



Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du
Diplôme d'Actuaire EURIA
et de l'admission à l'Institut des Actuaire

le 6 septembre 2023

Par : Alexis LAM

Titre : *L'Open Data* dans la tarification des mutuelles communales

Confidentialité : Non

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

***Membres présents du jury de l'Institut
des Actuaire :***

Romain LAILY

Arthur MAILLART

Saman NEJAD

Signature :

Membre présent du jury de l'EURIA :

Jérémy ROUOT

Signature :

Entreprise :

Actuelia

Signature :

Directeur de mémoire en entreprise :

Frank BOUKOBZA

Signature :

Invité :

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 :

REMERCIEMENTS

Pour débiter, je tiens à exprimer ma gratitude envers David Fitouchi, Louis-Anselme De Lamaze et Frank Boukobza, associés du cabinet Actuelia, pour m'avoir chaleureusement accueilli au sein de leur équipe. Une mention particulière revient à Frank Boukobza, qui a également été l'un de mes encadrants pour ce mémoire. Sa disponibilité constante pour répondre à toutes mes questions et ses précieux conseils m'ont permis d'améliorer mon mémoire.

J'adresse mes remerciements les plus sincères à Sandrine HUYNH, consultante senior au sein du cabinet Actuelia et tutrice en entreprise. Sa précieuse implication, même au sein d'un emploi du temps chargé, son dévouement lors de ses congés et même jusqu'aux heures tardives de la nuit, ont été d'un très grand soutien. Ses nombreuses relectures et ses commentaires constructifs m'ont guidé tout au long de la rédaction de ce mémoire.

Je souhaite également remercier Suzanne Noll, manager au sein du cabinet EY et tutrice académique. Sa disponibilité et sa rigueur dans l'encadrement de ce projet ont été d'une grande importance. Je tiens également à la remercier pour ses relectures attentives de mon mémoire ainsi que pour les éclairages qu'elle m'a apportés. De plus, elle n'a pas hésité à solliciter ses contacts lorsque j'avais des questions particulières sur le sujet du mémoire, ce qui m'a été d'une grande aide.

Merci à l'ensemble des collaborateurs actuels et anciens d'Actuelia pour leur accueil chaleureux et leur bienveillance. Ma gratitude s'étend également à tous ceux qui ont contribué de près comme de loin à l'avancement de ce mémoire.

Enfin, mes remerciements se tournent vers l'équipe pédagogique de l'EURIA et de l'ESILV pour la qualité des enseignements délivrés tout au long de mon parcours académique.

RÉSUMÉ

Une mutuelle communale est un partenariat entre une commune et un organisme d'assurance. Son objectif est de fournir une couverture santé à tarif préférentiel pour l'ensemble des résidents de ladite commune. Ce mémoire examine la perspective d'un assureur opérant sur ce marché de niche. Cependant, il est important de noter que la plupart des souscripteurs de ce type d'offre sont en réalité des seniors âgés de plus de 65 ans. Cette composition démographique engendre un risque de sélection adverse, soulevant ainsi des questions quant à l'équilibre technique de l'organisme assureur. Face à la difficulté de la mutualisation au niveau communal, l'assureur entreprend une évaluation visant à déterminer si une expansion de l'offre à une échelle régionale serait avantageuse. Les résultats permettront à l'assureur de décider s'il est préférable d'approcher les élus locaux de la région dans son ensemble ou de cibler chaque commune individuellement dans le but d'établir de nouveaux partenariats.

Ainsi, l'objectif principal du mémoire est la comparaison entre la tarification d'une offre de mutuelle communale et d'une offre à une échelle plus large. Les travaux du mémoire reposent sur une base de données intitulée l'*Open DAMIR*, datant de 2019. Elle présente la particularité d'avoir des lignes d'informations agrégées telles que les dépenses de santé disponibles uniquement par tranche de région. Il faut dans un premier temps, réussir à ventiler l'information contenue dans la base à une maille plus fine puis mettre à jour les données (inflation, impact de la réforme 100% Santé).

Enfin, dans le cas où la décision de démarcher les élus d'une région est prise, il est important de prendre en compte l'existence préalable d'une mutuelle communale au sein de la région. Ainsi, la concurrence entre une mutuelle communale et une mutuelle régionale sera modélisée à l'aide d'un marché concurrentiel fictif, mettant en évidence les parts de marché potentielles accessibles grâce à ce type de partenariat.

Mots clefs: Tarification, Santé, Mutuelle communale, Mutualisation, Antisélection, Modèles linéaires généralisés, Assurance maladie, *Open DAMIR*, *Open Data*, INSEE

ABSTRACT

A municipal mutual insurance involves a partnership between a local municipality and an insurance entity, aiming to offer preferential health coverage to all residents of the specific municipality. This paper examines the perspective of an insurer operating in this niche market. However, it is important to note that most subscribers to this type of offer are actually senior citizens aged over the age of 65. This demographic composition raises the risk of adverse selection, thereby raising questions about the technical balance of the insurance organization. Faced with the challenge of mutualization at the municipal level, the insurer undertakes an evaluation to determine whether expanding the offering on a larger scale would be advantageous. The results will enable the insurer to decide whether it is preferable to approach local officials of the entire region or target each municipality individually to establish new partnerships.

Thus, the primary objective of this thesis is the comparison between the pricing of a municipal mutual insurance and that of a wider-ranging option. The thesis work relies on a database known as *Open DAMIR*, published 2019. It has the unique feature of having aggregated information lines such as healthcare expenses available only by regional category. In the initial stage, it is necessary to disaggregate the information in the database into a finer granularity and then update the data (inflation, impact of the 100% Health reform).

Finally, in the case where the decision to approach officials of a region is made, it is important to consider the prior existence of a municipal mutual insurance within the region. Thus, the competition between a municipal mutual insurance and a regional one will be modeled using a fictional competitive market, highlighting the potential market shares accessible through this type of partnership.

