelle totale par âge est obtenue.



FIGURE 3 – Étapes de calcul de la prime pure annuelle pour les personnes de 63 et 64 ans

Les primes pures annuelles s'élèvent à 745,84 € pour une personne de 63 ans et à 762,63 € pour une personne de 64 ans. Sur une base mensuelle, ces primes sont respectivement de 62,15 € pour une personne de 63 ans et de 63,55 € pour une personne de 64 ans.

Pour garantir la fiabilité des deux primes calculées, elles sont comparées à celles obtenues à l'aide d'un logiciel de tarification marché. Ce logiciel, en utilisant les mêmes grilles de couverture que celles des prestations de la base de données, calcule les primes pures pour une personne de 63 et 64 ans. Les primes pures fournies par le logiciel de tarification sont plus élevées, s'élevant à 865,59 € pour une personne de 63 ans et à 881,38 € pour une personne de 64 ans. Sur une base mensuelle, l'écart est de 9,98 € pour une personne de 63 ans et de 9,90 € pour une personne de 64 ans. Malgré les différentes bases et méthodologies, les résultats sont globalement satisfaisants.

Ces deux primes pures calculées sont à la base de l'analyse de l'impact de la réforme des retraites de 2023 sur le portefeuille de la mutuelle.

Impact sur le portefeuille de la mutuelle

Comme indiqué dans la première partie de cette note de synthèse, le décalage de l'âge de départ à la retraite est progressif. Par conséquent, l'impact est évalué chaque année de 2023 à 2030. Les données de 2022 sont utilisées pour cette analyse, car elles correspondent à l'année précédant l'application du décalage. L'étude se base sur l'idée que les personnes impactées par le décalage de l'âge légal de départ à la retraite travailleront plus longtemps avant de quitter leur emploi et leur contrat collectif. Par exemple, les personnes nées en 1962 travailleront 6 mois de plus, celles nées en 1963 travailleront 9 mois de plus, et ainsi de suite. Ces périodes supplémentaires d'activité impliquent que ces individus resteront plus longtemps dans le portefeuille de la mutuelle.

Dans le portefeuille, il est supposé que chaque personne atteignant l'âge légal de la retraite quitte effectivement son emploi et son contrat collectif pour partir à la retraite. En outre, une hypothèse d'homogénéité est posée pour les autres tranches d'âge, stipulant que les entrées et sorties de ces individus dans le portefeuille se produisent à un rythme similaire.

Chaque année, entre 2023 et 2030, un nombre croissant de personnes sera amené à travailler plus longtemps en raison du décalage de l'âge de départ à la retraite. Pour estimer cet impact, il est pertinent d'utiliser les données des personnes actives dont les dates de naissance correspondent aux années concernées. Ainsi, chaque année, un certain nombre de personnes sera ajouté au portefeuille, tandis qu'un autre nombre sera retiré en raison des départs à la retraite.

Il est important de noter que lorsque les personnes sont amenées à travailler plus d'une année supplémentaire, elles ne sont pas retirées du portefeuille l'année suivante. Par exemple, une personne qui travaille 1 an et 3 mois supplémentaires à partir de 2027 sera comptabilisée pour 1 an en 2027 et pour 3 mois en 2028. Ainsi, cette personne ne sera retirée du portefeuille qu'en 2029. Par conséquent, sur le graphique suivant, aucune personne n'est retirée en 2028. Les personnes retirées en 2029 correspondent donc à celles ajoutées en 2027, et ainsi de suite.

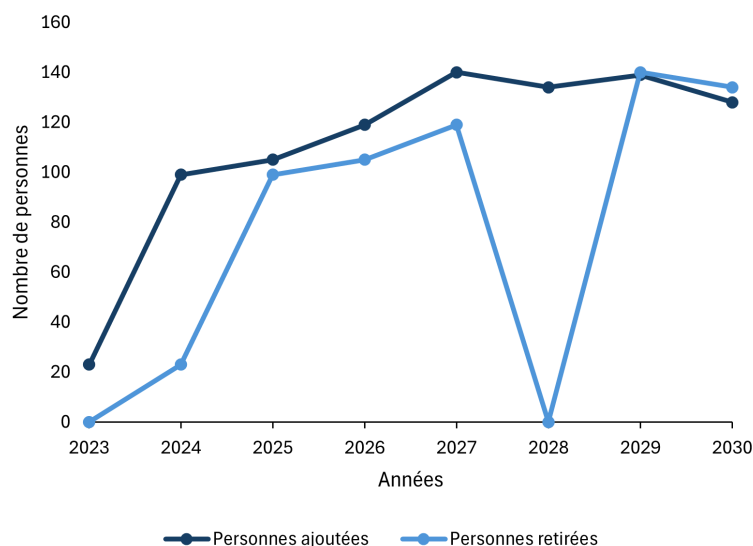


FIGURE 4 – Évolution des personnes ajoutées et retirées dans le portefeuille par année

Le calcul de l'impact prend en compte le nombre de personnes concernées, l'exposition supplémentaire (la période pendant laquelle elles restent dans le portefeuille) et la prime pure annuelle. Par exemple, en 2027, 140 personnes nées en 1965 travailleront 1 an et 3 mois de plus avant de partir à la retraite. Le montant ajouté au portefeuille pour cette année est déterminé en intégrant ces variables, ce qui donne un total de 104 417,17 € pour 2027. En revanche, le montant à retirer du portefeuille en 2027 correspond à celui ajouté en 2026, soit 88 754,59 €.

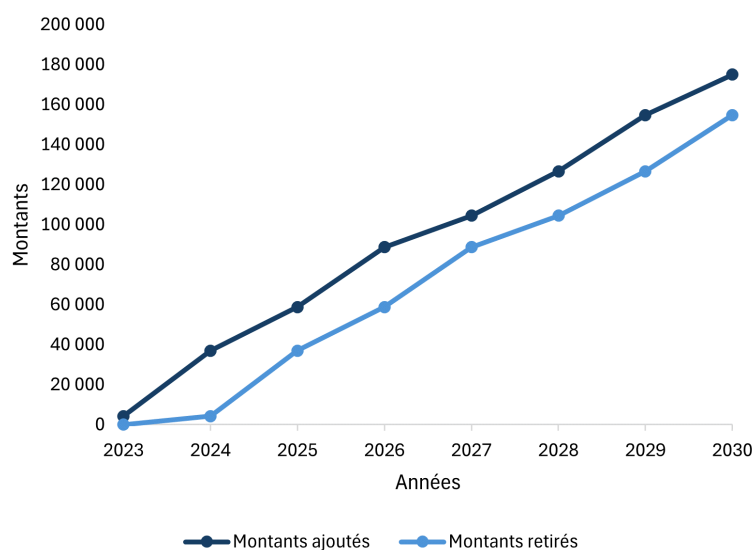


FIGURE 5 – Évolution des montants ajoutés et retirés dans le portefeuille par année

Chaque année, le portefeuille de la mutuelle est ajusté en fonction des nouvelles entrées (personnes travaillant plus longtemps) et des sorties (personnes partant à la retraite) et de leurs montants associés. Ainsi, le portefeuille est réévalué chaque année. En 2027, le montant total du portefeuille de la mutuelle est de 5 407 779,21€.

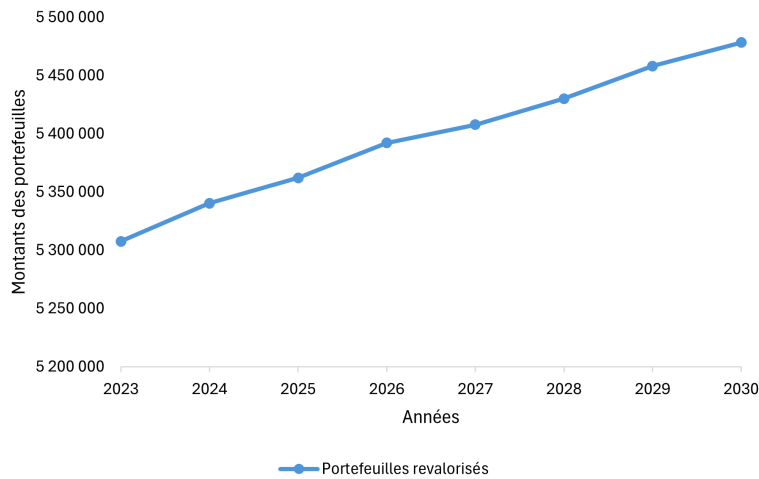


FIGURE 6 – Évolution des montants des portefeuilles par année

L'impact est mesuré par la différence en pourcentage entre le portefeuille d'une année donnée et celui de l'année précédente. Par exemple, en 2027, l'impact sur le portefeuille par rapport à 2026 est estimé à 0,29%.

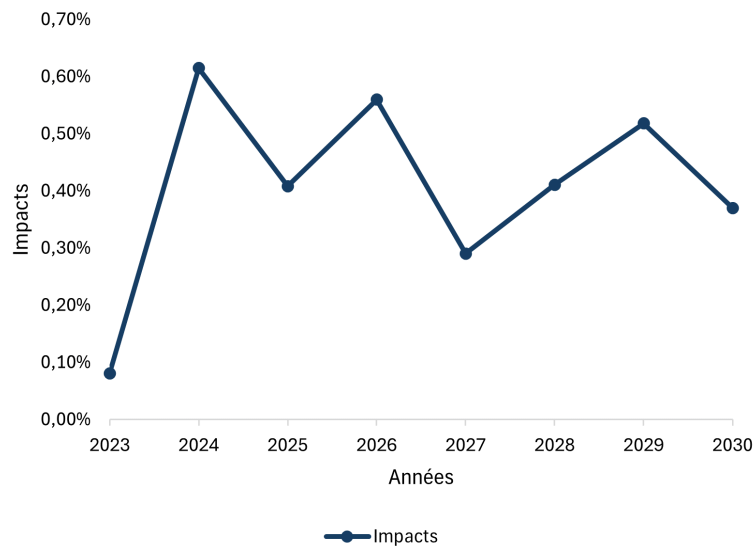


FIGURE 7 – Évolution des impacts sur le portefeuille par année

En moyenne, l'impact annuel de la réforme sur le portefeuille de la mutuelle entre 2023 et 2030 est estimé à 0,44%.

Conclusion et limites

Ce mémoire analyse l'impact de la réforme des retraites de 2023 sur les contrats collectifs d'une mutuelle généraliste en utilisant des modèles pour prédire les coûts de santé des non-cadres âgés de 63 et 64 ans. Les résultats indiquent que la réforme entraînera une légère augmentation des coûts pour la mutuelle, estimée à 0,44% par an jusqu'en 2030.

Des limites sont tout de même identifiées, notamment le manque de données. L'étude aurait pu bénéficier d'un plus grand volume de données provenant de diverses mutuelles, ce qui aurait permis de réduire les biais potentiels. Bien que les Modèles Linéaires Généralisés (GLM) aient été utilisés, d'autres approches, notamment celles basées sur le Machine Learning, pourraient fournir des résultats plus précis. De plus, les modèles réalisés prennent en compte seulement deux variables explicatives, alors que d'autres facteurs tels que le sexe, les régions ou les catégories socioprofessionnelles (agriculteurs, commerçants, etc.) pourraient affiner les résultats. Enfin, les changements de comportements liés aux autres conséquences de la réforme telles que la retraite progressive ou la pénibilité n'ont pas été pris en compte.

Bien que des limites telles que le manque de données et l'absence de certaines variables explicatives aient affecté la précision des résultats, l'étude demeure globalement pertinente. Les résultats obtenus sont considérés comme valables mais pourraient être améliorés avec des approches méthodologiques plus avancées et des données plus diversifiées dans les recherches futures.

Ce mémoire offre un aperçu de l'impact potentiel de la réforme des retraites de 2023 sur les contrats collectifs d'une mutuelle afin de développer des stratégies pour y faire face.

Synthesis Note

Context and issues

The pension reform of 2023 is responsible for raising the retirement age by two years, from from 62 to 64. This gradual shift came into effect on September 1, 2023, and applies to people born on or after September 1, 1961.

Years	Generations	Legal age + 3 months per year	Length of contribution for the full rate Acceleration of the Touraine timetable
2023	Between 01/09/1961 and 31/12/1961	62 years and 3 months	169 quarters
2024	1962	62 years and 6 months	169 quarters
2025	1963	62 years and 9 months	170 quarters
2026	1964	63 years	171 quarters
2027	1965	63 years and 3 months	172 quarters (43 years)
2028	1966	63 years and 6 months	172 quarters (43 years)
2029	1967	63 years and 9 months	172 quarters (43 years)
2030	1968	64 years	172 quarters (43 years)

TABLEAU 3 – Transitional phase from 62 to 64 years of age

This new measure has extended the length of careers, which could increase the insurer’s exposure to health-related risks. Indeed, this reform is leading to an aging workforce, forcing workers to stay longer in their jobs before retiring.

Extending working hours by two years means that more employees aged 63 and 64 will continue to benefit from group health contracts. As people age, their healthcare needs increase, as they are more likely to develop age-related health problems. These contracts, which cover employees’ medical expenses, will therefore have to cope with an increase in claims, due to the greater frequency of medical consultations, hospitalizations and treatments for this age group.

The aim of this report is to assess the potential impact of the 2023 pension reform on a portfolio of group contracts, and to quantify this impact for the coming years.

According to Garrouste.C and Perdrix.E’s article [1], which stresses that the behavior of working people differs from that of retirees, it is essential to base simulations on data from the working population. For this reason, the models presented below are based on data from working people.

Modeling

To do this, a model was run to estimate the average cost of healthcare and the frequency of healthcare use by people forced to work an extra two years as a result of the pension reform,

aged 63 and 64. Generalized linear models were applied to the data provided by the mutual insurer studied, including benefit bases for three different coverages as well as information on the beneficiaries of these benefits. It is important to note that the reimbursement amounts taken into account are limited to those paid by the mutual, excluding the portion covered by social security.

Schematics of modeling steps

The steps involved in average cost modelling are shown in the following diagram :

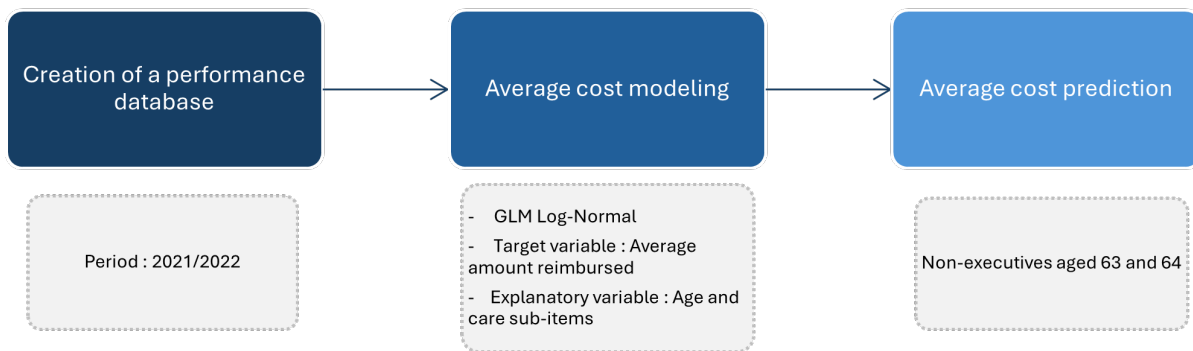


FIGURE 8 – Steps for predicting average cost

The steps involved in frequency modelling are similar to those for average cost and are shown in the following diagram :

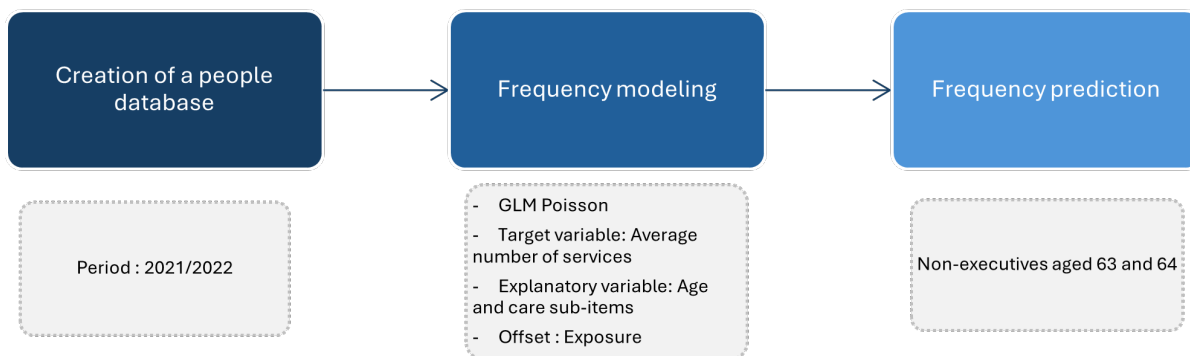


FIGURE 9 – Steps for frequency prediction

Description of modeling steps

Initially, the databases for benefits and associated persons are processed to obtain clean data for the models, in order to remove rows with missing data, for example. The study focuses on healthcare benefits and the people concerned for the years 2021 and 2022. Generalized regression models (GLM) are used to predict average cost and frequency. These two types of model are developed independently for each major healthcare category (optical, dental, outpatient, etc.).

Performance indicators, such as AIC and pseudo- R^2 , are used to select the model type. In particular, for the average cost model, the Log-Normal model is compared to the Gamma model with a link function « identity ». According to these indicators, the Log-Normal model is better suited to modeling average costs. In addition, the Poisson model is used to model frequency.

The same indicators are also used to assess the relevance of age in average cost and frequency GLMs. A comparison is made between models with and without age as an explanatory variable. The results show that the inclusion of age improves the performance of the models according to these metrics.

Next, a residual analysis is performed to examine the behavior of studentized residuals and detect any outliers. For studentized residuals, the aim is for 95% of them to lie within the range -2 to 2.

Stations	Proportions of residues between -2 and 2	
	Average cost	Frequency
Optical	94,55%	96,03%
Dental	96,15%	98,10%
Out-patient care	95,99%	70,84%
Pharmacy	96,28%	68,77%
Miscellaneous	93,42%	97,62%
Hospitalization	95,75%	84,21%
Hearing aids	97,35%	99,88%

TABLEAU 4 – Proportions of studentized residuals for average cost and frequency models for each care item

Most models have a proportion of residuals between the -2 and 2 thresholds of around 95%. However, the models for city care, pharmacy and hospitalization frequencies do not meet this standard. As a result, these models are less robust and their results less reliable. However, these models are retained for prediction purposes.

Finally, this analysis uses the models to predict the average cost and frequency for people aged 63 and 64. Predictions are made for various care sub-items, such as single lenses, dental prostheses, general medical consultations and so on.

Results

The predictions obtained are then used to calculate the pure premiums for these individuals for each care sub-item, by multiplying the average cost by the frequency for the two ages concerned. By summing the pure premiums for all care sub-items for both ages, the total annual premium per age is obtained.

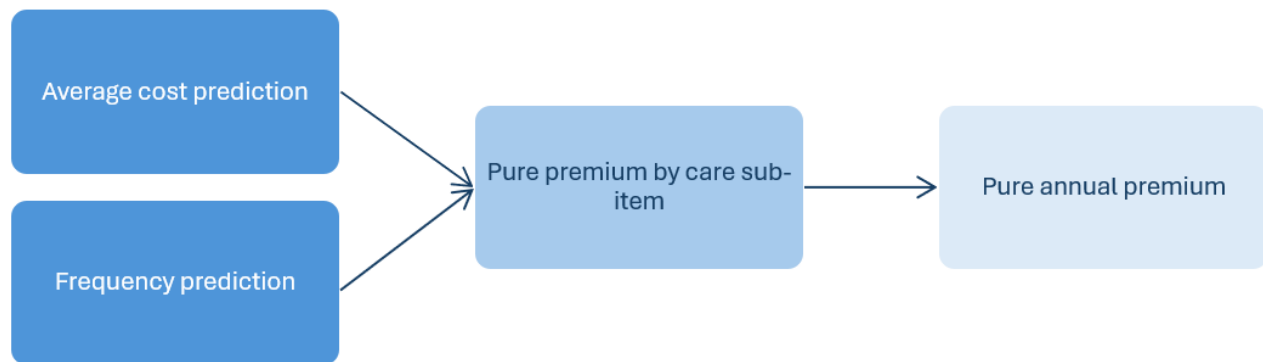


FIGURE 10 – Steps to calculate the annual pure premium for people aged 63 and 64

Pure annual premiums are €745.84 for a 63-year-old and €762.63 for a 64-year-old. On a monthly basis, these premiums are €62.15 for a 63-year-old and €63.55 for a 64-year-old.

To guarantee the reliability of the two premiums calculated, they are compared with those obtained using market pricing software. This software, using the same coverage grids as those for the benefits in the database, calculates the pure premiums for a person aged 63 and 64. The pure premiums provided by the pricing software are higher, amounting to €865.59 for a 63-year-old and €881.38 for a 64-year-old. On a monthly basis, the difference is €9.98 for a 63-year-old and €9.90 for a 64-year-old. Despite the different bases and methodologies, the results are satisfactory overall.

These two pure premiums are the basis for analyzing the impact of the 2023 pension reform on the mutual's portfolio.

Impact on the mutual's portfolio

As indicated in the first part of this summary, the shift in the retirement age is progressive. Consequently, the impact is assessed each year from 2023 to 2030. Data for 2022 are used for this analysis, as they correspond to the year preceding application of the shift. The study is based on the idea that people affected by the shift in retirement age will work longer before leaving their jobs and group contracts. For example, people born in 1962 will work 6 months longer, those born in 1963 will work 9 months longer, and so on. These additional periods of activity mean that these individuals will remain in the mutual's portfolio longer.

In the portfolio, it is assumed that each person reaching legal retirement age actually leaves his or her job and group contract to retire. In addition, a homogeneity assumption is made for the other age brackets, stipulating that these individuals enter and leave the portfolio at a similar rate.

Each year, between 2023 and 2030, a growing number of people will be required to work longer as a result of the shift in retirement age. To estimate this impact, it is appropriate to use data from working people whose dates of birth correspond to the years concerned. Thus, each year, a certain number of people will be added to the portfolio, while another number will be removed due to retirements.

It is important to note that when people work for more than one additional year, they are not removed from the portfolio the following year. For example, a person who works an additional 1 year and 3 months from 2027 will be counted for 1 year in 2027 and for 3 months in 2028. This person will not be removed from the portfolio until 2029. Consequently, in the following graph, no person is removed in 2028. The people removed in 2029 therefore correspond to those added in 2027, and so on.

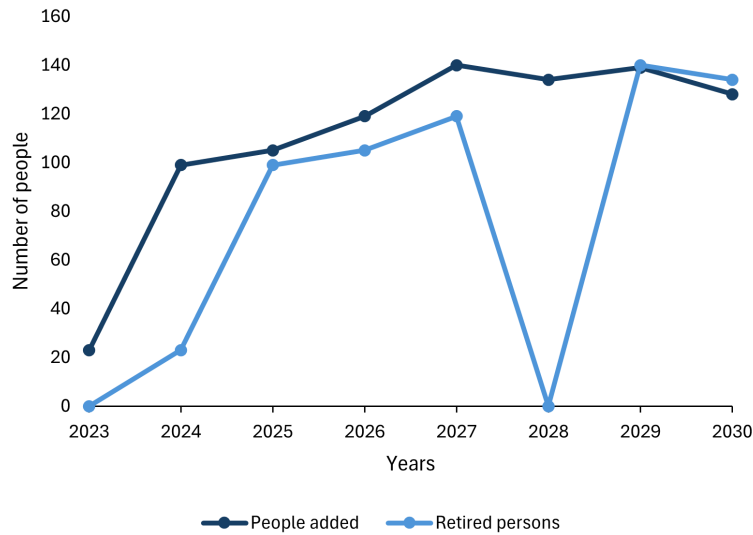


FIGURE 11 – Evolution of people added to and removed from the portfolio by year

The impact calculation takes into account the number of people concerned, the additional exposure (the period during which they remain in the portfolio) and the annual pure premium. For example, in 2027, 140 people born in 1965 will work 1 year and 3 months longer before retiring. The amount added to the portfolio for that year is determined by integrating these variables, giving a total of €104,417.17 for 2027. On the other hand, the amount to be withdrawn from the portfolio in 2027 corresponds to the amount added in 2026, i.e. €88,754.59.

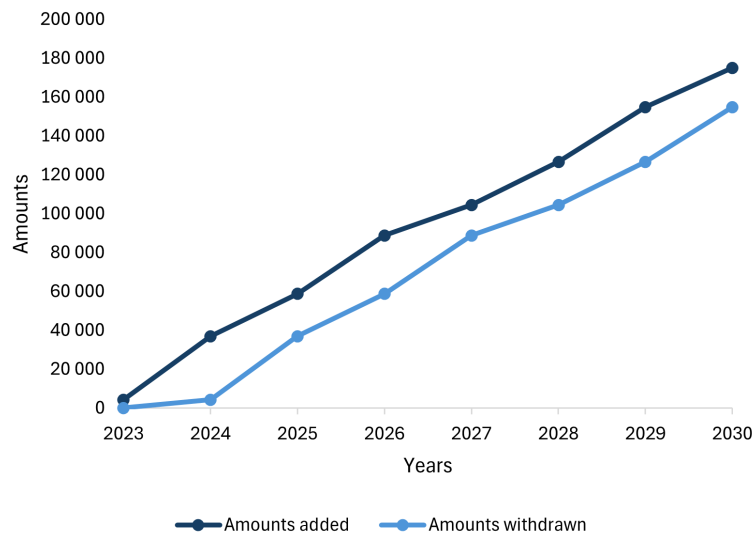


FIGURE 12 – Amounts added to and withdrawn from the portfolio by year

Each year, the mutual’s portfolio is adjusted to take account of new entries (people working longer) and exits (people retiring) and their associated amounts. In this way, the portfolio is revalued each year. In 2027, the total amount of the mutual’s portfolio is €5,407,779.21.

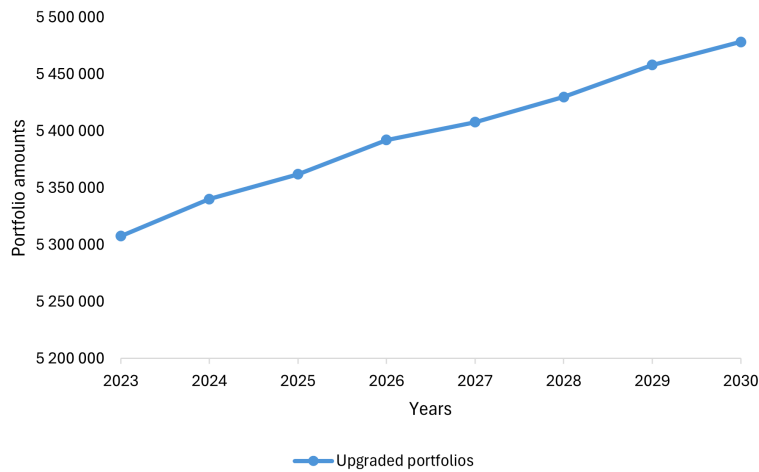


FIGURE 13 – Portfolio amounts by year

The impact is measured by the percentage difference between a given year’s portfolio and that of the previous year. For example, in 2027, the impact on the portfolio compared with 2026 is estimated at 0.29%.

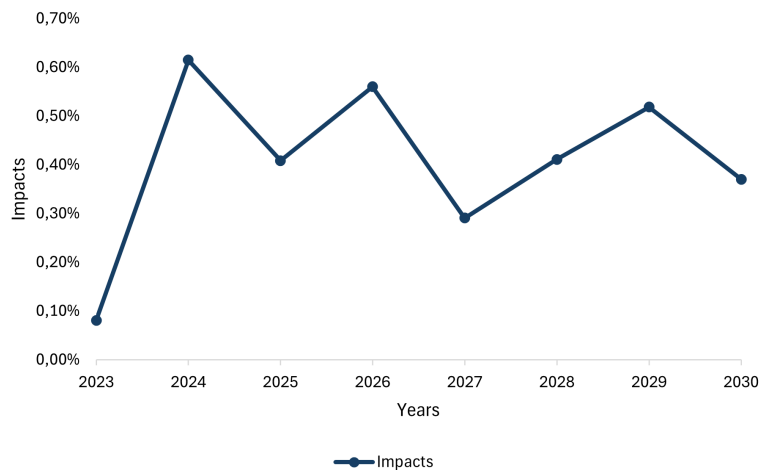


FIGURE 14 – Impact on the portfolio by year

On average, the annual impact of the reform on the mutual’s portfolio between 2023 and 2030 is estimated at 0.44%.

Conclusion and limits

This thesis analyzes the impact of the 2023 pension reform on the group contracts of a generalist mutual insurance company, using models to predict the healthcare costs of non-managers aged 63 and 64. The results indicate that the reform will lead to a slight increase in costs for the mutual, estimated at 0.44% per year until 2030.

Limitations were nevertheless identified, notably the lack of data. The study could have benefited

from a larger volume of data from various mutuals, which would have reduced potential biases. Although Generalized Linear Models (GLM) were used, other approaches, notably those based on Machine Learning, could provide more accurate results. What's more, the models carried out take into account only two explanatory variables, whereas other factors such as gender, regions or socio-professional categories (farmers, shopkeepers, etc.) could refine the results. Finally, changes in behavior linked to other consequences of the reform, such as phased retirement or arduous work, have not been taken into account.

Although limitations, such as lack of data and the absence of certain explanatory variables, affected the precision of the results, the study remains relevant overall. The results obtained are considered valid, but could be improved with more advanced methodological approaches and more diversified data in future research.

This brief provides an overview of the potential impact of the 2023 pension reform on a mutual's group contracts, with a view to developing strategies for dealing with it.

Remerciements

Tout d'abord, je souhaite exprimer ma gratitude à David Fitouchi, Frank Boukobza, Louis-Anselme de Lamaze et Camille Blanc-Vannet, les associés du cabinet Actuelia, pour m'avoir offert l'opportunité d'intégrer leur équipe et de mener à bien ce mémoire.

Je tiens également à remercier Chloé Blanchard et Ilan Assayag, mes tuteurs chez Actuelia, pour leur accompagnement tout au long de mon alternance et pour leur soutien précieux dans la réalisation de ce travail. Leur engagement et leurs conseils avisés ont été particulièrement appréciés.

Je remercie chaleureusement tous les collaborateurs du cabinet pour leur bonne humeur quotidienne et leurs conseils. Une mention particulière à mes collègues alternants pour leur soutien indéfectible durant les moments les plus difficiles, avec une pensée spéciale pour Paule Fomen, dont l'aide a été très précieuse. Un grand merci également à Basile Saisset pour son engagement et ses idées novatrices.

Je tiens également à exprimer ma reconnaissance à Guillaume Biessy, mon tuteur académique, pour ses conseils constructifs et ses suggestions de grande qualité. Mes remerciements vont aussi à l'ensemble des professeurs de l'ISUP pour la qualité de l'enseignement dispensé.

Enfin, du fond du cœur, je remercie ceux qui m'ont toujours soutenu tout au long de ma scolarité : mes parents, ma famille et mes amis. Mention spéciale à mes parents pour leur soutien constant, qui a été essentiel à la réalisation de ce mémoire.

Table des matières

Résumé	i
Abstract	i
Note de Synthèse	ii
Synthesis Note	x
Remerciements	xviii
Introduction	1
1 Le système de santé en France	3
1.1 La sécurité sociale	3
1.1.1 Définition et contexte	3
1.1.2 Mode de financement	3
1.1.3 Branches du régime général	3
1.2 Les organismes de santé privés	5
1.2.1 Définition	5
1.2.2 Mutuelles	5
1.2.3 Sociétés d'assurance	6
1.2.4 Institutions de prévoyance	6
1.3 Les régimes complémentaires	6
1.3.1 Contrats individuels	7
1.3.2 Contrats collectifs obligatoires	7
1.3.3 Contrats collectifs facultatifs des fonctionnaires	8
1.4 L'évolution récente des contrats collectifs	8
2 Organisation du système de retraite	9
2.1 Le système de retraite en France	9
2.1.1 Contexte	9
2.1.2 Définitions et principes	9
2.1.2.1 Caractère obligatoire	9
2.1.2.2 Système par répartition	9
2.1.2.3 Mode de financement	9
2.1.3 Les précédentes réformes de retraite en France	10
2.1.3.1 Réforme Balladur (1993)	10
2.1.3.2 Rapport Briet (1995) et rapport Charpin (1999)	11
2.1.3.3 Réforme Fillon (2003)	11

2.1.3.4	Réforme Woerth (2010)	11
2.1.3.5	Réforme Touraine (2014)	12
2.2	La réforme Macron (2023)	12
2.2.1	Pension minimale	12
2.2.2	Régimes spéciaux	12
2.2.3	Retraite progressive	13
2.2.4	Cumul emploi retraite	13
2.2.5	Mesures relatives aux droits des mères de famille	13
2.2.6	Pénibilité	13
2.2.7	Aidants	13
2.2.8	Suivi dans le temps	13
2.3	Le décalage de l'âge de départ à la retraite dans les autres pays	13
2.3.1	Viellissement des populations	14
2.3.2	Âge de départ à la retraite : de 62 à 67 ans	14
2.3.3	Taux de remplacement	14
2.4	Les conséquences de la réforme 2023 sur les complémentaires santé d'entreprise	15
3	Présentation des données utilisées et analyse descriptive	16
3.1	Présentation des bases de données	16
3.1.1	Bases de prestations	16
3.1.2	Bases de personnes	17
3.2	Traitement des données	17
3.2.1	Les bases de données étudiées	18
3.2.1.1	Base Prestations améliorée	18
3.2.1.2	Base Personnes améliorée	18
3.2.2	Variables corrigées	18
3.2.3	Variables créées	18
3.2.3.1	Base Prestations	18
3.2.3.2	Base Personnes	20
3.2.4	Prétraitement	21
3.2.4.1	Données manquantes	21
3.2.4.2	Valeurs aberrantes	22
3.3	Analyse descriptive	23
3.3.1	Présentation générale des données	23
3.3.2	Zoom sur les données utilisées pour l'étude	23
3.3.2.1	Évolution entre 2021 et 2022	23
3.3.2.2	Régions de provenance des assurés	25
3.3.2.3	Types de situations des assurés	25
3.3.2.4	Catégories Socioprofessionnelles des assurés	26
3.3.2.5	Répartition des âges des assurés	27
3.3.3	Présentation des différents postes de soins	30
3.3.4	Présentation des sous postes de soins	32
3.3.4.1	Optique	33
3.3.4.2	Dentaire	34
3.3.4.3	Soins de ville	35
3.3.4.4	Pharmacie	36
3.3.4.5	Divers	37

3.3.4.6	Hospitalisation	38
3.3.4.7	Aides auditives	39
3.3.5	Comparaison du comportement entre les personnes actives et retraitées	39
4	Projection du décalage de l'âge de départ à la retraite à partir d'un GLM	42
4.1	La tarification a priori	42
4.2	Modèles Linéaires Gaussiens	43
4.2.1	Définition	43
4.2.2	Estimation des paramètres	44
4.2.2.1	Estimation par Maximum de Vraisemblance	44
4.2.2.2	Estimation par la méthode des moindres carrés	44
4.3	Modèles Linéaires Généralisés (GLM)	44
4.3.1	Familles exponentielles	45
4.3.2	Composantes d'un GLM	46
4.3.2.1	Composante aléatoire	46
4.3.2.2	Composante déterministe	46
4.3.2.3	Fonction de lien	46
4.3.3	Définition	46
4.3.4	Estimation des paramètres	46
4.3.4.1	Estimation de β	47
4.3.5	Sélection de modèles et déviance	48
4.3.5.1	Déviance	49
4.3.5.2	Théorème de Wilks	50
4.3.5.3	Mesures de performances	50
4.3.6	Interprétation d'une sortie R d'un GLM	51
4.3.7	Importance de l'analyse des résidus	52
4.4	Théorie des modèles utilisés dans l'étude	53
4.4.1	GLM Log-Normal	53
4.4.1.1	Définition	53
4.4.1.2	Fonction de lien	54
4.4.2	GLM Poisson	54
4.4.2.1	Définition	54
4.4.2.2	Fonction de lien	54
4.5	Présentation des objectifs de modélisation	55
4.6	Construction des bases de données	55
4.6.1	Base de données pour le GLM de coût moyen	55
4.6.2	Base de données pour le GLM de fréquence	55
4.6.3	Séparation des bases de données	56
4.7	Mise en place des modèles GLM	56
4.7.1	GLM de coût moyen	56
4.7.1.1	Sélection des variables	56
4.7.1.2	Choix du modèle	56
4.7.1.3	Pertinence de l'âge	58
4.7.2	GLM de fréquence	59
4.7.2.1	Sélection des variables	59
4.7.2.2	Choix du modèle	59
4.7.2.3	Pertinence de l'âge	60

4.8	Diagnostic des modèles	61
4.8.1	Analyse des p-values	61
4.8.1.1	Optique	61
4.8.1.2	Dentaire	62
4.8.1.3	Soins de ville	62
4.8.1.4	Pharmacie	63
4.8.1.5	Divers	63
4.8.1.6	Hospitalisation	63
4.8.1.7	Aides auditives	64
4.8.2	Analyse des résidus	64
4.8.2.1	Résidus studentisés du poste de soins Optique	65
4.8.2.2	Résultat des proportions de résidus studentisés pour tous les postes de soins	65
4.9	Prédictions	66
4.9.1	Résultats des prédictions pour des actifs de 50 ans	66
4.9.2	Résultats des prédictions pour des actifs de 63 et 64 ans	67
4.9.2.1	Optique	67
4.9.2.2	Dentaire	68
4.9.2.3	Soins de ville	68
4.9.2.4	Pharmacie	69
4.9.2.5	Divers	69
4.9.2.6	Hospitalisation	70
4.9.2.7	Aides auditives	70
5	Application à la tarification	72
5.1	Présentation des résultats appliqués à la tarification	72
5.2	Présentation de l'outil de tarification marché	73
5.2.1	Fonctionnement de l'outil	73
5.2.1.1	Démographie	73
5.2.1.2	Remplissage d'une grille de garantie	74
5.2.1.3	Visualisation des résultats	74
5.2.2	Grille de tarification utilisée	75
5.3	Analyse des résultats et comparaison avec la tarification du logiciel	76
5.3.1	63 ans	76
5.3.2	64 ans	77
5.4	Impact sur le portefeuille de la mutuelle	78
5.4.1	Données utilisées	79
5.4.2	Hypothèses	79
5.4.2.1	Hypothèses globales	79
5.4.2.2	Hypothèses sur le nombre de personnes ajouté par an	80
5.4.3	Impact sur le portefeuille à horizon 2030	82
5.5	Limites de l'étude et ouverture	83
5.5.1	Données utilisées	83
5.5.1.1	Données manquantes	84
5.5.1.2	Retraitement des montants de l'hospitalisation	84
5.5.1.3	Exclusion des enfants	84
5.5.1.4	Distinction des années des bases de données	84

5.5.1.5	Régions de provenance des assurés	84
5.5.1.6	Type de situations des assurés	85
5.5.1.7	Catégories socioprofessionnelles des assurés	85
5.5.1.8	Postes de soins	85
5.5.1.9	Comparaison du comportement des actifs et des retraités	85
5.5.2	Modèles utilisés	86
5.5.2.1	Choix des modèles	86
5.5.2.2	Indicateurs de performances des modèles	86
5.5.2.3	Prédictions	86
5.5.3	Impact sur le portefeuille	86
5.5.4	Conclusion sur les limites	87
Conclusion		87
Annexes		90
A Les taux de cotisations de droit commun au 1er janvier 2023		91
B Graphiques des résidus studentisés		92
B.1	Dentaire	92
B.2	Soins de ville	92
B.3	Pharmacie	93
B.4	Divers	93
B.5	Hospitalisation	93
B.6	Aides auditives	94
C Détail de la comparaison avec le logiciel de tarification par sous-postes de soins		95
Liste des figures		97
Liste des tableaux		99
Liste des équations		101
Bibliographie		104

Introduction

La réforme des retraites de 2023, qui repousse l'âge de départ de 62 à 64 ans, engendre des changements significatifs à plusieurs niveaux. À partir du 1^{er} septembre 2023, l'âge légal de départ à la retraite sera progressivement reculé à raison de trois mois par année de naissance, jusqu'à atteindre l'objectif de 64 ans en 2030. Ce recul de l'âge de départ ne se limite pas à un impact individuel, il affecte également les dynamiques au sein des entreprises.