Keywords: Pricing, Health, Municipality mutual insurance company, Mutualisation, Antiselection, Generalized linear model, Health insurance, *Open DAMIR*, *Open Data*, INSEE

NOTE DE SYNTHÈSE

Contexte et problématique

Une mutuelle communale naît de la collaboration entre une commune et un organisme d'assurance, avec l'objectif de fournir des couvertures santé à tarif avantageux à tous les habitants de cette commune. Ce mémoire examine la perspective d'un assureur opérant sur le marché des mutuelles communales. Cependant, la majorité des souscripteurs sont des seniors de plus de 65 ans, créant un risque de sélection adverse et soulevant des inquiétudes concernant l'équilibre technique de l'assureur.

Face aux difficultés inhérentes à la mutualisation des risques à l'échelle communale, l'intuition de l'assureur le pousse à explorer la possibilité d'une mutualisation au niveau régional, jugée potentiellement plus viable selon le principe de mutualisation. Par conséquent, une étude approfondie est entreprise pour déterminer la pertinence de maintenir l'échelle communale par rapport à une démarche commerciale ciblant les élus locaux des régions.

Après avoir exposé les fondements des mutuelles communales, notamment les conditions de leur mise en place et les aspects juridiques, le processus de construction de la base de données à partir d'*Open Data* a été présenté en détail. Ensuite, trois méthodes de tarification ont été introduites, accompagnées de justifications pour leur sélection. En dernier lieu, un critère a été défini pour évaluer la pertinence d'une mutuelle régionale dans les régions étudiées. Si la décision d'approcher les élus d'une région est prise, il est important de prendre en compte la présence préalable d'une mutuelle communale dans la même région. Ainsi, une simulation concurrentielle fictive sera réalisée afin d'évaluer la concurrence entre ces deux types de mutuelles. Cette simulation mettra en évidence les parts de marché potentielles accessibles grâce à ce type de partenariat.

La création de la base de données

Dans un premier temps, la constitution de la base de données utilisée pour la tarification s'appuie sur les jeux de données *Open DAMIR* (Dépenses d'Assurance Maladie Inter-Régimes) des années 2018 et 2019. Ces données ont été fusionnées et ont fait l'objet de plusieurs étapes de retraitement au sein de l'entreprise dans le cadre d'un précédent mémoire. Néanmoins, la base fournie était une version de sauvegarde intermédiaire. Ainsi, des vérifications ont été entreprises quant à la qualité des données de la base, ce qui a entraîné la mise en place de traitements pour les montants de remboursements de la Sécurité Sociale qui étaient incomplets, nécessitant ainsi leur recalcul. La base de données se caractérise par la présence de données agrégées, notamment les dépenses de santé regroupées par région et par catégorie de soins (soins courants, dentaire, optique, etc.). Il fallait donc répartir ces informations en une granularité plus fine, correspondant à celle des départements pour espérer une tarification à ce niveau d'agrégation. Ce choix a été motivé par l'absence de données au niveau communal. En conséquence, dans le cadre de ce mémoire, la notion de commune a été assimilée à celle de département. Cela se traduit par une ventilation des indicateurs tels que les dépenses de santé, les quantités d'actes, et d'autres, au sein des différents départements de la France.

À cet effet, deux méthodes ont été mises en œuvre. Pour la méthode principale, des coefficients de ventilation sont déduits à partir d'une autre *Open Data* en santé, ce qui permet de prendre en compte les tendances de consommation propres à chaque département. Néanmoins, la source externe employée ne possède pas la même exhaustivité que la base *Open DAMIR* et omet certains postes de soins. Par conséquent, pour ces postes omis, une seconde approche est adoptée : les indicateurs sont distribués en fonction de la population de chaque département. En d'autres termes, la pondération des indicateurs régionaux dépendra de la taille démographique respective de chaque département.

Une fois que les indicateurs ont été répartis, ils ont été ajustés pour prendre en compte l'effet de l'inflation et l'impact de la réforme 100% Santé. De plus, l'hypothèse a été formulée selon laquelle le portefeuille de l'assureur englobait la totalité de la population française. En conséquence, la variable d'exposition, représentant le nombre de bénéficiaires, a été intégrée en calculant le nombre d'habitants par tranche d'âge et par département, en croisant les données avec une base de l'INSEE. Par la suite, les montants de remboursements correspondant à deux niveaux de garantie d'une mutuelle communale ont été calculés au sein de la base de données.

En dernier lieu, cette base de données a été enrichie par l'intégration de variables externes dans le but d'identifier des tendances de consommation au sein des différents départements. Par exemple, une analyse a été menée pour déterminer si une disponibilité accrue en termes d'établissements de santé était associée à une augmentation de la fréquence des consultations médicales. Cependant, les études réalisées n'ont pas permis d'établir une relation précise quant à l'impact de ces variables externes.

Statistiques de la nouvelle base de données

Des analyses statistiques ont été effectuées pour s'appropriier les données. L'objectif était d'identifier de nouvelles anomalies potentielles résultant de la ventilation des données, laquelle présente certaines limitations. Dans la sous-catégorie d'actes "Forfaits journaliers", des montants de dépenses unitaires de 1800 € ont été enregistrés en Haute-Corse, et de 1200 € en Corse-du-Sud, pour la classe d'âge des 0-19 ans. Toutefois, ces montants ne sont pas en accord avec le forfait journalier hospitalier standard de 20 € par jour d'hospitalisation. Cela suggérerait qu'en moyenne, un résident de la Corse âgé entre 0 et 19 ans passerait trois mois à l'hôpital par an. Cependant, même pour les nourrissons faisant face à des problèmes de santé, ce scénario ne semble pas être suffisamment récurrent pour influencer les montants de cette tranche d'âge. Compte tenu de ces anomalies et du fait que d'autres sous-familles d'actes montrent également des montants excessivement élevés dans ces deux départements, la décision a été prise d'exclure la Haute-Corse et la Corse-du-Sud de l'étude. Par conséquent, les tarifs seront établis pour l'ensemble de la France continentale.

Une cartographie a été réalisée dans le but d'étudier la possible corrélation entre les montants annuels de dépenses de santé des Français par département et la disponibilité de l'offre de soins. D'autres tests ont été réalisés, mais seule la disponibilité des soins est présentée à titre d'exemple.

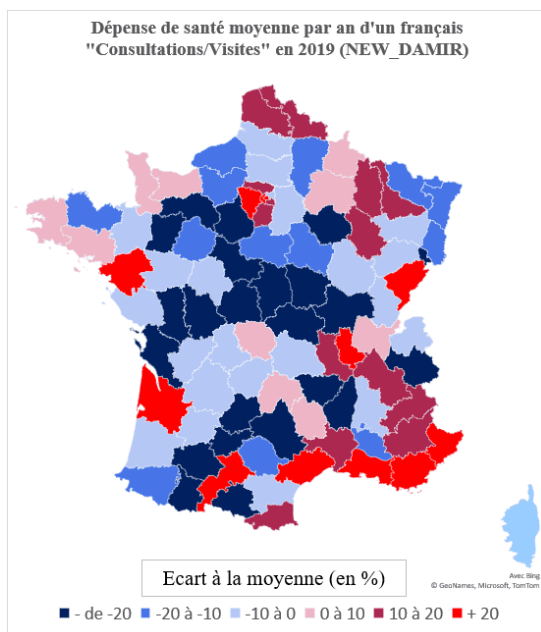


FIGURE 1 – Dépense annuelle moyenne des "Consultations/Visites"

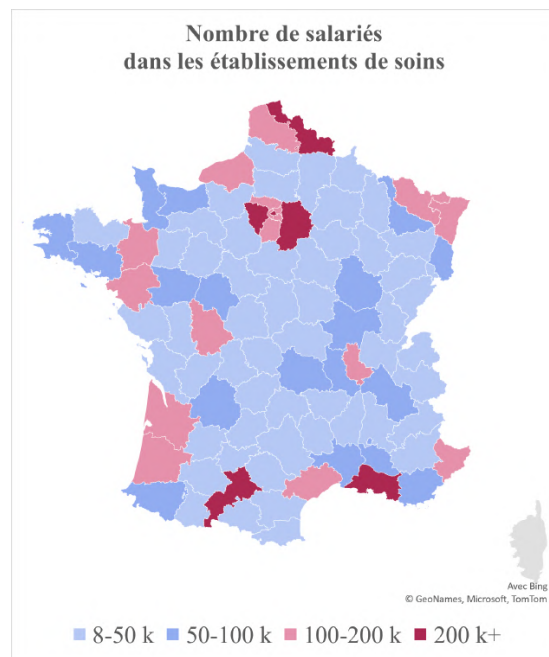


FIGURE 2 – Nombre de salariés dans les établissements de santé par département

Les régions d'Ile-de-France, de Provence-Alpes-Côte d'Azur ainsi que du nord affichent une consommation plus élevée dans la sous-catégorie de soins "Consultations/Visites". Cette observation suggère une tendance où les dépenses de santé augmentent en corrélation avec la disponibilité de l'offre de soins dans chaque département, ce qui pourrait impliquer qu'une hausse des montants remboursés par les complémentaires santé soit également présente. Cependant, cette tendance ne trouve pas de confirmation dans les analyses univariées ultérieures du mémoire.

La mise en place de la tarification

Dans le domaine de la santé il est possible que selon le niveau de garantie, la fréquence ne soit pas totalement indépendante du coût. En effet, il s'agit de la problématique de l'aléa moral, un patient lorsqu'il sait qu'il est bien remboursé par sa complémentaire santé peut avoir plus tendance à se rendre chez le médecin. C'est la raison pour laquelle ce mémoire adopte une approche de tarification basée sur deux niveaux de garantie (V1 : entrée de gamme et V2 : moyenne gamme), afin de légitimer l'utilisation de l'approche fréquence \times coût. Dans le cadre de ce mémoire, il est admis qu'il puisse exister une légère corrélation entre la fréquence et le coût, toutefois, celle-ci ne doit pas dépasser un certain seuil d'acceptabilité.

Ainsi la base de données a été subdivisée en plusieurs bases distinctes, de manière à ne contenir que des segments tarifaires homogènes, permettant de calculer des "sous-primés". Chaque base est ensuite soumise à un triple test de corrélation afin de sélectionner la méthode de tarification la plus appropriée parmi les trois méthodes mises en place dans ce mémoire. Si une absence de corrélation entre la fréquence et les coûts est observée, la préférence est donnée à l'utilisation d'un modèle linéaire généralisé (GLM) appliqué à la variable de fréquence ainsi qu'à la variable de coût. En cas de corrélation, un autre modèle linéaire généralisé appliqué directement à la prime pure est utilisé. Enfin, lorsque les deux premières approches ne se sont pas avérées concluantes ou adaptées aux données, une méthode de calcul direct a été mise en place en se basant sur les statistiques extraites des données. Ainsi, toutes ces "sous-primés" ont été additionnées afin d'obtenir le tarif global du département ou de la région.

En cas d'utilisation des GLM, les vérifications usuelles ont été réalisées, notamment en ce qui concerne l'adéquation des lois mathématiques et la cohérence du modèle, laquelle a été évaluée au moyen d'un test de déviance. Par ailleurs, la sélection des variables par la méthode *AIC forward* n'a pas conduit à la conservation des variables externes lors du calcul du tarif. Cette exclusion est due à leur faible significativité et à la présence de multicollinéarité par rapport aux variables d'âge et de localisation. Un exemple d'une "sous-prime" est présenté pour la sous-famille "Actes d'imagerie et de radiologie".

Variables explicatives	Entité	Base test		
		Réel	Prédit	Écart
CLASSE_AGE + DPT	Fréquence moyenne (Unitaire)	1,95	1,97	0,84%
	Coût moyen unitaire (Garantie V1)	15,11	15,68	3,79%
CLASSE_AGE + REGION	Fréquence moyenne (Unitaire)	1,95	2,03	3,96%
	Coût moyen unitaire (Garantie V1)	15,11	15,46	2,31%

FIGURE 3 – Résultats modélisation par 2 GLM (Actes d'imagerie et de radiologie) Garantie V1

Ainsi, pour cette sous-famille, il est observé que le modèle semble prédire à la hausse la fréquence et le coût, mais que les écarts relatifs ne dépassent pas 4 % en valeur absolue. Cet écart a été considéré comme satisfaisant en sollicitant l'avis d'expert.

Les résultats de la tarification

Les tarifs sont présentés sous forme de primes pures mensuelles pour la couverture globale. Elles correspondent à la somme totale des sous-primes calculées pour chaque sous-famille. La figure ci-dessous affiche donc tarif par région et par classe d'âge de la garantie entrée gamme V1.