Les entreprises offrant des contrats de santé collectifs devront adapter leurs dispositifs en réponse à la réforme. Avec l'allongement de la durée de vie active, un plus grand nombre de salariés restera couvert par ces contrats pendant une période prolongée, ce qui pourrait entraîner une augmentation du nombre d'assurés notamment les plus consommant et une évolution des conditions de couverture. Les entreprises pourraient ainsi être amenées à ajuster les garanties de leurs contrats pour mieux répondre aux besoins spécifiques d'une population active vieillissante. La réforme pourrait également entraîner une augmentation des coûts associés à ces contrats de santé, en raison de la hausse potentielle des dépenses de santé des salariés âgés. Ces implications soulèvent des enjeux importants tant pour les employeurs et les organismes qui devront gérer ces nouvelles réalités, que pour les salariés.

Selon un communiqué de presse de Verspieren [2], la proportion d'assurés actifs de plus de 60 ans devrait croître progressivement, avec une augmentation prévue de 65% d'ici 2030. Actuellement, les plus de 60 ans représentent 7% de la population couverte, mais cette proportion pourrait atteindre 10% à terme. Cette croissance du nombre de bénéficiaires de plus de 60 ans au cours des huit prochaines années devrait s'accompagner d'une hausse de la consommation médicale. Toujours d'après ce communiqué de presse, la consommation des personnes de plus de 60 ans est supérieure de 63% à la moyenne, atteignant 750 € contre 460 € en moyenne pour l'ensemble des tranches d'âge. Cette augmentation devrait entraîner une hausse de 2,70% des coûts associés aux contrats de frais de santé sur une période de 8 ans, soit 0,34% par an. Pour faire face à cette tendance, les entreprises devront s'adapter en envisageant soit une hausse des cotisations des assurés soit une révision des garanties proposées.

Avant de réaliser ce mémoire, plusieurs recherches ont été menées sur des travaux d'actuariat antérieurs portant sur les impacts des réformes des retraites précédentes. De nombreux mémoires évaluent en effet les conséquences de ces réformes, principalement sur la prévoyance, à savoir l'invalidité et l'incapacité, ainsi que, dans une moindre mesure, sur la santé. Les mémoires de Ballo H. [3], Maury A. [4], Lopez D. [5] et Lim R. [6] constituent des exemples significatifs qui ont renforcé l'intérêt d'approfondir, dans ce mémoire, l'analyse de l'impact des réformes des retraites sur les prestations de santé.

L'objectif de ce mémoire est donc d'évaluer l'impact de la réforme des retraites de 2023 sur les contrats collectifs de santé, en se basant sur les données d'une mutuelle standard.

Dans un premier temps, il s'agira de calculer les primes pures pour les personnes travaillant deux années supplémentaires. Ensuite, une analyse sera effectuée pour évaluer l'impact de cette réforme sur le portefeuille de la mutuelle concernée, jusqu'en 2030. L'objectif est de déterminer l'effet annuel de la réforme sur la mutuelle, depuis sa mise en œuvre en 2023 jusqu'en 2030.

Pour ce faire, une méthode classique de modélisation actuarielle, la régression logistique, sera employée. Cette technique est largement utilisée pour identifier les relations entre divers facteurs de données et pour prédire des valeurs en fonction de ces facteurs. Elle permettra de modéliser et de calculer les primes pures nécessaires pour analyser l'impact.

Les données utilisées proviennent d'une mutuelle offrant des contrats collectifs avec diverses couvertures de soins. Ces données seront d'abord traitées et nettoyées avant d'être intégrées dans les modèles mathématiques élaborés.

Chapitre 1

Le système de santé en France

1.1 La sécurité sociale

Cette section s'appuie sur plusieurs sources, notamment les sites du Service Public [7], de la Sécurité Sociale [8], du portail Ameli (Assurance Maladie en ligne) [9] et de la Cour des Comptes [10].

1.1.1 Définition et contexte

Créée en 1945, la Sécurité sociale constitue le pilier central du système de santé en France. Elle regroupe un ensemble d'organismes visant à protéger les résidents du territoire français contre les aléas des risques sociaux tels que la maladie, l'incapacité de travail, l'invalidité, la vieillesse, le décès, les charges familiales et le chômage. Cette protection est assurée par l'affiliation obligatoire des assurés sociaux et de leurs ayants droit à l'un des régimes de la Sécurité sociale, déterminé en fonction de leur situation professionnelle.

1.1.2 Mode de financement

Le financement de la Sécurité sociale repose principalement sur les cotisations sociales, détaillées en Annexe [A]. Ces cotisations sont prélevées sur les revenus du travail salarié, avec une part à la charge des employés et une part à la charge des employeurs (cotisations sociales patronales). Les impôts et les cotisations sur d'autres types de revenus, notamment la Contribution Sociale Généralisée (CSG), constituent l'essentiel de l'autre part du financement. La Sécurité sociale distribue ensuite les prestations prévues par le Code de la Sécurité sociale.

1.1.3 Branches du régime général

Les prestations couvertes par la Sécurité sociale englobent les domaines de la famille, de la santé, des retraites et de l'autonomie. Ainsi, la Sécurité sociale en France se divise en plusieurs branches au sein du régime général :

- **La branche maladie**, administrée par la Caisse nationale de l'assurance maladie (Cnam), prend en charge les frais d'hospitalisation, les dépenses liées aux médicaments, les consultations chez les professionnels de santé, ainsi que le versement d'indemnités journalières en cas d'arrêt de travail ;
- **La branche accident du travail**, également gérée par la Caisse nationale de l'assurance maladie (Cnam), couvre les frais liés aux maladies professionnelles et aux accidents du travail ;

- **La branche famille**, administrée par la Caisse des allocations familiales (CAF) et ses organismes locaux, verse divers types de prestations, notamment celles liées à la naissance, à la garde d’enfants, au logement, ainsi qu’à la solidarité, comme la gestion du Revenu de Solidarité Active (RSA) ;
- **La branche retraite**, gérée par la Caisse nationale d’assurance vieillesse (Cnav) et ses organismes locaux, assure l’enregistrement des revenus sur le compte vieillesse de chaque individu tout au long de sa vie active, à partir des déclarations de son employeur. Cela permet ensuite de calculer le montant des pensions de retraite et de les verser aux bénéficiaires ;
- **La branche autonomie**, au sein des Caisses nationales de solidarité pour l’autonomie, veille à garantir l’équité territoriale, la qualité et l’efficacité de l’accompagnement des personnes âgées et des personnes handicapées, quel que soit leur lieu de vie ;
- **La branche recouvrement**, responsable de la collecte des recettes, est gérée par l’Agence centrale des organismes de sécurité sociale. Au niveau local, cette tâche est confiée aux Unions de recouvrement des cotisations de sécurité sociale et des allocations familiales (URSSAF).

Ces branches composent le régime général de la Sécurité sociale. À cela s’ajoutent le régime agricole, géré par la Mutualité Sociale Agricole (MSA), ainsi que divers régimes spéciaux, qui complètent la couverture de la branche maladie.

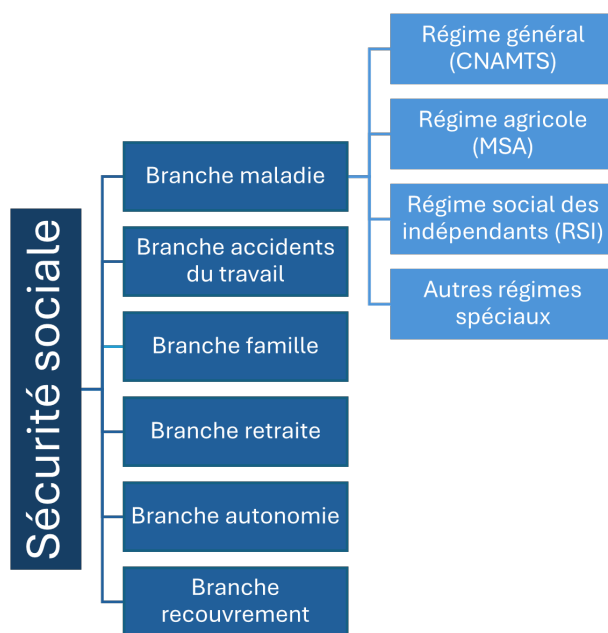


FIGURE 1.1 – L’organisation de la sécurité sociale

Dans ce mémoire, seulement la branche maladie sera étudiée.

Tous les résidents en France participent au financement de la Sécurité sociale, qui assure ensuite la couverture de l’ensemble de la population. La contribution de chaque individu est déterminée en fonction de ses revenus et non de ses risques personnels. Les trois principes fondamentaux de ce système sont les suivants :

- L’égalité d’accès aux soins ;
- La qualité des soins ;

- La solidarité (principe du pot commun).

Selon l'article de la DREES [11], la France est le pays d'Europe qui consacre le plus grand pourcentage de son produit intérieur brut (PIB) à la protection sociale pour la 7e année consécutive. En 2022, ce taux s'élève à 32,2%, devant l'Italie (29,6%), l'Allemagne (29,2%) et le Danemark (27%).

1.2 Les organismes de santé privés

Cette partie s'inspire de plusieurs sources, notamment du site du ministère de l'Économie et des Finances [12], ainsi que des fiches 6 [13] et 7 [14] de la DREES.

1.2.1 Définition

Pour compléter les remboursements assurés par la Sécurité sociale, il est possible de souscrire une assurance complémentaire auprès d'un organisme privé, appelé régime complémentaire de santé. Ces assurances constituent le deuxième financeur des dépenses de santé en France.

L'adhésion aux régimes d'assurance complémentaire maladie peut être facultative sur une base individuelle ou obligatoire ou facultative sur une base collective via leur entreprise. En général, ces organismes d'assurance couvrent au minimum l'intégralité du ticket modérateur (la part restant à la charge du patient après remboursement par l'assurance maladie et avant déduction des participations forfaitaires) ainsi que le forfait hospitalier. Cependant, la couverture des dépassements d'honoraires pratiqués par les médecins varie en fonction des contrats et n'est pas systématiquement incluse.

Il existe trois types d'organismes d'assurance privés, chacun ayant des modes de fonctionnement distincts et une couverture des risques différente :

- Les mutuelles relevant du code de la Mutualité ;
- Les sociétés d'assurance relevant du code des Assurances ;
- Les institutions de prévoyance relevant du code de la Sécurité sociale.

1.2.2 Mutuelles

Les mutuelles et groupements mutualistes, également appelés mutuelles de « 45 », comptent 272 organismes complémentaires santé sur environ 400. Selon l'article de la DREES [15], elles restent dominantes sur le marché de l'assurance santé, représentant 47% des cotisations collectées en 2021.

Toutefois, leur part de marché a diminué depuis 2001 au profit des compagnies d'assurance. En 2021, la masse des cotisations collectées pour la santé a augmenté de 1,1 milliard d'euros (+5,8%), malgré une perte de parts de marché des mutuelles depuis 2012.

Originaires de la charte de la mutualité de 1898, ces mutuelles et groupements, qui fonctionnent selon le principe d'autogestion, sont des organismes privés à but non lucratif. Elles ne peuvent pas pratiquer de discrimination en fonction de l'état de santé des adhérents.

Elles financent une couverture sociale complémentaire à celle de la Sécurité sociale ainsi que des

actions de prévoyance, de solidarité et d'entraide, en utilisant des cotisations dont le montant ne dépend pas des caractéristiques individuelles des adhérents. Les remboursements maladie constituent les deux tiers de leur activité, mais elles peuvent également gérer des risques à long terme (vieillesse, invalidité, décès) et proposer des opérations de prévoyance d'entreprise.

Dans ce mémoire, les données utilisées proviennent d'une mutuelle.

1.2.3 Sociétés d'assurance

Les sociétés d'assurance sont des organismes complémentaires à but lucratif, contrairement aux mutuelles et aux institutions de prévoyance qui ont une mission sociale. Ces entreprises, axées sur le capital, visent la réalisation de bénéfices et la redistribution de ceux-ci à leurs actionnaires. Elles sont donc exclues des comptes de la protection sociale.

Toutefois, les mêmes règles prudentielles s'appliquent à tous les régimes complémentaires. Selon la DREES [15], environ 100 de ces sociétés, sur un total de 397, collectent 36% des cotisations santé en 2022, soit en moyenne 165 millions d'euros.

Elles offrent des tarifs variés en fonction des risques individuels tels que l'âge, le lieu de résidence et l'état de santé. Une société d'assurance, sous réserve d'avoir obtenu les agréments nécessaires, peut exercer des activités d'assurance vie ou non vie. Toutefois, en vertu du principe de spécialité, elle ne peut combiner ces deux types d'assurance au sein de la même entreprise et doit alors confier l'activité non liée à l'assurance vie à une société distincte.

1.2.4 Institutions de prévoyance

Les institutions de prévoyance sont généralement de grandes entités, au nombre d'environ 25. Selon la DREES [15], elles représentent en moyenne 283 millions d'euros de cotisations dans le marché des complémentaires santé en 2022, contre 84 millions d'euros pour les mutuelles.

Ces organismes de droit privé à but non lucratif offrent une couverture complémentaire et sont gérés de manière paritaire par des représentants des salariés adhérents et des entreprises employeurs. Elles se spécialisent dans la protection complémentaire au niveau de la prévoyance et de la retraite, et exercent exclusivement dans le domaine des assurances de personnes, couvrant les risques tels que la retraite, le décès, l'incapacité et l'invalidité.

1.3 Les régimes complémentaires

Cette section s'appuie sur diverses sources, notamment les sites des assurances Alan [16] et APICIL [17], ainsi que le portail de la fonction publique [18].

Il existe trois types de régimes complémentaires. La complémentaire individuelle, la complémentaire collective obligatoire et la complémentaire collective facultative.

1.3.1 Contrats individuels

Les contrats individuels sont principalement conçus pour les étudiants, les travailleurs indépendants et les retraités. Le souscripteur doit être une personne physique. Ce contrat offre des garanties à la fois au souscripteur et éventuellement à ses ayants droit, moyennant le paiement d'une cotisation qu'il finance lui-même auprès d'un organisme assureur.

Il existe une grande variété de contrats d'assurance santé individuels, permettant à chacun de choisir une mutuelle adaptée à ses capacités financières et aux besoins de sa famille à couvrir.

1.3.2 Contrats collectifs obligatoires

Le contrat collectif de mutuelle santé est établi entre un groupe de personnes et un assureur. Dans ce cas, ce n'est pas l'individu mais une entreprise ou une association qui choisit et négocie le contrat. Le souscripteur, généralement un employeur ou une organisation, peut offrir à ses salariés et à leurs ayants droit les garanties du contrat d'assurance, financées par une cotisation partagée entre l'employeur et le salarié.

Depuis l'entrée en vigueur de la loi ANI en 2016, les entreprises du secteur privé sont tenues de proposer un contrat collectif de mutuelle à l'ensemble de leurs salariés. Dans le cadre de cette législation, les employeurs doivent se conformer à plusieurs obligations :

- Mettre en place une mutuelle collective obligatoire pour tous les salariés. Cependant, la loi prévoit certaines exceptions permettant aux salariés de ne pas adhérer : ceux déjà couverts par le contrat collectif de leur conjoint, les salariés à temps partiel, les apprentis, ainsi que les contrats à durée déterminée (CDD) de moins de 3 mois, entre autres.
- Financer au moins 50% des cotisations des salariés. L'employeur a la possibilité de contribuer davantage, jusqu'à 100% des cotisations, s'il le souhaite.
- Proposer des garanties minimales, appelées le « panier de soins ». Le contrat collectif d'assurance santé doit couvrir les éléments suivants :
 - L'intégralité du ticket modérateur sur les consultations, les actes et les prestations remboursables par l'assurance maladie ;
 - Le forfait journalier hospitalier en cas d'hospitalisation ;
 - Les frais dentaires, y compris les prothèses et l'orthodontie, jusqu'à hauteur de 125% du tarif conventionnel ;
 - Les frais d'optique, remboursés de manière forfaitaire tous les 2 ans (ou annuellement pour les enfants et en cas de modification de la vue).
- Maintenir les droits de garantie en cas de départ. Lorsqu'un salarié quitte l'entreprise, il peut conserver sa complémentaire santé collective grâce à un mécanisme appelé portabilité de la mutuelle. Toutefois, plusieurs conditions doivent être remplies :
 - La rupture du contrat de travail ne doit pas résulter d'une faute lourde ;
 - Le motif de la cessation du contrat de travail doit permettre de bénéficier des indemnités de France Travail, comme en cas de rupture conventionnelle, de démission légitime, de licenciement sans faute grave ou de fin de contrat à durée déterminée (CDD) ;
 - Le salarié avait adhéré à la complémentaire santé d'entreprise.

Si le salarié remplit ces conditions, il peut conserver ses droits de garantie pendant une durée maximale de 12 mois. Les ayants droit bénéficient également du maintien des garanties. La portabilité de la mutuelle est gratuite pour l'ex-salarié ainsi que pour l'entreprise.

Par ailleurs, la protection sociale complémentaire d'entreprise offre des tarifs collectifs généralement moins élevés que ceux des contrats individuels, grâce à l'effet de mutualisation des risques. De plus, chaque salarié est tenu d'adhérer au régime, même s'il n'a pas de besoins spécifiques en matière de santé.

1.3.3 Contrats collectifs facultatifs des fonctionnaires

Le fonctionnaire titulaire ne bénéficie pas d'un dispositif collectif comme dans le secteur privé et n'est donc pas obligé de souscrire une complémentaire santé, mais cela est fortement recommandé. Bien que la règle soit celle du contrat individuel, certaines dispositions la rendent plus flexible. En effet, certains employeurs publics apportent leur soutien pour aider leurs agents à souscrire des mutuelles. Les organismes référencés dans la fonction publique d'État ou labellisés par les collectivités territoriales peuvent offrir des solutions avantageuses.

Cependant, avec l'annonce de la réforme de la Protection Sociale Complémentaire (PSC) des fonctionnaires, la participation financière de l'employeur deviendra obligatoire pour tous les fonctionnaires, quel que soit leur statut. Les contrats de PSC santé seront progressivement instaurés à partir du 1er janvier 2025, en fonction des différents employeurs.

1.4 L'évolution récente des contrats collectifs

Selon un article de FIDUCIAL [19], le coût des mutuelles santé a constamment augmenté ces dernières années, en raison de l'incapacité de la Sécurité Sociale à couvrir l'ensemble des dépenses de santé, ce qui se traduit par une hausse du reste à charge pour les assurés. En 2024, cette situation se manifeste par une augmentation des franchises médicales et des participations forfaitaires, afin de maintenir une prise en charge totale des cas les plus graves. Les entreprises ne sont pas épargnées et doivent assumer une part importante des coûts des complémentaires santé pour leurs employés. Une enquête de la Mutualité Française prévoit une hausse moyenne des cotisations de 8,1% en 2024.

Cette hausse continue des cotisations place de nombreux assurés face à des différences financières croissantes pour accéder à des soins médicaux de qualité. Cette pression accrue sur le budget santé pousse les entreprises à réévaluer leurs politiques d'assurance collective, ce qui peut entraîner une diminution des avantages offerts aux employés ou un transfert de charges vers ces derniers. De plus, l'augmentation des charges sociales peut impacter la rentabilité des entreprises, notamment des Petites et Moyennes Entreprises (PME) et des Très Petites Entreprises (TPE), et affecter leur compétitivité en raison de politiques d'assurance santé plus avantageuses. Les entreprises doivent donc trouver un équilibre entre la qualité de la couverture santé offerte et son coût, tout en restant compétitives dans le domaine des avantages sociaux.

L'impact de la réforme des retraites de 2023 sur les contrats collectifs étudié dans ce mémoire pourrait venir s'ajouter à cette hausse, entraînant une augmentation encore plus significative des coûts.

Chapitre 2

Organisation du système de retraite

Ce chapitre s'inspire du cours d'Actuariat Retraite de Gautron N. dispensé à l'ISUP [20].

2.1 Le système de retraite en France

2.1.1 Contexte

Le système de retraite a été instauré en 1945 avec la création des régimes de sécurité sociale. Cependant, bien avant cette date, des pensions de retraite existaient déjà pour certaines professions, notamment dans la fonction publique et pour quelques travailleurs qualifiés du secteur privé. Ce dispositif permet aux travailleurs de cesser leur activité tout en continuant à percevoir un revenu appelé « pension ». À l'origine, la retraite était destinée aux personnes trop âgées pour continuer à travailler et donc considérées comme « inaptes ». Depuis, ce système a beaucoup évolué, avec plusieurs réformes visant à mieux l'adapter aux changements démographiques.

2.1.2 Définitions et principes

Trois éléments principaux fondent le système de retraite français.

2.1.2.1 Caractère obligatoire

Toute personne qui entre dans la vie active contribue à sa retraite. Ainsi, chaque année, l'employeur verse des cotisations salariales et patronales, destinées à financer la future pension de retraite de ses employés.

2.1.2.2 Système par répartition

Le système par répartition repose sur un principe de solidarité inter-générationnelle : les cotisations versées chaque année par les actifs servent à financer les pensions des retraités de cette même année. Ainsi, les travailleurs d'aujourd'hui financent les retraites actuelles tout en accumulant des droits pour leur propre retraite future.

2.1.2.3 Mode de financement

La pension de retraite est calculée en fonction du montant total des cotisations versées tout au long de la carrière professionnelle de la personne, en proportion de ses revenus. En général, une

personne active cotise automatiquement à deux caisses de retraite, déterminées par son statut : la caisse de retraite de base et la caisse de retraite complémentaire obligatoire. Les deux caisses prélèvent des cotisations sur le revenu, et au moment de la retraite, chacune verse une pension. Ces caisses fonctionnent selon un système de répartition.

Le système de retraite français intègre une dimension familiale significative, prenant en compte certains événements de la vie dans le calcul des pensions, comme les congés de maternité. Il offre également des mécanismes tels que les rentes de réversion, permettant au conjoint survivant de bénéficier, sous certaines conditions, d'une partie de la retraite du conjoint décédé. De plus, ce système tient compte des aléas de la vie active, comme les périodes d'invalidité ou de chômage, qui pourraient limiter la capacité d'une personne à cotiser pour sa retraite.

Il est possible de se constituer un complément de retraite en rejoignant un régime de retraite supplémentaire d'entreprise, proposé par l'employeur dans le cadre d'un contrat collectif, ou en souscrivant à un contrat de retraite individuelle auprès d'un organisme d'assurance, par exemple. Ce régime fonctionne sur le principe de la capitalisation, ce qui signifie que les prestations sont financées par les réserves financières accumulées au fil des années.

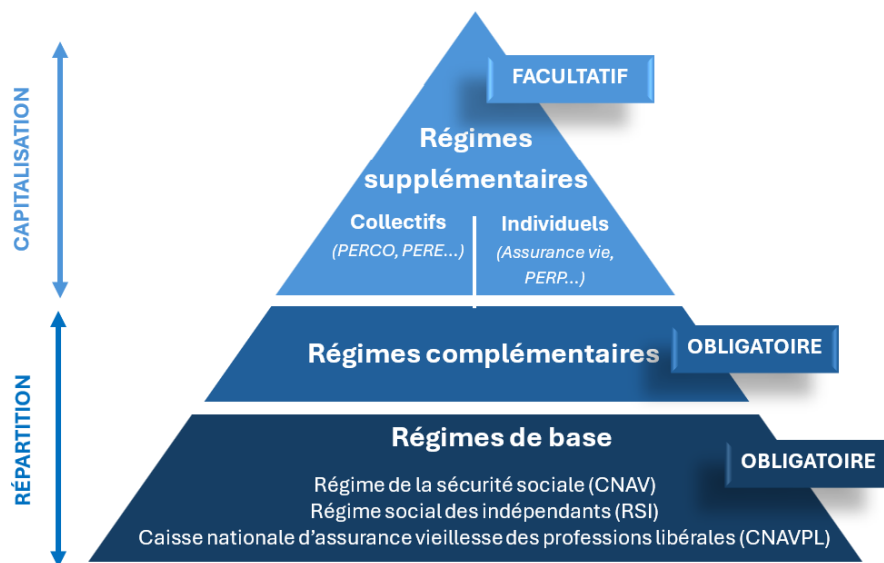


FIGURE 2.1 – Organisation du système de retraite en France

2.1.3 Les précédentes réformes de retraite en France

Conçu après la Seconde Guerre mondiale, le système de retraite a subi plusieurs réformes au cours des trente dernières années pour garantir sa pérennité financière face au vieillissement de la population.

2.1.3.1 Réforme Balladur (1993)

À la fin des années 1980, le système de retraite rencontre d'importantes difficultés financières en raison de la combinaison de la crise économique et du vieillissement de la population. Le déficit

atteint alors un niveau sans précédent, s'élevant à 40 milliards de francs. La réforme de 1993, connue sous le nom de réforme Balladur, introduit plusieurs mesures pour rétablir l'équilibre financier, mais ces mesures ne concernent que le système de retraite du secteur privé. Désormais, les pensions sont indexées sur les prix plutôt que sur les salaires, la durée d'assurance requise pour obtenir le taux plein passe de 150 à 160 trimestres, et le calcul du salaire de référence est modifié (en passant des 10 aux 25 meilleures années). Fin 1995, le « plan Juppé » échoue dans son objectif d'étendre ces mesures aux fonctionnaires et aux entreprises publiques, notamment le relèvement de la durée d'assurance à 40 ans.

2.1.3.2 Rapport Briet (1995) et rapport Charpin (1999)

Entre 1990 et 2000, plusieurs rapports alertent sur la situation financière du système de retraite. Le rapport Briet (1995) met en évidence que, en raison notamment d'une détérioration rapide des ratios démographiques, le déficit du régime général pourrait passer d'environ 20 milliards à 107 milliards de francs d'ici 2015. Pour remédier à cette situation et assurer l'équité entre les générations, il propose d'allonger la durée de cotisation.

Le rapport Charpin (1999) formule des constats similaires, prévoyant des déficits de 220 milliards de francs d'ici 2020 et de 600 milliards en 2040. Pour y faire face, il recommande d'allonger la durée de cotisation à 42,5 ans.

2.1.3.3 Réforme Fillon (2003)

La réforme Fillon de 2003 poursuit l'augmentation de la durée de cotisation. Face à une estimation d'un besoin de financement de 43 milliards d'euros par an à partir de 2020, elle introduit plusieurs mesures importantes. Elle aligne les règles de retraite des fonctionnaires sur celles des salariés du secteur privé : la durée de cotisation nécessaire pour une retraite à taux plein est harmonisée avec celle du secteur privé, et les pensions des fonctionnaires sont désormais indexées sur les prix, comme celles des salariés depuis 1993, plutôt que sur le point de la fonction publique.

La réforme introduit également plus de flexibilité dans les conditions de départ, avec la mise en place des mécanismes de décote et de surcote. En outre, un dispositif de départ anticipé est créé pour les carrières longues, permettant à ceux qui ont commencé à travailler tôt et ont cotisé la durée requise de partir à la retraite avant 60 ans. Entre 2007 et 2008, les principales évolutions de la loi de 2003 relatives à la fonction publique sont étendues aux régimes spéciaux (SNCF, RATP, IEG...), avec une augmentation progressive de la durée de cotisation de 37,5 ans à 40 ans, l'instauration d'un mécanisme de décote/surcote et l'indexation des pensions sur l'évolution des prix.

2.1.3.4 Réforme Woerth (2010)

La réforme de 2010 est motivée par le constat de déficits excessifs, estimés à 70 milliards d'euros en 2030 et à 100 milliards d'euros en 2050. Elle modifie l'âge de départ à la retraite : l'âge légal pour ouvrir des droits est désormais fixé à 62 ans (contre 60 ans auparavant), et l'âge pour bénéficier d'une retraite à taux plein sans décote est repoussé à 67 ans (contre 65 ans). Cette réforme introduit également diverses mesures visant à accroître les ressources du système de retraite en durcissant ou en supprimant certains dispositifs existants.

2.1.3.5 Réforme Touraine (2014)

La réforme Touraine introduit un allongement progressif de la durée de cotisation, qui sera portée à 43 ans pour les personnes nées à partir de 1973. Par ailleurs, elle crée un compte personnel de prévention de la pénibilité, destiné à permettre aux salariés exposés à des risques professionnels (tels que le travail de nuit ou dans des environnements pénibles) de bénéficier d'une retraite anticipée.

2.2 La réforme Macron (2023)

La réforme Macron prévoit un recul progressif de l'âge légal de départ à la retraite, qui passera de 62 à 64 ans. De plus, elle accélère l'augmentation de la durée de cotisation requise pour bénéficier d'une retraite sans décote, issue de la réforme Touraine. À partir du 1^{er} septembre 2023, l'âge légal de départ est avancé de 3 mois et la durée de cotisation nécessaire est augmentée d'un trimestre.

Il est toujours possible de partir à la retraite avant l'âge légal dans certains cas, tels que les carrières longues, les problèmes de santé ou le handicap, le travail pénible, la retraite progressive, ou l'exposition à l'amiante. En revanche, il est également possible de prolonger son activité au-delà de l'âge légal par choix volontaire, ce qui entraînera l'application de surcotes. L'âge auquel la décote est annulée, permettant d'obtenir le taux plein, reste fixé à 67 ans.

Années	Génération	Âge légal + 3 mois par an	Durée de cotisation pour le taux plein Accélération du calendrier Touraine
2023	Entre le 01/09/1961 et 31/12/1961	62 ans et 3 mois	169 trimestres
2024	1962	62 ans et 6 mois	169 trimestres
2025	1963	62 ans et 9 mois	170 trimestres
2026	1964	63 ans	171 trimestres
2027	1965	63 ans et 3 mois	172 trimestres (43 ans)
2028	1966	63 ans et 6 mois	172 trimestres (43 ans)
2029	1967	63 ans et 9 mois	172 trimestres (43 ans)
2030	1968	64 ans	172 trimestres (43 ans)

TABLEAU 2.1 – Phase transitoire du décalage de 62 à 64 ans

La réforme 2023 possède d'autres dispositions.

2.2.1 Pension minimale

La pension minimale augmente de 100 € par mois. Indexée sur le SMIC, elle sera d'environ 1 200 € brut mensuel pour une carrière complète.

2.2.2 Régimes spéciaux

À partir de septembre 2023, les nouveaux entrants ne seront plus couverts par les régimes spéciaux (clause du grand-père) pour des secteurs tels que la RATP, les industries électriques et gazières, ainsi que les clercs et employés de notaires. Les fonctionnaires en catégories actives (comme les pompiers, policiers et militaires) conserveront un âge de départ spécifique, mais celui-ci sera majoré de 2 ans. Certains régimes spéciaux autonomes, comme celui de l'Opéra de Paris, seront maintenus.

2.2.3 Retraite progressive

La retraite progressive permet à un salarié de commencer à percevoir une partie de sa pension de retraite tout en réduisant son temps de travail. Le montant de la pension est proportionnel à la réduction du temps de travail. Depuis la réforme, l'employeur ne peut refuser une demande de retraite progressive, sauf en cas d' « incompatibilité avec l'activité économique de l'entreprise ».

2.2.4 Cumul emploi retraite

Les retraités qui poursuivent une activité professionnelle pourront bénéficier d'une seconde pension de retraite, calculée uniquement sur les nouvelles périodes de cotisation.

2.2.5 Mesures relatives aux droits des mères de famille

Une surcote pouvant atteindre 5% est instaurée pour les mères de famille qui atteignent le nombre d'annuités requis et prennent leur retraite entre 63 et 64 ans. Dans le cadre du partage, bien que peu fréquent en pratique, de 4 trimestres entre les parents pour l'éducation d'un enfant, au moins 2 trimestres doivent obligatoirement être attribués à la mère.

2.2.6 Pénibilité

Le « Compte Professionnel de Prévention » est renforcé pour mieux prendre en compte les droits des travailleurs, offrant notamment la possibilité de bénéficier de congés de reconversion. De plus, un « fonds d'investissement dans la prévention de l'usure professionnelle » est créé pour financer des actions de reconversion.

2.2.7 Aidants

Les aidants familiaux, contraints de réduire leur activité pour s'occuper d'un proche, bénéficieront de validations de trimestres pour leur retraite.

2.2.8 Suivi dans le temps

Un rapport sur les mesures concernant l'emploi des seniors devra être présenté avant le 1^{er} octobre 2027. De plus, des objectifs sont fixés pour réduire l'écart entre les montants des pensions perçues par les femmes et les hommes, avec pour ambition de le diminuer de moitié d'ici 2037 et de l'éliminer totalement d'ici 2050.

Ces changements potentiels pourraient avoir des effets positifs ou négatifs sur la santé des individus, mais ils ne seront pas abordés dans ce mémoire.

2.3 Le décalage de l'âge de départ à la retraite dans les autres pays

Selon l'article de Vie publique [21], la France n'est pas le seul pays où la réforme du système des retraites fait débat. Les droits à la retraite diffèrent d'un pays européen à l'autre, mais une hybri-

dation des systèmes est observée, fondée sur des principes de contributivité ou de redistributivité. Les financements sont organisés par répartition et/ou par capitalisation.

2.3.1 Vieillesse des populations

La population européenne est vieillissante. Selon le même article, en 2020, 20,6% de la population de l'Union Européenne avait 65 ans et plus, soit une augmentation de trois points en dix ans. Ce chiffre reflète une tendance générale au vieillissement, bien que cette évolution varie en fonction des années (par exemple, en raison de la surmortalité liée au Covid-19) et des pays au sein de l'UE, certains États de l'Est ayant une population moins âgée que d'autres. Ce pourcentage diffère également entre les pays : en 2020, il était de 20,4% en France, 21,8% en Allemagne et 18,2% en Pologne. Les personnes âgées de 65 ans et plus étaient proportionnellement les plus nombreuses en Italie (23,2%), en Grèce (22,3%), en Finlande (22,3%) et au Portugal (22,1%).

En matière de retraites, le nombre absolu de personnes âgées est certes important, mais il doit être mis en relation avec le nombre d'actifs. Le ratio de dépendance, c'est-à-dire le rapport entre le nombre de personnes âgées de 65 ans et plus et le nombre de personnes âgées de 15 à 64 ans, est particulièrement déterminant, notamment dans les régimes financés par répartition. Ce ratio est en augmentation générale : pour l'ensemble des pays de l'UE, il est passé de 28,3% à 32,5% entre 2014 et 2021. En France, cet indicateur est passé de 28,3% en 2014 à 33,6%. À l'avenir, ce ratio devrait se détériorer encore en raison du vieillissement de la population : Eurostat estime qu'il atteindra 45,3% en 2046 pour la France.

2.3.2 Âge de départ à la retraite : de 62 à 67 ans

L'âge de départ à la retraite varie considérablement selon les pays, les sexes, la prise en compte des enfants et les carrières, avec des âges allant de 62 à 67 ans.

Au moment de la rédaction de ce même article, en janvier 2023, les âges de départ à la retraite varient selon les pays : en Suède, Norvège et Slovaquie, il est fixé à 62 ans ; aux Pays-Bas et en Italie, il est de 67 ans. En Espagne, il sera de 67 ans à partir de 2027. En Autriche, les hommes partent à 65 ans, tandis que les femmes nées après 1964 verront leur âge de départ passer progressivement de 60 à 65 ans. Au Danemark, l'âge légal de départ est de 67 ans, avec une augmentation prévue à 68 ans en 2030. En France, l'âge légal de départ sera progressivement relevé à 64 ans d'ici 2030.

De nombreux pays augmentent progressivement l'âge de départ à la retraite, et la France suit également cette évolution, illustrant ainsi une tendance de fond.

Il est important de noter que les élections législatives de juin 2024 pourraient entraîner un changement de l'âge légal de départ à la retraite. Cependant, même si cet âge est abaissé, il est possible que la situation se répète dans quelques années.

2.3.3 Taux de remplacement

Les pensions de retraite varient considérablement d'un pays à l'autre. Selon l'Organisation de Coopération et de Développement Économiques (OCDE) [22], la moyenne dans les pays de l'UE correspond à un taux de remplacement de 54% du salaire brut. Ce taux mesure la capacité d'un

système de retraite à fournir un revenu aux retraités en remplacement de leur salaire, qui constitue leur principale source de revenus avant la retraite. Le taux de remplacement brut est calculé en divisant les prestations de retraite brutes par le salaire brut avant la retraite. Selon le rapport annuel du Conseil d’Orientation des Retraites (COR), les pensions représentent entre 44 et 73% des derniers revenus. L’Italie, l’Espagne et la Suède se situent en haut de l’échelle, tandis que les Pays-Bas, la Belgique et l’Allemagne figurent en bas.