GARANTIE V1	CLASSE AGE							
REGION	0	20	30	40	50	60	70	80
Île-de-France	17,5	21,2	26,8	31,1	41,1	54,8	68,5	104,9
Centre-Val de Loire	14,1	16,2	19,9	23,8	32,8	42,8	58,4	88,0
Bourgogne-Franche-Comté	13,2	16,3	20,0	23,7	31,2	42,1	58,2	87,1
Normandie	14,4	16,5	19,7	23,7	32,7	41,6	57,5	81,7
Hauts-de-France	14,7	16,2	20,8	25,7	34,6	44,7	64,1	101,2
Grand Est	15,0	17,6	23,1	26,7	36,0	48,1	57,4	91,2
Pays de la Loire	14,9	17,1	21,2	25,4	32,6	41,2	63,1	84,0
Bretagne	14,8	16,5	21,6	24,3	32,9	42,9	58,0	88,8
Nouvelle-Aquitaine	13,7	16,6	22,8	25,9	34,0	43,6	58,6	91,3
Occitanie	16,4	20,7	25,0	28,8	38,3	49,0	65,7	112,0
Auvergne-Rhône-Alpes	14,3	17,5	23,1	26,7	34,9	44,7	60,7	89,8
Provence-Alpes-Côte d'Azur	18,5	23,6	27,2	31,7	44,2	57,4	76,6	134,3

FIGURE 4 – Primes pures mensuelles Garantie V1

En analysant les tarifs obtenus par régions, deux effets principaux se manifestent. Tout d'abord, le tarif connaît une augmentation significative en fonction de l'âge. De plus, un effet lié à la région est observé, bien que son impact ne soit pas considérable. En d'autres termes, les tarifs varient d'une région à l'autre, mais cette variation n'est pas excessive. Cette situation peut être attribuée au niveau de la garantie (entrée de gamme), où les plafonds de remboursement sont atteints, restreignant ainsi les possibilités de remboursement par la complémentaire santé et, par conséquent, limitant la variation des tarifs. L'analyse des tarifs de la garantie supérieure a révélé des écarts plus significatifs entre les régions, suggérant que l'utilisation de plafonds de garantie plus élevés a permis de mettre en lumière de manière plus précise les tendances de consommation.

Pour l'analyse des tarifs départementaux, l'approche s'est concentrée sur la comparaison d'une région avec les départements qui la composent. Par exemple, les tarifs des départements au sein de la région Île-de-France, principalement constituée de zones urbaines, ont affiché une similitude pour les deux niveaux de garantie. En revanche, dans la région Nouvelle-Aquitaine, qui englobe à la fois des zones urbaines et rurales, les tarifs présentent des écarts plus marqués, lesquels s'accroissent avec l'augmentation du niveau de garantie.

La comparaison entre le tarif d'une région et les tarifs des départements qui la composent a permis d'établir un critère de pertinence pour la mise en place d'une mutuelle régionale. Ainsi, le critère permettant de décider s'il faut solliciter les élus de la région consistera à examiner la répartition des départements présentant des tarifs supérieurs et inférieurs à ceux de la région. Par conséquent, si plus de la moitié des départements affichent des tarifs plus élevés que celui de la région, il sera considéré comme opportun d'initier une démarche auprès des élus de la région.

Toutefois, quelques résultats surprenants ont été observés. Pour les mettre en évidence, les départements ont été classés en catégories en fonction de l'écart de leur prime pure par rapport à la moyenne. Les tendances sont présentées à travers des cartographies pour la classe d'âge 60-69 ans, avec une prime pure moyenne de 46,39 € pour la garantie V1 et 62,66 € pour la garantie V2.

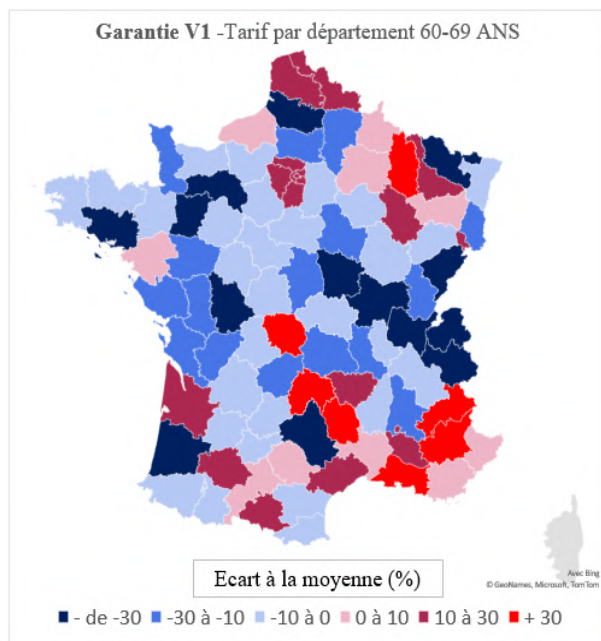


FIGURE 5 – Tarif par département (V1)

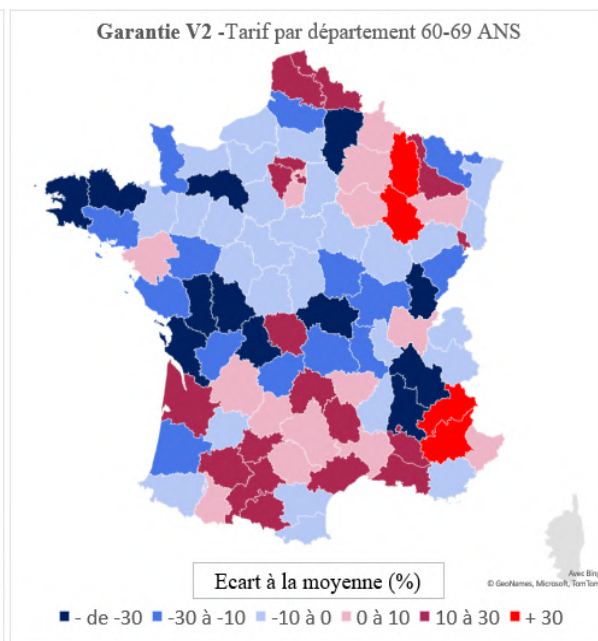


FIGURE 6 – Tarif par département (V2)

Les tarifs des départements d'Île-de-France ainsi que de la région Provence-Alpes-Côte-d'Azur se révèlent supérieurs à la moyenne, en accord avec leurs tendances de consommation également supérieures à la moyenne. Néanmoins, pour les deux niveaux de garanties, des pics tarifaires se manifestent dans les départements marqués en rouge, indiquant un écart par rapport à la moyenne de plus de 30 %. Tous ces départements se caractérisent par une exposition très limitée en termes de bénéficiaires, c'est-à-dire une faible population dans la tranche d'âge 60-69 ans. De plus, il semble que les départements éligibles au régime de l'Alsace-Moselle semblent être sous-tarifés. Ceci peut s'expliquer par le fait que les données dans la base pour ces départements sont celles du régime de la sécurité sociale. Cependant, sur ce territoire, il existe un régime local auquel une partie des habitants adhèrent. Néanmoins, l'exposition suppose que l'intégralité de la population adhère au régime de la sécurité sociale, créant ainsi une disparité. Il aurait fallu prendre en compte cette différence d'exposition lors de l'étape de tarification.

Le marché concurrentiel

Dans la réalité, l'introduction d'une mutuelle régionale ne signifierait pas la suppression des mutuelles communales ou départementales déjà existantes. Ainsi, en se basant sur le critère défini précédemment, il a été décidé que la région Île-de-France semble favorable à la mise en place d'une mutuelle régionale, tandis que la région Nouvelle-Aquitaine ne l'est pas. Ce postulat a donc été mis à l'épreuve grâce à la simulation d'un marché concurrentiel fictif qui oppose un assureur cherchant à mutualiser ses risques au niveau départemental à un autre optant pour la

mutualisation régionale. Une base de données représentant une population fictive a été créée pour les deux régions testées, sa composition démographique ayant pour objectif de se rapprocher de celle généralement intéressée par des mutuelles communales. Cette composition inclut 60 % de seniors âgés de plus de 60 ans. Il est supposé que les individus opteront pour une assurance auprès de l'assureur offrant le tarif le moins cher correspondant à leur profil. Il est supposé la marge de frais est neutre par rapport aux frais que l'assureur pourrait rencontrer. Ainsi, seule une marge de bénéfice de 3 % a été choisie et sera appliquée aux primes pures afin de constituer une prime commerciale simplifiée. Les résultats sont présentés ci-dessous :

Garantie V1 Nouvelle-Aquitaine	3 % de marge	
	Résultat	Part de marché
Assureur départemental	12 M€	60%
Assureur régional	-35 M€	40%
Garantie V2 Nouvelle-Aquitaine	3 % de marge	
	Résultat	Part de marché
Assureur départemental	14 M€	60%
Assureur régional	-45 M€	40%

FIGURE 7 – Résultats du marché concurrentiel pour la Nouvelle-Aquitaine

Garantie V1 Île-de-France	3 % de marge	
	Résultat	Part de marché
Assureur départemental	8 M€	47%
Assureur régional	-8 M€	53%
Garantie V2 Île-de-France	3 % de marge	
	Résultat	Part de marché
Assureur départemental	11 M€	47%
Assureur régional	-15 M€	53%

FIGURE 8 – Résultats du marché concurrentiel pour l'Île-de-France

Pour la région Nouvelle-Aquitaine, il est observé que la mutuelle régionale présente des défaillances par rapport à la mutuelle départementale pour les deux niveaux de garanties. Ces défaillances sont attribuées à l'apparition d'un problème d'antisélection, ce qui se traduit par l'obtention d'un résultat de -35 M€ pour la garantie V1 et -45 M€ pour la garantie V2. En effet, l'adhésion à l'offre régionale n'a pas été souscrite par les habitants des départements où les tarifs étaient inférieurs à ceux de la région. Simultanément, une sous-tarifcation de l'offre a été constatée pour les autres départements, ce qui a entraîné l'attraction des résidents de ces zones.

Cependant, pour la région Île-de-France, il n'était pas anticipé que l'assureur régional afficherait également un résultat négatif. En effet, dans la réalité, la région Île-de-France a effectivement mis en place une mutuelle régionale qui affiche une bonne santé financière, selon les professionnels travaillant au sein des organismes assureurs. Cette différence peut s'expliquer par le fait que dans le cadre de ce mémoire, les adhérents optent automatiquement pour l'assureur offrant les tarifs les plus bas. Cependant, dans la réalité, il est envisageable que certaines personnes ne prennent pas nécessairement le temps de comparer les tarifs, tandis que d'autres ont davantage confiance en une entité régionale plutôt qu'en une entité départementale. Il est également possible que certaines personnes ne soient informées que de l'existence de la mutuelle régionale, sans connaître celle de la mutuelle communale/départementale.

Afin de s'aligner davantage avec la réalité, il serait nécessaire de prendre en compte, pour chaque ligne de la base fictive, un groupe d'individus susceptibles de ne pas opter pour le prix le plus bas.

Conclusion

Les résultats de cette étude ont montré la possibilité de détailler les informations extraites de la base de données *Open DAMIR* à une échelle plus fine. Malgré certaines limitations, les tarifs établis ont démontré une cohérence relative avec les tendances observées dans le domaine de l'assurance santé.

Pour approfondir ces recherches, il serait opportun de partir de la version la plus récente de la base de données *Open DAMIR*. Cette mise à jour plus récente pourrait offrir une meilleure saisie des effets du 100 % Santé, pris en compte uniquement pour les coûts dans cette étude, mais non pour la fréquence.

Concernant les tarifs, des écarts significatifs ont été observés entre les tranches d'âge seniors. L'application de techniques de lissage pourrait être pertinente, soit au niveau départemental, en fonction de l'âge, voire les deux simultanément.

Les variables externes n'ont pas été intégrées dans les modèles de tarification en raison de leur faible signification par rapport aux autres variables explicatives, à savoir l'âge et la localisation. Une alternative serait d'exclure la variable de localisation des modèles et de baser la tarification exclusivement sur l'âge et les variables externes. De plus, la création d'un zonier basé sur les primes pures obtenues dans cette étude pourrait être envisagée.