Pour rétablir l’équilibre entre le nombre de pensionnés et les cotisations, plusieurs pays ont choisi d’engager des réformes. Les options disponibles comprennent :

- D’augmenter les recettes par les cotisations ou les impôts ;
- De réduire les prestations en diminuant les montants des pensions et leur nombre ;
- De modifier le système en redéfinissant les objectifs.

2.4 Les conséquences de la réforme 2023 sur les complémentaires santé d’entreprise

Selon l’article du Cercle de l’Épargne [23], le recul de l’âge de départ de la retraite de 62 à 64 ans et l’accélération de la durée de cotisation auront des répercussions sur d’autres régimes sociaux, tels que ceux liés au chômage, à la santé et aux accidents du travail, ainsi que sur les régimes complémentaires, incluant la santé, la prévoyance et la retraite supplémentaire.

En reportant l’âge légal de la retraite de 62 à 64 ans, la réforme de 2023 accroît le nombre potentiel de personnes en activité. Selon l’article, le surplus pourrait dépasser 350 000 personnes par an d’ici 2032, représentant près de la moitié d’une génération. Cette augmentation concernerait principalement les personnes âgées de 60 à 64 ans. Ainsi, la population active croîtrait de 1,2%, tandis que celle des retraités diminuerait de 2%.

Le report de l’âge de départ à la retraite de deux ans devrait entraîner des surcoûts sur les complémentaires de santé contractées par les entreprises. En effet, les dépenses de santé augmentent significativement après 60 ans, et les arrêts maladie tendent à être plus longs en raison de problèmes de santé potentiellement graves.

C’est pourquoi ce mémoire concentre son analyse sur l’impact de la réforme des retraites sur les prestations de santé, en particulier pour les personnes qui travailleront deux années supplémentaires et bénéficiant d’un contrat collectif de mutuelle. Une projection de l’évolution du portefeuille de cette mutuelle jusqu’en 2030 sera également effectuée.

Chapitre 3

Présentation des données utilisées et analyse descriptive

Cette section présente les données utilisées pour mener ces travaux. Elle détaille les bases de données d'origine, les ajustements effectués pour rendre les informations exploitables, ainsi qu'une première analyse de celles-ci.

3.1 Présentation des bases de données

Les bases de données utilisées proviennent d'une mutuelle généraliste donc il n'y a pas de biais de région ou de catégorie socioprofessionnelle. Cette mutuelle propose des contrats individuels et collectifs en complémentaire santé. Des personnes actives et retraitées sont présentes dans cette base.

Pour l'étude, trois bases de données regroupant des contrats collectifs seront exploitées. Ces bases correspondent à trois garanties distinctes, chacune offrant des couvertures différentes. Pour chaque garantie, deux bases de données sont disponibles : l'une contenant les prestations et les caractéristiques et l'autre les informations sur les personnes assurées. Au total, six bases de données devront être analysées et traitées, couvrant les prestations des assurés et des ayants droits sur deux années, 2021 et 2022.

Étant donné que les données proviennent de trois garanties distinctes, les tarifs calculés dans le chapitre 5 refléteront des tarifs « moyens » de ces trois grilles.

3.1.1 Bases de prestations

Dans ces bases, les variables sont axées sur les prestations. Plusieurs variables sont présentes :

- **Date de la prestation** ;
- **Grands postes de soins** : Dentaire, Optique, Pharmacie... ;
- **Montant remboursé éventuellement par d'autres mutuelles** ;
- **Montant remboursé par la mutuelle** ;

- **Montant remboursé par la Sécurité Sociale** ;
- **Montant restant à charge de l'assuré** ;
- **Montant total de la prestation** ;
- **Numéro associé à chaque prestation** ;
- **Numéro de l'assuré** ;
- **Numéro du bénéficiaire** ;
- **Numéro du contrat** ;
- **Sous-postes de soins** : Soins dentaires, verres simples, verres complexes,... ;
- **Taux de remboursement de la Sécurité Sociale** ;
- **Type de secteur** : Conventionné ou non.

Dans la suite lorsqu'il est mentionné « montants remboursés », cela correspond au montant que rembourse la mutuelle seulement.

3.1.2 Bases de personnes

Dans ces bases, les variables se concentrent sur les personnes protégées (assurés, conjoints/concubins, enfants et ayants droits).

Plusieurs variables sont présentes :

- **Catégories socioprofessionnelles des assurés** ;
- **Code du département de l'assuré** ;
- **Date de début du contrat** ;
- **Date de fin du contrat** ;
- **Date de naissance du bénéficiaire** ;
- **Numéro de l'assuré** ;
- **Numéro du bénéficiaire** ;
- **Numéro du contrat** ;
- **Rôle du bénéficiaire** : Assuré, Enfant, Conjoint/Concubin ou Ayant Droit ;
- **Sexe du bénéficiaire** ;
- **Statut des assurés** : Actif, Retraité, Sortie de groupe... ;
- **Type de contrat** : Collectif obligatoire ou collectif facultatif ;
- **Type de régime suivi par la Sécurité Sociale** : Général, MGEN, SNCF...

Les assurés sont les titulaires du contrat d'assurance, tandis que les bénéficiaires sont associés à chacun de ces assurés. Certaines variables se rapportent spécifiquement aux assurés, tandis que d'autres concernent les bénéficiaires.

3.2 Traitement des données

Le traitement des bases de données est réalisé sur R, permettant de créer une base unique sur laquelle seront menées les analyses de l'étude. La période d'observation couvre les prestations et les personnes protégées sur une durée de deux ans, allant du 01/01/2021 au 31/12/2022.

3.2.1 Les bases de données étudiées

3.2.1.1 Base Prestations améliorée

L'objectif initial est de créer une base de données unique intégrant les prestations et les personnes pour les trois couvertures. Les étapes suivantes ont été réalisées pour chaque couverture avant d'obtenir la base finale.

Lors de l'importation des bases de données, il est crucial de définir le type de chaque colonne (chaîne de caractères, nombre ou date) afin d'éviter les erreurs de transcription des valeurs. Ensuite, il est nécessaire de fusionner, pour chaque couverture, la base des prestations avec la base des personnes assurées afin d'associer chaque prestation aux assurés correspondants. Cette fusion se réalise en utilisant la variable Numéro Bénéficiaire et l'année de chaque prestation, puisque les prestations des assurés couvrent une période de deux ans.

Ainsi, trois nouvelles bases de données sont obtenues, une pour chaque couverture, chacune regroupant les prestations par année ainsi que les personnes couvertes.

Pour finaliser cette étape de traitement, les trois bases obtenues sont concaténées afin de créer la base finale requise pour l'analyse. La concaténation est possible car les trois bases ont la même structure.

3.2.1.2 Base Personnes améliorée

Les informations sur les personnes issues des trois bases de données sont également regroupées. Dans la suite de l'étude, pour analyser la fréquence des différents postes de soins, il est essentiel de disposer d'une base comprenant l'ensemble des personnes couvertes par ces garanties. Divers retraitements et ajouts de variables ont également été réalisés dans la base des personnes ce qui fait l'objet de la partie suivante.

3.2.2 Variables corrigées

Pour améliorer la pertinence des variables, plusieurs d'entre elles ont été corrigées. En effet, certaines variables étaient dans des formats peu adaptés pour une analyse descriptive précise. Par exemple, les dates étaient au format "01JAN2022 :00 :00 :00" sous forme de chaîne de caractères. Elles ont donc été reformatées en "01/01/2022", avec l'heure supprimée, car elle n'est pas nécessaire pour l'étude.

3.2.3 Variables créées

3.2.3.1 Base Prestations

Pour améliorer le traitement de la base de données finale des prestations, plusieurs variables ont été ajoutées à celle-ci.

Variable Âge

La base de données contient les dates de naissance des personnes. Cependant, il est plus pertinent d'utiliser l'âge des personnes plutôt que leur date de naissance, étant donné que l'étude

porte principalement sur la réforme des retraites. L'âge est donc calculé à partir de la date de naissance de chaque assuré et du milieu de l'année de leur date de prestation. En effet, pour obtenir une estimation précise de l'âge, la date du 01/07/N est utilisée comme référence pour chaque année, car elle correspond au milieu de l'année, offrant une répartition équilibrée des individus avant et après cette date. Étant donné que les prestations couvrent une période de deux ans, les personnes sont également concernées sur cette période.

Variable Région

Les numéros de départements figurent dans la base de prestations. Pour l'analyse descriptive, il est pertinent d'examiner quelle région enregistre le plus grand nombre de prestations et celle où les dépenses sont les plus élevées et d'évaluer si la région est une variable discriminante du tarif. Même si l'analyse révèle que cette variable a un impact sur le tarif, elle ne fait pas l'objet de ce mémoire. Afin de travailler avec plus de données, les départements ont été regroupés par région. Ainsi, une colonne Région a été ajoutée, contenant les codes INSEE des régions correspondants aux départements présents dans la base.

Variable Nombre de jours

Dans la suite de l'analyse, chaque poste de soins sera présenté. Toutefois, un poste a nécessité un retraitement préalable : il s'agit de l'hospitalisation, qui inclut divers sous-postes telles que les chambres particulières, les chambres en ambulatoire, les chambres en psychiatrie, etc. Les montants remboursés par la mutuelle pour ces chambres correspondent au total des jours d'hospitalisation. Par exemple, si une personne est restée 4 jours consécutifs, le montant affiché dans la base est pour l'ensemble des 4 jours. Ainsi, le nombre exact de jours passés dans chaque type de chambre n'est pas précisé. Ce manque de précision est problématique, car il est souhaité que le montant remboursé par la mutuelle soit normalisé.

Pour remédier à cela, une colonne « NB jours » a été ajoutée, indiquant le nombre de jours d'hospitalisation des personnes en fonction du montant total associé à chaque prestation. Cela permet de mieux refléter la durée réelle d'hospitalisation et d'obtenir une analyse plus précise de la fréquentation.

Pour créer cette colonne, des recherches ont été effectuées chez différents assureurs tels que Malakoff Humanis [24], AG2R La Mondiale [25] et également le site Ameli [9] et d'après celles-ci il a été supposé que les chambres en médecine, chirurgie, maternité et ambulatoire coûtent 50 € par jour, tandis que les chambres en psychiatrie coûtent 100 € par jour. Quant au frais de séjour, il est supposé qu'ils sont égaux à 20€. Le montant total de la prestation est donc divisé par ces coûts journaliers, selon le type de chambre, pour déterminer le nombre de jours d'hospitalisation pour chaque prestation. Par exemple, pour une prestation de chambre en médecine d'un montant de 200 €, il est estimé que la personne a séjourné 4 jours à l'hôpital. Il est important de noter qu'un biais potentiel peut exister, car les coûts moyens pour ces sous-postes de soins sont basés sur des hypothèses.

Pour les autres sous-postes de soins hospitaliers comme les forfaits hospitaliers et les honoraires et autres frais qui ne sont pas concernés par cette méthode de calcul, le nombre de jours est fixé à 1. Cette colonne sera utilisée ultérieurement pour prédire la fréquence du poste Hospitalisation.

Variable Montants Hospitalisation

Dans la même logique que la colonne NB jours, il est nécessaire de corriger le montant associé aux chambres hospitalières, étant donné que les montants des sous-postes de soins, tels que les chambres en médecine, chirurgie, maternité, ambulatoire et psychiatrie et les frais de séjour, sont calculés pour la totalité des jours consécutifs passés à l'hôpital. Pour ce faire, une nouvelle colonne, Montants Hospitalisation, a été créée pour ajuster les montants des remboursements.

Le montant correct des chambres hospitalières est obtenu en divisant le montant total des remboursements de la mutuelle par le nombre de jours d'hospitalisation, tel que calculé dans la colonne NB jours. Pour déterminer le montant total remboursé par la mutuelle à l'aide des deux colonnes créées, il suffit de multiplier les valeurs de la colonne Montants Hospitalisation par celles de la colonne NB Jours.

Pour les autres sous-postes de soins hospitaliers, pour lesquels le nombre de jours est fixé à 1, le montant de remboursement ne nécessite pas d'ajustement et reste inchangé.

3.2.3.2 Base Personnes

Dans cette base, les dates de début et de fin de contrat sont disponibles. Étant donné que les prestations et les personnes couvrent une période de deux ans, certaines personnes sont présentes uniquement la première année (2021), d'autres seulement la deuxième année (2022), et la plupart sont présentes pendant les deux années.

Pour analyser la consommation des prestations par personne et par année, chaque personne a été dupliquée, ce qui donne deux parties : une pour l'année 2021 et une autre pour l'année 2022. Cette duplication est essentielle pour calculer l'exposition et le nombre de prestations pour chaque personne.

Variable Exposition

Cette variable permet de mesurer l'exposition des personnes pour chaque année. Elle est égale à :

$$\text{Exposition} = \max\left(\frac{\min(01/01/2022; \text{Date fin contrat}) - \max(01/01/2021; \text{Date début contrat})}{365}; 0\right) \quad (3.1)$$

Si la date de fin de contrat n'est pas spécifiée, la date du 31/12/2022 est automatiquement appliquée pour l'étude, car la période de couverture de ces bases s'étend du 01/01/2021 au 31/12/2022.

Ainsi, si une personne est présente toute l'année, du 01/01/N au 31/12/N, son exposition est de 1. Si elle est présente uniquement une partie de l'année, son exposition sera comprise entre]0; 1[. Si elle n'est pas présente durant l'année, son exposition est de 0. Cette variable d'exposition sera prise en compte dans le modèle de prédiction de la fréquence.

Variable NB prestations

Pour prédire la fréquence, il est nécessaire de comptabiliser combien de prestations sont consommées par les personnes protégées en fonction des sous-postes de soins. Une colonne intitulée « NB prestations » a été créée dans la base des personnes pour dénombrer, pour chaque sous-poste de soins, le nombre de prestations consommées par chaque personne au cours de l'année.

Pour ce faire, chaque personne a été dupliquée 53 fois, correspondant aux 53 sous-postes de soins répertoriés. Cela permet de disposer d'une ligne par personne, par année, et par sous-poste de soins. Le doublage initial des personnes pour les deux années ayant déjà été effectué, il suffit donc de les dupliquer 53 fois dans la base.

Ensuite, le calcul de cette colonne s'appuie sur la base des prestations, comptabilisant, en fonction du Numéro de bénéficiaire, de l'année, et du sous-poste de soins, le nombre de prestations consommées par chaque personne. Par exemple, si une personne a bénéficié de 5 prestations de soins dentaires en 2021, la colonne NB prestations affichera 5. Si une personne n'était pas présente durant l'année ou n'a pas consommé de soins dentaires, la colonne indiquera 0.

Naturellement, le nombre de prestations correspond au nombre de jours pour le poste d'hospitalisation notamment pour les chambres hospitalières et les frais de séjour.

Au début de l'analyse descriptive, la base initiale, sans aucune duplication (les 53 duplications énoncées précédemment), est utilisée pour identifier précisément le nombre de personnes distinctes présentes, ainsi que pour examiner leur statut et d'autres caractéristiques.

3.2.4 Prétraitement

L'analyse descriptive se déroule en deux étapes. La première est consacrée à l'ensemble des individus présents dans la base, tandis que la seconde se concentre sur une version filtrée de la base, après avoir retiré certains éléments qui seront mentionnés par la suite. Dans les graphiques suivants, tous les montants sont exprimés en euros.

La base de prestations comprend 830 118 observations, tandis que la base de personnes en compte 10 905 avant le prétraitement des données. Cela constitue une quantité de données suffisante pour poursuivre l'analyse. Il est important de noter que plus le volume de données est élevé, meilleure sera la fiabilité des résultats.

3.2.4.1 Données manquantes

Dans la base de prestations, certaines variables sont essentielles, telles que le montant remboursé par la mutuelle, la date de naissance, le numéro du bénéficiaire, les postes de soins et les sous-postes de soins. Le nombre d'observations manquantes pour ces variables représente 0,081% de l'ensemble de la base. Par mesure de précaution et pour éviter de biaiser les résultats, les observations contenant des données manquantes dans ces variables ont été supprimées.

L'étude se concentre exclusivement sur la France métropolitaine. Les prestations des régions d'outre-mer et des pays étrangers représentent 0,031% de la base totale, tandis que les personnes

de ces régions représentent 0,073% de la base totale. En conséquence, les prestations et les personnes issues des régions d’outre-mer et des pays étrangers ont été supprimées. Par ailleurs, les observations comportant des régions manquantes, que ce soit dans la base de prestations ou la base de personnes, ont été supprimées. Heureusement, aucune prestation ni aucune personne dans ces bases ne présentait de région manquante.

Certaines prestations sont associées à des personnes dont l’exposition est égale à 0, c’est-à-dire que la date de soin de la prestation est postérieure à la date de fin de garantie. Cela est probablement dû à une anomalie dans la base de données. Ces valeurs représentent 0,093% de la base totale des prestations et ont donc été supprimées.

3.2.4.2 Valeurs aberrantes

La variable représentant le montant des prestations remboursées par la mutuelle est l’une des plus importantes de l’étude et est de nature continue. Il est donc essentiel de vérifier qu’elle ne contienne pas de valeurs extrêmes ou aberrantes.

Le boxplot suivant permet de visualiser l’éventail des différents montants présents.

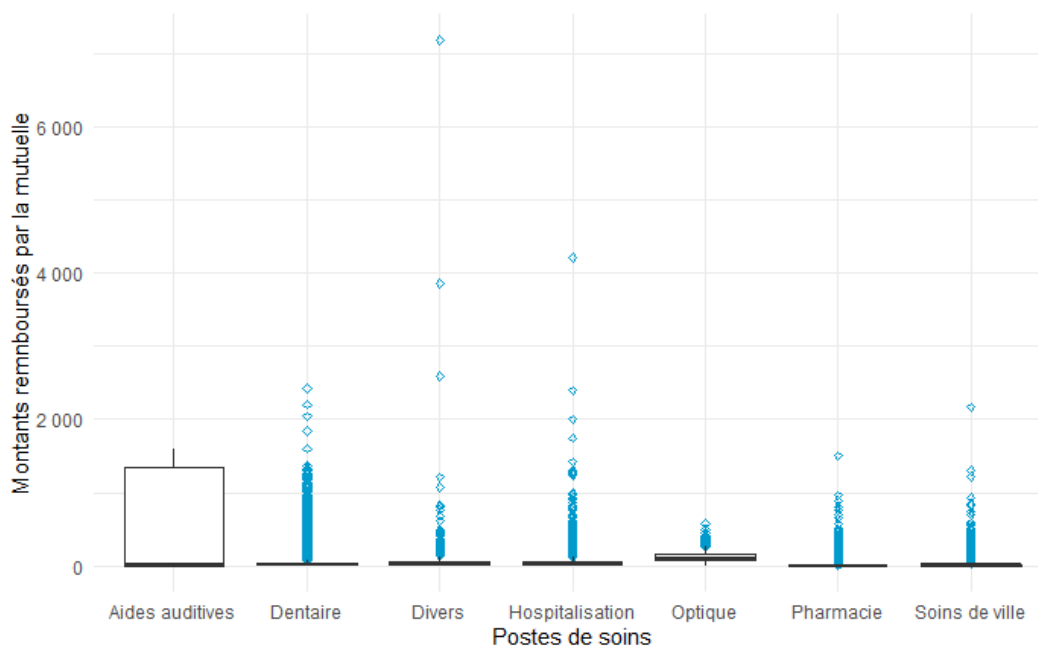


FIGURE 3.1 – Boxplot des montants de remboursements de la mutuelle par poste de soins

Certains montants remboursés par la mutuelle sont élevés, avec un montant maximum de 7 185 € pour une prestation de grand appareillage du poste divers. Bien que ce montant soit élevé, il n’est pas considéré comme une valeur extrême dans le domaine de la santé. Par conséquent, tous les montants de la base de prestations sont conservés pour l’étude.

3.3 Analyse descriptive

3.3.1 Présentation générale des données

Après les modifications précédentes, la base de prestations compte désormais 828 410 observations, tandis que la base de personnes est composée de 10 897 individus distincts couvrant les deux années.

Parmi ces observations, il y a 10 897 personnes protégées, réparties comme suit : 5 068 Assurés, 1 991 Conjoints/Concubins, 3 822 Enfants et 16 Ayants Droits. Les contrats sont détenus par les Assurés, et chaque Assuré peut avoir des personnes rattachées à son contrat, telles que des Conjoints/Concubins, des Enfants ou des Ayants Droits.

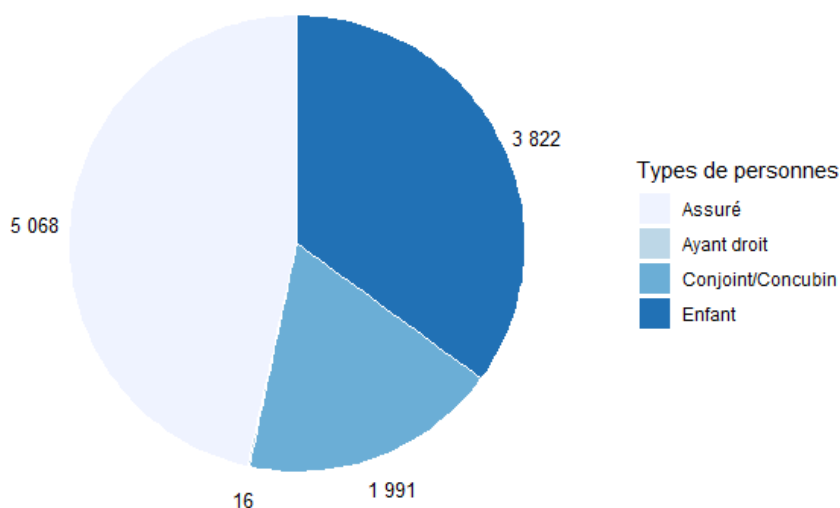


FIGURE 3.2 – Répartition des personnes

Par la suite, les Ayants Droits et les Conjoints/Concubins sont inclus dans l'analyse étant donné que les informations indispensables les concernant sont complètes. En revanche, les Enfants ne seront plus inclus. En effet, étant donné que le projet examine l'impact du report de l'âge de départ à la retraite sur les dépenses de santé des personnes qui travailleront deux ans de plus, il n'est pas pertinent d'inclure les enfants dans cette étude.

La base de prestations contient désormais 603 420 prestations, tandis que la base de personnes comprend 7 074 individus distincts.

3.3.2 Zoom sur les données utilisées pour l'étude

3.3.2.1 Évolution entre 2021 et 2022

La base de prestations couvre les années 2021 et 2022. Le nombre de prestations, le montant total des remboursements et l'exposition des personnes protégées peuvent varier d'une année à l'autre en raison de facteurs tels que la Covid-19 ou l'implémentation du 100% santé notamment en 2020

et 2021. Il est donc crucial de comparer et d'analyser l'évolution du coût moyen par prestation et par personne, ainsi que la fréquence des prestations, pour chaque année.

Les variables Nombre de prestations, Somme des expositions des personnes et Somme des montants de prestations sont directement extraites de la base de données. Les autres variables sont calculées à partir de ces données.

$$\text{Coût moyen par prestation} = \frac{\text{Somme des montants de prestations}}{\text{Nombre de prestations}} \quad (3.2)$$

où comme énoncé dans le chapitre précédent, pour le poste d'hospitalisation, le nombre de prestation inclut le nombre de jours.

$$\text{Coût moyen par personne} = \frac{\text{Somme des montants de prestations}}{\text{Nombre de personnes pondéré par l'exposition}} \quad (3.3)$$

$$\text{Fréquence} = \frac{\text{Nombre de prestations}}{\text{Nombre de personnes pondéré par l'exposition}} \quad (3.4)$$

Les résultats sont données dans le tableau suivant :

	2021	2022	Variations entre 2021 et 2022
Nombres de prestations	330 387	330 369	-0,01%
Nombre de personnes pondéré par l'exposition	5 325,98	5 344,47	0,35%
Somme des montants de prestations	4 045 804	3 822 583	-5,52%
Coût moyen par prestation	12,25	11,57	-5,51%
Coût moyen par personne	759,64	715,24	-5,84%
Fréquence	62,03	61,82	-0,35%

TABLEAU 3.1 – Évolution du coût moyen et de la fréquence entre 2021 et 2022

Le coût moyen par prestation et par personne ont diminué de 5,51% et 5,84% respectivement en 2022 par rapport à 2021. Cette baisse est principalement due à une réduction de 5,52% du montant total des prestations en 2022 par rapport à 2021.

	2021	2022	Variations entre 2021 et 2022
Optique	169	160	-5,65%
Dentaire	177,20	149,10	-15,86%
Soins de ville	208	203	-2,24%
Pharmacie	61,59	60,92	-1,10%
Divers	78,68	79,98	1,65%
Hospitalisation	47,40	46,02	-2,90%
Aides auditives	17,74	16,42	-7,43%

TABLEAU 3.2 – Évolution du coût moyen par personne et par grand poste entre 2021 et 2022

Ce tableau montre que la diminution des coûts moyens par prestation et par personne est principalement attribuable au poste dentaire. En examinant les variations entre 2021 et 2022 du coût moyen par personne, on observe que la plupart des postes restent relativement stables, avec des

changements allant de -7,43% pour les aides auditives à +1,65% pour le divers. En revanche, le poste de soins dentaires affiche une baisse significative de près de 15,86% entre les deux années.

Plus en détail, les sous-postes dentaires suivants connaissent des baisses de coût moyen par personne de plus de 15% : les Inlay Core (dans toutes les catégories : Panier libre, modéré et RAC 0) et les prothèses dentaires (Panier libre et RAC 0). Ces réductions sont dues à une diminution des montants totaux des prestations pour ces sous-postes.

Enfin, la fréquence des prestations entre 2021 et 2022 est globalement stable (+0,35%).

En combinant la diminution du coût moyen par prestation et la légère augmentation de la fréquence, la baisse totale est de 5,86%.

3.3.2.2 Régions de provenance des assurés

Suite à la suppression des prestations et personnes provenant des régions d'outre-mer et de l'étranger, il est pertinent de représenter la proportion des montants des prestations par région en France pour observer où les dépenses sont les plus élevées et où il y a le plus grand nombre de prestations.

Régions	Nombre de prestations	Répartition des montants
Île-de-France	274 608	53,11%
Centre-Val de Loire	130 915	16,47%
Bourgogne-Franche-Comté	890	0,15%
Normandie	23 489	2,89%
Hauts-de-France	17 462	2,22%
Grand Est	38 506	4,81%
Pays de la Loire	14 438	1,84%
Bretagne	5 382	0,45%
Nouvelle-Aquitaine	58 848	6,54%
Occitanie	55 200	5,59%
Auvergne-Rhône-Alpes	15 373	2,54%
Provence-Alpes-Côte d'Azur	24 783	3,20%
Corse	862	0,19%

TABLEAU 3.3 – Proportion des montants de prestations par région

L'Île-de-France est la région avec le nombre de prestations et les dépenses les plus élevées, représentant respectivement 274 608 prestations et 53,11% du montant total des dépenses, en raison de sa forte densité de population. La région Centre-Val de Loire se distingue également avec 130 915 prestations et environ 16,47% du montant total des dépenses. En revanche, la Bourgogne-Franche-Comté et la Corse présentent un nombre de prestations et une proportion des dépenses relativement faibles.

3.3.2.3 Types de situations des assurés

Dans la base de personnes, le statut d'activité des Assurés, des Conjoints/Concubins et des Ayants Droit est connu. Cette étude permet d'observer la proportion de personnes actives et retraitées présente dans la base.

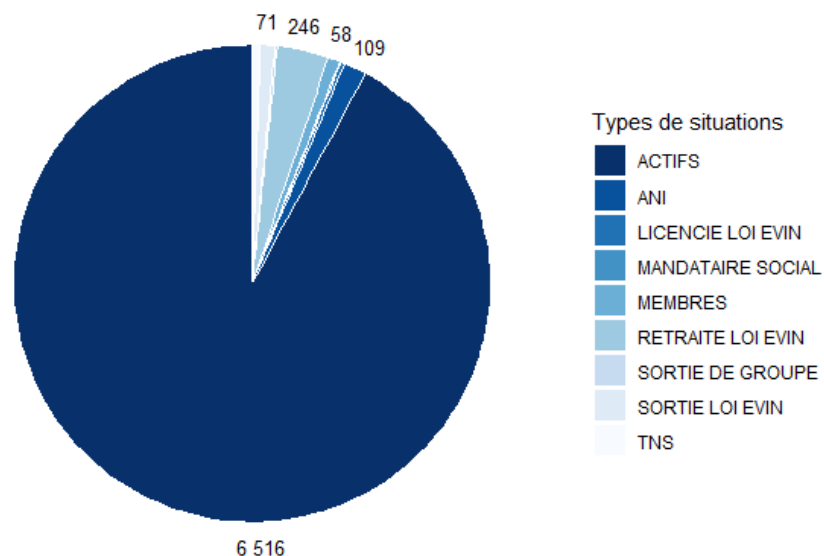


FIGURE 3.3 – Répartition des statuts d’activités des personnes protégées

La majorité des personnes, soit 6 516, sont des actifs. Parmi les autres, 328 sont considérées comme retraitées, dont 246 sous la loi Evin, 71 en sortie de loi Evin, et 11 en sortie de groupe. Les personnes en sortie de groupe et en sortie de loi Evin ont quitté l’entreprise mais restent couvertes par la mutuelle associée. La part restante est constituée de 109 personnes couvertes par un Accord National Interprofessionnel (ANI), ainsi que de quelques licenciés loi Evin, mandataires sociaux, membres et Travailleurs Non-Salariés (TNS).

3.3.2.4 Catégories Socioprofessionnelles des assurés

Dans la base de personnes, la Catégorie Socioprofessionnelle des Assurés, des Conjoints/Concubins et des Ayants Droit est renseignée.

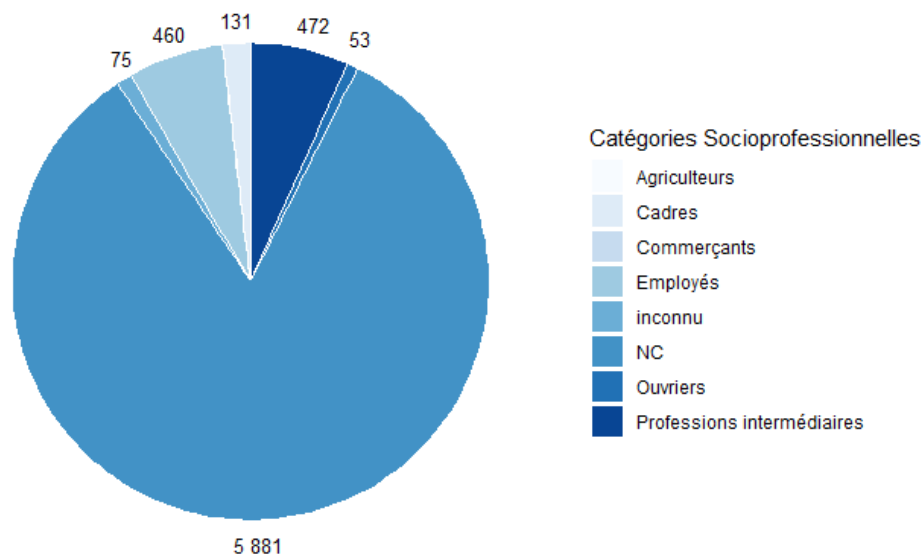


FIGURE 3.4 – Répartition des catégories socioprofessionnelles

La majorité des personnes sont des non-cadres, représentant 97,09% de l'ensemble, incluant les agriculteurs, commerçants, employés, ouvriers et professions intermédiaires. La proportion de personnes cadres est très faible, à 1,85%, tandis que 1,06% des personnes ont une catégorie socioprofessionnelle inconnue. Étant donné que cette proportion est minime, ces individus sont conservés pour la suite de l'étude et dans les modèles qui vont suivre. En effet, puisque la base de données est principalement constituée de personnes non-cadres, les modèles qui suivront se concentreront sur cette catégorie de personnes.

3.3.2.5 Répartition des âges des assurés

La base de données comprend un large éventail d'âges parmi les personnes protégées. Il est pertinent de tracer le nombre de personnes par âge pour vérifier si les dépenses de soins sont cohérentes avec la répartition des individus par âge. Comme mentionné précédemment, les enfants ne sont pas inclus dans cette analyse.

Les prestations sont sur deux années, 2021 et 2022. Par conséquent, pour les observations suivantes, la base de personnes a été doublée. Par exemple dans la base de personnes, un individu qui est présent les deux années va être répertorié une seule fois au même titre qu'un individu présent 6 mois. Il est donc important de doubler la base afin d'inclure l'ensemble des individus sur les deux années. De plus, les personnes avec des expositions égales à 0 ont été supprimées.

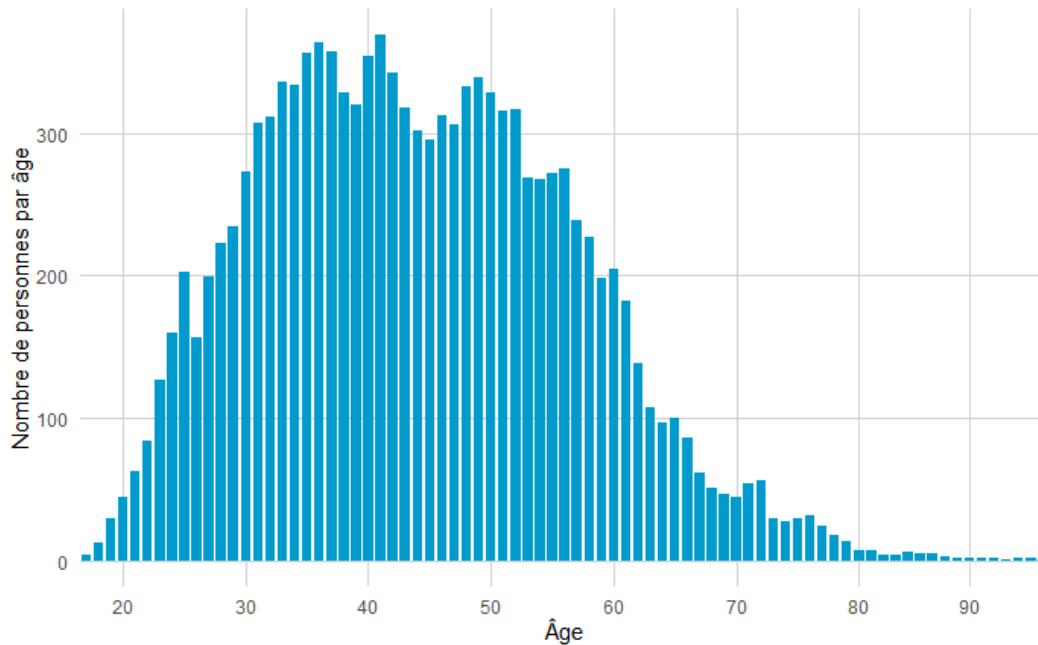


FIGURE 3.5 – Répartition du nombre de personnes par âge

Le graphique ci-dessus illustre la répartition du nombre de personnes par âge pour les années 2021 et 2022. Les âges varient de 16 à 95 ans comprenant les personnes actives et retraitées, avec une distribution croissante de 16 à 40 ans, suivie d’une décroissance jusqu’à 95 ans. La majorité des personnes dans la base ont un âge compris entre 28 et 60 ans.

Les graphiques suivants comportent deux échelles : une à gauche pour le graphique en barres bleu clair, et une à droite pour la courbe bleue foncée.

La répartition du nombre de prestations en fonction des âges permet d’identifier les tranches d’âge qui bénéficient le plus de soins de santé. Pour une comparaison fiable, il est essentiel de confronter cette répartition avec celle des âges dans la base de données.

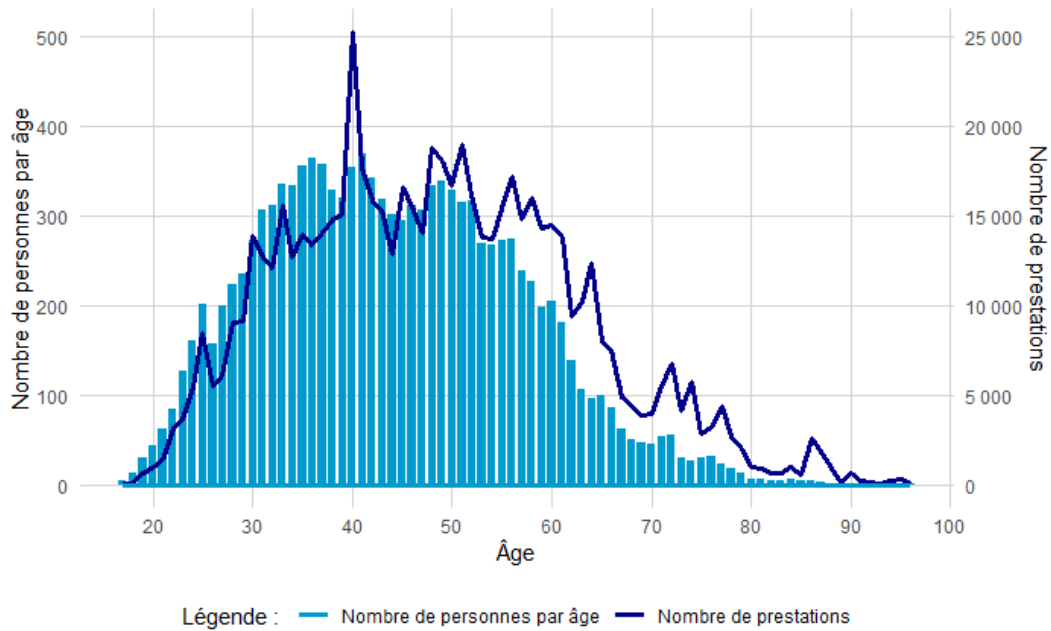


FIGURE 3.6 – Répartition du nombre de personnes par âge et du nombre de prestations par âge

Par rapport au nombre de personnes âgées de 20 à 40 ans dans la base, le nombre de prestations paraît faible. En revanche, l'effet inverse est observable pour la tranche des 40 ans et plus, où le nombre de prestations semble élevé par rapport au nombre de personnes. Un écart significatif entre les deux courbes apparaît autour de 55 ans et se maintient jusqu'à environ 90 ans. Cet écart se réduit après 90 ans en raison du faible nombre de personnes présentes, entraînant une diminution du nombre de prestations.