Ainsi, les mutuelles communales, bien qu'ayant gagné en popularité ces dernières années, continuent de représenter un marché de niche. Cette forme de partenariat entre les collectivités locales et les assureurs demeure largement méconnue du grand public. De plus, malgré le ton pessimiste des conclusions de ce mémoire quant à l'avenir des mutuelles régionales, il a été montré qu'il existait des écarts possibles entre les comportements des résidents réels et le contexte spécifique examiné ici. Par conséquent, même si cela n'a pas été observé dans ce mémoire, il reste envisageable de considérer les mutuelles régionales comme des opportunités commerciales tout en améliorant l'accès aux soins pour les populations vulnérables.

La région Île-de-France illustre cette dynamique avec la création de la première mutuelle régionale en France, suscitant un intérêt croissant parmi les jeunes actifs d'après les témoignages des professionnels de ces organismes. Cette évolution suggère la possibilité d'une adoption plus étendue de ces offres, offrant potentiellement une voie encourageante pour d'autres régions à suivre cet exemple.

SYNTHESIS NOTE

Context and Issue

A municipal mutual insurance arises from the collaboration between a municipality and an insurance organization, with the objective of providing affordable health coverage to all the inhabitants of that municipality. This paper examines the perspective of an insurer operating in the municipal mutuals market. However, the majority of policyholders are seniors over 65 years old, creating a risk of adverse selection and raising concerns about the insurer's technical balance.

Given the challenges linked to risk sharing at the municipal level, the insurer's intuition compels them to investigate the potential for risk sharing at the departmental or regional level. Consequently, a study is being conducted to determine the relevance of maintaining the municipal scale compared to a business approach targeting local officials from regions.

After outlining the foundations of municipal mutual insurance, including the conditions of their establishment and legal aspects, the process of building the database from Open Data was presented in detail. Subsequently, three pricing methods were introduced, along with justifications for their selection. Finally, a criterion was defined to assess the relevance of a regional mutual insurance scheme in the studied areas. If the decision to approach officials of a region is made, it's important to consider the prior presence of a municipal mutual insurance in the same region. Thus, a fictitious competitive simulation will be carried out to evaluate the competition between these two types of mutual insurance. This simulation will highlight the potential market shares accessible through this type of partnership.

The creation of the database

First and foremost, the establishment of the database used for pricing in this study relies on the Open DAMIR datasets from the years 2018 and 2019. These datasets were merged and underwent multiple reprocessing steps within the company as part of a previous thesis. However, the provided database was an intermediate backup version; therefore, checks were carried out regarding the data quality in the database. This led to the implementation of processes to address incomplete Social Security reimbursement amounts, necessitating their recalculation. As a result, the database is characterized by the presence of aggregated data, specifically healthcare expenses grouped by region and by category of care (general care, dental, optical, etc.). Consequently, it was necessary to disaggregate this information to a finer granularity, corresponding to the level of departments, in order to achieve pricing at this level of aggregation. This choice was motivated by the absence of data at the municipal level. As a result, within the scope of this study, the concept of a commune has been equated with that of a department. This translates to a breakdown of indicators such as healthcare expenses, quantities of procedures, and others, across the various departments of France.

With this objective in mind, two methodologies have been implemented. For the primary approach, distribution coefficients are extrapolated from another health-related Open Data source, which enables the incorporation of consumption trends specific to each individual department. Nonetheless, the external source employed does not have the same amount of information as the Open DAMIR database and omits certain care categories. Consequently, for these omitted categories, a secondary approach is embraced : indicators are allocated based on each department's population. As a result, the weighting of regional indicators will depend on the respective demographic size of each department.

Once the indicators were allocated, they were adjusted to account for the effect of inflation and the impact of the 100% Health reform. Additionally, the assumption was made that the insurer's portfolio encompassed the entire French population. As a result, the exposure variable, representing the number of beneficiaries, was incorporated by calculating the population count per age group and department, cross-referencing the data with an INSEE database. Subsequently, the reimbursement amounts corresponding to two levels of insurance coverage for a municipal mutual insurance were computed within the database.

Lastly, this database was enhanced through the incorporation of external variables aimed at identifying consumption trends within various departments. For instance, an analysis was conducted to ascertain whether increased availability of healthcare facilities was linked to a rise in the frequency of medical consultations. However, the conducted studies did not highlight a precise relationship regarding the impact of these external variables.

Statistics from the new database

Statistical analyses were conducted to scrutinize all categories of medical procedures. The aim was to uncover potential new anomalies stemming from data allocation, which comes with certain limitations. In the subcategory of "Forfait journaliers", unit expenditure amounts of 1800 € were recorded in Haute-Corse, and 1200 € in Corse-du-Sud, for the age group 0-19 years. However, these amounts do not align with the standard hospital daily package fee of 20 € per day of hospitalization. This would suggest that on average, a resident of Corsica aged 0-19 would spend three months in the hospital per year. Nonetheless, even for infants facing health issues, this scenario doesn't seem to be recurrent enough to significantly influence the expenditure in this age group. Given these anomalies and the fact that other subcategories of procedures also exhibit excessively high amounts in these two departments, the decision was made to exclude Haute-Corse and Corse-du-Sud from the study. Consequently, the pricing will be established for the entirety of metropolitan France.

A mapping was carried out to investigate the potential correlation between the annual healthcare expenditure amounts of French citizens by the department and the availability of healthcare services.