En divisant le nombre total de prestations pour tous les postes de soins par le nombre de personnes dans chaque tranche d'âge, il est possible de déterminer le nombre moyen de prestations par âge. Par exemple, une personne de 20 ans reçoit en moyenne 19,77 prestations par an, tandis qu'une personne de 60 ans en reçoit en moyenne 67,02 par an. Cette tendance montre une augmentation du nombre de prestations moyen avec l'âge. Ce phénomène est en accord avec les observations générales selon lesquelles les personnes plus âgées ont tendance à bénéficier de plus de prestations de santé.

Dans la même optique, il est pertinent de visualiser la somme des dépenses de prestations en fonction des âges pour déterminer quelles tranches d'âge engagent le plus de coûts. La répartition des dépenses de soins est également comparée au nombre de personnes par âge afin d'assurer une comparaison optimale.

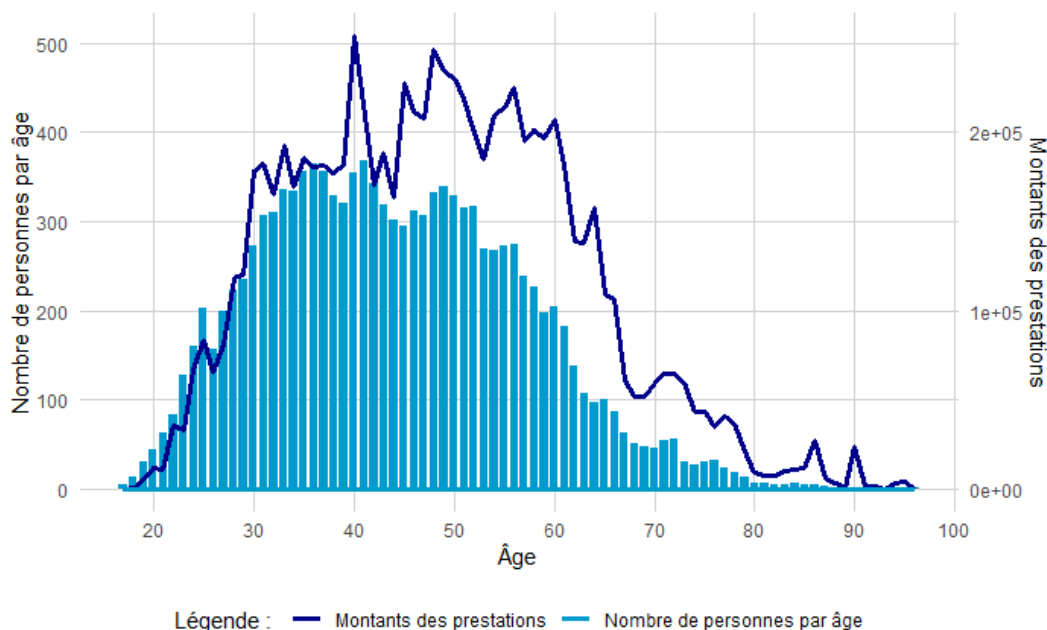


FIGURE 3.7 – Répartition du nombre de personnes par âge et de la somme des montants de prestations par âge

Par rapport au nombre de personnes âgées de 20 à 30 ans dans la base, les montants des prestations semblent faibles. En revanche, l'effet inverse est observable chez les personnes de 40 ans et plus, où les montants des prestations sont élevés par rapport au nombre de personnes. Un écart significatif entre les deux courbes apparaît autour de 40 ans et se maintient jusqu'à environ 90 ans. Cet écart diminue après 90 ans en raison du faible nombre de personnes, ce qui entraîne une réduction du montant total des dépenses de prestations.

En divisant la somme des montants des prestations pour chaque tranche d'âge par le nombre de personnes dans cette tranche, le montant moyen des prestations pour cet âge est obtenu. Par exemple, une personne de 20 ans reçoit en moyenne 258,51 € de prestations par an, tandis qu'une personne de 60 ans reçoit en moyenne 1009,15 € par an. Ainsi, les montants de prestations augmentent avec l'âge. Cette tendance est cohérente avec le graphique précédent : d'une part, les personnes plus âgées bénéficient d'un plus grand nombre de prestations, ce qui entraîne des coûts plus élevés ; d'autre part, les prestations nécessaires avec l'âge, telles que les prothèses dentaires, tendent à être plus coûteuses. La répartition des coûts pour le poste d'hospitalisation pourrait également varier, étant donné que les personnes âgées ont tendance à utiliser davantage de services hospitaliers, qui sont généralement plus coûteux.

3.3.3 Présentation des différents postes de soins

La suite de l'analyse descriptive se concentre sur les postes de soins, en évaluant la fréquence et le coût moyen par poste et sous-poste de soins. Seules les données pertinentes pour les modèles futurs sont conservées. Ainsi, les retraités, y compris les retraités loi Evin, les sorties de groupe et les sorties loi Evin, sont exclus de l'analyse. De plus, les modèles sont réalisés en considérant uniquement les personnes âgées de 62 ans ou moins ; les personnes de 63 ans et plus sont donc supprimées des bases de prestations et de personnes.

Pour cette partie de l'analyse, la base de prestations comprend 495 686 observations.

La base de données des prestations comprend plusieurs postes de soins. Chaque prestation est assignée à un type de soin spécifique, ce qui permet de les classer et d'analyser la répartition des prestations parmi ces types. La base de prestations couvre toujours une période de 2 ans.

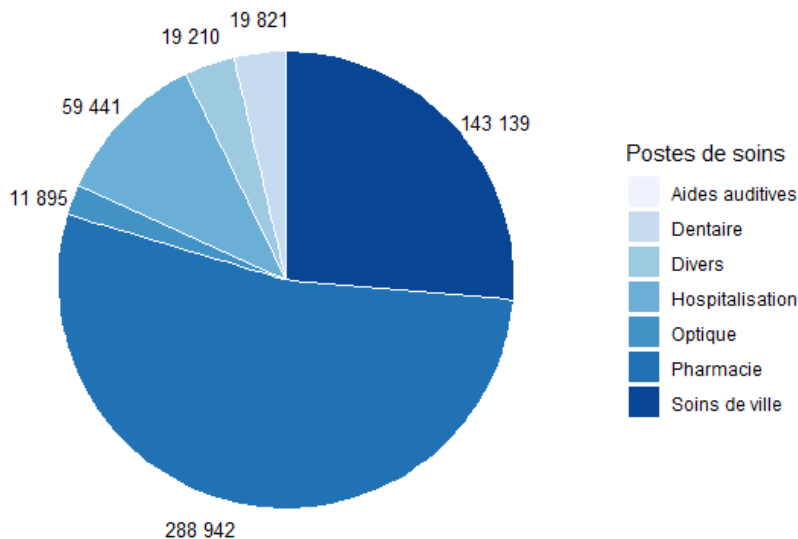


FIGURE 3.8 – Répartition du nombre de prestations par poste de soins

La majorité des prestations concernent des soins pharmaceutiques, principalement des médicaments. Les soins de ville composés des consultations chez les médecins généralistes ou spécialistes ou des analyses en laboratoire sont également significatifs, représentant environ un tiers des prestations, avec une part importante dédiée aux consultations médicales. Le poste hospitalisation est également assez présent. Les autres postes de soins comprennent l'optique, le dentaire, les soins divers (telles que les contraceptions et les cures thermales) et les aides auditives. Bien que les prestations pour les aides auditives soient relativement faibles, avec un total de 142 prestations, ce poste sera conservé dans l'étude pour évaluer son coût pour la mutuelle.

Le graphique circulaire précédent illustre la répartition du nombre de prestations par poste de soins. Il est également important d'examiner le coût moyen et la fréquence de chaque poste de soins. Dans la suite de l'étude, ces mesures seront prédites pour les personnes de 63 et 64 ans, qui seront amenées à travailler deux années supplémentaires.

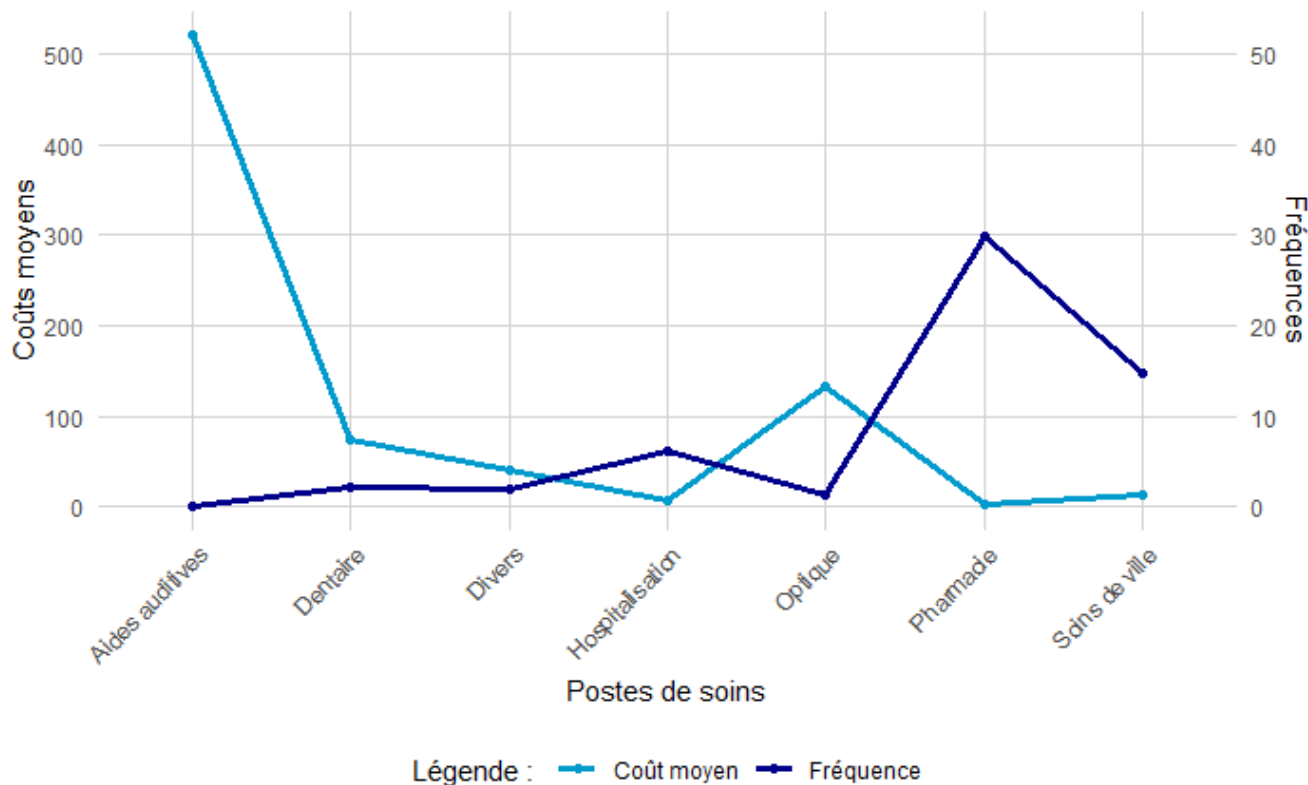


FIGURE 3.9 – Coût moyen et fréquence par poste de soins

D’après le graphique circulaire précédent, les soins de pharmacie comptent le plus grand nombre de prestations. Ce poste affiche la fréquence la plus élevée, à 29,92 par an, avec un coût moyen relativement faible de 1,84€. Les soins de ville suivent avec une fréquence élevée de 14,82 et un coût moyen également modéré de 12,82€. Il est logique que ces deux postes présentent les fréquences les plus élevées et les coûts moyens les plus bas, car ils sont à la fois les plus fréquemment utilisés et les moins coûteux pour les assurés.

En revanche, les postes tels que l’optique, le dentaire, le divers et l’hospitalisation montrent des fréquences plus faibles, variant entre 1,23 et 6,15. Les aides auditives, avec une fréquence particulièrement basse de 0,01, sont aussi très peu courantes. Cela semble logique, car le nombre de personnes concernées est relativement faible et, parmi celles-ci, les garanties sont souvent limitées. Cependant, elles présentent le coût moyen le plus élevé, à 520,29€ par an, suivies par l’optique, le dentaire et le divers.

Il convient de noter que 60% des coûts moyens associés au poste d’hospitalisation ont été déterminés de manière arbitraire dans la section 3.2.3.1.

3.3.4 Présentation des sous postes de soins

Chaque poste de soins se divise en plusieurs sous-postes, qui détaillent les prestations spécifiques fournies. Les modèles ultérieurs se baseront sur ces sous-postes. Les sous-postes de soins sont analysés en fonction de leur coût moyen et de leur fréquence annuelle.

3.3.4.1 Optique

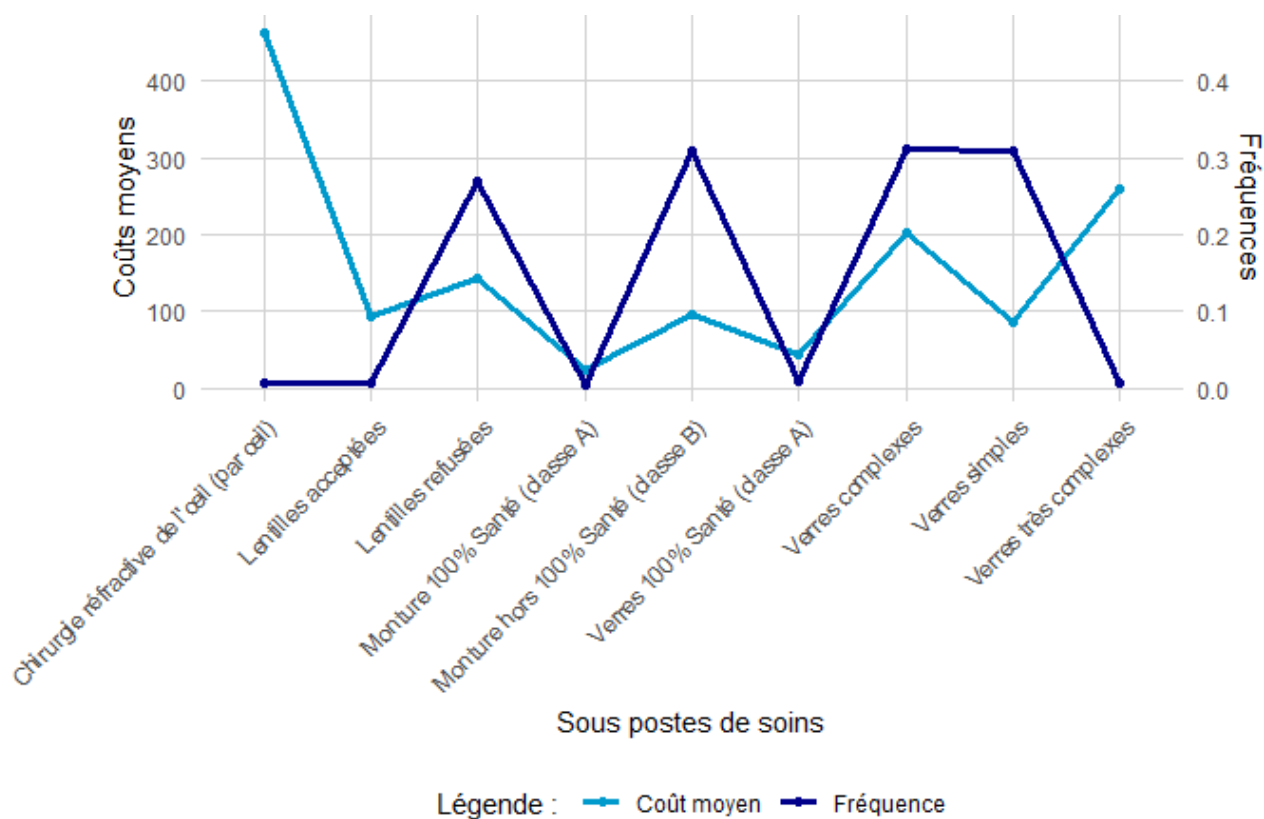


FIGURE 3.10 – Coût moyen et fréquence de l'optique

La chirurgie réfractive de l'œil affiche le coût moyen le plus élevé, à 463,02€, avec une fréquence très faible de 0,0057, ce qui est logique étant donné le coût élevé et la rareté de cette procédure. Les lentilles refusées et les montures hors 100% santé classe B présentent des fréquences assez élevées.

En ce qui concerne les verres, les verres complexes ont la fréquence la plus élevée et un coût moyen intermédiaire, cela est cohérent, car en général, les ophtalmologues prescrivent plus fréquemment ce type de verre. Les verres très complexes, quant à eux, présentent un coût moyen élevé pour une fréquence très faible. Enfin, les sous-postes de 100% santé classe A se caractérisent par les coûts moyens les plus bas, étant limité par la sécurité sociale.

3.3.4.2 Dentaire

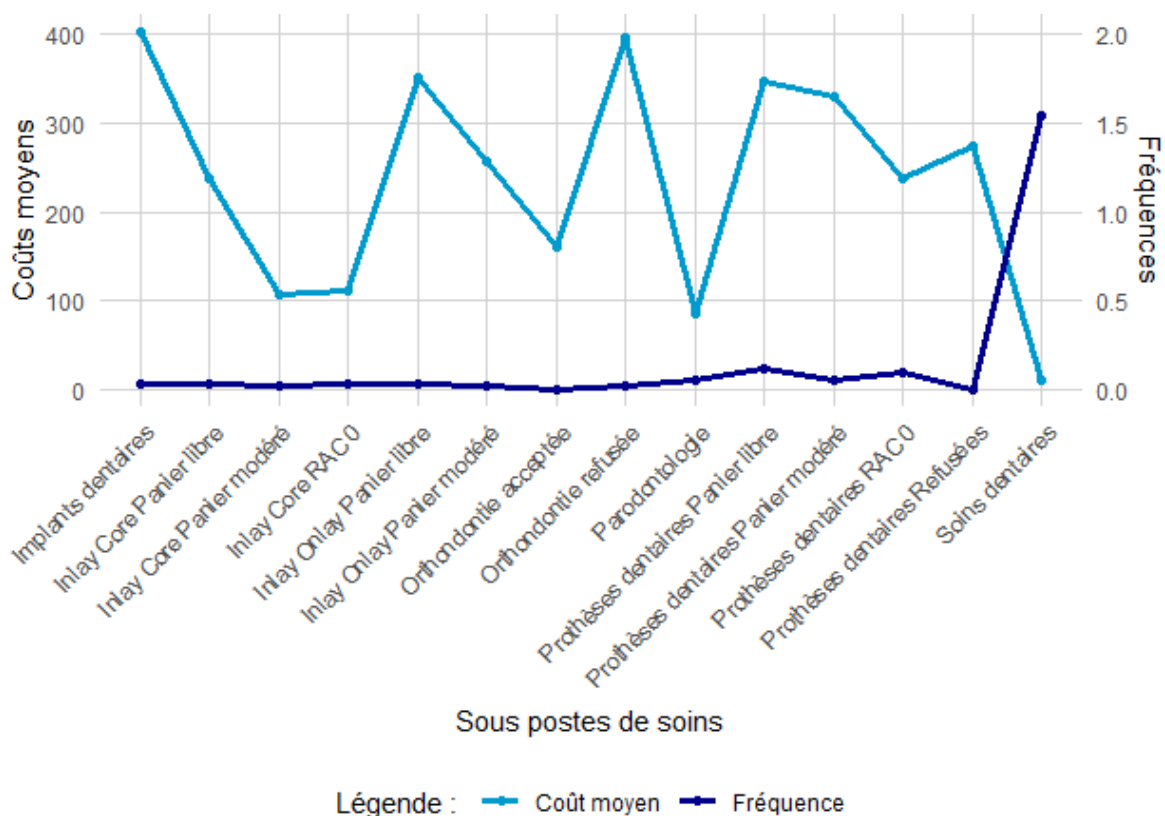


FIGURE 3.11 – Coût moyen et fréquence du dentaire

Tous les sous-postes dentaires affichent des coûts moyens compris entre 85 et 402€, à l'exception des soins dentaires, dont le coût moyen remboursé est relativement faible, proche de 10€. Cette différence est logique, car les soins dentaires sont généralement moins coûteux que les autres prestations dentaires.

En ce qui concerne les fréquences, elles sont inférieures à 1 pour tous les sous-postes dentaires, sauf pour les soins dentaires, qui présentent une fréquence de 1,55. Cela est cohérent, car les soins dentaires est le sous-poste le plus fréquemment utilisé dans le domaine dentaire. À cela s'ajoutent d'éventuelles limites du contrat, qui peuvent dissuader les personnes de recourir à des prestations coûteuses.

3.3.4.3 Soins de ville

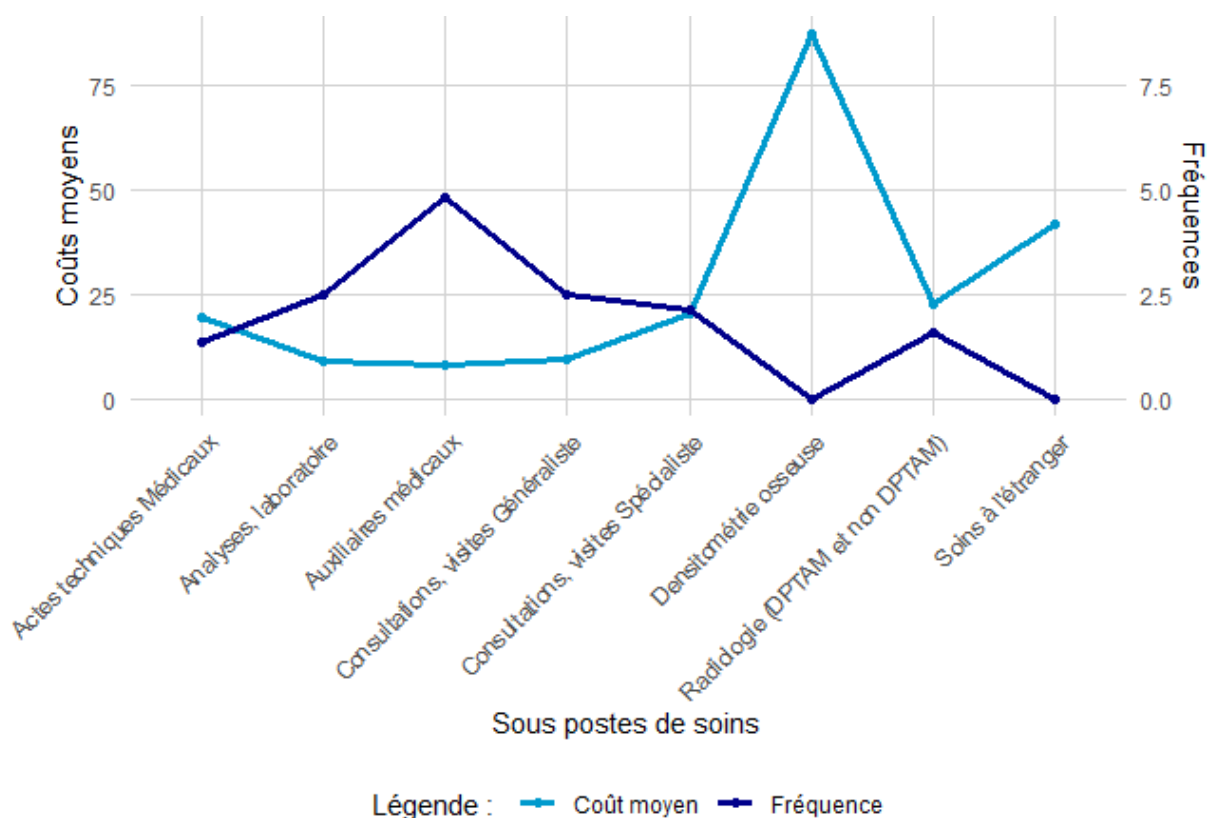


FIGURE 3.12 – Coût moyen et fréquence des soins de ville

La densitométrie osseuse (examen médical qui permet de mesurer la densité minérale osseuse) affiche le coût moyen le plus élevé, à 87€, et une fréquence très faible de 0,0003, ce qui est cohérent avec sa nature coûteuse et peu fréquente. Les autres coûts moyens des prestations varient entre 7,97 et 42€. Par exemple, le coût moyen d'une consultation de spécialiste est plus élevé que celui d'une consultation généraliste, avec des valeurs respectives de 20,40€ et 9,25€, ce qui est attendu puisque les montants des consultations de médecins générales sont globalement encadrés par la sécurité sociale.

En ce qui concerne les fréquences, les auxiliaires médicaux, qui incluent des actes telles que la kinésithérapie et l'orthoptie, ont la fréquence la plus élevée, à 4,79. Ce sous-poste présente également le coût moyen le plus bas, à 7,97€.

3.3.4.4 Pharmacie

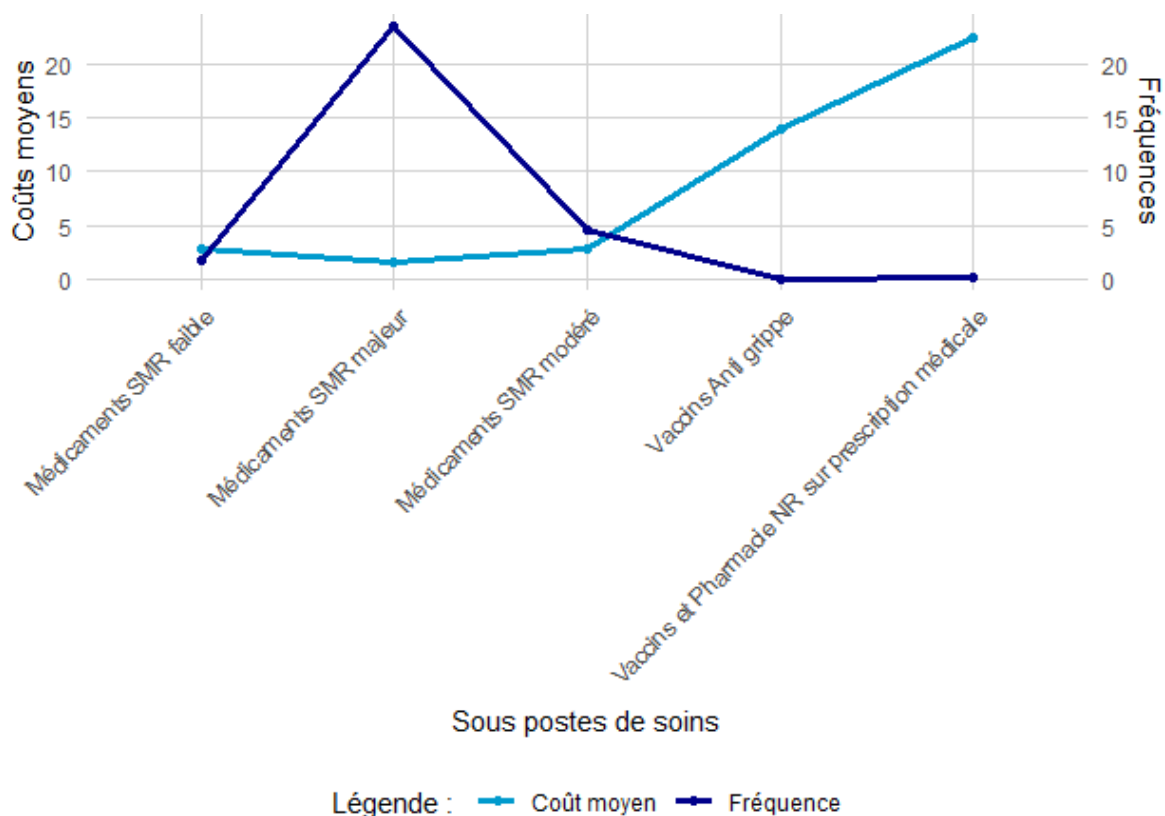


FIGURE 3.13 – Coût moyen et fréquence des soins pharmaceutiques

Les médicaments avec un Service Médical Rendu (SMR) majeur présentent le coût moyen le plus bas, à 1,48€, et la fréquence la plus élevée, à 23,45. Cela est logique, car ces médicaments sont très courants et peu coûteux. En revanche, les autres prestations médicamenteuses affichent des coûts moyens avoisinant 2,85€. Les médicaments SMR faible ont une fréquence de 1,73, tandis que ceux avec un SMR modéré ont une fréquence de 4,62.

Les vaccins, en revanche, affichent les coûts moyens les plus élevés, variant entre 13,91 et 22,49€, avec des fréquences inférieures à 1. Cela est cohérent, étant donné que les vaccins courants sont généralement administrés une fois par an maximum et ont souvent un coût plus élevé.

3.3.4.5 Divers

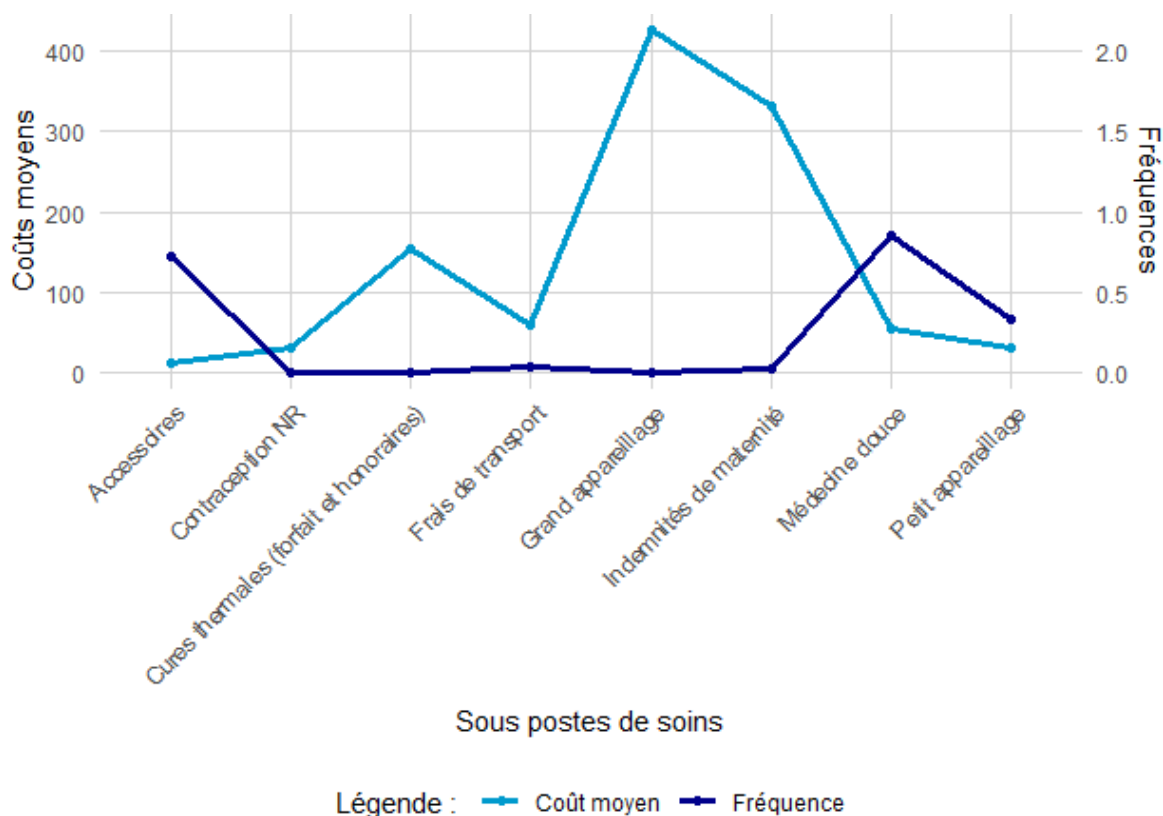


FIGURE 3.14 – Coût moyen et fréquence des soins divers

Le grand appareillage présente le coût moyen le plus élevé, à 424,61€, et une fréquence très faible de 0,001. Cette combinaison est cohérente en raison du coût élevé et de la rareté de cette prestation. Les indemnités de maternité et les cures thermales affichent également des coûts moyens élevés, respectivement de 332,13€ et 154,37€, avec des fréquences très faibles. Ce constat est logique car les prestations de ce poste de soins représentent une petite partie des dépenses globales des personnes protégées. Les autres sous-postes du poste divers ont des coûts moyens compris entre 13 et 60€.

La médecine douce présente une fréquence relativement élevée, d'environ 0,8. Une tendance à la hausse de la consommation de ces prestations est observable au cours des dernières années, ce qui confirme la fréquence obtenue.

3.3.4.6 Hospitalisation

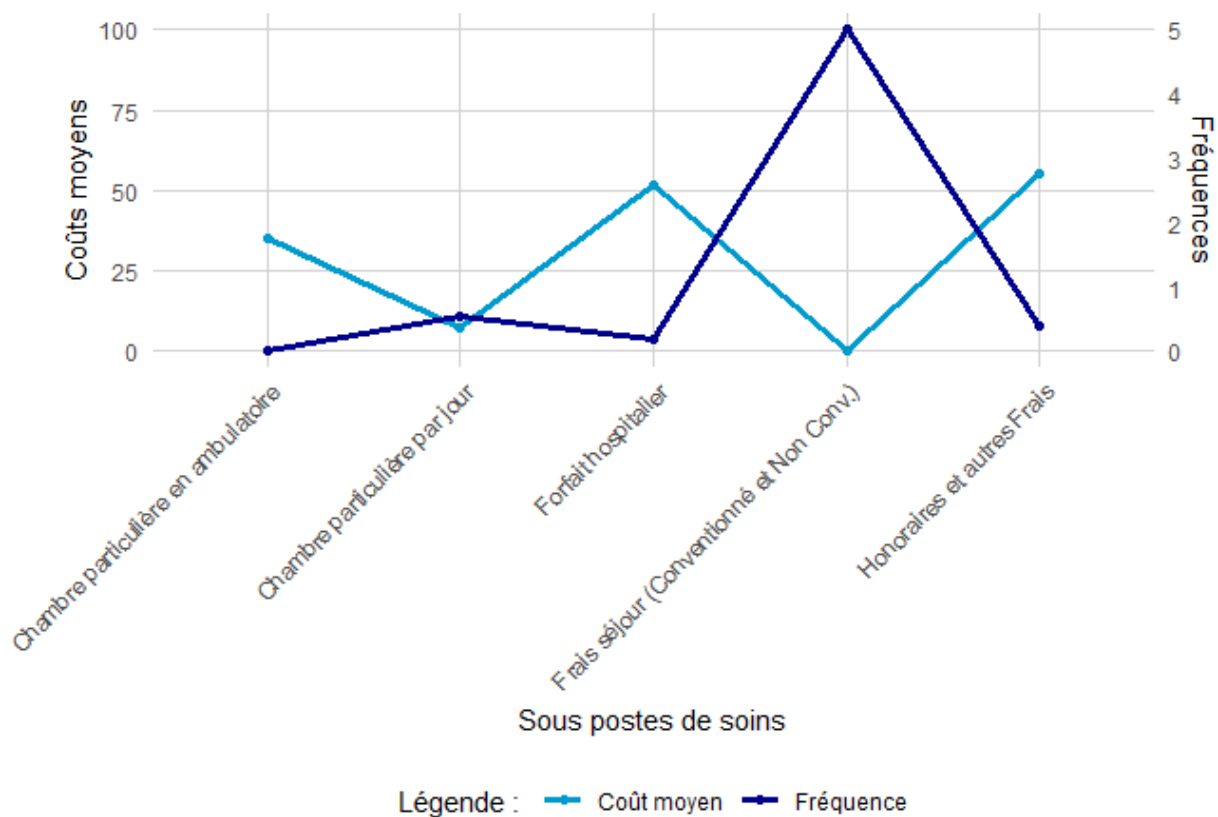


FIGURE 3.15 – Coût moyen et fréquence de l'hospitalisation

Les prestations des honoraires et autres frais présentent le coût moyen le plus élevé, à 55,26€. Les autres sous-postes de l'hospitalisation affichent des coûts moyens variant entre 7,50€ et 51,65€. Les fréquences de ces sous-postes sont toutes inférieures à 0,6 hormis les frais de séjour qui ont une fréquence à 5, ce qui est logique étant donné que ces frais sont consommés à chaque intervention hospitalière.

3.3.4.7 Aides auditives

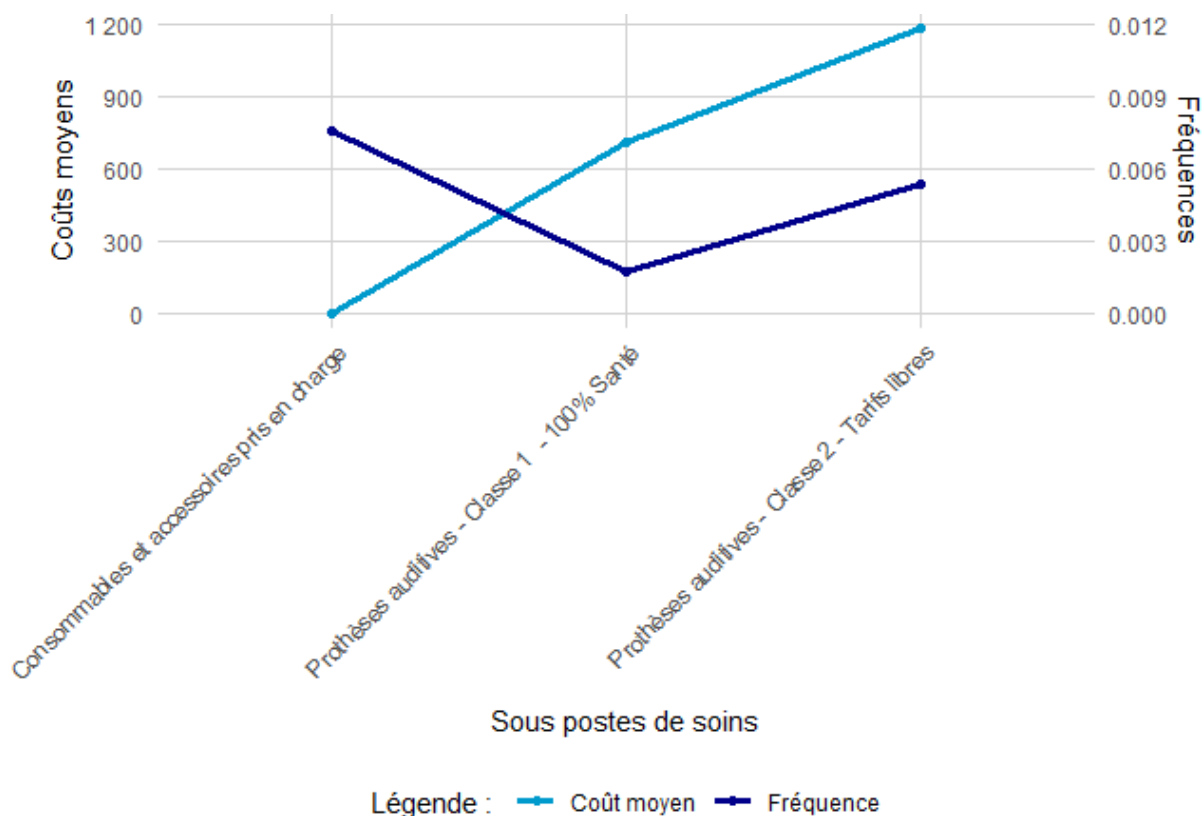


FIGURE 3.16 – Coût moyen et fréquence des aides auditives

Le poste des aides auditives est le plus coûteux dans le domaine de la santé, principalement en raison des prothèses auditives. Le coût moyen des prothèses auditives est élevé, atteignant 710€ pour les tarifs 100% santé et 1 184,39€ pour les tarifs libres. Cela est logique, car les tarifs libres sont généralement plus chers que les tarifs 100% santé car celui-ci permet de limiter les coûts. Les fréquences associées à ces prestations sont toutes inférieures à 0,006, ce qui est cohérent avec le coût élevé, la relative rareté de ces soins et la limite de la garantie. En revanche, les accessoires d'appareillage présentent un coût moyen beaucoup plus bas, à 3,05€, avec une fréquence faible de 0,008, ce qui est également en accord avec leur nature moins coûteuse et peu courante.

3.3.5 Comparaison du comportement entre les personnes actives et retraitées

Pour les modèles présentés ultérieurement, les données utilisées concernent des actifs âgés de 62 ans et moins. L'objectif est de prédire le coût pour les personnes de 63 et 64 ans. Afin de justifier l'importance de ne pas utiliser les données des retraités directement, il est essentiel de comparer le comportement des personnes de 63/64 ans actives avec celui des retraités du même âge. En effet si les comportements et la consommation de soins sont similaires entre les deux groupes, il serait suffisant de tarifier les actifs de 63/64 ans de la même manière que les retraités.

Une comparaison entre les personnes actives de 63/64 ans et les retraités de ces mêmes âges a

été réalisée. Cette comparaison a été effectuée à la fois sur le coût total pour la mutuelle, tous postes confondus, et sur chaque poste de soins individuel.

	Nombre de prestations		Coût X Fréquence		Écart entre actifs et retraités
	Actifs	Retraités	Actifs	Retraités	
Au global	16 624	5 943	1 434,86	2 678,81	86,70%
Optique	128	36	118,75	139,62	17,58%
Dentaire	290	184	187,62	554,81	195,71%
Soins de ville	4 048	1 907	328,92	573,88	74,47%
Pharmacie	9 600	2 776	91,68	129,88	41,67%
Divers	370	147	82,01	79,69	-2,83%
Hospitalisation	2 166	893	1 970,75	3 424,86	73,78%
Aides auditives	22	0	197,91	NaN	NaN

TABLEAU 3.4 – Comparaison du comportement entre les actifs de 63/64 ans et les retraités de 63/64 ans

Tous les montants indiqués dans ce tableau correspondent à la prime pure pour ces personnes, c'est-à-dire le coût pour la mutuelle sur une année. Plus précisément, il s'agit du coût moyen multiplié par la fréquence moyenne des personnes présentes dans la base. Il y a 149 personnes actives de 63/64 ans et 55 personnes retraitées de ces mêmes âges.

En général, les personnes retraitées affichent une prime pure plus élevée que les personnes actives. L'augmentation de la consommation chez les retraités par rapport aux actifs est notable, atteignant presque 90%. La prime pure pour le poste de dentaire chez les retraités est près de trois fois supérieure à celle des actifs, avec une hausse de 195,71%. Les postes d'optique, de soins de ville, d'hospitalisation et de pharmacie montrent des augmentations pour les retraités allant de 17,58% à 74,47%. En revanche, le poste divers, qui englobe la médecine douce, les indemnités de maternité, les contraceptifs, etc., enregistre une légère diminution de 2,83% chez les retraités par rapport aux actifs.

Il est important de considérer l'importance du nombre de prestations pour évaluer ces résultats. En effet, de manière générale, une plus grande quantité de données rend les résultats plus fiables. Il est observable que le nombre de prestations des retraités est inférieur à celui des actifs. Par exemple, le poste d'aides auditives compte seulement 22 prestations pour les actifs et aucune pour les retraités, rendant la comparaison impossible. Même si les retraités avaient un nombre de prestations pour ce poste, 22 prestations pour les actifs restent relativement faibles, nécessitant un nombre plus élevé pour une comparaison optimale.

Selon cette analyse, les actifs et les retraités présentent, de manière générale, des comportements distincts en matière de consommation de soins de santé. Dans l'article de Garrouste.C et Perdrix.E [1], elles affirment que : « *Le passage de la vie active à la retraite entraîne de nombreux bouleversements. [...] Cette transition vers la retraite peut conduire à une modification de l'exposition de l'individu à certaines pathologies. Par exemple, le risque de dépression peut être lié à l'exposition au stress (en particulier dans la période travaillée), mais il peut également avoir pour origine une diminution de la vie sociale de la personne qui survient, parfois, au passage à la retraite. La retraite peut également conduire à une modification de l'exposition aux risques professionnels (accidents du travail et maladies professionnelles). Mais cette exposition n'est pas similaire en fonction de la*

catégorie socioprofessionnelle des individus. »

Bien que de nombreux facteurs, telles que la catégorie socioprofessionnelle, la nature du métier et les caractéristiques individuelles interviennent, la transition vers la retraite peut modifier les comportements des personnes. Ces changements peuvent se traduire par une augmentation ou une diminution de la consommation, en grande partie influencée par l'état de santé physique et l'exposition au risque. Par ailleurs, des facteurs psychologiques jouent également un rôle important dans cette transition de la vie active à la retraite, ces effets dépendant de l'état mental de l'individu.

Chapitre 4

Projection du décalage de l'âge de départ à la retraite à partir d'un GLM

L'objectif est d'estimer les dépenses de santé des individus actifs âgés de 62 à 64 ans dans la base de données, afin d'évaluer l'impact du report de l'âge de départ à la retraite sur les prestations de santé. Pour cela, des méthodes statistiques seront employées, en tenant compte de certaines caractéristiques des assurés. Cette étape est cruciale pour pouvoir ensuite effectuer une tarification et donc évaluer la prime nécessaire supplémentaire qui devra être demandé aux assurés si la mutuelle souhaite couvrir le coût additionnel lié à la réforme des retraites de 2023.

Cette section comprendra d'abord une présentation théorique des modèles utilisés, puis d'une partie pratique en les appliquant aux variables de la base de données.

La partie théorique présentée dans ce chapitre repose sur le cours d'Économétrie de l'assurance non-vie de Thomas M., donné à l'ISUP [26] et sur le site de l'Institut des Actuaire [27].

4.1 La tarification a priori

Le modèle présenté ci-dessous fait partie du système de tarification dit « a priori ». Cette méthode consiste à déterminer la prime d'un assuré en se basant sur ses caractéristiques individuelles. Lorsqu'un client souscrit à un contrat d'assurance, l'assureur doit calculer la prime que l'assuré devra payer, en tenant compte de certains facteurs. Cette prime doit refléter le coût estimé des prestations de l'assuré sur une période donnée et être suffisamment élevée pour que la compagnie d'assurance puisse couvrir d'éventuels sinistres. Pour estimer ce coût de manière optimale, il est nécessaire d'utiliser des méthodes statistiques en s'appuyant sur les caractéristiques spécifiques de l'assuré.

Il est important de distinguer la prime pure et la prime commerciale. La prime pure correspond au montant que l'assureur doit prévoir à $t = 0$ (au début du contrat) pour couvrir le coût estimé de la garantie accordée, c'est-à-dire le montant que l'assuré est susceptible de lui coûter. La prime commerciale correspond à la prime pure additionnée aux frais qu'a l'assureur soit :

$$\text{Prime commerciale} = \text{Prime pure} + \text{Frais de gestion} + \text{Frais d'acquisition} \quad (4.1)$$

D'un point de vue statistique, estimer les montants de dépenses futurs revient à effectuer une régression d'une variable réponse Y , représentant le coût du bénéficiaire, sur un ensemble de variables explicatives X , représentant les caractéristiques de celui-ci. Pour simplifier, les tarifs présentés par la suite sont individuels et non familiaux.

Le principe de la régression est de modéliser $\mathbb{E}[Y|X]$ comme une fonction g des variables explicatives X , soit

$$\mathbb{E}[Y|X] = g(X) \tag{4.2}$$

La variable Y s'écrit alors

$$Y = g(X) + \varepsilon \tag{4.3}$$

où ε est un bruit aléatoire. Il représente l'écart entre Y et son espérance conditionnelle, soit l'erreur obtenue lorsque Y est remplacé par son espérance conditionnelle.

4.2 Modèles Linéaires Gaussiens

4.2.1 Définition

L'objectif des modèles linéaires gaussiens est d'expliquer une variable quantitative à partir de variables explicatives où celles-ci pourront être quantitatives ou qualitatives.

Soit un modèle avec n observations et p variables explicatives. Les variables suivantes sont valables pour la suite :

- $Y = (Y_1, \dots, Y_n)'$ le vecteur colonne représentant les variables réponses ;
- $X_i = (X_{i,1}, \dots, X_{i,p})'$ le vecteur colonne représentant les variables explicatives pour la variable réponse Y_i ;
- X la matrice de taille $n \times p$ dont les lignes sont des vecteurs lignes X_i' ;
- $\beta = (\beta_1, \dots, \beta_p)'$ le vecteur colonne correspondant aux paramètres du modèle ;
- ε le terme d'erreur, c'est à dire $\varepsilon = Y - \mathbb{E}[Y|X]$.

Un modèle est un modèle linéaire gaussien s'il s'écrit sous la forme :

$$Y = X\beta + \varepsilon \tag{4.4}$$

et s'il vérifie les hypothèses suivantes :

- (\mathcal{H}_1) la matrice X est de plein rang, soit $\text{rang}(X) = p$;
- (\mathcal{H}_2) le bruit ε est gaussien centré et ses éléments sont non corrélés,

$$\varepsilon \sim \mathcal{N}_p(0, \sigma^2 I) \tag{4.5}$$

4.2.2 Estimation des paramètres

4.2.2.1 Estimation par Maximum de Vraisemblance

L'estimation par maximum de vraisemblance est basée sur la vraisemblance du modèle linéaire gaussien :

$$\mathcal{L}(\beta, Y) = \prod_{i=1}^n f(y_i, \beta) \quad (4.6)$$

où $f(y_i, \beta)$ la densité de la loi normale sur Y .

Pour obtenir l'estimateur du maximum de vraisemblance $\hat{\beta}$, il faut maximiser sa log-vraisemblance en fonction de β . Pour cela, il faut résoudre le système d'équation suivant, pour lequel $\hat{\beta}$ est solution :

$$\frac{\partial}{\partial \beta_j} \ln (\mathcal{L}(\beta_1, \dots, \beta_k; Y)) = 0 \text{ pour } j = 1, \dots, p \quad (4.7)$$

Il est également possible d'obtenir l'estimateur du maximum de vraisemblance de σ^2 en maximisant la log-vraisemblance selon σ^2 .

4.2.2.2 Estimation par la méthode des moindres carrés

Cette méthode vise à minimiser la somme des carrés des écarts pondérés entre les observations et les projections du modèle. Sous les hypothèses précédentes, cette méthode est appliquée pour estimer β par :

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|^2 \quad (4.8)$$

De plus, $\hat{\beta}$ est unique, sans biais et est donné par la formule :

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (4.9)$$

Par ailleurs, un estimateur sans biais de la variance σ^2 est donné par :

$$\hat{\sigma}^2 = \frac{1}{n-p} \|Y - X\hat{\beta}\|^2 \quad (4.10)$$

4.3 Modèles Linéaires Généralisés (GLM)

Initialement développés en 1972 par Nelder et Wedderburn, les Modèles Linéaires Généralisés (GLM) permettent d'analyser des données discrètes et continues, pour lesquelles la loi normale n'est plus appropriée. Ces modèles étendent les modèles linéaires classiques en élargissant les types de distributions possibles, ce qui permet de surmonter certaines limites des modèles traditionnels. Contrairement aux modèles linéaires classiques, les GLM n'exigent pas que la variable à expliquer suive une distribution normale ; il suffit qu'elle appartienne à une famille exponentielle. Cependant,

une condition clé des GLM est qu'ils supposent que les observations soient indépendantes, ou du moins non corrélées.

Les GLM permettront de prédire les coûts moyens et les fréquences futures pour les personnes âgées de 63 et 64 ans, en se basant sur diverses caractéristiques qui leur sont associées.

4.3.1 Familles exponentielles

Un modèle statistique $(\Omega, \mathcal{F}, (\mathbb{P}_{\theta, \phi})_{\theta \in \Theta, \phi > 0})$ est appelé famille exponentielle si les probabilités $\mathbb{P}_{\theta, \phi}$ admettent une densité f par rapport à une mesure dominante avec

$$f_{\theta, \phi}(y) = c_{\phi}(y) \exp\left(\frac{y\theta - a(\theta)}{\phi}\right) \quad (4.11)$$

où θ est le paramètre canonique et ϕ le paramètre de dispersion, souvent considéré comme un paramètre de nuisance. Un paramètre de nuisance est un paramètre qui n'est pas d'un intérêt immédiat mais qui doit être pris en compte dans l'analyse des paramètres d'intérêt. $a(\theta)$ est de classe C^2 et est convexe et $c_{\phi}(y)$ ne dépend pas de θ .

Si Y est distribuée selon une loi appartenant à une famille exponentielle, alors

$$\mathbb{E}[Y] = a'(\theta) \text{ et } \text{Var}[Y] = \phi a''(\theta) \quad (4.12)$$

où $\mathbb{E}[Y]$ et $\text{Var}[Y]$ sont respectivement l'espérance et la variance de Y .

Plusieurs distributions appartiennent à une famille exponentielle notamment :

- **Loi normale** de paramètres m et σ^2 avec σ^2 connu

$$\theta = m, \quad a(\theta) = \frac{\theta^2}{2}, \quad \phi = \sigma^2; \quad (4.13)$$

- **Loi exponentielle ou Gamma** de paramètres k et λ , avec k connu

$$\theta = -\frac{1}{\lambda}, \quad a(\theta) = -k \log(-\theta), \quad \phi = 1; \quad (4.14)$$

- **Loi de Poisson** de paramètre λ

$$\theta = \log \lambda, \quad a(\theta) = e^{\theta}, \quad \phi = 1; \quad (4.15)$$

- **Loi de Bernoulli ou Binomiale** de paramètre n et p , avec n connu

$$\theta = \log\left(\frac{p}{1-p}\right), \quad a(\theta) = n \log(1 + e^{\theta}), \quad \phi = 1; \quad (4.16)$$

- **Loi Binomiale négative** de paramètres p et r , avec r connu

$$\theta = \log(1-p), \quad a(\theta) = -r \log(1 - e^{\theta}), \quad \phi = 1. \quad (4.17)$$

4.3.2 Composantes d'un GLM

Un GLM est constitué de trois composantes :

4.3.2.1 Composante aléatoire

Cette composante est caractérisée par la loi de probabilité de la variable d'intérêt Y . Cette loi appartient à la famille exponentielle et est de la forme :

$$f_{\theta,\phi}(y) = c_{\phi}(y) \exp\left(\frac{y\theta - a(\theta)}{\phi}\right) \quad (4.18)$$

comme vu précédemment.

4.3.2.2 Composante déterministe

La composante déterministe représente le prédicteur linéaire qui est défini par le vecteur de taille n suivant :

$$\eta = X\beta = \beta_0 + \sum_{i=1}^p X_i\beta_i = \beta_0 + \beta_1x_1 + \dots + \beta_px_p \quad (4.19)$$

où β est le vecteur des p paramètres à estimer et X la matrice des X_i variables explicatives.

4.3.2.3 Fonction de lien

La fonction de lien permet d'établir un lien entre la composante déterministe c'est à dire le prédicteur linéaire et l'espérance de Y . Cette fonction, notée g , est bijective et elle peut être notée de la façon suivante :

$$g(\mu(X)) = g(\mathbb{E}[Y|X]) = X\beta = \eta. \quad (4.20)$$

4.3.3 Définition

En reprenant les notations utilisées précédemment, un modèle est un modèle linéaire généralisé s'il vérifie les hypothèses suivantes :

1. $Y|X = x \sim \mathbb{P}_{\theta,\phi}$ appartient à une famille exponentielle ;
2. $g(\mu(X)) = g(\mathbb{E}[Y|X]) = X\beta$ pour une certaine fonction g bijective, appelé fonction de lien.

Si plusieurs fonctions de lien peuvent convenir vis-à-vis d'une variable Y , le choix se portera toujours la fonction de lien qui permet de minimiser les écarts entre les valeurs prédites par le modèle et les valeurs réelles.

4.3.4 Estimation des paramètres

Les paramètres à estimer sont β et ϕ .

4.3.4.1 Estimation de β

Les principales méthodes d'estimation du vecteur des paramètres β sont les suivantes :

- Méthode du maximum de vraisemblance ;
- Méthode des moindres carrés ordinaires ;
- Méthode des moments.

En général, la méthode du maximum de vraisemblance fournit la meilleure estimation en offrant une plus grande flexibilité quant au type de données et en réduisant les hypothèses à respecter concernant les résidus. Cependant, il est important de souligner que cette estimation peut être biaisée dans le cas d'un échantillon fini. En raison de ce biais potentiel, certains praticiens préfèrent utiliser les estimations non biaisées fournies par la méthode des moindres carrés ordinaires, bien qu'elles nécessitent un cadre théorique plus spécifique.

β est ainsi estimé par maximum de vraisemblance. Ces estimations correspondent à la résolution d'un système d'équations des dérivées de la fonction de log-vraisemblance par rapport à chacun des paramètres.

Soit :

$$\begin{cases} \eta_i = X_i' \beta \\ \mu_i = \mathbb{E}[Y_i | X_i] = g^{-1}(X_i' \beta) = g^{-1}(\eta_i) \\ \theta_i = (a')^{-1}(\mu_i) = (a')^{-1}(g^{-1}(X_i' \beta)) = (a')^{-1}(g^{-1}(\eta_i)) \end{cases} \quad (4.21)$$

Les Y_i sont supposés indépendants, la log-vraisemblance s'écrit :

$$\ell(\beta) = \sum_{i=1}^n \log f(Y_i; \beta, \phi) = \sum_{i=1}^n \underbrace{\left\{ \log c_\phi(Y_i) + \frac{Y_i \theta_i - a(\theta_i)}{\phi} \right\}}_{:=\ell_i(\theta_i)} \quad (4.22)$$

Soit l'estimateur du maximum de vraisemblance de β_j ,

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i(\theta_i)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta_j} \quad (4.23)$$

D'une part,

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{Y_i - a'(\theta_i)}{\phi} = \frac{Y_i - \mu_i}{\phi} \quad (4.24)$$

D'autre part, comme $\eta_i = X_i' \beta$ cela donne,

$$\frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \eta_i} x_{i,j} \quad (4.25)$$

où $x_{i,j}$ est la j^e coordonnées de X'_i .

Enfin,

$$\frac{\partial \theta_i}{\partial \eta_i} = \left(\frac{\partial \eta_i}{\partial \theta_i} \right)^{-1} \quad (4.26)$$

Et,

$$\frac{\partial \eta_i}{\partial \theta_i} = \frac{\partial \eta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \theta_i} = g'(\mu_i) a''(\theta_i) \quad (4.27)$$

Ainsi,

$$\frac{\partial \ell}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^n \frac{x_{i,j}(Y_i - \mu_i)}{g'(\mu_i) a''(\theta_i)} \quad (4.28)$$

Soit D la matrice diagonale dont les coefficients sont égaux à $\frac{1}{g'(\mu_i) a''(\theta_i)}$, alors :

$$\frac{\partial \ell}{\partial \beta_j} = 0 \text{ pour tout } j = 1, \dots, p \iff X'D(Y - \mu) = 0 \quad (4.29)$$

Ainsi l'estimateur du maximum de vraisemblance noté $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ est solution de l'équation :

$$X'D(Y - g^{-1}(X\beta)) = 0 \quad (4.30)$$

Dans le cas général, cette équation ne peut pas se résoudre explicitement. L'estimateur du maximum de vraisemblance est calculé numériquement, souvent grâce à l'algorithme de Newton-Ramphson.

4.3.5 Sélection de modèles et déviance

Une fois les paramètres estimés, il est crucial de s'assurer que le modèle représente fidèlement la réalité et qu'il est valide. Toutefois, dans le cas des modèles linéaires généralisés, les informations sur les résidus sont limitées. Notamment, ces résidus n'ont pas de raison d'avoir une variance constante, et leur distribution reste inconnue.

Une façon classique de procéder est de comparer le modèle ajusté \mathcal{M} à un modèle plus général qui a autant de paramètres que de variables réponses et donc estime parfaitement les données. Ce modèle est appelé modèle saturé et est noté $\widetilde{\mathcal{M}}$. C'est un modèle linéaire généralisé avec la même distribution et la même fonction de lien que le modèle \mathcal{M} . En effet, s'il y a n observations Y_1, \dots, Y_n , alors le modèle saturé est décrit avec n paramètres en supposant que X est de rang n .

L'estimateur du maximum de vraisemblance de β sous le modèle saturé est noté $\tilde{\beta}$. Il est solution de l'équation :

$$X'D(Y - g^{-1}(X\beta)) = 0. \quad (4.31)$$

Or, dans le modèle saturé, la matrice X' est de rang n donc elle est inversible, donc l'estimateur du maximum de vraisemblance $\tilde{\beta}$ est tel que :

$$Y - g^{-1}(X\tilde{\beta}) = 0. \quad (4.32)$$

Ainsi, pour tout i ,

$$Y_i = g^{-1}(X'_i\tilde{\beta}) = \tilde{Y}_i \quad (4.33)$$

où \tilde{Y}_i est la valeur ajustée du modèle saturé. Chaque valeur ajustée est donc égale à une observation et le modèle saturé est parfaitement ajusté.

Dans la suite, la log-vraisemblance du modèle saturé $\tilde{\mathcal{M}}$ est notée $\tilde{\ell}$ et celle du modèle estimé \mathcal{M} est notée $\hat{\ell}$. Il est possible de comparer ces deux modèles à partir de leur log-vraisemblance et ensuite de faire un test de rapports de vraisemblance. Soit :

$$\log(\lambda) = \tilde{\ell} - \hat{\ell}. \quad (4.34)$$

Si les valeurs de $\log(\lambda)$ sont grandes, cela signifie que le modèle estimé décrit mal les données par rapport au modèle saturé. Pour déterminer ce que représente de grandes valeurs pour $\log(\lambda)$, il est nécessaire de définir une région critique. Pour cela la loi, éventuellement asymptotique, de λ est nécessaire. La définition de la *déviante* et le *théorème de Wilks* sont essentiels à la suite.

4.3.5.1 Déviante

La déviante est définie comme

$$\Delta = 2 \log(\lambda) = 2(\tilde{\ell} - \hat{\ell}) \quad (4.35)$$

c'est à dire

$$\Delta = 2 \sum_{i=1}^n \frac{Y_i(\tilde{\theta}_i - \hat{\theta}_i) + a(\tilde{\theta}_i) - a(\hat{\theta}_i)}{\phi} \quad (4.36)$$

Les déviances résiduelles (deviance residuals) δ_i sont aussi définis par :

$$\delta_i = \text{sign}(y_i - \mu_i) \sqrt{\delta_i^2} \quad (4.37)$$

où

$$\delta_i^2 = 2 \frac{Y_i(\tilde{\theta}_i - \hat{\theta}_i) + a(\tilde{\theta}_i) - a(\hat{\theta}_i)}{\phi} \quad (4.38)$$

δ_i correspond à la racine carrée de la contribution de l'observation i sur la déviante.

4.3.5.2 Théorème de Wilks

Si les hypothèses du GLM sont vérifiées, alors

$$\Delta \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_{n-p}^2 \quad (4.39)$$

avec p le nombre de paramètres à estimer.

D'après le théorème de Wilks, la déviance peut être approchée par une loi du χ_{n-p}^2 où p est le nombre de paramètres à estimer dans le modèle estimé. Ainsi, la comparaison du modèle estimé avec le modèle saturé revient à comparer la déviance à une loi du χ^2 au bon nombre de degrés de liberté. La déviance peut être écrite de la façon suivante :

$$\Delta = \sum_{i=1}^n \delta_i^2 \quad (4.40)$$

Si δ_i est trop grand, alors l'assuré i contribue au fait que la déviance soit grande. Comme l'espérance de Δ est $n-p$, chaque δ_i^2 devrait valoir approximativement $(n-p)/n \approx 1$. Une valeur de δ_i^2 éloignée de 1 indique que l'observation i contribue au mauvais ajustement.

4.3.5.3 Mesures de performances

AIC et BIC

D'autres méthodes existent pour comparer des modèles qui ne seraient pas emboîtés, notamment les critères d'AIC et de BIC qui se basent sur la vraisemblance.

- L'**AIC (Akaike's Information Criterion)** qui vaut

$$AIC(\mathcal{M}) = -2 \log \mathcal{L}_{\mathcal{M}}(\beta) + 2p \quad (4.41)$$

- Le **BIC (Bayesian Information Criterion)**

$$BIC(\mathcal{M}) = -2 \log \mathcal{L}_{\mathcal{M}}(\beta) + p \log(n) \quad (4.42)$$

où \mathcal{M} est le modèle estimé.

Étant donné que le BIC impose une pénalisation plus forte aux modèles de grande dimension, il tend à sélectionner des modèles plus simples que ceux choisis par le critère AIC. Dans les deux cas, l'objectif est de minimiser la valeur du critère d'information.

Pour réduire la dimension des modèles à l'aide de ces critères, il convient d'utiliser une méthode dite *pas à pas* :

- **Ascendante** ou **forward** : Il suffit de commencer avec le modèle le plus simple, qui ne contient que la constante, puis d'ajouter progressivement les variables qui améliorent le critère. L'algorithme s'arrête lorsque plus aucune variable ne peut améliorer le critère.

- **Descendante** ou **backward** : Il est nécessaire de commencer avec le modèle complet, incluant toutes les variables, puis de supprimer progressivement la variable la moins informative. L'algorithme s'arrête lorsque toutes les variables restantes sont jugées significatives.
- **Stepwise** : À chaque étape, il est possible soit d'ajouter, soit d'éliminer une variable.

Pseudo- R^2

Les pseudo- R^2 sont des indicateurs de performances permettant de déterminer si un modèle est suffisamment concluant. Ils sont basés sur la comparaison des vraisemblances ou des log-vraisemblances du modèle ajusté \mathcal{M} et du modèle nul \mathcal{M}_0 c'est à dire le modèle composé uniquement de la constante.

Plusieurs types de pseudo- R^2 existent :

- R^2 de McFadden :

$$R_{MF}^2 = 1 - \frac{\hat{\ell}}{\ell_0} \quad (4.43)$$

où $\hat{\ell}$ est la log-vraisemblance du modèle estimé et ℓ_0 la log-vraisemblance du modèle nul.

- R^2 de Cox et Snell :

$$R_{CS}^2 = 1 - \left(\frac{\mathcal{L}_0}{\mathcal{L}_{\mathcal{M}}} \right)^{\frac{2}{n}} \quad (4.44)$$

où $\mathcal{L}_{\mathcal{M}}$ est la vraisemblance du modèle estimé et \mathcal{L}_0 la vraisemblance du modèle nul.

- R^2 de Nagelkerke :

$$R_N^2 = \frac{R_{CS}^2}{\max[R_{CS}^2]} \quad (4.45)$$

où R_{CS}^2 est le R^2 de Cox et Snell.

Pour tous les pseudo- R^2 , si $R^2 = 0$, il n'est pas possible de faire mieux que le modèle trivial et si $R^2 = 1$, le modèle est parfaitement ajusté et est donc parfait.

Dans ce mémoire, seul le pseudo- R^2 de McFadden sera utilisé. Il est calculé directement à partir de la déviance du modèle ajusté et la déviance du modèle nul :

$$R^2 = 1 - \frac{D_{\mathcal{M}}}{D_0} \quad (4.46)$$

4.3.6 Interprétation d'une sortie R d'un GLM

Lors de l'affichage du modèle GLM construit, plusieurs valeurs sont présentées. Quatre colonnes principales apparaissent :

- **Estimate** : Il s’agit des estimations qui correspondent aux coefficients des variables indépendantes dans le modèle ;
- **Std Error** : Les erreurs standards des estimations des coefficients sont également affichées. Elles mesurent la variabilité des estimations. Une petite erreur standard indique que l’estimation du coefficient est précise et que les données fournissent une bonne information sur la relation entre la variable explicative et la variable dépendante ;
- **t value** : Il s’agit du rapport entre l’estimation du coefficient et son erreur standard (Estimate / Std Error). Ce rapport, souvent appelé statistique t , permet de tester l’hypothèse nulle selon laquelle un coefficient de régression est égal à zéro. Plus la valeur t est grande (en termes absolus), plus l’estimation est éloignée de zéro, ce qui suggère que le coefficient est statistiquement significatif ;
- **$\Pr(> |t|)$** : Il s’agit des p-valeurs associées à la statistique de test t . Elles indiquent la probabilité d’obtenir un résultat aussi extrême que celui observé si l’hypothèse nulle est vraie. Une p-valeur faible (généralement $< 0,05$) indique que le coefficient est statistiquement significatif, ce qui permet de rejeter l’hypothèse nulle.

D’autre part, plusieurs lignes sont présentes :

- **L’intercept** représente l’estimation du modèle lorsque toutes les variables explicatives sont égales à zéro ;
- **L’ensemble des variables explicatives du modèle.**

Pour finir, plusieurs indicateurs sont présents :

- **Dispersion parameter for gaussian family taken to be** : Il indique le paramètre de dispersion associé à la famille de distribution choisie (comme Gaussienne, Poisson, etc.). Dans le modèle de Poisson, ce paramètre est généralement supposé être égal à 1 ;
- **Null deviance** : Il s’agit de la déviance d’un modèle avec uniquement une constante. Elle mesure la mauvaise adéquation du modèle nul avec les données ;
- **Residual deviance** : Il s’agit de la déviance du modèle ajusté. Elle évalue la mauvaise adéquation du modèle ajusté aux données ;
- **AIC (Akaike Information Criterion)** : Il mesure la qualité du modèle ajusté en tenant compte à la fois de la déviance et du nombre de paramètres dans le modèle. Des valeurs plus faibles de l’AIC indiquent un meilleur modèle, car l’AIC pénalise la complexité du modèle (c’est-à-dire le nombre de paramètres) tout en récompensant un bon ajustement aux données.

4.3.7 Importance de l’analyse des résidus

L’analyse des résidus a pour objectif de tester la validité d’un modèle de régression. Elle permet d’identifier les éventuelles défaillances du modèle, ce qui en fait une étape essentielle du processus d’évaluation.

Différents résidus existent :

Dans la suite $i = 1, \dots, n$.

– **Les résidus bruts :**

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i \quad (4.47)$$

où Y_i correspondent aux valeurs observées et \hat{Y}_i aux valeurs estimées par le modèle. L'inconvénient de ces résidus est que, même sous l'hypothèse d'homoscédasticité, ils ne présentent pas nécessairement la même variance. Pour rappel, l'homoscédasticité signifie que la variance des erreurs stochastiques de la régression est constante pour chaque observation i .

La variance des résidus brutes est définie tel que $\text{Var}[\hat{\varepsilon}_i] = \sigma^2(1 - h_{ii})$ où σ^2 est la variance des erreurs du modèle et h_{ii} est le coefficient diagonal de la matrice de projecteur H qui est définie par $H = X(X'X)^{-1}X'$ où X est la matrice des variables explicatives.

– **Les résidus standardisés :**

$$\tilde{\varepsilon}_i = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2}} \quad (4.48)$$

où $\hat{\varepsilon}_i$ sont les résidus brutes et le dénominateur $\hat{\sigma}^2$ est l'estimation de la variance des erreurs résiduelles.

L'inconvénient de ces résidus est qu'il n'y a pas forcément d'indépendance entre $\hat{\varepsilon}_i$ et $\hat{\sigma}^2$, donc la loi de $\tilde{\varepsilon}_i$ n'est pas connue.

– **Les résidus studentisés :**

$$\varepsilon_i^* = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}_{(i)}^2(1 - h_{ii})}} \quad (4.49)$$

où $\hat{\sigma}_{(i)}^2$ est l'estimateur de σ^2 dans le modèle privé de l'observation i . Ces résidus sont particulièrement utiles pour détecter des valeurs aberrantes car ils ajustent l'écart-type estimé pour exclure l'effet de l'observation en question. Ils offrent donc une analyse plus précise.

Plusieurs types de graphiques permettent d'analyser les résidus, chacun fournissant des informations différentes : résidus bruts, résidus studentisés, distance de Cook, points leviers et quantiles empiriques des résidus. Les vérifications et les conclusions dépendent des graphiques utilisés.

Le graphique présenté dans la suite représente les résidus studentisés en fonction de l'index des observations. Ce type de graphique permet de détecter d'éventuelles valeurs aberrantes. Pour identifier ces valeurs, il suffit de repérer les points situés au-delà des lignes pointillées, qui correspondent généralement à un seuil autour de ± 2 . Pour que les erreurs suivent une distribution normale, environ 95% des résidus devraient se situer entre ces deux lignes de seuil, ce qui signifie qu'il devrait y avoir moins de 5% de valeurs considérées comme aberrantes.

4.4 Théorie des modèles utilisés dans l'étude

4.4.1 GLM Log-Normal

4.4.1.1 Définition

La distribution du modèle log-normal s'écrit :

$$\ln(Y) = X\beta + \varepsilon \quad (4.50)$$

où :

- $Y = (Y_1, \dots, Y_n)'$ le vecteur colonne représentant les variables réponses. Cette variable est transformée par la fonction logarithme ;
- $X_i = (X_{i,1}, \dots, X_{i,p})'$ le vecteur colonne représentant les variables explicatives pour la variable réponse Y_i ;
- X la matrice de taille $n \times p$ dont les lignes sont des vecteurs lignes X_i' ;
- $\beta = (\beta_1, \dots, \beta_p)'$ le vecteur colonne correspondant aux paramètres du modèle ;
- ε le terme d'erreur.

4.4.1.2 Fonction de lien

La fonction de lien du modèle log-normal est la fonction logarithme et s'écrit :

$$g(\mu(X)) = \ln(\mu) \quad (4.51)$$

où $\mu = \exp(X\beta)$.

4.4.2 GLM Poisson

4.4.2.1 Définition

Le modèle de Poisson modélise la loi de Y par une loi de Poisson de paramètre $\lambda = \lambda(X)$ telle que :

$$\ln(\lambda(X)) = X\beta \quad (4.52)$$

où :

- $Y = (Y_1, \dots, Y_n)'$ le vecteur colonne représentant les variables réponses. Cette variable est transformée par la fonction logarithme ;
- $X_i = (X_{i,1}, \dots, X_{i,p})'$ le vecteur colonne représentant les variables explicatives pour la variable réponse Y_i ;
- X la matrice de taille $n \times p$ dont les lignes sont des vecteurs lignes X_i' ;
- $\beta = (\beta_1, \dots, \beta_p)'$ le vecteur colonne correspondant aux paramètres du modèle.

La moyenne et la variance de Y sont égales à λ , le paramètre de la loi de Poisson.

4.4.2.2 Fonction de lien

La fonction de lien du modèle Poisson est également la fonction logarithme et s'écrit :

$$g(\mu(X)) = \ln(\mu) \quad (4.53)$$

où $\mu = \exp(X\beta)$.

4.5 Présentation des objectifs de modélisation

L'objectif de cette étude est de prédire le coût pour une mutuelle des personnes de 63 et 64 ans qui continueront à travailler deux années supplémentaires.

L'étude visera à élaborer des modèles mathématiques qui visent à calculer les primes pures des personnes qui travailleront deux années supplémentaires. Les modèles choisis sont des Modèles Linéaires Généralisés (GLM). Les étapes de réalisation des modèles seront d'abord définies, puis chaque modèle sera construit pour chaque poste de soins : optique, dentaire, soins de ville, pharmacie, divers, hospitalisation, et aides auditives. Étant donné qu'une grille de tarification détaille chaque poste de soins par sous-poste, l'objectif est d'obtenir un montant pour chaque sous-poste de soins. Enfin, une prédiction sera effectuée pour les personnes de 63 et 64 ans pour chaque sous-poste de soins, permettant de calculer la prime pure, c'est-à-dire le coût de ces personnes pour la mutuelle.

Pour établir une tarification précise, deux variables essentielles sont prises en compte : le coût moyen et la fréquence. En effet, le coût pour une mutuelle d'une personne est déterminé par le produit du coût moyen des prestations et de la fréquence des prestations. Pour refléter fidèlement la réalité, il est crucial de réaliser deux modèles GLM : l'un pour prédire le coût moyen des prestations des personnes de 63 et 64 ans, et l'autre pour prédire la fréquence de ces personnes.

Chaque GLM est élaboré séparément. Ensuite, des prédictions sont effectuées pour les personnes de 63 ans concernant le coût moyen et la fréquence. De manière similaire, des prédictions sont également réalisées pour les personnes de 64 ans.

4.6 Construction des bases de données

Comme mentionné précédemment, un GLM est effectué pour le coût moyen et un autre pour la fréquence, et cela pour chaque poste de soins. Ces deux GLM sont construits de manière différente.

4.6.1 Base de données pour le GLM de coût moyen

Le GLM de coût moyen est élaboré à partir des données de la base de prestations. Il inclut les prestations de toutes les personnes, à l'exception des enfants, des personnes considérées comme retraitées (retraités loi Evin, sortie de groupe et sortie loi Evin), ainsi que des personnes âgées de 63 ans et plus. Par ailleurs, étant donné qu'un GLM de coût moyen est effectué pour chaque poste de soins, la base de données est filtrée en fonction du poste concerné.

4.6.2 Base de données pour le GLM de fréquence

Le GLM de fréquence est constitué des données provenant de la base de personnes. Il inclut toutes les personnes, à l'exception des enfants, des personnes considérées comme retraitées (retraités loi Evin, sortie de groupe et sortie loi Evin), ainsi que des individus âgés de 63 ans et plus. De plus, étant donné qu'un GLM de fréquence est réalisé pour chaque poste de soins, la base de données est filtrée en fonction du poste concerné.