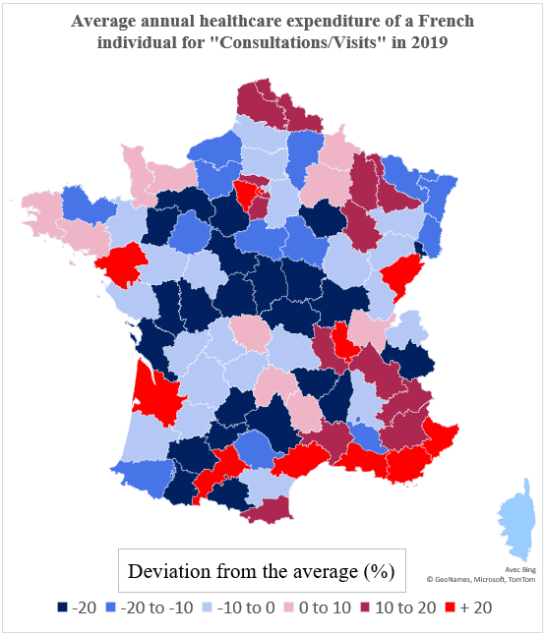


FIGURE 9 – Average annual expenditure for "Consultations/Visits"

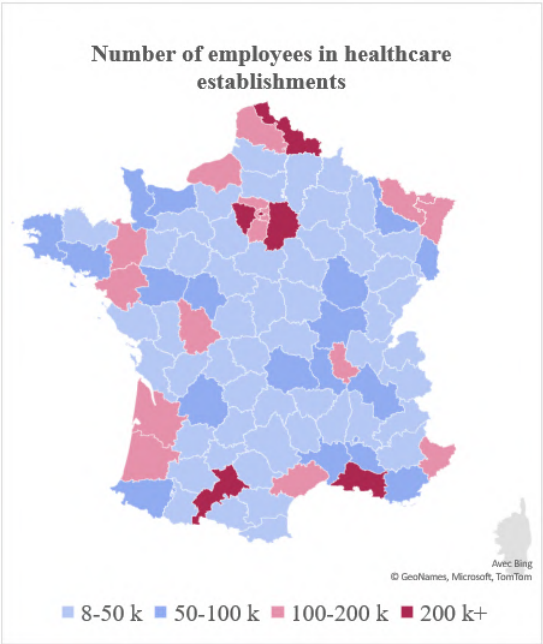


FIGURE 10 – Number of employees in healthcare establishments

The Île-de-France, Provence-Alpes-Côte d'Azur, and northern regions exhibit higher consumption in the "Consultations/Visits" sub-category of healthcare. This observation suggests a trend where healthcare expenses increase in correlation with the availability of healthcare services in each department, which could imply that higher amounts reimbursed by supplementary health insurance might also be present. However, this trend is not confirmed by subsequent univariate analyses of the study.

The establishment of the pricing

In the realm of healthcare, it's possible that, based on the level of insurance coverage, frequency might not be entirely independent of the cost. This is indeed the issue of moral hazard, where a patient who knows they are well reimbursed by their supplementary health insurance might be more inclined to visit the doctor. This is why this study adopts a pricing approach based on two levels of coverage (V1 : basic and V2 : mid-range), in order to justify the utilization of the frequency * cost approach. Within the scope of this study, it is acknowledged that there might exist a slight correlation between frequency and cost ; however, this correlation should not surpass a certain threshold of acceptability.

Thus, the database was subdivided into several distinct subsets, each containing homogeneous rate segments, facilitating the calculation of "sub-premiums." Each subset then underwent a triple correlation test to select the most appropriate pricing method among the three methods established in this study. In case of an absence of correlation between frequency and costs, preference was given to employing a Generalized Linear Model (GLM) applied to both the frequency and cost variables. If a correlation was observed, an alternative Generalized Linear Model applied directly to the pure premium was utilized. Lastly, when the first two approaches were not proven to be conclusive or suitable for the data, a direct calculation method was implemented based on statistics extracted from the data. Consequently, all these "sub-premiums" were summed to obtain the overall premium for the department or region.

In the case of using GLMs, conventional checks were performed, particularly concerning the appropriateness of mathematical laws and model coherence, which was evaluated through a deviance test. Unfortunately, the variable selection process using the forward AIC method did not retain the external variables in the premium calculation, owing to their low significance compared to age and location variables. An example of a "sub-premium" is presented in the sub-category "Imaging and Radiology Procedures".

Explanatory variables	Entité	Base test		
		Real	Predict	Difference
AGE_CLASS + DPT	Average frequency (Unit)	1,95	1,97	0,84%
	Average unit cost (V1 coverage)	15,11	15,68	3,79%
AGE_CLASS + REGION	Average frequency (Unit)	1,95	2,03	3,96%
	Average unit cost (V1 coverage)	15,11	15,46	2,31%

FIGURE 11 – Results of modeling using 2 GLMs (Imaging and Radiology Procedures) Coverage V1

Thus, in this sub-category, it's observed that the model appears to predict an increase in both frequency and cost, with relative differences not exceeding 4% in absolute value. This difference has been deemed satisfactory by seeking expert opinion.

The results of the pricing

The premiums are presented as pure monthly premiums for comprehensive coverage, which means it is the total sum of sub-premiums calculated for each sub-category. The figure below displays the premium per region and per age group for the entry-level coverage V1.

COVERAGE V1	AGE							
REGION	0	20	30	40	50	60	70	80
Île-de-France	17,5	21,2	26,8	31,1	41,1	54,8	68,5	104,9
Centre-Val de Loire	14,1	16,2	19,9	23,8	32,8	42,8	58,4	88,0
Bourgogne-Franche-Comté	13,2	16,3	20,0	23,7	31,2	42,1	58,2	87,1
Normandie	14,4	16,5	19,7	23,7	32,7	41,6	57,5	81,7
Hauts-de-France	14,7	16,2	20,8	25,7	34,6	44,7	64,1	101,2
Grand Est	15,0	17,6	23,1	26,7	36,0	48,1	57,4	91,2
Pays de la Loire	14,9	17,1	21,2	25,4	32,6	41,2	63,1	84,0
Bretagne	14,8	16,5	21,6	24,3	32,9	42,9	58,0	88,8
Nouvelle-Aquitaine	13,7	16,6	22,8	25,9	34,0	43,6	58,6	91,3
Occitanie	16,4	20,7	25,0	28,8	38,3	49,0	65,7	112,0
Auvergne-Rhône-Alpes	14,3	17,5	23,1	26,7	34,9	44,7	60,7	89,8
Provence-Alpes-Côte d'Azur	18,5	23,6	27,2	31,7	44,2	57,4	76,6	134,3

FIGURE 12 – Pure monthly premiums Coverage V1

By analyzing the premiums obtained by regions, two main effects become apparent. Firstly, the premium experiences a significant increase based on age. Additionally, a regional effect is observed, although its impact is not considerable. In other words, premiums vary from one region to another, but this variation is not excessive. This situation can be attributed to the level of coverage (entry-level), where reimbursement limits are reached, thus constraining the possibilities for reimbursement by supplementary health insurance and, consequently, limiting the variation of premiums. The analysis of premiums for the higher coverage level revealed more significant disparities between regions, suggesting that the utilization of higher coverage limits has allowed for a more precise illumination of consumption trends.