4.6.3 Séparation des bases de données

Chaque base de données mentionnée précédemment est divisée en un jeu d'apprentissage et un jeu de test. Le jeu d'apprentissage représente 80% des données initiales, tandis que le jeu de test en comprend 20%. Ce découpage est une étape cruciale qu'il ne faut pas négliger, car il pourrait entraîner une surévaluation ou, au contraire, une sous-évaluation du modèle. Le jeu d'apprentissage sert à entraîner le modèle et est utilisé par l'algorithme d'apprentissage. Le jeu de test, quant à lui, permet de mesurer l'erreur du modèle final sur des données inédites, optimisant ainsi la prédiction du coût moyen en fonction des sous-postes de soins.

Le modèle GLM est ensuite créé en tenant compte de la loi souhaitée, de la variable cible, des variables explicatives sélectionnées, ainsi que de la base de données d'entraînement.

Enfin, une analyse des résidus est effectuée pour chaque modèle.

4.7 Mise en place des modèles GLM

4.7.1 GLM de coût moyen

4.7.1.1 Sélection des variables

Une étape clé dans la réalisation d'un GLM est la sélection des variables, à savoir la variable cible et les variables explicatives. Pour le GLM de coût moyen, la variable cible correspond au montant remboursé par la mutuelle.

Pour les variables explicatives, l'objectif est d'obtenir un coût moyen par sous-poste de soins, ce qui justifie la sélection de la variable correspondant à ces sous-postes. De plus, l'âge est une variable essentielle, car l'objectif final est de prédire le coût moyen en fonction de l'âge. Il est également crucial d'analyser l'effet de l'âge sur les estimations des montants des prestations, en évaluant l'impact de cette variable sur la performance du modèle. Ainsi, pour chaque poste de soins, une comparaison des performances est effectuée entre le GLM sans l'âge comme variable explicative et le GLM avec l'âge. Dans chaque cas, les mesures de performance, telles que l'AIC et le pseudo- R^2 , sont calculées afin de comparer les deux modèles et évaluer leur efficacité.

Dans le but de construire des modèles simples en se basant uniquement sur les sous-postes de soins et l'âge, les autres variables présentes dans la base de données tels que les régions ou les types de situations des personnes couvertes ne sont prises en compte.

La catégorie socioprofessionnelle n'est pas retenue comme variable explicative, car la base de données est principalement constituée de personnes non-cadres. Ainsi, les résultats obtenus seront essentiellement représentatifs de cette population.

4.7.1.2 Choix du modèle

Les coûts des sinistres des complémentaires santé suivent généralement une distribution non négative et asymétrique à gauche, car ils ont une borne inférieure (coût > 0) tout en présentant une large variabilité dans la partie inférieure. Les distributions log-normale et Gamma sont souvent adaptées pour modéliser ces coûts moyens. De plus, la loi Gamma peut être utilisée avec diverses

fonctions de lien. En conséquence, ces deux distributions ont été testées, et la meilleure est choisie.

Six modèles sont testés : Log-Normal, Gamma avec les fonctions de lien inverse, log et identité, et Gaussien avec les fonctions de lien identité et log. Chaque modèle est évalué et comparé aux autres à l'aide de mesures de performance. Pour simplifier, seules les mesures de performance des deux modèles présentant les résultats et les tests les plus cohérents sont présentées et comparées, et le meilleur des deux est retenu pour l'étude. Les performances des modèles sont comparées à l'aide de deux mesures : l'AIC et le pseudo- R^2 .

Les deux modèles ayant les meilleurs résultats et tests sont le Log-Normal et le Gamma avec la fonction de lien identité. Il reste à comparer les mesures de performance de chacun de ces modèles et à conserver celui qui s'avère le plus satisfaisant.

Les deux types de modèles sont réalisés pour chaque poste de soins. Par la suite, le modèle utilisant la loi log-normale sera désigné par « GLM log-normale », tandis que le modèle utilisant la loi Gamma avec la fonction de lien identité sera appelé « GLM Gamma ».

AIC

	Optique	Dentaire	Soins de ville	Pharmacie	Divers	Hospitalisation	Aides auditives
GLM Log-Normale AVEC AGE	63 521	128 326	499 086	808 006	78 228	63 425	1 010
GLM Gamma identité AVEC AGE	101 113	126 052	788 156	704 246	132 973	58 308	1 519

TABLEAU 4.1 – Comparaison des AIC des modèles GLM Log-Normale et Gamma identité

Pour rappel, parmi deux modèles utilisant des lois différentes mais avec les mêmes variables cibles et explicatives, le modèle ayant l'AIC le plus faible est préféré.

Dans le tableau ci-dessus, les valeurs de l'AIC pour les modèles GLM log-normale des postes de soins optiques, soins de ville, divers et aides auditives sont inférieures à celles des modèles GLM Gamma. L'écart le plus notable se trouve pour le poste de soins de ville, avec une différence d'environ 289 000. Pour les autres postes de soins, les AIC sont plus faibles pour le modèle GLM Gamma. Par conséquent, étant donné qu'un plus grand nombre de postes de soins présente des AIC inférieurs pour le modèle GLM log-normal et que l'écart est significatif pour le poste de soins de ville, la loi log-normale apparaît comme le modèle le plus approprié.

Pseudo- R^2

	Optique	Dentaire	Soins de ville	Pharmacie	Divers	Hospitalisation	Aides auditives
GLM Log-Normale AVEC AGE	36,47%	75,48%	14,77%	15,42%	58,06%	31,88%	95,93%
GLM Gamma identité AVEC AGE	44,38%	81,64%	18,05%	7,84%	47,11%	31,57%	81,26%

TABLEAU 4.2 – Comparaison des pseudo- R^2 des modèles GLM Log-Normale et Gamma identité

Pour rappel, parmi deux modèles utilisant des lois différentes mais avec les mêmes variables cibles et explicatives, le modèle ayant le pseudo- R^2 le plus élevé est préféré.

Pour les postes optique, dentaire et les soins de ville, le pseudo- R^2 est supérieur pour le GLM

Gamma, avec une différence d'environ 8% pour l'optique, 6% pour le dentaire et 3% pour les soins de ville. En revanche, pour les postes pharmacie, divers, hospitalisation et aides auditives, le pseudo- R^2 est supérieur pour le GLM log-normale, avec une différence d'environ 8% pour la pharmacie, 9% pour le divers, 0,31% pour l'hospitalisation et 15% pour les aides auditives. Tous les écarts sont relativement faibles, à l'exception du poste des aides auditives, où l'écart dépasse 10%. Étant donné que le modèle log-normal présente un pseudo- R^2 supérieur pour un plus grand nombre de postes, il est considéré comme plus adapté. Ainsi, d'après cet indicateur, il est préférable de conserver le modèle log-normale.

Conclusion

En résumé, les AIC des modèles Gamma sont très élevés par rapport à ceux du modèle log-normale. Les écarts entre les résultats des deux modèles pour les pseudo- R^2 sont faibles, mais ils sont globalement en faveur du modèle log-normale. Il est donc préférable de conserver la loi log-normale pour estimer le coût moyen.

4.7.1.3 Pertinence de l'âge

L'effet de l'âge est un facteur crucial qui doit être étudié pour évaluer son impact sur le modèle GLM. Ainsi, l'AIC et le pseudo- R^2 sont calculés pour le modèle Log-Normal dans deux configurations : d'une part, avec une seule variable explicative, les sous-postes de soins, et d'autre part, avec les sous-postes de soins et l'âge.

AIC

	Optique	Dentaire	Soins de ville	Pharmacie	Divers	Hospitalisation	Aides auditives
GLM Log-Normale SANS AGE	63 524	128 332	499 176	808 020	78 233	63 435	1 007
GLM Log-Normale AVEC AGE	63 521	128 326	499 086	808 006	78 228	63 425	1 010

TABLEAU 4.3 – Comparaison des AIC des modèles GLM Log-Normale et Gamma identité

Pour rappel, parmi deux modèles ayant la même loi, la même variable cible, mais des variables explicatives différentes, le modèle avec l'AIC le plus faible est préféré.

L'écart entre les AIC des modèles avec et sans la variable âge est très faible. L'ajout de la variable âge entraîne une diminution de l'AIC pour tous les postes de soins, à l'exception des aides auditives. En effet, pour ce poste de soins, l'AIC augmente très légèrement avec l'ajout de la variable âge. Étant donné que l'AIC diminue pour presque tous les postes de soins avec l'ajout de la variable âge, et que l'augmentation de celui-ci pour les aides auditives est très faible, cet indicateur suggère que l'âge a un effet positif sur le modèle.

Pseudo- R^2

	Optique	Dentaire	Soins de ville	Pharmacie	Divers	Hospitalisation	Aides auditives
GLM Log-Normale SANS AGE	36,46%	75,47%	14,74%	15,32%	57,99%	31,09%	95,93%
GLM Log-Normale AVEC AGE	36,47%	75,48%	14,77%	15,42%	58,06%	31,88%	95,93%

TABLEAU 4.4 – Comparaison des pseudo- R^2 des modèles GLM Log-Normale et Gamma identité

Pour rappel, parmi deux modèles ayant la même loi, la même variable cible, mais des variables explicatives différentes, le modèle avec le pseudo- R^2 le plus élevé est préféré.

Le pseudo- R^2 reste constant ou augmente légèrement pour tous les postes de soins, avec des variations allant de 0,01% à 0,78% lorsque l'âge est inclut comme variable explicative. Ces augmentations étant très faibles, la différence entre les deux modèles est négligeable, ce qui est cohérent.

Conclusion

D'après ces deux mesures de performance, l'âge a un impact positif sur la performance des modèles. Cela confirme que l'âge influence les coûts moyens des prestations.

Le modèle sélectionné pour estimer le coût moyen est le GLM log-normale, avec le montant du remboursement de la mutuelle comme variable cible et les sous-postes de soins ainsi que l'âge des personnes comme variables explicatives.

4.7.2 GLM de fréquence

4.7.2.1 Sélection des variables

Pour le GLM de fréquence, la variable cible est le nombre de prestations reçues par chaque personne.

Pour les variables explicatives, l'objectif est d'obtenir la fréquence par sous-poste de soins, ce qui justifie la sélection de la variable correspondant à ces sous-postes. De plus, l'âge est une variable essentielle pour prédire la fréquence en fonction de cet âge. Comme pour le GLM de coût moyen, il est crucial d'examiner l'effet de l'âge sur les estimations du nombre de prestations afin d'évaluer son impact sur la performance du modèle. Ainsi, pour chaque poste de soins, une comparaison des performances est effectuée entre le GLM sans l'âge comme variable explicative et le GLM avec l'âge.

Les mesures de performance, telles que l'AIC et le pseudo- R^2 , sont calculées pour comparer les deux modèles et évaluer leur efficacité. En outre, chaque GLM prend en compte l'exposition des personnes pendant la durée du contrat. Cette variable est importante dans les GLM de fréquence, car la fréquence des prestations varie en fonction de la durée de l'exposition : une personne assurée pendant un an n'a pas la même fréquence qu'une personne assurée pendant un mois. Cette variable est donc intégrée comme offset dans les modèles.

4.7.2.2 Choix du modèle

Le GLM utilisé pour la fréquence est un GLM Poisson. En effet, la fréquence est une variable de comptage, et le GLM Poisson est particulièrement adapté à ce type de variable. La fonction de lien la plus couramment employée est la fonction de lien logarithmique, qui permet de modéliser une relation multiplicative entre les variables explicatives et le nombre de prestations. Cette fonction sera donc utilisée ici.

Bien que la sélection de la loi du GLM pour estimer la fréquence soit plus simple que pour le coût moyen, il est également important d'étudier la performance des modèles en fonction de l'âge. Ainsi, comme pour les GLM de coût moyen, la performance des deux GLM Poisson est compa-

rée : l'un utilise uniquement les sous-postes de soins comme variable explicative, tandis que l'autre inclut à la fois les sous-postes de soins et l'âge.

4.7.2.3 Pertinence de l'âge

AIC

	Optique	Dentaire	Soins de ville	Pharmacie	Divers	Hospitalisation	Aides auditives
GLM Poisson SANS AGE	54 077	76 301	435 665	466 632	77 780	358 756	1 790
GLM Poisson AVEC AGE	54 078	76 195	435 603	459 426	77 781	358 643	1 677

TABLEAU 4.5 – Comparaison des AIC des modèles GLM Poisson avec et sans la variable explicative AGE

Pour rappel, parmi deux modèles ayant la même loi, la même variable cible, mais des variables explicatives différentes, le modèle avec l'AIC le plus faible est préféré.

L'ajout de la variable âge entraîne une diminution des AIC pour les postes de soins dentaire, soins de ville, pharmacie, hospitalisation et aides auditives. Cette amélioration observée dans la majorité des postes de soins indique que l'âge contribue positivement à la performance du modèle. Ainsi, selon l'AIC, l'inclusion de l'âge améliore le modèle.

Pseudo- R^2

	Optique	Dentaire	Soins de ville	Pharmacie	Divers	Hospitalisation	Aides auditives
GLM Poisson SANS AGE	23,68%	43,72%	21,24%	51,02%	29,57%	19,65%	2,03%
GLM Poisson AVEC AGE	23,68%	43,82%	21,26%	51,88%	29,58%	19,65%	8,64%

TABLEAU 4.6 – Comparaison des pseudo- R^2 des modèles GLM Poisson avec et sans la variable explicative AGE

Pour rappel, parmi deux modèles ayant la même loi, la même variable cible, mais des variables explicatives différentes, le modèle avec le pseudo- R^2 le plus élevé est préféré.

L'ajout de la variable âge entraîne une stabilité ou une augmentation des pseudo- R^2 pour tous les postes de soins. En particulier, l'âge a un effet notable sur le poste des aides auditives, avec une augmentation du pseudo- R^2 de plus de 6%. L'âge améliore ou maintient la performance du modèle pour tous les postes de soins. Ainsi, selon le pseudo- R^2 , l'âge a un impact positif sur le modèle.

Conclusion

D'après ces deux mesures de performance, l'âge a un impact positif sur les modèles. Cela confirme que l'âge influence la fréquence des prestations.

Le modèle choisi pour estimer la fréquence est le GLM Poisson, avec le nombre de prestations comme variable cible et les sous-postes de soins ainsi que l'âge des personnes comme variables explicatives. L'exposition est intégrée dans l'offset de ce modèle.

4.8 Diagnostic des modèles

Cette section est dédiée à l'analyse des p-values des modèles de coût moyen et de fréquence. Elle comprend également une étude des résidus.

4.8.1 Analyse des p-values

Les p-values permettent d'évaluer l'influence des variables sur la variable cible. Plus la p-value est faible, plus la variable est significative pour la variable cible. Les variables ayant des p-values supérieures à 0,05 ne sont pas considérées comme significatives. Les p-values des modèles GLM de coût moyen et de fréquence sont présentées pour chaque poste de soins.

4.8.1.1 Optique

		P-values	
		Coût moyen	Fréquence
Optique	Sous-postes de soins		
	Intercept	≈ 0	≈ 0
	AGE	0,638	0,596
	Lentilles acceptées	≈ 0	0,970
	Lentilles refusées	≈ 0	≈ 0
	Monture 100% Santé (classe A)	≈ 0	0,181
	Monture hors 100% Santé (classe B)	≈ 0	≈ 0
	Verres 100% Santé (classe A)	≈ 0	0,002
	Verres complexes	≈ 0	≈ 0
	Verres simples	≈ 0	≈ 0
	Verres très complexes	≈ 0	0,395

TABLEAU 4.7 – P-values du GLM de coût moyen et de fréquence pour le poste optique

Dans le GLM de coût moyen pour l'optique, toutes les variables sont significatives sauf l'âge. Ce résultat n'est pas surprenant, car bien que la vue puisse se détériorer avec l'âge, de nombreuses personnes jeunes utilisent déjà des lunettes avec des corrections élevées, entraînant des prestations coûteuses.

Dans le GLM de fréquence pour l'optique, presque toutes les variables sont significatives, à l'exception de l'âge, des lentilles acceptées, des montures 100% santé (classe A) et des verres très complexes. Le manque de significativité de l'âge s'explique de manière similaire à celle observée pour le coût moyen : de nombreux jeunes portent des lunettes dont la correction évolue, entraînant une consommation fréquente.

Pour les sous-postes de soins moins significatifs, il y a eu très peu de prestations dans la base de données, ce qui limite la capacité à détecter un effet significatif. Par exemple, il y a eu seulement 22 prestations pour les lentilles acceptées, 35 pour les montures 100% santé (classe A), et 47 pour les verres complexes, tandis que 983 personnes ont bénéficié de prestations pour les lentilles refusées.

4.8.1.2 Dentaire

	Sous-postes de soins	P-values	
		Coût moyen	Fréquence
Dentaire	Intercept	≈ 0	≈ 0
	AGE	0,007	≈ 0
	Inlay Core Panier libre	≈ 0	0,017
	Inlay Core Panier modéré	≈ 0	≈ 0
	Inlay Core RAC 0	≈ 0	0,008
	Inlay Onlay Panier libre	0,007	0,006
	Inlay Onlay Panier modéré	0,013	≈ 0
	Orthodontie acceptée	≈ 0	≈ 0
	Orthodontie refusée	0,011	≈ 0
	Parodontologie	≈ 0	≈ 0
	Prothèses dentaires Panier libre	≈ 0	≈ 0
	Prothèses dentaires Panier modéré	≈ 0	≈ 0
	Prothèses dentaires RAC 0	≈ 0	≈ 0
	Prothèses dentaires Refusées	0,050	≈ 0
	Soins dentaires	≈ 0	≈ 0

TABLEAU 4.8 – P-values du GLM de coût moyen et de fréquence pour le poste dentaire

Dans le GLM de coût moyen pour le dentaire, toutes les variables sont significatives, ce qui indique qu'elles influencent le montant du remboursement de la mutuelle. L'âge est également significatif, ce qui est logique, car il est bien connu que les personnes ont tendance à consommer des prestations dentaires plus coûteuses, telles que les prothèses, à mesure qu'elles vieillissent.

Dans le GLM de fréquence pour le dentaire, toutes les variables sont également significatives. Cela signifie que chacune d'elles a un impact sur le nombre de prestations reçues par les personnes.

4.8.1.3 Soins de ville

	Sous-postes de soins	P-values	
		Coût moyen	Fréquence
Soins de ville	Intercept	≈ 0	≈ 0
	AGE	≈ 0	≈ 0
	Analyses, laboratoire	≈ 0	≈ 0
	Auxiliaires médicaux	≈ 0	≈ 0
	Consultations, visites Généraliste	≈ 0	≈ 0
	Consultations, visites Spécialiste	≈ 0	≈ 0
	Densitométrie osseuse	0,001	≈ 0
	Radiologie (DPTAM et non DPTAM)	≈ 0	≈ 0
	Soins à l'étranger	≈ 0	≈ 0

TABLEAU 4.9 – P-values du GLM de coût moyen et de fréquence pour le poste soins de ville

Dans les modèles GLM de coût moyen et de fréquence pour les soins de ville, toutes les variables sont significatives. Cela signifie que chaque variable a un effet réel sur le montant du remboursement de la mutuelle ainsi que sur le nombre de prestations reçues par les personnes.

4.8.1.4 Pharmacie

		P-values	
		Coût moyen	Fréquence
Pharmacie	Sous-postes de soins		
	Intercept	≈ 0	≈ 0
	AGE	≈ 0	≈ 0
	Médicaments SMR majeur	≈ 0	≈ 0
	Médicaments SMR modéré	≈ 0	≈ 0
	Vaccins Anti grippe	≈ 0	≈ 0
Vaccins et Pharmacie NR sur prescription médicale		≈ 0	≈ 0

TABLEAU 4.10 – P-values du GLM de coût moyen et de fréquence pour le poste pharmacie

Dans les modèles GLM de coût moyen et de fréquence pour la pharmacie, toutes les variables sont significatives. Cela indique que chaque variable exerce un effet réel sur le montant du remboursement de la mutuelle et sur le nombre de prestations reçues par les personnes.

4.8.1.5 Divers

		P-values	
		Coût moyen	Fréquence
Divers	Sous-postes de soins		
	Intercept	≈ 0	≈ 0
	AGE	≈ 0	0,520
	Contraception NR	≈ 0	≈ 0
	Cures thermales (forfait et honoraires)	≈ 0	≈ 0
	Frais de transport	≈ 0	≈ 0
	Grand appareillage	≈ 0	≈ 0
	Indemnités de maternité	≈ 0	≈ 0
	Médecine douce	≈ 0	≈ 0
Petit appareillage	≈ 0	≈ 0	

TABLEAU 4.11 – P-values du GLM de coût moyen et de fréquence pour le poste divers

Dans les modèles GLM de coût moyen et de fréquence pour le poste divers, toutes les variables sont significatives, à l'exception de l'âge dans le GLM de fréquence. Ce résultat n'est pas surprenant, car les consommations pour ce type de soins peuvent varier considérablement entre les individus, indépendamment de leur âge. D'autres facteurs, tels que l'éducation et le style de vie, peuvent avoir un impact plus significatif.

4.8.1.6 Hospitalisation

		P-values	
		Coût moyen	Fréquence
Hospitalisation	Sous-postes de soins		
	Intercept	≈ 0	≈ 0
	AGE	≈ 0	≈ 0
	Chambre particulière par jour	0,185	≈ 0
	Forfait hospitalier	0,006	≈ 0
	Frais séjour (Conventionné et Non Conv)	≈ 0	≈ 0
Honoraires et autres Frais		≈ 0	≈ 0

TABLEAU 4.12 – P-values du GLM de coût moyen et de fréquence pour le poste hospitalisation

Dans le GLM de coût moyen pour l’hospitalisation, une variable n’est pas significative : les chambres particulières par jour. Ce manque de significativité est compréhensible, car les montants remboursés pour ce sous-poste varient considérablement de 15 à 5700 €. Cette grande variation peut biaiser les tests de significativité des variables. En comparaison, la plage des montants pour les forfaits hospitaliers est beaucoup plus restreinte, allant de 9 à 1280 €.

Dans le GLM de fréquence, toutes les variables sont significatives pour le nombre de prestations reçues par les personnes.

4.8.1.7 Aides auditives

		P-values	
		Coût moyen	Fréquence
Aides auditives	Sous-postes de soins		
	Intercept	0,755	≈ 0
	AGE	0,669	≈ 0
	Prothèses auditives - Classe 1 - 100% Santé	≈ 0	≈ 0
Prothèses auditives - Classe 2 - Tarifs libres	≈ 0	0,003	

TABLEAU 4.13 – P-values du GLM de coût moyen et de fréquence pour le poste aides auditives

Dans le GLM de coût moyen pour les aides auditives, la moitié des variables ne sont pas significatives, notamment l’intercept et l’âge. Les montants associés aux consommables et accessoires, inclus dans l’intercept, sont très faibles par rapport aux autres montants de la base de données, variant entre 0,6 et 42 €, tandis que les montants pour les prothèses vont de 582,76 à 1600 €. Cette différence peut diluer l’effet global de la variable, car les faibles montants ne fournissent pas suffisamment de variation pour détecter statistiquement un effet significatif. De plus, l’âge n’est pas significatif, ce qui est attendu, car les prestations d’aides auditives restent coûteuses indépendamment de l’âge.

Dans le GLM de fréquence, toutes les variables sont significatives pour le nombre de prestations reçues par les personnes.

Pour conclure sur les p-values, la majorité des variables sont significatives pour les différents postes de soins, à l’exception de certaines dont la non-significativité est justifiée par des facteurs spécifiques, tels que des variations de montants ou des caractéristiques particulières des prestations.

4.8.2 Analyse des résidus

Cette section est dédiée à l’analyse des résidus pour tous les modèles de coûts moyens et de fréquence. Pour cette analyse, des graphiques spécifiques sont utilisés pour représenter les résidus. Le graphique réalisé présente les résidus studentisés en fonction de l’index des observations. L’index représente la position des observations dans la base de données, permettant ainsi de visualiser l’ordre des données.

Ce type de graphique est particulièrement utile pour détecter les valeurs aberrantes. Il est essentiel que 95% des résidus studentisés se situent entre les seuils de -2 et 2 pour que les résidus suivent une distribution normale et qu’il y ait peu de valeurs aberrantes.

Seuls les graphiques des résidus pour les modèles de coût moyen et de fréquence du poste optique seront présentés dans cette section. Les autres graphiques de résidus pour chaque poste de soins sont disponibles en annexe [B].

4.8.2.1 Résidus studentisés du poste de soins Optique

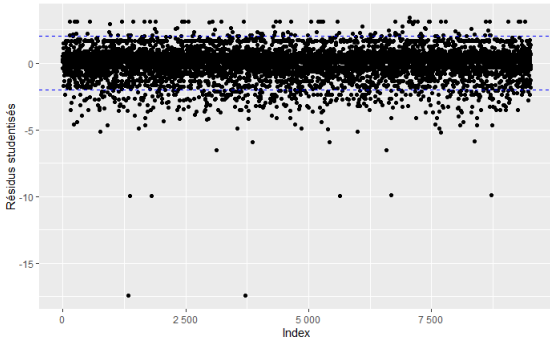


FIGURE 4.1 – Résidus studentisés du modèle de coût moyen pour l’optique

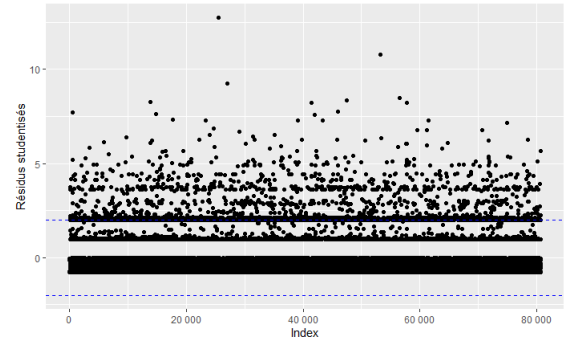


FIGURE 4.2 – Résidus studentisés du modèle de fréquence pour l’optique

Environ 95% des résidus studentisés du poste de soins optique se situent entre les deux seuils, aussi bien pour le modèle de coût moyen que pour le modèle de fréquence. Cela indique que les erreurs suivent une distribution normale.

4.8.2.2 Résultat des proportions de résidus studentisés pour tous les postes de soins

Postes	Proportions de résidus entre -2 et 2	
	Coût moyen	Fréquence
Optique	94,55%	96,03%
Dentaire	96,15%	98,10%
Soins de ville	95,99%	70,84%
Pharmacie	96,28%	68,77%
Divers	93,42%	97,62%
Hospitalisation	95,75%	84,21%
Aides auditives	97,35%	99,88%

TABEAU 4.14 – Proportions de résidus studentisés pour les modèles de coût moyen et de fréquence pour chaque poste de soins

Pour la plupart des modèles, la majorité des résidus studentisés se situent entre les deux lignes bleues, notamment pour les modèles de coût moyen des postes de dentaire, de soins de ville, de pharmacie, d’hospitalisation, et d’aides auditives, ainsi que pour les modèles de fréquence des postes d’optique, de dentaire, de divers et d’aides auditives. La limite des 95% des résidus situés en dehors des seuils est respectée, et d’après cette analyse, les erreurs suivent une distribution normale.

Pour les modèles de coût moyen des postes d’optique et divers, la proportion de résidus compris entre les seuils est presque égale à 95%. La limite de 95% des résidus situés en dehors des

seuils est donc presque respectée. Les erreurs suivent globalement une distribution normale.

En revanche, les modèles de fréquence des postes de soins de ville, de pharmacie et d'hospitalisation présentent des proportions de résidus compris entre les deux lignes plus faibles, respectivement de 70,84% pour les soins de ville, 68,77% pour la pharmacie et 84,21% pour l'hospitalisation. Ces pourcentages sont inférieurs d'environ 25% pour la pharmacie et les soins de ville au seuil généralement considéré comme acceptable et de 10% environ pour l'hospitalisation. Les résidus ne semblent pas suivre une loi normale, ce qui indique la présence probable d'un nombre significatif de valeurs aberrantes. En outre, ces résultats ne sont pas réellement surprenants, étant donné qu'il s'agit d'un modèle de fréquence qui représente un nombre de sinistres. Ces observations suggèrent que les erreurs de prédiction pour ces postes de soins peuvent être plus importantes que pour les autres. Ce biais potentiel sera pris en compte dans l'interprétation des résultats à venir.

4.9 Prédictions

Cette section présente les résultats des modèles développés précédemment. Pour rappel, les GLM ont été élaborés en fonction de l'âge et des sous-postes de soins, en se basant sur une population de personnes actives âgées de 62 ans ou moins.

4.9.1 Résultats des prédictions pour des actifs de 50 ans

Dans un premier temps, il est essentiel de vérifier si les modèles établis prédisent correctement les données. Pour cela, une prédiction est réalisée pour des individus actifs âgés de 50 ans, puis comparée aux montants observés dans la base de données initiale. Concrètement, chaque montant prédit par poste de soins est comparé au produit du coût moyen et de la fréquence des prestations extraits de la base de données pour ce même poste. Cette comparaison est effectuée à la fois sur le coût total pour la mutuelle, tous postes confondus, ainsi que sur chaque poste de soins de manière individuelle. Tous les montants sont exprimés en euros.

	Coût X Fréquence		Écarts entre le montant observé et le montant prédit
	Montants observés	Montants prédits	
Optique	175,46	149,16	-14,99%
Dentaire	127,45	111,35	-12,63%
Soins de ville	151,55	143,32	-5,43%
Pharmacie	59,88	30,13	-49,69%
Divers	58,30	69,80	19,72%
Hospitalisation	87,60	64,51	-26,35%
Aides auditives	15,26	7,88	-48,36%
Total par an	675,50	576,15	-14,71%
Total par mois	56,29	48,01	-14,71%

TABLEAU 4.15 – Comparaison des montants observés et des montants prédits pour un actif de 50 ans

Dans le tableau ci-dessus, les montants prédits sont inférieurs aux montants observés pour tous les postes de soins, à l'exception du poste divers. Les écarts les plus importants concernent les postes

pharmacie et aides auditives, avec des différences respectives de 49,69% et 48,36%.

L'écart mensuel est de 8,28 € en faveur des montants observés. Cet écart, jugé acceptable, permet de conclure que les modèles élaborés prédisent globalement bien les données. Il est toutefois important de noter que, les montants prédits étant légèrement inférieurs aux montants observés, il est possible que les résultats des prédictions pour les personnes de 63 et 64 ans soient également légèrement en deçà des attentes.

4.9.2 Résultats des prédictions pour des actifs de 63 et 64 ans

L'objectif de cette étude est de prédire le coût moyen et la fréquence des prestations pour les personnes âgées de 63 et 64 ans. Chaque prédiction est effectuée séparément pour chaque sous-poste de soins, en tenant compte de l'âge spécifique, soit 63 ans d'un côté et 64 ans de l'autre.

Pour la prédiction de la fréquence, l'exposition est intégrée dans le GLM, donc elle doit également être considérée lors des prédictions. Ainsi, pour chaque estimation de fréquence, une exposition de 1 an est choisie, ce qui signifie que la personne pour laquelle la fréquence est prédite a été couverte par le contrat pendant un an.

Les résultats finaux de la tarification sont obtenus en multipliant le coût moyen de chaque sous-poste de soins par la fréquence associée. Ces résultats seront détaillés dans le chapitre 5.

Tous les coûts moyens sont exprimés en euros, tandis que les fréquences représentent le nombre de fois qu'une personne a consommé une prestation. Tous les résultats sont annualisés.

4.9.2.1 Optique

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Optique	Chirurgie réfractive de l'œil (par œil)	418,62	0,0059	418,73	0,0059
	Lentilles acceptées	66,04	0,0059	66,06	0,0059
	Lentilles refusées	111,47	0,2718	111,50	0,2719
	Monture 100% Santé (classe A)	23,02	0,0043	23,02	0,0043
	Monture hors 100% Santé (classe B)	94,63	0,3087	94,66	0,3089
	Verres 100% Santé (classe A)	37,95	0,0105	37,96	0,0105
	Verres complexes	192,60	0,3149	192,65	0,3151
	Verres simples	82,34	0,3098	82,36	0,3100
	Verres très complexes	342,90	0,0070	392,97	0,0120

TABLEAU 4.16 – Résultats des prédictions de coût moyen et de fréquence de l'optique pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes d'optique présentent une légère augmentation entre 63 et 64 ans. Ces augmentations varient de 0,005 à 50,07€, 50,07€ pour les verres très complexes, qui connaissent la plus forte augmentation. Cela peut s'expliquer par le fait que ces verres sont généralement coûteux et que leur prix a tendance à augmenter avec l'âge. Cela s'explique par le fait que la vision se détériore fréquemment en vieillissant, en raison de la presbytie ou d'autres problèmes de santé oculaire. Seul le coût moyen des montures 100% santé classe A reste stable.

De même, les fréquences augmentent ou restent stables entre ces deux âges, avec des variations comprises entre 0,0001 et 0,005 en fonction des sous-postes de soins.

Ces augmentations, tant pour le coût moyen que pour la fréquence, sont cohérentes avec les attentes, car il est bien établi que les coûts et/ou les fréquences tendent à augmenter avec l'âge.

4.9.2.2 Dentaire

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Dentaire	Implants dentaires	316,01	0,0406	316,52	0,0409
	Inlay Core Panier libre	224,47	0,0559	224,83	0,0562
	Inlay Core Panier modéré	109,36	0,0409	109,54	0,0411
	Inlay Core RAC 0	113,46	0,0446	113,64	0,0449
	Inlay Onlay Panier libre	308,75	0,0614	309,25	0,0617
	Inlay Onlay Panier modéré	260,45	0,0531	260,86	0,0533
	Orthodontie acceptée	70,49	0,0012	70,60	0,0012
	Orthodontie refusée	387,75	0,0217	388,38	0,0219
	Parodontologie	53,42	0,0561	53,51	0,0566
	Prothèses dentaires Panier libre	312,49	0,1765	312,83	0,1776
	Prothèses dentaires Panier modéré	274,82	0,1044	275,10	0,1049
	Prothèses dentaires RAC 0	156,92	0,1355	157,18	0,1363
	Prothèses dentaires Refusées	209,05	0,0024	209,38	0,0024
	Soins dentaires	8,03	2,3169	8,05	2,3316

TABLEAU 4.17 – Résultats des prédictions de coût moyen et de fréquence du dentaire pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes dentaires augmentent légèrement entre 63 et 64 ans, avec des hausses variant de 0,02 € à 0,63 €.

De même, les fréquences d'utilisation du dentaires augmentent ou restent stables entre ces deux âges, avec des variations allant de 0,002 à 0,0147, 0,0147 pour les soins dentaires. Cette dernière représente la plus forte augmentation de fréquence parmi les sous-postes dentaires.

Ces tendances sont cohérentes avec les attentes, car les coûts et la fréquence du poste dentaire tendent à augmenter avec l'âge.

4.9.2.3 Soins de ville

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Soins de ville	Actes techniques Médicaux	12,23	1,4256	12,25	1,4287
	Analyses, laboratoire	4,93	2,5520	4,94	2,5575
	Auxiliaires médicaux	5,42	5,5722	5,43	5,5844
	Consultations, visites Généraliste	8,42	2,7234	8,44	3,7294
	Consultations, visites Spécialiste	15,95	3,2139	15,98	3,2209
	Radiologie (DPTAM et non DPTAM)	10,20	1,6019	10,22	1,6053
	Soins à l'étranger	34,86	0,0084	34,93	0,0084
	Densitométrie osseuse	87,09	0,0004	87,26	0,0004

TABLEAU 4.18 – Résultats des prédictions de coût moyen et de fréquence des soins de ville pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes des soins de ville augmentent légèrement entre 63 et 64 ans, avec des hausses variant de 0,01 € à 0,17 €.

Les fréquences d'utilisation montrent également une tendance à augmenter ou rester stable entre ces âges, avec des variations allant de 0,0031 à 1,006, 1,006 pour les consultations de médecine générale, représentant la plus forte augmentation de fréquence parmi les sous-postes de soins. Il est important de noter que les résidus du modèle de fréquence pour les soins de ville ne suivent pas une loi normale, ce qui suggère que ces résultats doivent être interprétés avec prudence.

Pour évaluer la fiabilité des prévisions, une comparaison avec un logiciel de tarification sera effectuée dans la suite de l'étude.

4.9.2.4 Pharmacie

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Pharmacie	Médicaments SMR faible	2,06	2,4480	2,07	2,4910
	Médicaments SMR majeur	0,67	32,4715	0,67	33,0428
	Médicaments SMR modéré	1,63	6,4034	1,64	6,5161
	Vaccins Anti grippe	14,19	0,0069	14,24	0,0070
	Vaccins et Pharmacie NR sur prescription médicale	17,00	0,1338	17,06	0,1361

TABLEAU 4.19 – Résultats des prédictions de coût moyen et de fréquence de la pharmacie pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes de pharmacie augmentent légèrement entre 63 et 64 ans, avec des hausses comprises entre 0,01 € et 0,06 €. Cependant, le coût moyen des médicaments SMR majeur reste stable.

En parallèle, les fréquences d'utilisation augmentent entre ces deux âges, avec des variations allant de 0,0001 à 0,5713, 0,5713 pour les médicaments SMR majeur. Ce constat montre que, bien que le coût moyen pour ce sous-poste ne change pas, la fréquence d'utilisation des médicaments augmente avec l'âge. Il est important de noter que les résidus du modèle de fréquence pour la pharmacie ne suivent pas une loi normale. Ainsi, ces résultats doivent être interprétés avec prudence.

4.9.2.5 Divers

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Divers	Accessoires	7,55	0,7411	7,57	0,7415
	Contraception NR	27,18	0,0064	27,27	0,0064
	Cures thermales (forfait et honoraires)	83,18	0,0072	83,44	0,0072
	Frais de transport	43,02	0,0379	43,15	0,0379
	Médecine douce	55,07	0,8519	55,24	0,8524
	Petit appareillage	23,50	0,3249	23,57	0,3250
	Grand appareillage	280,56	0,0011	281,45	0,0011
	Indemnités de maternité	356,94	0,0290	358,08	0,0291

TABLEAU 4.20 – Résultats des prédictions de coût moyen et de fréquence du divers pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes divers augmentent légèrement entre 63 et 64 ans, avec des hausses allant de 0,02 € à 1,13 €, 1,13€ pour les indemnités de maternité, qui représentent l'augmentation la plus importante.

De même, les fréquences augmentent ou restent stables entre ces deux âges, avec des variations comprises entre 0,0002 et 0,0004.

4.9.2.6 Hospitalisation

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Hospitalisation	Chambre particulière en ambulatoire	48,69	0,0792	49,27	0,0792
	Chambre particulière par jour	56,02	0,5818	56,69	0,5830
	Forfait hospitalier	36,81	0,9945	37,25	0,9949
	Frais séjour (Conventionné et Non Conv.)	5,29	4,9362	5,35	4,9456
	Honoraires et autres Frais	24,33	0,4248	24,62	0,4256

TABLEAU 4.21 – Résultats des prédictions de coût moyen et de fréquence de l'hospitalisation pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes d'hospitalisation augmentent légèrement entre 63 et 64 ans. Les augmentations varient de 0,29 € à 0,67 €, 0,67€ pour les chambres particulières par jour.

De même, les fréquences augmentent entre ces deux âges, avec des hausses allant de 0,0001 à 0,0025, 0,0094 pour les frais de séjour. Ce sous-poste enregistre les plus fortes augmentations en termes de fréquence. Ce résultat est attendu, car les personnes sont généralement plus susceptibles d'être hospitalisées à mesure qu'elles vieillissent et donc d'avoir des frais de séjour.

4.9.2.7 Aides auditives

Postes	Sous-postes de soins	63 ans		64 ans	
		Coûts moyens	Fréquences	Coûts moyens	Fréquences
Aides auditives	Consommables et accessoires pris en charge	1,58	0,0463	1,58	0,0518
	Prothèses auditives - Classe 1 - 100% Santé	766,61	0,0010	770,02	0,0012
	Prothèses auditives - Classe 2 - Tarifs libres	1 118,85	0,0150	1 123,83	0,0160

TABLEAU 4.22 – Résultats des prédictions de coût moyen et de fréquence des aides auditives pour les personnes de 63 et 64 ans

Tous les coûts moyens des sous-postes d'aides auditives augmentent légèrement entre 63 et 64 ans. Les augmentations varient de 0,01 € pour les accessoires à près de 5 € pour les prothèses auditives de classe 2 avec tarifs libres. Ce phénomène est cohérent avec le fait que les prothèses auditives sont coûteuses et que, avec l'âge, la dégradation de l'audition peut entraîner des coûts plus élevés pour ces prothèses.

Les fréquences augmentent également entre ces deux âges, avec des hausses allant de 0,0001 à 0,0055, 0,0055 pour les accessoires. Bien que le coût moyen des accessoires augmente très légèrement, la fréquence de leur utilisation connaît la plus forte augmentation. Cela est explicable par le fait que les accessoires, tels que les piles pour prothèses auditives, ont un coût relativement fixe,

indépendamment de la complexité des prothèses utilisées.

Pour conclure cette section, les coûts moyens et les fréquences des sous-postes de soins augmentent légèrement ou restent stables entre 63 et 64 ans. Cette tendance est attendue, car il est généralement observé qu'avec l'âge, les coûts moyens et le nombre de prestations augmentent.

Chapitre 5

Application à la tarification

Ce chapitre est dédié à la tarification. Les modèles développés précédemment ont permis de prédire les coûts moyens et les fréquences pour les personnes de 63 et 64 ans. Il est désormais pertinent de présenter le montant de la prime pure, obtenu en multipliant le coût moyen par la fréquence. Ces résultats seront ensuite comparés à ceux fournis par un logiciel de tarification marché afin de vérifier leur cohérence.

Enfin, l'impact sur la mutuelle étudiée sera évalué pour répondre à la problématique du mémoire : Quel est l'impact de la réforme des retraites de 2023 sur la tarification des contrats collectifs ?

5.1 Présentation des résultats appliqués à la tarification

Les résultats suivants correspondent au produit du coût moyen et de la fréquence pour chaque sous-poste de soins. Afin de simplifier la présentation, les résultats ont été agrégés par poste de soins. Pour plus de détails, le calcul individuel par sous-poste est disponible en annexe [C].

Tous les résultats présentés sont en euros.

Postes	Coûts moyens X Fréquences
Optique	151,40
Dentaire	203,45
Soins de ville	151,10
Pharmacie	39,47
Divers	73,21
Hospitalisation	109,51
Aides auditives	17,68
Total par an	745,84
Total par mois	62,15

TABLEAU 5.1 – Coût moyen X Fréquence pour une personne de 63 ans

Postes	Coûts moyens X Fréquences
Optique	153,84
Dentaire	204,92
Soins de ville	160,15
Pharmacie	40,30
Divers	73,48
Hospitalisation	110,97
Aides auditives	18,97
Total par an	762,63
Total par mois	63,55

TABLEAU 5.2 – Coût moyen X Fréquence pour une personne de 64 ans

Les résultats du produit coût moyen \times fréquence pour les personnes de 64 ans montrent une hausse sur tous les postes de soins par rapport aux personnes de 63 ans. Les augmentations varient de 0,27 à 9,05 €, 9,05€ observée pour les soins de ville, qui affichent la plus forte hausse. Cette augmentation significative est notamment due à la forte progression de la fréquence des

consultations de médecine générale, qui a augmenté d'environ 1,0059 points entre les deux âges. Bien que tous les sous-postes montrent une légère augmentation tant en termes de coût moyen que de fréquence, c'est l'augmentation marquée de la fréquence des consultations de médecine générale qui contribue de manière significative à la hausse globale.

La somme des résultats pour chaque poste de soins donne la prime pure annuelle : 745,84 € pour les personnes de 63 ans et 762,63 € pour celles de 64 ans. Convertie en montants mensuels, cela correspond à 62,15 € et 63,55 € respectivement. Ainsi, une augmentation de 2,25% est observée entre les deux âges.

Ces résultats semblent cohérents mais il peut être pertinent de les comparer aux montants fournis par un logiciel de tarification pour valider leur exactitude. La comparaison et l'évaluation de la fiabilité de ces montants seront abordées dans la Partie 3, permettant de déterminer si ces résultats sont réalistes ou non.

5.2 Présentation de l'outil de tarification marché

L'outil utilisé est un logiciel de tarification de marché. Cet outil permet de tarifier des contrats collectifs ou individuels en se basant sur une démographie et une grille de garanties, incluant des couvertures de base et/ou des options supplémentaires. Pour cette étude, seules les tarifications des garanties de base seront considérées.

Deux tarifications seront réalisées : une pour une personne de 63 ans et une pour une personne de 64 ans.

5.2.1 Fonctionnement de l'outil

5.2.1.1 Démographie

Pour effectuer une tarification, il est essentiel de disposer d'une démographie. Dans la plupart des cas pour un contrat collectif, cette démographie inclut les informations suivantes :

- **L'ensemble des personnes couvertes par le contrat** : il s'agit de tous les employés ou membres de l'entreprise ;
- **Moyenne d'âge** : l'âge moyen des assurés ;
- **Situation matrimoniale** : le nombre de personnes célibataires, veuves, divorcées, mariées ou en union libre ;
- **Répartition par sexe** : le pourcentage de femmes ;
- **Nombre d'enfants à charge** : le nombre d'enfants dépendants des assurés ;
- **Âge moyen des enfants** : l'âge moyen des enfants à charge ;
- **Salaire annuel** : le revenu moyen des assurés ;
- **Catégorie socioprofessionnelle** : la répartition entre cadres, non-cadres, etc ;
- **Région des personnes couvertes** : il est possible de mettre « Autres » pour ne pas mettre de régions spécifiques.

Ces informations permettent d'adapter la tarification en fonction des caractéristiques spécifiques de la population assurée.

Dans le cadre de l'étude, les prédictions réalisées et présentées précédemment ont été effectuées pour les profils suivants :

- Une personne âgée de 63 ans, non cadre et sans enfant à charge ;
- Une personne âgée de 64 ans, non cadre et sans enfant à charge.

Ces profils démographiques sont utilisés pour les deux tarifications de l'outil afin de garantir la cohérence des comparaisons avec les résultats obtenus à partir des modèles GLM.

5.2.1.2 Remplissage d'une grille de garantie

Lors de la réalisation de la tarification, il est nécessaire de sélectionner une démographie préalablement créée.

Ensuite, il convient de remplir la grille de garantie dans le logiciel. Cette grille est divisée en postes de soins (hospitalisation, soins de ville, etc.), en correspondance avec ceux utilisés pour les tarifications basées sur les GLM. Chaque poste est également subdivisé en sous-postes de soins, permettant de spécifier le niveau de garantie pour chacun d'eux.

Le niveau de garantie est divisé en deux éléments :

1. Niveaux de garanties : Ces niveaux correspondent aux plafonds de remboursement de la couverture, par exemple, 50, 400, etc ;
2. Assiettes de remboursement : Les assiettes déterminent la manière dont les remboursements sont calculés, c'est à dire l'unité du remboursement. Elles peuvent inclure :
 - Pourcentage des frais réels ;
 - Pourcentage de la base de remboursement ;
 - Pourcentage du ticket modérateur ;
 - Pourcentage du PMSS (Plafond Mensuel de la Sécurité Sociale) ;
 - Montants fixes en euros.

Il est également possible de mixer ces assiettes, par exemple, en combinant 50% du PMSS avec un montant fixe de 100€.

Pour cette étude, les primes pures calculées à partir des GLM se basent uniquement sur les remboursements de la mutuelle. Les remboursements de la sécurité sociale n'ont pas été intégrés dans la grille de garantie.

Chaque ligne de sous-postes de soins doit être remplie avec le niveau de garantie et l'assiette correspondante de la grille de couverture sélectionnée.

5.2.1.3 Visualisation des résultats

Une fois la grille de garantie complétée, le logiciel calcule le taux technique de la cotisation pour chaque ligne de sous-postes de soins, en fonction de la démographie sélectionnée et de la grille de

garantie. Le logiciel fournit uniquement ce pourcentage.

En revanche, le logiciel inclut des taxes dans le taux technique, qu'il est important de soustraire. Ces taxes représentent 34,07% du taux technique. Ainsi, le taux technique réel, sans les taxes, est calculé comme suit :

$$\begin{aligned} \text{Taux technique réel} &= (1 - 34,07\%) \times \text{Taux technique avec taxes} \\ &= 65,93\% \times \text{Taux technique avec taxes} \end{aligned} \quad (5.1)$$

Pour déterminer le montant de la prime pure pour chaque sous-poste de soins, il faut multiplier le taux technique réel par le PASS (Plafond Annuel de la Sécurité Sociale) de l'année concernée par la tarification. Les contrats associés à ces grilles de garanties se rapportent aux années 2021 et 2022, et le PASS pour ces années est resté constant à 41 136 €.

Ainsi, le montant de la cotisation pure pour chaque sous-poste de soins est calculé comme suit :

$$\text{Prime pure réelle} = \text{Taux technique réel} \times \text{PASS} \quad (5.2)$$

Une fois les primes pures réelles calculées pour chaque sous-poste de soins, elles sont additionnées pour obtenir la prime pure annuelle.

5.2.2 Grille de tarification utilisée

Au début du mémoire, il a été mentionné que les personnes utilisées pour cette étude provenaient de trois garanties différentes, chacune correspondant à une grille de couverture distincte. Ces grilles sont des grilles « standards », sans options supplémentaires. Étant donné les différences entre ces grilles, il est complexe de sélectionner une grille unique pour l'analyse. Pour obtenir une grille de garantie la plus représentative, la proportion de personnes couvertes par chaque garantie est calculée. Ensuite, une grille de garantie composite est élaborée, prenant en compte chaque grille en fonction de la proportion de personnes couvertes par chacune d'elles.

	Garantie 1	Garantie 2	Garantie 3
% de personnes couvertes	56%	17%	27%

TABEAU 5.3 – Proportion de personnes couvertes par garantie

Plus de la moitié des personnes, soit 56%, sont couvertes par la garantie 1. La garantie 2 couvre 17% des personnes, tandis que la garantie 3 en couvre 27%. Il est à noter que la garantie 1 offre la couverture la plus complète, avec plusieurs lignes de garanties couvrant jusqu'à 100% des frais réels, ce qui en fait la garantie la plus coûteuse. En revanche, la garantie 3 offre la couverture la moins étendue, et est donc la moins chère. Pour la tarification, un mélange de ces trois garanties est utilisé, en tenant compte des proportions de personnes couvertes par chacune d'elles.

5.3 Analyse des résultats et comparaison avec la tarification du logiciel

L'étape de comparaison des résultats des primes pures est importante. En effet, elle permet de vérifier si les résultats obtenus à partir des modèles sont fiables et représentatifs de la réalité. Il est important de noter que l'analyse des résidus des GLM de fréquence des soins de ville et de la pharmacie n'était pas concluante, ce qui pourrait entraîner des divergences entre les résultats des modèles et ceux fournis par le logiciel de tarification. L'objectif est donc d'identifier les éventuels écarts entre les deux approches, en déterminant les sources potentielles de ces différences. Les résultats sont présentés par poste de soins, avec un détail de la comparaison pour chaque sous-poste disponible en annexe [C].

Tous les résultats sont présentés en euros, tandis que les écarts sont exprimés en pourcentage pour permettre une meilleure compréhension des différences en tenant compte de l'ordre de grandeur de chaque poste. Les écarts sont calculés selon la formule suivante :

$$\text{Écart (\%)} = \frac{\text{Prime pure GLM} - \text{Prime pure logiciel}}{\text{Prime pure logiciel}} \times 100 \quad (5.3)$$

5.3.1 63 ans

Postes	Primes pures GLM	Primes pures logiciel	Variations
Optique	151,40	151,99	-0,38%
Dentaire	203,45	284,06	-28,38%
Soins de ville	151,10	170,79	-11,53%
Pharmacie	39,47	103,14	-61,73%
Divers	73,21	33,08	121,33%
Hospitalisation	109,51	113,04	-3,12%
Aides auditives	17,68	9,50	86,18%
Total par an	745,84	865,59	-13,84%
Total par mois	62,15	72,13	-13,84%

TABLEAU 5.4 – Comparaison des tarifications pour une personne de 63 ans par poste de soins

Deux postes de soins, l'optique et l'hospitalisation, présentent des primes pures très proches de celles obtenues avec le logiciel de tarification. Les écarts observés sont de -0,38% et -3,12% respectivement par rapport aux primes calculées par le logiciel. Ces faibles différences indiquent que les résultats des prédictions de coût moyen et de fréquence des GLM pour ces deux postes sont fiables et proches de la réalité.

Le poste de soins de ville montre une prime pure relativement proche de celle obtenue avec le logiciel de tarification. En effet, la prime pure calculée par le GLM est inférieure de 11,53% par rapport à celle du logiciel. Compte tenu de cet écart modéré, les résultats de coût moyen et de fréquence des GLM pour ce poste de soins peuvent être considérés comme corrects.

Le poste de soins du dentaire présente une prime pure moyennement proche de celle obtenue avec le logiciel de tarification, avec un écart de -28,38 %. Cet écart significatif indique que les

résultats de coût moyen et de fréquence des GLM pour ce poste sont moyennement fiables. L'importante variation de ce poste entre 2021 et 2022, constatée dans l'analyse descriptive, pourrait expliquer cet écart avec le logiciel. Par conséquent, les prédictions de coût moyen et de fréquence pour ce poste doivent être considérées avec plus de prudence.

En revanche, trois postes de soins montrent un écart important par rapport aux primes pures obtenues avec le logiciel de tarification. Ces écarts sont de -61,73% pour la pharmacie, +86,18% pour les aides auditives, et +121,33% pour les soins divers. L'analyse des résidus de la fréquence pour la pharmacie avait déjà révélé des anomalies, rendant un écart notable attendu pour ce poste. De plus, le poste divers englobe plusieurs prestations différentes, ce qui peut compliquer la tâche du logiciel pour appliquer une tarification précise, entraînant ainsi un risque de sous-tarification. Par ailleurs, étant donné que ce poste de soins est de plus en plus consommé, un logiciel qui n'est pas régulièrement mis à jour peut également entraîner des erreurs de sous-tarification. Bien que les écarts pour les postes divers et aides auditives soient significatifs, il est important de souligner que ces postes représentent les primes pures les plus faibles. Par conséquent, leur impact global sur la tarification est relativement limité.

Au global, les écarts entre certaines primes pures tendent à se compenser, ce qui entraîne une diminution de la prime pure annuelle de -13,84% par rapport à celle obtenue avec le logiciel de tarification. Cet écart se traduit par une différence mensuelle de 9,98€. Il peut provenir d'un mixe des grilles de garanties étant donné que les personnes présentes dans la base ont des couvertures différentes. Bien que l'écart entre la prime pure mensuelle du GLM et celle du logiciel ne soit pas excessif, il reste significatif et mérite une attention particulière. Cela rassure quant à la fiabilité globale des modèles, mais souligne l'importance de surveiller ces écarts.

5.3.2 64 ans

Postes	Primes pures GLM	Primes pures logiciel	Variations
Optique	153,84	154,83	-0,64%
Dentaire	204,92	289,25	-29,15%
Soins de ville	160,15	173,94	-7,93%
Pharmacie	40,30	105,01	-61,62%
Divers	73,48	33,71	117,95%
Hospitalisation	110,97	115,10	-3,59%
Aides auditives	18,97	9,53	99,02%
Total par an	762,63	881,38	-13,47%
Total par mois	63,55	73,45	-13,47%

TABLEAU 5.5 – Comparaison des tarifications pour une personne de 64 ans par poste de soins

Les primes pures issues des GLM pour une personne de 64 ans suivent globalement la même tendance que celles calculées pour une personne de 63 ans.

L'optique et l'hospitalisation montrent des primes pures très proches de celles obtenues via le logiciel de tarification, avec des écarts légèrement plus élevés que pour une personne de 63 ans, soit respectivement -0,64% et -3,59% par rapport aux primes pures du logiciel. Les résultats des prédictions de coût moyen et de fréquence des GLM pour ces deux postes restent donc fiables et proches de la réalité.

Pour une personne de 64 ans, l'écart entre la prime pure du GLM et celle du logiciel est légèrement réduit pour le poste de soins de ville, atteignant -7,93%. Les résultats des prédictions de coût moyen et de fréquence des GLM pour ce poste sont donc jugés corrects, compte tenu de cet écart.

Pour une personne de 64 ans, l'écart pour le poste de soins dentaires est légèrement plus élevé que pour une personne de 63 ans, atteignant -29,15%. Cet écart significatif suggère que les résultats de coût moyen et de fréquence des GLM pour ce poste sont moyennement fiables. Par conséquent, les prédictions de coût moyen et de fréquence pour ce poste doivent être interprétées avec une légère prudence.

En revanche, les trois postes de soins présentant un écart important par rapport aux primes pures du logiciel de tarification sont les suivants : -61,62% pour la pharmacie, +99,02% pour les aides auditives, et +117,95% pour les soins divers. Les écarts pour la pharmacie et les soins divers sont légèrement plus petits que pour une personne de 63 ans, tandis que l'écart pour les aides auditives est légèrement plus grand. Les raisons de ces écarts restent les mêmes que celles observées pour une personne de 63 ans.

Au total, les écarts entre certaines primes pures se compensent, entraînant une prime pure annuelle qui est inférieure de -13,47% par rapport à celle du logiciel, ce qui représente une légère amélioration par rapport à la personne de 63 ans. Par mois, cela correspond à un écart de 9,90 €. Bien que l'écart entre la prime pure mensuelle du GLM et celle du logiciel ne soit pas aberrant, il est important de le prendre en compte dans l'évaluation globale.

De manière générale, les résultats obtenus ne sont pas aberrants et sont majoritairement satisfaisants. Les écarts observés, tant pour les personnes de 63 que de 64 ans, sont principalement dus aux limites des données et des modèles utilisés. Ces aspects seront examinés en détail dans la dernière partie.

5.4 Impact sur le portefeuille de la mutuelle

L'objectif initial est d'évaluer l'impact du report de l'âge de départ à la retraite sur les contrats collectifs, et par conséquent sur le portefeuille d'une mutuelle gérant ces contrats. Cet impact est observé chaque année jusqu'à l'horizon 2030. Le report de l'âge de la retraite est progressif et concerne les personnes nées à partir du 1er septembre 1961. Pour rappel :

Dates de naissance	Durées dans le portefeuille
Entre le 01/09/1961 et 31/12/1961	3 mois
1962	6 mois
1963	9 mois
1964	1 an
1965	1 an et 3 mois
1966	1 an et 6 mois
1967	1 an et 9 mois
1968	2 ans

TABLEAU 5.6 – Phase transitoire du décalage de 62 à 64 ans

Cette partie permet donc de répondre à la problématique.

5.4.1 Données utilisées

Pour rappel, les primes pures annuelles en euros pour les personnes de 63 et 64 ans, calculées à partir des modèles GLM, sont respectivement égales à :

Âge	Primes pures par an
63 ans	745,84
64 ans	762,63

TABLEAU 5.7 – Primes pures annuelles calculées des personnes de 63 et 64 ans

La mise en place du décalage commence le 01/09/2023, c'est pourquoi le choix a été fait de se baser uniquement sur le portefeuille de l'année 2022 afin d'optimiser la pertinence des résultats.

Pour observer l'impact sur le portefeuille, il est nécessaire de connaître le portefeuille initial de l'année 2022. Il représente la somme totale des prestations remboursées par la mutuelle aux personnes protégées, et il s'élève à 5 303 362,04€.

5.4.2 Hypothèses

5.4.2.1 Hypothèses globales

L'objectif pour mesurer l'impact est d'inclure les personnes qui seront affectées par le décalage progressif de l'âge de départ à la retraite. Selon le tableau retraçant la phase transitoire de ce décalage de 62 à 64 ans, les personnes nées en 1962 travailleront 6 mois de plus, celles nées en 1963, 9 mois de plus, et ainsi de suite. Par conséquent, ces personnes resteront couvertes par leur contrat collectif pendant cette durée supplémentaire.

Années	Dates de naissance	Durées dans le portefeuille	63 ans	64 ans
			Durées dans le portefeuille	Durées dans le portefeuille
2023	Entre le 01/09/1961 et 31/12/1961	3 mois	0,25	0
2024	1962	6 mois	0,50	0
2025	1963	9 mois	0,75	0
2026	1964	1 an	1	0
2027	1965	1 an et 3 mois	1	0,25
2028	1966	1 an et 6 mois	1	0,50
2029	1967	1 an et 9 mois	1	0,75
2030	1968	2 ans	1	1

TABLEAU 5.8 – Récapitulatif des expositions supplémentaires en fonction des dates de naissance

Le tableau ci-dessus résume les durées supplémentaires pendant lesquelles les personnes resteront dans le portefeuille de la mutuelle entre 2023 et 2030. Par exemple, une personne née en 1965 qui travaillera 1 an et 3 mois de plus à partir de 2027 aura une exposition de 1 sur l'année 2027, puis de 0,25 sur l'année 2028. Après cela, elle atteindra l'âge légal de la retraite et quittera son contrat collectif.

Il est supposé que chaque personne ayant atteint l'âge légal de la retraite quitte effectivement son emploi et son contrat collectif pour partir à la retraite. De plus, une hypothèse d'homogénéité est posée pour les autres tranches d'âge, stipulant que les entrées et les sorties de ces individus dans le portefeuille se déroulent à un rythme identique.

5.4.2.2 Hypothèses sur le nombre de personnes ajouté par an

Chaque année, entre 2023 et 2030, un nombre croissant de personnes sera amené à travailler plus longtemps en raison du décalage de l'âge de départ à la retraite. Pour estimer cet impact, il est réaliste d'utiliser les personnes actives de la base de données dont les dates de naissance correspondent aux années concernées. Par exemple, pour l'année 2023, les personnes nées entre le 01/09/1961 et le 31/12/1961 sont prises en compte. Le nombre de ces personnes est ensuite multiplié par l'exposition supplémentaire dans le portefeuille (0,25 pour cet exemple) et par la prime pure annuelle des 63 ans calculée à partir des modèles GLM.

Ce calcul se complique lorsque ces individus travaillent au-delà d'un an supplémentaire. Par exemple, une personne née en 1965 restera 1 an dans le portefeuille en 2027 et 3 mois dans celui de 2028. Ainsi, l'ensemble des personnes nées en 1965 devra être ajouté en 2028 pendant une durée plus courte, c'est donc l'exposition qui entre en compte. Ce principe s'applique de la même manière pour les années suivantes. La formule pour illustrer ces principes est présentée ci-dessous.

$$\begin{aligned}
 & \text{Montant à ajouter au portefeuille}_N \\
 & = \\
 & \text{Nombre de personnes ajouté}_N \times \text{Exposition sur l'année des 63 ans} \times \text{Prime pure des 63 ans} \\
 & + \\
 & \text{Nombre de personnes ajouté}_{N-1} \times \text{Exposition sur l'année des 64 ans} \times \text{Prime pure des 64 ans}
 \end{aligned}
 \tag{5.4}$$

où N est l'année étudiée et $N - 1$ l'année précédente.

Chaque année, un nombre de personnes est ajouté au portefeuille en raison du décalage de l'âge de départ à la retraite, mais un nombre est également retiré, à l'exception de la première et de la sixième année. La raison pour laquelle aucune personne n'est retirée la sixième année est développée dans le paragraphe suivant. Par exemple, les personnes ajoutées en 2023, qui travaillent 3 mois de plus, ne seront plus présentes l'année suivante, car il est supposé qu'elles prennent leur retraite à l'âge légal. Ainsi, en 2024, les personnes ajoutées en 2023 seront retirées du portefeuille, avec les prestations associées à ces individus également supprimées.

Il est important de noter que lorsque les personnes sont amenées à travailler exclusivement plus d'une année supplémentaire, elles ne sont pas retirées du portefeuille l'année suivante. Par exemple, une personne qui travaille 1 an et 3 mois supplémentaires à partir de 2027 sera comptée pour 1 an en 2027 et pour 3 mois en 2028. En conséquence, cette personne sera retirée du portefeuille seulement en 2029. C'est pourquoi, dans le tableau suivant, aucune personne n'est retirée en 2028. Les personnes retirées en 2029 correspondent donc à celles ajoutées en 2027 et ainsi de suite.

Il est important de distinguer le nombre de personnes ajoutées du montant ajouté. Les personnes ajoutées peuvent rester dans le portefeuille pendant plus d'une année et donc apparaître sur plusieurs années de revalorisation. En revanche, le montant ajouté chaque année est spécifique à l'année en question et ne se reporte pas d'une année sur l'autre. En effet, si la même personne mentionnée précédemment reste 3 mois en 2028, le montant correspondant à cette période est inclus dans le montant ajouté pour l'année 2028 (d'après la formule précédente).

Le tableau suivant synthétise ces explications. Les montants de chaque année sont en euros.

Année	A ajouter au portefeuille		A retirer du portefeuille	
	Nombre de personnes	Montants	Nombre de personnes	Montants
2023	23	4 288,56	0	0,00
2024	99	36 918,93	23	4 288,56
2025	105	58 734,66	99	36 918,93
2026	119	88 754,59	105	58 734,66
2027	140	104 417,17	119	88 754,59
2028	134	126 634,09	0	104 417,17
2029	139	154 767,34	140	126 634,09
2030	128	174 970,99	134	154 767,34

TABLEAU 5.9 – Tableau des hypothèses pour le calcul de l'impact

Pour l'année 2027, par exemple, 140 personnes travailleront 1 an et 3 mois de plus avant d'atteindre l'âge légal de la retraite. Pour déterminer le montant à ajouter au portefeuille, procédez comme suit :

$$\begin{aligned}
 &= 140 \times 1 \times 745,84 + 119 \times 0 \times 762,63 \\
 &= 104\,417,17
 \end{aligned}
 \tag{5.5}$$

Le montant à ajouter au portefeuille pour l'année 2027 est 104 417,17€.

Pour les autres années, le montant à ajouter au portefeuille est calculé de la même manière.

Le montant à retirer du portefeuille en 2027 correspond au montant ajouté en 2026, soit 88 754,59 €.

5.4.3 Impact sur le portefeuille à horizon 2030

Chaque année, le portefeuille est ajusté en tenant compte de l'ajout de nouvelles personnes en raison d'une prolongation de leur activité professionnelle, ainsi que du retrait de celles qui partent à la retraite. Ainsi, le portefeuille est revalorisé annuellement selon le calcul suivant :

$$\text{Portefeuille}_N = \text{Portefeuille}_{N-1} + \text{Montant ajouté}_N - \text{Montant retiré}_{N-1} \quad (5.6)$$

où N est l'année du portefeuille calculé, $N-1$ l'année précédente, $\text{Montant retiré}_N = \text{Montant ajouté}_{N-1}$ comme expliqué précédemment.

Pour la première année de calcul, en 2023, le portefeuille $_{N-1}$ correspond au portefeuille initial, c'est-à-dire celui de 2022, et aucun montant n'est à retirer.

L'impact correspond à l'écart en pourcentage entre le portefeuille $_N$ et le portefeuille $_{N-1}$. Il est calculé de la manière suivante :

$$\text{Impact}_N = \frac{\text{Portefeuille}_N - \text{Portefeuille}_{N-1}}{\text{Portefeuille}_{N-1}} \times 100 \quad (5.7)$$

De même que pour l'estimation du portefeuille, pour calculer l'impact de l'année 2023, le Portefeuille $_{N-1}$ correspond au portefeuille initial de 2022.

Le tableau suivant présente un récapitulatif de l'ensemble des portefeuilles estimés et de l'impact calculé pour chaque année.

Année	Portefeuilles	Impacts
2023	5 307 650,60	0,08%
2024	5 340 280,97	0,61%
2025	5 362 096,70	0,41%
2026	5 392 116,63	0,56%
2027	5 407 779,21	0,29%
2028	5 429 996,13	0,41%
2029	5 458 129,38	0,52%
2030	5 478 333,03	0,37%

TABLEAU 5.10 – Tableau des portefeuilles estimés et des impacts

Pour l'exemple sur l'année 2027, le portefeuille estimé est :

$$\begin{aligned} \text{Portefeuille}_{2027} &= \text{Portefeuille}_{2026} + \text{Montant ajouté}_{2027} - \text{Montant retiré}_{2026} \\ &= 5\,392\,116,63 + 104\,417,17 - 88\,754,59 \\ &= 5\,407\,779,21 \end{aligned} \quad (5.8)$$

Ainsi, le portefeuille estimé en 2027 pour cette mutuelle, basé sur les hypothèses et les données prédites par les modèles GLM, s'élève à 5 407 779,21 €.

Quant à l'impact, il est égal à :

$$\begin{aligned}
 \text{Impact}_{2027} &= \frac{\text{Portefeuille}_{2027} - \text{Portefeuille}_{2026}}{\text{Portefeuille}_{2026}} \\
 &= \frac{5\,407\,779,21 - 5\,392\,116,63}{5\,392\,116,63} \\
 &= 0,29\%
 \end{aligned}
 \tag{5.9}$$

L'impact de la réforme des retraites sur le portefeuille de l'année 2027 par rapport à celui de 2026 est de 0,29%.

Les impacts de chaque année sont relativement proches mais varient légèrement. Il est donc pertinent de calculer l'impact moyen entre 2023 et 2030, en tenant compte du nombre de personnes ajoutées chaque année dans le portefeuille.

Ainsi, l'impact moyen de la réforme des retraites sur les contrats collectifs d'une mutuelle, de 2023 à l'horizon 2030, en fonction du nombre de personnes ajoutées, est de 0,44% par an. La problématique de l'étude est donc résolue.

D'après un extrait qui présente les résultats du cabinet de conseil Galea [28], Gautron N. dit : « *En tenant compte des arrêts de travail, certaines études tablent sur une hausse de 0,5% par an sur huit ans des coûts d'indemnisation pour les assureurs pouvant être répercutés sur les primes. Les chiffres de progression sont un peu plus faibles, autour de 0,3% par an, pour ce qui concerne la dérive des dépenses de santé à horizon 2030* ».

En ce qui concerne les soins de santé, d'après l'article, l'impact est estimé à environ 0,3% par an sur une période de 8 ans. La moyenne de l'impact annuel sur 8 ans, d'après les analyses et études réalisées dans ce mémoire, est de 0,44%. Étant donné que le résultat obtenu n'est pas aberrant et que le taux trouvé est relativement proche de celui mentionné dans l'article, cela confirme la fiabilité de l'étude ainsi que les étapes de l'analyse réalisées.

5.5 Limites de l'étude et ouverture

5.5.1 Données utilisées

Pour réaliser une étude optimale, il est nécessaire de disposer d'un volume conséquent de données. Plus le nombre de données est élevé, plus il est possible de les traiter de manière détaillée, qu'elles soient analysées séparément ou globalement, et de minimiser les insuffisances. Pour obtenir des résultats encore plus fiables, il aurait été avantageux d'avoir davantage de données, ainsi que des données provenant de différentes mutuelles.

5.5.1.1 Données manquantes

Outre le manque général de données, l'étude se concentre exclusivement sur les personnes et les prestations en France métropolitaine, ce qui limite la validité des résultats à ce territoire. Pour obtenir des résultats applicables aux territoires d'outre-mer et aux pays étrangers, il aurait été nécessaire d'avoir davantage de données spécifiques à ces régions et, éventuellement, de réaliser une étude distincte pour ensuite comparer les résultats avec ceux de la France métropolitaine afin d'identifier d'éventuelles différences significatives dans les primes pures.

5.5.1.2 Retraitement des montants de l'hospitalisation

Pour le poste hospitalisation, les montants des prestations ont été recalculés en fonction de la colonne Nombre de jours, créée pour les sous-postes correspondant aux chambres. Les hypothèses utilisées sont que les chambres en médecine, chirurgie, maternité et ambulatoire coûtent 50 € par jour, tandis que les chambres en psychiatrie coûtent 100 € par jour et les frais de séjour coûtent 20€. Ces montants représentent des valeurs moyennes applicables uniquement aux hôpitaux publics. Ils peuvent varier considérablement si les prestations sont fournies par des cliniques privées.

5.5.1.3 Exclusion des enfants

Dans l'ensemble de l'étude, les enfants n'ont pas été pris en compte. Étant donné que le projet se concentre sur l'impact du décalage de l'âge de départ à la retraite, leur inclusion n'était pas pertinente pour les objectifs de cette étude, qui visait à obtenir des primes pures calculées par personne.

Cependant, il serait pertinent d'inclure les enfants dans une autre étude visant à prédire les coûts moyens et les fréquences pour des familles dont l'un des membres a 63 ou 64 ans. L'objectif serait d'observer l'impact de ce décalage de l'âge de départ sur des tarifications de couvertures de famille comprenant dans le même contrat les enfants et les conjoints/concubins.

5.5.1.4 Distinction des années des bases de données

Les bases de données utilisées pour l'étude couvrent les années 2021 et 2022. L'analyse descriptive révèle une légère diminution du coût moyen par personne et par prestation en 2022 par rapport à 2021. Les modèles ont été élaborés en utilisant les données de ces deux années pour en assurer un nombre suffisant. Cependant, les primes pures calculées pourraient varier selon que l'on utilise les données de 2021 ou celles de 2022. Il serait pertinent d'examiner si cette évolution observée entre les deux années se reflète également dans les primes pures obtenues.

De plus, la baisse du coût moyen entre les deux années est attribuée à une diminution des montants des prestations pour certains sous-postes du dentaire. Il aurait été intéressant d'explorer les raisons de cette baisse et d'analyser spécifiquement les sous-postes dentaires concernés pour comprendre les facteurs sous-jacents à cette évolution. Il est également noté une inflation de la consommation des soins de santé au cours des dernières années, ce qui pourrait amplifier l'impact d'ici 2030.

5.5.1.5 Régions de provenance des assurés

Les régions d'origine des assurés sont présentées dans l'analyse descriptive mais ne sont pas intégrées dans les modèles de l'étude. Le choix fait dans ce mémoire était de créer des modèles

simples en se basant uniquement sur les sous-postes de soins et l'âge comme variables explicatives, afin d'obtenir des résultats généraux pour l'ensemble de la France métropolitaine. En effet, les prestations sont réparties de manière inégale entre les régions, avec une forte concentration en Île-de-France et en Centre-Val de Loire. Pour intégrer les régions dans les modèles, il aurait été nécessaire d'avoir une répartition uniforme des prestations entre toutes les régions, car une région avec un faible nombre de prestations pourrait produire des résultats peu représentatifs. Si la variable région avait été incluse, les montants des primes pures pour les personnes de 63 et 64 ans pourraient varier en fonction de la région de France.

5.5.1.6 Type de situations des assurés

Tous les assurés utilisés dans les modèles sont des personnes actives, comprenant divers types comme les mandataires sociaux, les licenciés, etc. Cependant, ces différents types d'actifs n'ont pas été distingués dans les modèles, en raison de l'objectif d'une étude plus globale. Il est possible que les primes pures calculées pour les personnes de 63 et 64 ans présentent des variations en fonction des types de situations.

5.5.1.7 Catégories socioprofessionnelles des assurés

Les modèles ont été réalisés sans inclure la catégorie socioprofessionnelle, car la majorité des personnes dans la base de données sont des non-cadres. Cependant, les non-cadres se divisent en plusieurs sous-catégories, telles que les agriculteurs, les employés, etc. Par conséquent, cette variable pourrait introduire des différences significatives dans les primes pures des personnes de 63 et 64 ans.

De plus, les primes pures obtenues concernent uniquement les personnes non-cadres, en raison du nombre relativement faible de personnes cadres dans la base de données, qui ont donc été négligées. Il serait intéressant de réaliser une étude similaire en utilisant exclusivement les données des personnes cadres (avec plus de données), afin de déterminer si les primes pures pour les personnes de 63 et 64 ans présentent des différences notables.

5.5.1.8 Postes de soins

Le nombre de données est particulièrement limité pour certains postes de soins. En effet, les prestations pour la pharmacie et les soins de ville sont les plus nombreuses, ce qui est attendu puisque ce sont les postes les plus couramment consommés. En revanche, le poste des aides auditives ne compte que 142 observations, ce qui est insuffisant. Ce manque de données considérable peut affecter la fiabilité et la pertinence des résultats. Pour une étude optimale, il aurait été nécessaire d'avoir un plus grand nombre de prestations pour les aides auditives et, de manière générale, pour l'ensemble des postes de soins.

5.5.1.9 Comparaison du comportement des actifs et des retraités

De manière similaire aux postes de soins, la base de données contient seulement 149 personnes actives de 63/64 ans et 55 personnes retraitées de ces mêmes âges. Pour observer de manière plus fiable les différences de comportement entre ces deux catégories, il serait nécessaire de disposer d'un nombre plus important de personnes dans chaque groupe.

5.5.2 Modèles utilisés

5.5.2.1 Choix des modèles

Les modèles choisis sont des modèles linéaires généralisés (GLM) en raison de leur méthode de modélisation bien établie et de la solidité de la théorie qui les sous-tend. Cependant, il serait également pertinent de tester d'autres modèles, tels que ceux basés sur le machine learning. Comparer la performance et les résultats des GLM avec ceux de modèles de machine learning pourrait offrir des perspectives intéressantes.

En outre, les types les plus courants de GLM ont été testés, notamment pour le modèle de coût moyen. Cependant, il existe des modèles GLM plus complexes qui n'ont pas été explorés. Il pourrait être intéressant d'évaluer ces modèles plus sophistiqués et de comparer leurs résultats afin de déterminer s'ils offrent des améliorations significatives en termes de précision ou de performance.

5.5.2.2 Indicateurs de performances des modèles

L'AIC et le pseudo- R^2 ont été utilisés pour comparer plusieurs modèles. Les résultats des pseudo- R^2 montrent que certains sont très faibles, notamment pour les postes de soins tels que les soins de ville, la pharmacie, l'hospitalisation, et l'optique dans le GLM de coût moyen, ainsi que pour l'optique, les soins de ville, le divers, l'hospitalisation, et les aides auditives dans le GLM de fréquence. Ces faibles valeurs de pseudo- R^2 indiquent que les modèles ont une capacité limitée à prédire les données, ce qui constitue une limitation importante à prendre en compte.

Pour évaluer la fiabilité des modèles, une analyse des résidus a été réalisée. Cette analyse montre des résultats plutôt satisfaisants, à l'exception de la fréquence des soins de ville et de la pharmacie. Cependant, cela n'est pas très problématique, car il s'agit de modèles de fréquence qui prédisent le nombre de sinistres. Idéalement, il serait utile d'utiliser un autre indicateur de performance pour compléter l'évaluation. À noter que, les primes pures calculées à partir des coûts moyens et des fréquences prédites ne présentent pas d'écart très aberrant par rapport à celles du logiciel de tarification, avec une différence d'environ 10 €. Cela confirme que les résultats des résidus pour ces modèles de fréquence ne soulèvent pas de préoccupations.

5.5.2.3 Prédiction

Les primes pures obtenues à partir des prédictions de coûts moyens et de fréquences ont été comparées avec celles fournies par un logiciel de tarification. Bien que certaines différences aient été observées, il est difficile de déterminer avec précision si ces écarts sont dus aux coûts moyens ou aux fréquences. Pour une comparaison plus précise et pour identifier les principales sources d'erreurs, il serait utile de disposer d'un logiciel de tarification qui fournit les coûts moyens et les fréquences pour chaque sous-poste de soins.

5.5.3 Impact sur le portefeuille

L'analyse de l'impact sur le portefeuille a été réalisée jusqu'à horizon 2030 pour l'ensemble du portefeuille de la mutuelle. Les personnes affectées par la réforme des retraites de 2023 ont été ajoutées à la base de données au fur et à mesure des années pour modéliser l'évolution du portefeuille. Cependant, les autres personnes présentes n'ont pas été projetées dans le temps ; par

exemple, les personnes de 30 ans restent les mêmes dans les portefeuilles de 2023 et de 2027.

Pour obtenir des résultats plus réalistes, il serait nécessaire de projeter l'évolution de l'ensemble des assurés en introduisant des hypothèses sur de nouveaux entrants (personnes souscrivant à un contrat collectif) et des partants (assurés partant à la retraite avant l'âge légal, invalidité, décès, etc.).

Les limites mentionnées ci-dessus pourraient expliquer pourquoi l'impact moyen obtenu en fonction du nombre de personnes ajoutées est de 0,44%, supérieur à celui de l'article de Gautron.N [28], qui est de 0,3%. Ces limites incluent notamment le manque de données pour certains postes de soins, la non-considération des différentes catégories socioprofessionnelles (cadres), ainsi que l'absence de projection des personnes dans le temps.

De plus, les résultats dépendent de la grille de garantie et de la population couverte d'une mutuelle. Donc pour une mutuelle ayant des grilles de garanties et une population radicalement différente, l'impact sera probablement différent.

5.5.4 Conclusion sur les limites

Idéalement, une plus grande quantité de données, provenant de plusieurs mutuelles, aurait été souhaitable pour cette étude. Le manque de données a limité la création des modèles et a contribué aux écarts observés dans les résultats finaux, ainsi qu'aux tests de fiabilité des modèles. Cette insuffisance de données constitue une limite importante à prendre en compte.

Les variables explicatives utilisées dans les modèles se limitent à l'âge et aux sous-postes de soins, afin de simplifier l'analyse. Pour obtenir des résultats plus précis, une prochaine étude pourrait inclure d'autres variables telles que le sexe, les régions, les catégories socioprofessionnelles (CSP) ou le métier spécifique des personnes. Analyser l'impact de la réforme des retraites de 2023 en fonction des types de métiers pourrait permettre de déterminer si les dépenses de santé varient selon ces différentes catégories.

Néanmoins, les limites identifiées ne remettent pas en cause la pertinence et la justesse du résultat global de l'étude.

Conclusion

Ce mémoire a exploré les dynamiques complexes liées à la tarification de contrats collectifs dans le contexte de la réforme des retraites de 2023. L'objectif principal était d'évaluer l'impact potentiel de cette réforme sur une mutuelle détenant des contrats collectifs.

L'étude s'est focalisée sur la modélisation des coûts de santé et des fréquences d'utilisation des soins en utilisant des Modèles Linéaires Généralisés (GLM). Ces modèles ont permis d'élaborer des prédictions de coût moyen et de fréquence pour les personnes âgées de 63 et 64 ans non-cadres, sur la base des données fournies par une mutuelle généraliste. Les prédictions ont été réalisées par poste de soins (optique, dentaire, soins de ville, etc.), aboutissant à des primes pures calculées en multipliant les coûts moyens par les fréquences obtenues pour ces deux tranches d'âge. Les primes pures calculées sont respectivement égales à 745,84€ pour les personnes de 63 ans et à 762,63€ pour les personnes de 64 ans par an.

La comparaison des primes pures calculées avec celles fournies par un logiciel de tarification du marché a confirmé la fiabilité des résultats, mensuellement la prime calculée étant d'environ 10 € inférieure à celle du logiciel.

Ces analyses ont permis d'évaluer l'impact de la réforme de 2023 sur le portefeuille de la mutuelle étudiée. Il en ressort que l'intégration progressive des personnes affectées par cette réforme entraînera une légère augmentation des coûts pour les mutuelles, avec **un impact moyen de 0,44% par an d'ici 2030**. Toutefois, ce chiffre doit être interprété avec précaution en raison des limites de l'étude, telles que l'hypothèse de ne pas considérer les nouveaux entrants et sortants du portefeuille pour évaluer cet impact, l'hypothèse selon laquelle toutes les personnes prennent leur retraite à l'âge légal et également l'hypothèse où il n'y a pas de nouveau changement réglementaire impactant la santé.

Le mémoire révèle plusieurs limites ayant influencé les résultats, notamment le manque de données riches et diversifiées, qui a restreint la robustesse des modèles et la fiabilité des prédictions. L'étude s'est principalement concentrée sur des variables explicatives simples, tels que l'âge et les sous-postes de soins, ce qui a contribué à des résultats moins précis.

Enfin, pour pallier ces limites et enrichir la recherche, il serait pertinent de recueillir des données provenant de plusieurs mutuelles, ce qui permettrait d'obtenir une modélisation plus solide et de valider les résultats de manière croisée. L'intégration de variables explicatives supplémentaires enrichirait l'analyse, offrant une perspective plus complète et aidant à mieux comprendre les variations observées dans les comportements de consommation de soins en fonction des différentes catégories d'assurés.

Ce mémoire donne un aperçu des impacts de la réforme des retraites de 2023 sur les contrats collectifs, en quantifiant les effets potentiels sur une mutuelle. Il fournit également des indications utiles pour anticiper l'évolution d'un portefeuille similaire à celui étudié, offrant ainsi des informations intéressantes à toute mutuelle souhaitant évaluer les impacts potentiels de la réforme.

Annexes

Annexe A

Les taux de cotisations de droit commun au 1er janvier 2023

Les taux de cotisations de droit commun au 1er janvier 2023				
Contribution	Assiette	Taux part employeur	Taux part salariale	Total
CSG (charges sociales généralisées) imposable	98,25% du salaire jusqu'à 4 PASS* et 100% au-delà		2,40%	2,40%
CSG (charges sociales généralisées) non imposable	98,25% du salaire jusqu'à 4 PASS* et 100% au-delà		6,80%	6,80%
CRDS (contribution pour le remboursement de la dette sociale)	98,25% du salaire jusqu'à 4 PASS et 100% au-delà		0,50%	0,50%
Contribution solidarité autonomie	Salaire total	0,30%		0,30%
Assurance maladie, maternité, invalidité, décès	Salaire total	7% ou 13%		7% ou 13%
Assurance vieillesse plafonnée	Jusqu'à 1 fois le PASS	8,55%	6,90%	15,45%
Assurance vieillesse déplafonnée	Salaire total	1,90%	0,40%	2,30%
Allocations familiales	Salaire total	5,25 ou taux réduit à 3.45%*		5,25 ou taux réduit à 3.45%*
Accident du travail	Salaire total	Variable		Variable
FNAL (entreprise < 50 salariés)	Jusqu'à 1 fois le PASS	0,10%		0,10%
FNAL (entreprise > 50 salariés)	Au-delà de 1 fois le PASS	0,50%		0,50%
Cotisation chômage	Jusqu'à 4 PASS	4,05%		4,05%
Cotisation AGS	Jusqu'à 4 PASS	0,15%		0,15%
Forfait social	Salaire total	8 ou 20%		8 ou 20%
Contribution au dialogue social	Salaire total	0,016%		0,016%

TABLEAU A.1 – Les taux de cotisations de droit commun au 1er janvier 2023

Annexe B

Graphiques des résidus studentisés

B.1 Dentaire

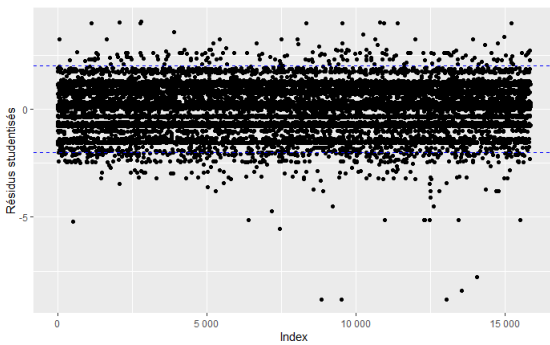


FIGURE B.1 – Résidus studentisés du modèle de coût moyen pour le dentaire

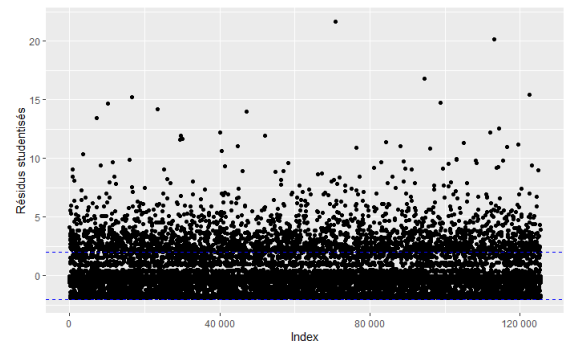


FIGURE B.2 – Résidus studentisés du modèle de fréquence pour le dentaire

B.2 Soins de ville

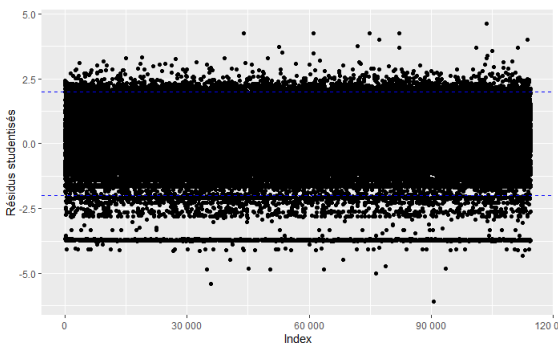


FIGURE B.3 – Résidus studentisés du modèle de coût moyen pour les soins de ville

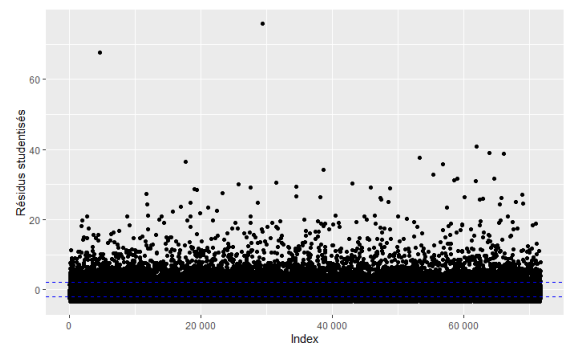


FIGURE B.4 – Résidus studentisés du modèle de fréquence pour les soins de ville

B.3 Pharmacie

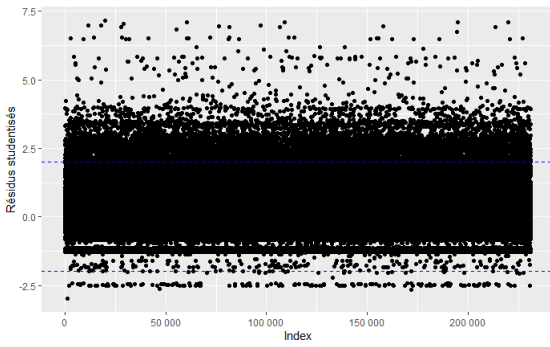


FIGURE B.5 – Résidus studentisés du modèle de coût moyen pour la pharmacie

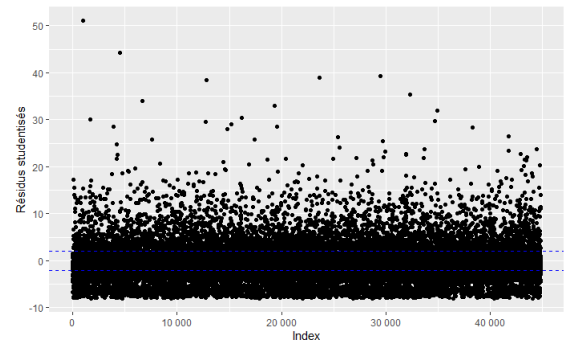


FIGURE B.6 – Résidus studentisés du modèle de fréquence pour la pharmacie

B.4 Divers

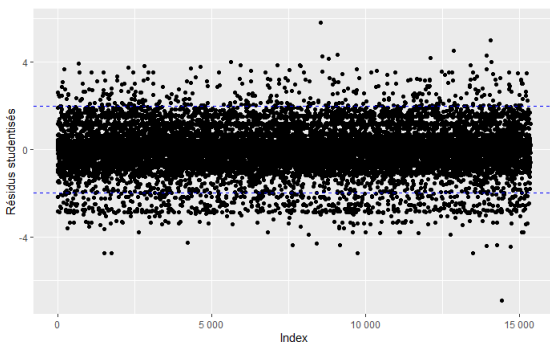


FIGURE B.7 – Résidus studentisés du modèle de coût moyen pour le divers

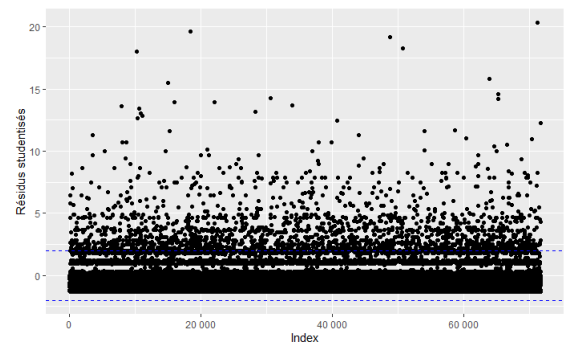


FIGURE B.8 – Résidus studentisés du modèle de fréquence pour le divers

B.5 Hospitalisation

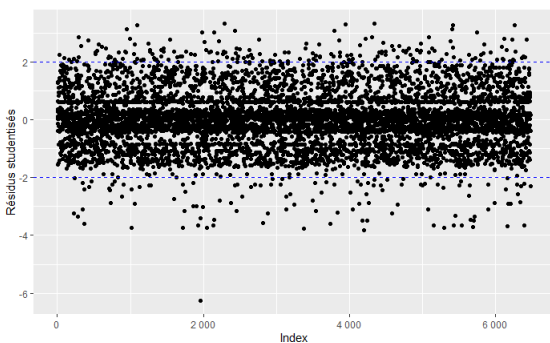


FIGURE B.9 – Résidus studentisés du modèle de coût moyen pour l'hospitalisation

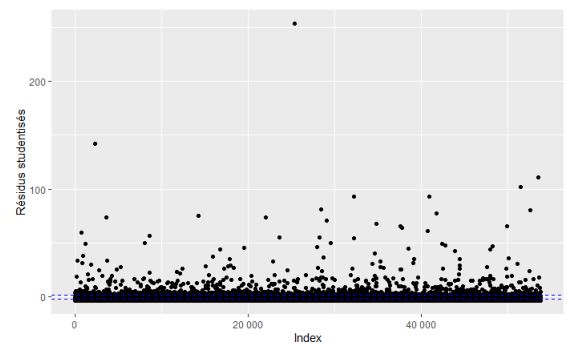


FIGURE B.10 – Résidus studentisés du modèle de fréquence pour l'hospitalisation

B.6 Aides auditives

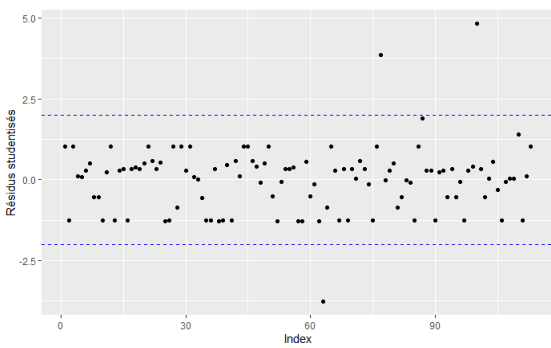


FIGURE B.11 – Résidus studentisés du modèle de coût moyen pour les aides auditives

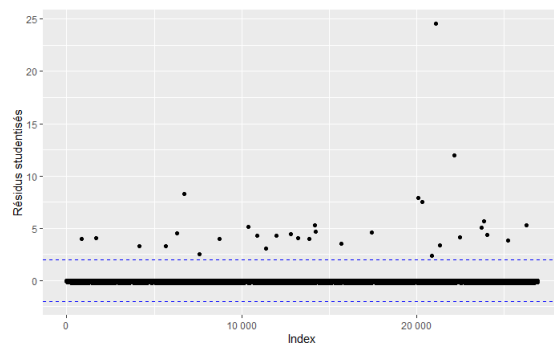


FIGURE B.12 – Résidus studentisés du modèle de fréquence pour les aides auditives

Annexe C

Détail de la comparaison avec le logiciel de tarification par sous-postes de soins

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Optique	Chirurgie réfractive de l'œil (par œil)	2,45	1,03	2,46	1,06
	Lentilles acceptées	0,39	0,19	0,39	0,19
	Lentilles refusées	30,30	18,36	30,32	18,69
	Monture 100% Santé (classe A)	0,10	0,00	0,10	0,00
	Monture hors 100% Santé (classe B)	29,21	20,37	29,24	20,75
	Verres 100% Santé (classe A)	0,40	0,00	0,40	0,00
	Verres complexes	60,66	29,10	60,70	29,67
	Verres simples	25,51	23,87	25,53	24,33
	Verres très complexes	2,39	59,07	4,70	60,15
	Total par an		151,40	151,99	153,84

FIGURE C.1 – Détail des comparaisons des tarifications de l'optique

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Dentaire	Implants dentaires	12,83	0,43	12,95	0,43
	Inlay Core Panier libre	12,55	31,32	12,63	31,89
	Inlay Core Panier modéré	4,48		4,50	
	Inlay Core RAC 0	5,06	3,58	5,10	3,63
	Inlay Onlay Panier libre	18,97	51,37	19,08	52,32
	Inlay Onlay Panier modéré	13,83		13,90	
	Orthodontie acceptée	0,09	1,14	0,09	1,17
	Orthodontie refusée	8,42	0,46	8,50	0,46
	Parodontologie	3,00	0,11	3,03	0,11
	Prothèses dentaires Panier libre	55,17	133,81	55,57	136,28
	Prothèses dentaires Panier modéré	28,68		28,86	
	Prothèses dentaires RAC 0	21,26	32,71	21,43	33,30
	Prothèses dentaires Refusées	0,51	0,00	0,51	0,00
	Soins dentaires	18,61	29,13	18,76	29,64
	Total par an		203,45	284,06	204,92

FIGURE C.2 – Détail des comparaisons des tarifications du dentaire

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Soins de ville	Actes techniques Médicaux	17,43	4,66	17,50	4,72
	Analyses, laboratoire	12,57	23,62	12,62	24,06
	Auxiliaires médicaux	30,22	29,91	30,34	30,48
	Consultations, visites Généraliste	22,94	48,49	31,47	49,41
	Consultations, visites Spécialiste	51,27	51,45	51,48	52,37
	Radiologie (DPTAM et non DPTAM)	16,34	11,50	16,41	11,72
	Soins à l'étranger	0,29	0,25	0,29	0,26
	Densitométrie osseuse	0,04	0,89	0,04	0,92
Total par an		151,10	170,79	160,15	173,94

FIGURE C.3 – Détail des comparaisons des tarifications des soins de ville

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Pharmacie	Médicaments SMR faible	5,05	10,22	5,16	10,41
	Médicaments SMR majeur	21,60	52,45	22,05	53,40
	Médicaments SMR modéré	10,45	40,03	10,67	40,76
	Vaccins Anti grippe	0,10	0,00	0,10	0,00
	Vaccins et Pharmacie NR sur prescription médicale	2,28	0,43	2,32	0,43
	Total par an		39,47	103,14	40,30

FIGURE C.4 – Détail des comparaisons des tarifications de la pharmacie

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Divers	Accessoires	5,60	1,84	5,62	1,90
	Contraception NR	0,17	0,10	0,17	0,11
	Cures thermales (forfait et honoraires)	0,60	1,84	0,60	1,87
	Frais de transport	1,63	1,82	1,64	1,84
	Médecine douce	46,92	16,11	47,09	16,38
	Petit appareillage	7,63	11,07	7,66	11,28
	Grand appareillage	0,29		0,30	
	Indemnités de maternité	10,37	0,30	10,41	0,33
	Total par an		73,21	33,08	73,48

FIGURE C.5 – Détail des comparaisons des tarifications du divers

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Hospitalisation	Chambre particulière en ambulatoire	3,86	4,34	3,90	4,42
	Chambre particulière par jour	32,60	4,18	33,05	4,23
	Forfait hospitalier	36,61	51,23	37,06	52,18
	Frais séjour (Conventionné et Non Conv.)	26,12	44,37	26,48	45,18
	Honoraires et autres Frais	10,33	8,92	10,48	9,09
Total par an		109,51	113,04	110,97	115,10

FIGURE C.6 – Détail des comparaisons des tarifications de l'hospitalisation

Postes	Sous postes	63 ans		64 ans	
		Primes pures GLM	Primes pures logiciel	Primes pures GLM	Primes pures logiciel
Aides auditives	Consommables et accessoires pris en charge	0,07	0,03	0,08	0,04
	Prothèses auditives - Classe 1 - 100% Santé	0,79	0,00	0,89	0,00
	Prothèses auditives - Classe 2 - Tarifs libres	16,82	9,47	17,99	9,49
	Total par an		17,68	9,50	18,97

FIGURE C.7 – Détail des comparaisons des tarifications des aides auditives

Table des figures

1	Étapes pour la prédiction du coût moyen	iv
2	Étapes pour la prédiction de la fréquence	v
3	Étapes de calcul de la prime pure annuelle pour les personnes de 63 et 64 ans	vi
4	Évolution des personnes ajoutées et retirées dans le portefeuille par année	viii
5	Évolution des montants ajoutés et retirés dans le portefeuille par année	viii
6	Évolution des montants des portefeuilles par année	ix
7	Évolution des impacts sur le portefeuille par année	ix
8	Steps for predicting average cost	xii
9	Steps for frequency prediction	xiii
10	Steps to calculate the annual pure premium for people aged 63 and 64	xiv
11	Evolution of people added to and removed from the portfolio by year	xvi
12	Amounts added to and withdrawn from the portfolio by year	xvi
13	Portfolio amounts by year	xvii
14	Impact on the portfolio by year	xvii
1.1	L'organisation de la sécurité sociale	4
2.1	Organisation du système de retraite en France	10
3.1	Boxplot des montants de remboursements de la mutuelle par poste de soins	22
3.2	Répartition des personnes	23
3.3	Répartition des statuts d'activités des personnes protégées	26
3.4	Répartition des catégories socioprofessionnelles	27
3.5	Répartition du nombre de personnes par âge	28
3.6	Répartition du nombre de personnes par âge et du nombre de prestations par âge	29
3.7	Répartition du nombre de personnes par âge et de la somme des montants de prestations par âge	30
3.8	Répartition du nombre de prestations par poste de soins	31
3.9	Coût moyen et fréquence par poste de soins	32
3.10	Coût moyen et fréquence de l'optique	33
3.11	Coût moyen et fréquence du dentaire	34
3.12	Coût moyen et fréquence des soins de ville	35
3.13	Coût moyen et fréquence des soins pharmaceutiques	36
3.14	Coût moyen et fréquence des soins divers	37
3.15	Coût moyen et fréquence de l'hospitalisation	38
3.16	Coût moyen et fréquence des aides auditives	39
4.1	Résidus studentisés du modèle de coût moyen pour l'optique	65
4.2	Résidus studentisés du modèle de fréquence pour l'optique	65

B.1	Résidus studentisés du modèle de coût moyen pour le dentaire	92
B.2	Résidus studentisés du modèle de fréquence pour le dentaire	92
B.3	Résidus studentisés du modèle de coût moyen pour les soins de ville	92
B.4	Résidus studentisés du modèle de fréquence pour les soins de ville	92
B.5	Résidus studentisés du modèle de coût moyen pour la pharmacie	93
B.6	Résidus studentisés du modèle de fréquence pour la pharmacie	93
B.7	Résidus studentisés du modèle de coût moyen pour le divers	93
B.8	Résidus studentisés du modèle de fréquence pour le divers	93
B.9	Résidus studentisés du modèle de coût moyen pour l'hospitalisation	93
B.10	Résidus studentisés du modèle de fréquence pour l'hospitalisation	93
B.11	Résidus studentisés du modèle de coût moyen pour les aides auditives	94
B.12	Résidus studentisés du modèle de fréquence pour les aides auditives	94
C.1	Détail des comparaisons des tarifications de l'optique	95
C.2	Détail des comparaisons des tarifications du dentaire	95
C.3	Détail des comparaisons des tarifications des soins de ville	96
C.4	Détail des comparaisons des tarifications de la pharmacie	96
C.5	Détail des comparaisons des tarifications du divers	96
C.6	Détail des comparaisons des tarifications de l'hospitalisation	96
C.7	Détail des comparaisons des tarifications des aides auditives	96

Liste des tableaux

1	Phase transitoire du décalage de 62 à 64 ans	iii
2	Proportions de résidus studentisés pour les modèles de coût moyen et de fréquence pour chaque poste de soins	vi
3	Transitional phase from 62 to 64 years of age	xi
4	Proportions of studentized residuals for average cost and frequency models for each care item	xiv
2.1	Phase transitoire du décalage de 62 à 64 ans	12
3.1	Évolution du coût moyen et de la fréquence entre 2021 et 2022	24
3.2	Évolution du coût moyen par personne et par grand poste entre 2021 et 2022	24
3.3	Proportion des montants de prestations par région	25
3.4	Comparaison du comportement entre les actifs de 63/64 ans et les retraités de 63/64 ans	40
4.1	Comparaison des AIC des modèles GLM Log-Normale et Gamma identité	57
4.2	Comparaison des pseudo- R^2 des modèles GLM Log-Normale et Gamma identité	57
4.3	Comparaison des AIC des modèles GLM Log-Normale et Gamma identité	58
4.4	Comparaison des pseudo- R^2 des modèles GLM Log-Normale et Gamma identité	58
4.5	Comparaison des AIC des modèles GLM Poisson avec et sans la variable explicative AGE	60
4.6	Comparaison des pseudo- R^2 des modèles GLM Poisson avec et sans la variable explicative AGE	60
4.7	P-values du GLM de coût moyen et de fréquence pour le poste optique	61
4.8	P-values du GLM de coût moyen et de fréquence pour le poste dentaire	62
4.9	P-values du GLM de coût moyen et de fréquence pour le poste soins de ville	62
4.10	P-values du GLM de coût moyen et de fréquence pour le poste pharmacie	63
4.11	P-values du GLM de coût moyen et de fréquence pour le poste divers	63
4.12	P-values du GLM de coût moyen et de fréquence pour le poste hospitalisation	63
4.13	P-values du GLM de coût moyen et de fréquence pour le poste aides auditives	64
4.14	Proportions de résidus studentisés pour les modèles de coût moyen et de fréquence pour chaque poste de soins	65
4.15	Comparaison des montants observés et des montants prédits pour un actif de 50 ans	66
4.16	Résultats des prédictions de coût moyen et de fréquence de l'optique pour les personnes de 63 et 64 ans	67
4.17	Résultats des prédictions de coût moyen et de fréquence du dentaire pour les personnes de 63 et 64 ans	68
4.18	Résultats des prédictions de coût moyen et de fréquence des soins de ville pour les personnes de 63 et 64 ans	68

4.19	Résultats des prédictions de coût moyen et de fréquence de la pharmacie pour les personnes de 63 et 64 ans	69
4.20	Résultats des prédictions de coût moyen et de fréquence du divers pour les personnes de 63 et 64 ans	69
4.21	Résultats des prédictions de coût moyen et de fréquence de l'hospitalisation pour les personnes de 63 et 64 ans	70
4.22	Résultats des prédictions de coût moyen et de fréquence des aides auditives pour les personnes de 63 et 64 ans	70
5.1	Coût moyen X Fréquence pour une personne de 63 ans	72
5.2	Coût moyen X Fréquence pour une personne de 64 ans	72
5.3	Proportion de personnes couvertes par garantie	75
5.4	Comparaison des tarifications pour une personne de 63 ans par poste de soins . . .	76
5.5	Comparaison des tarifications pour une personne de 64 ans par poste de soins . . .	77
5.6	Phase transitoire du décalage de 62 à 64 ans	79
5.7	Primes pures annuelles calculées des personnes de 63 et 64 ans	79
5.8	Récapitulatif des expositions supplémentaires en fonction des dates de naissance . .	80
5.9	Tableau des hypothèses pour le calcul de l'impact	81
5.10	Tableau des portefeuilles estimés et des impacts	82
A.1	Les taux de cotisations de droit commun au 1er janvier 2023	91

Liste des Équations

3.1	Définition de l'exposition	20
3.2	Définition du coût moyen par prestation	24
3.3	Définition du coût moyen par personne	24
3.4	Définition de la fréquence	24
4.1	Définition de la prime commerciale	42
4.2	Définition de $\mathbb{E}[Y X]$	43
4.3	Définition de la variable Y	43
4.4	Définition d'un modèle linéaire gaussien	43
4.5	Définition du bruit gaussien	43
4.6	Vraisemblance du modèle	44
4.7	Log vraisemblance	44
4.8	Équation des moindres carrés	44
4.9	$\hat{\beta}$	44
4.10	Estimateur sans biais de la variance σ^2	44
4.11	Définition des familles exponentielles	45
4.12	Définition de l'espérance et de la variance pour Y distribuée selon une loi appartenant à une famille exponentielle	45
4.13	Définition des paramètres de la loi normale sous une famille exponentielle	45
4.14	Définition des paramètres de la loi exponentielle ou Gamma sous une famille exponentielle	45
4.15	Définition des paramètres de la loi Poisson sous une famille exponentielle	45
4.16	Définition des paramètres de la loi Bernoulli ou Binomiale sous une famille exponentielle	45
4.17	Définition des paramètres de la loi Binomiale négative sous une famille exponentielle	45
4.18	Fonction de densité d'une famille exponentielle	46
4.19	Prédicteur linéaire	46
4.20	Fonction de lien	46
4.21	Maximum de vraisemblance dans le cas du GLM	47
4.22	Log-vraisemblance dans le cas du GLM	47
4.23	Dérivée EMV 1	47
4.24	Dérivée EMV 2	47
4.25	Dérivée EMV 3	47
4.26	Dérivée EMV 4	48
4.27	Dérivée EMV 5	48
4.28	Dérivée EMV 6	48
4.29	Dérivée EMV 7	48
4.30	Dérivée EMV 8	48
4.31	Équation de laquelle $\hat{\beta}$ est solution	49
4.32	Estimateur du maximum de vraisemblance $\tilde{\beta}$	49
4.33	Définition de \tilde{Y}_i valeur ajustée du modèle saturé	49

4.34	Différence entre la vraisemblance du modèle saturé $\widetilde{\mathcal{M}}$ et du modèle estimé \mathcal{M}	49
4.35	Définition de la déviance	49
4.36	Définition de la déviance détaillée	49
4.37	Définition des déviations résiduelles	49
4.38	Définition de la contribution de l'observation i sur la déviance	49
4.39	Loi vers laquelle tend la déviance	50
4.40	Nouvelle écriture de la déviance	50
4.41	Définition de l'AIC	50
4.42	Définition du BIC	50
4.43	R^2 de McFadden	51
4.44	R^2 de Cox et Snell	51
4.45	R^2 de Nagelkerke	51
4.46	Formule R^2 de McFadden à partir de la déviance	51
4.47	Définition des résidus bruts	53
4.48	Définition des résidus standardisés	53
4.49	Définition des résidus studentisés	53
4.50	Définition du modèle log-normal	54
4.51	Fonction de lien du modèle log-normal	54
4.52	Définition du modèle Poisson	54
4.53	Fonction de lien du modèle Poisson	54
5.1	Définition du taux technique réel du logiciel de tarification	75
5.2	Définition de la prime pure réelle du logiciel de tarification	75
5.3	Définition de l'écart entre la prime pure du GLM et la prime pure du logiciel de tarification	76
5.4	Formule du montant à ajouter au portefeuille par an	80
5.5	Exemple de calcul du montant à ajouter au portefeuille par an	81
5.6	Définition du portefeuille revalorisé	82
5.7	Définition de l'impact par an	82
5.8	Exemple de calcul d'un portefeuille revalorisé en 2027	83
5.9	Exemple de calcul de l'impact en 2027	83

Bibliographie

- [1] Médecine/sciences par GARROUSTE C. et PERDRIX E. (2020). *L'effet de la retraite sur la santé.*
- [2] VERSPIEREN Courtier en assurance (2023). *L'étude menée par VERSPIEREN révèle une légère hausse du reste à charge en 2022 et mesure l'impact de la réforme des retraites sur la consommation médicale.*
- [3] BALLO H. (2021). *Mémoire d'actuariat : Impact de la réforme des retraites de 2010 sur le risque arrêt de travail et projection d'impacts.*
- [4] MAURY A. *Mémoire d'actuariat : Conséquences d'une réforme des retraites sur un portefeuille d'assurances de personnes.*
- [5] LOPEZ D. (2017). *Mémoire d'actuariat : Etude des impacts du recul de trois ans de l'âge légal de départ à la retraite sur les régimes de prévoyance collective et santé collectif.*
- [6] LIM R. *Mémoire d'actuariat : Provisionnement en arrêt de travail et impact de la réforme des retraites.*
- [7] Service Public.
- [8] Sécurité Sociale.
- [9] Ameli (Assurance maladie en ligne).
- [10] Cour des comptes (2023). *Rapport sur l'application des lois de financement de la sécurité sociale.*
- [11] DREES (Direction de la recherche des études de l'évolution et des statistiques) sous la direction de DHERBECOURT C. DIDIER M. et LEFEBRE G. (2023). *La protection sociale en France et en Europe en 2022.*
- [12] Economie.gouv (2022). *Tout savoir sur les complémentaires santé.*
- [13] DREES sous la direction de BARLET M. GAINI M. GONZALEZ L. et LEGAL R. (2019). *La complémentaire santé : acteurs, bénéficiaires, garanties - édition 2019, Fiche 6 : L'activité des différents types d'organismes de complémentaire santé.*
- [14] DREES sous la direction de BARLET M. GAINI M. GONZALEZ L. et LEGAL R. (2019). *La complémentaire santé : acteurs, bénéficiaires, garanties - édition 2019, Fiche 7 : Un marché de plus en plus concentré et structuré par des groupes.*
- [15] DREES sous la direction de LAPINTE A. POLLAK C. et SOLOTAREFF R. (2024). *La complémentaire santé : acteurs, bénéficiaires, garanties - édition 2024.*
- [16] Alan assurance (2022). *Qu'est-ce qu'un contrat collectif de mutuelle ?*
- [17] APICIL (2024). *Fonctionnaire titulaire : la mutuelle est-elle obligatoire en 2024 ?*
- [18] Le portail de la fonction publique (2022). *La réforme de la protection sociale complémentaire des agents publics.*

- [19] FIDUCIAL (2024). *Mutuelle santé : les changements attendus sur 2024*.
- [20] GAUTRON N. (2023). Cours d'actuariat retraite, ISUP.
- [21] Vie publique par ELETA-DE FILIPPIS R. Maitre de conférences à l'université du Havre (2023). *Système de retraite : comment font les autres pays de l'UE ?*
- [22] Éditions OCDE (2018). *Taux de remplacement bruts*, dans Pensions at a Glance 2017 : OECD and G20 indicators.
- [23] Le cercle de l'épargne (2023). *Report de l'âge de la retraite, santé, prévoyance et chômage*.
- [24] Malakoff Humanis. *La prise en charge des chambres individuelles par la mutuelle et la CPAM*.
- [25] AG2R La Mondiale. *Chambre individuelle à l'hôpital : quelle prise en charge ?*
- [26] THOMAS M. (2023). Cours d'assurance non-vie, ISUP.
- [27] Institut des Actuaires (2017). *Tarifcation : Prime pure / Prime commerciale*.
- [28] Le comparateur assurance par NAIMI JC. (2023). *La réforme des retraites pourrait renchérir le coût des contrats d'assurance santé et de prévoyance*.