For the analysis of departmental premiums, the approach centered on comparing a region with the constituent departments. For example, within the Île-de-France region, predominantly composed of urban zones, the premiums of departments displayed uniformity across both coverage levels. In contrast, in the Nouvelle-Aquitaine region, which encompasses both urban and rural zones, premiums exhibited more pronounced divergences that escalated with the elevation of coverage levels.

Comparing the premium of a region with the premiums of the constituent departments enabled the establishment of a relevance criterion for the implementation of a regional mutual insurance. Thus, the criterion to decide whether to approach the officials of the region will involve examining the distribution of departments with premiums higher and lower than that of the region. Therefore, if more than half of the departments display premiums higher than that of the region, it will be deemed opportune to initiate an approach to the officials of the region.

However, a few surprising results were observed. To highlight these, the departments were categorized based on the deviation of their pure premium from the average. The trends are presented through maps for the age group 60-69 years, with an average pure premium of 46.39 € for coverage V1 and 62.66 € for coverage V2.

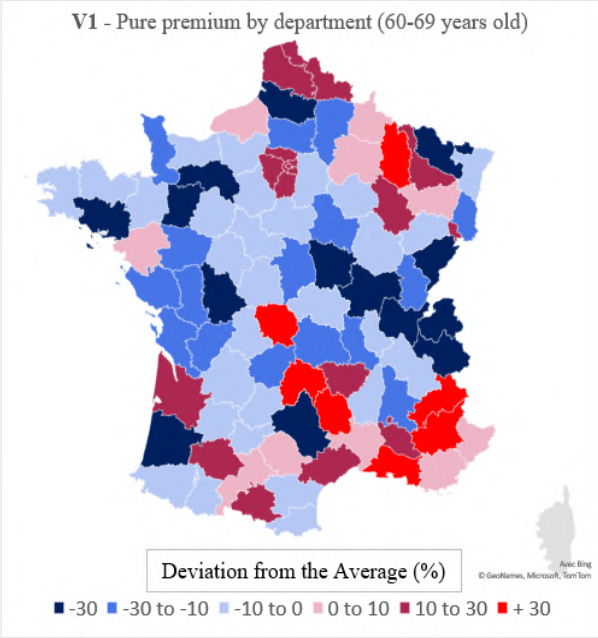


FIGURE 13 – Premium by department Coverage V1

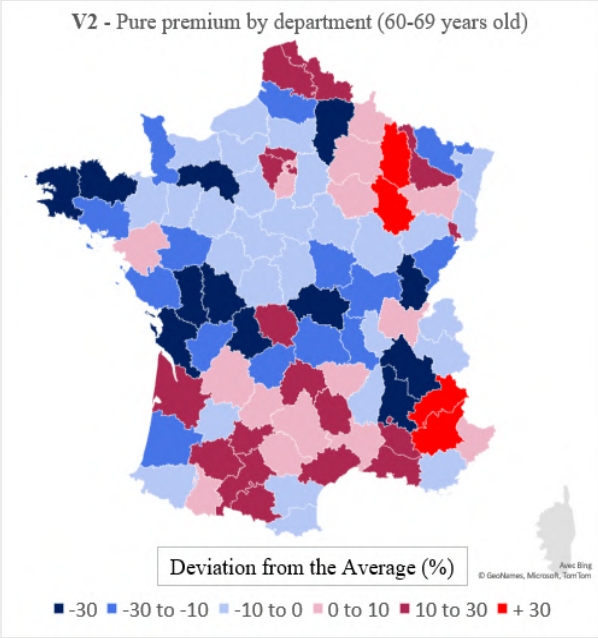


FIGURE 14 – Premium by department Coverage V2

The premiums for the departments of Île-de-France and the Provence-Alpes-Côte d’Azur region turn out to be higher than the average, in line with their consumption trends that are also above average. Nonetheless, some surprising observations emerge. Indeed, for both coverage levels, tariff spikes are visible in the departments marked in red, indicating a deviation from the average by over 30 %. All these departments are characterized by very limited exposure in terms of beneficiaries, meaning a low population within the 60-69 age group. Furthermore, it seems that the departments eligible for the Alsace-Moselle regime appear to be under-priced. This could be explained by the fact that the data in the database for these departments is that of the national social security scheme. However, in this territory, there exists a local scheme to which a portion of the inhabitants adhere. Nevertheless, the exposure assumes that the entire population adheres to the national social security scheme, creating a disparity. This difference in exposure should have been taken into account during the pricing step.

The competitive market

In reality, the introduction of a regional mutual insurance company would not entail the elimination of existing municipal or departmental mutual insurance entities. Therefore, based on the previously defined criterion, it was determined that the Île-de-France region seemed conducive to establishing a regional mutual insurance scheme, while the Nouvelle-Aquitaine region was not in favor of it. This assumption was put to the test through the simulation of a fictional competitive market, pitting an insurer seeking to pool its risks at the departmental level against another

opting for regional pooling. For this purpose, a database representing a fictional population was constructed for the two regions under examination. The demographic composition of this database aimed to closely resemble the profile typically interested in municipal mutual insurance. This composition included 60 % senior citizens aged over 60. It is assumed that individuals will choose an insurance plan from the insurer offering the most affordable rate corresponding to their profile. The assumption is that the cost margin remains neutral in relation to the expenses the insurer might encounter. Thus, only a 3 % profit margin has been selected and will be applied to the base premiums to create a simplified commercial premium. The results are presented below :

Coverage V1 Nouvelle-Aquitaine		3 % profit margin	
	Result	Market share	
Departmental insurer	12 M€	60%	
Regional insurer	-35 M€	40%	
Coverage V2 Nouvelle-Aquitaine		3 % profit margin	
	Result	Market share	
Departmental insurer	14 M€	60%	
Regional insurer	-45 M€	40%	

FIGURE 15 – Results of the competitive market for Nouvelle-Aquitaine

Coverage V1 Île-de-France		3 % profit margin	
	Result	Market share	
Departmental insurer	8 M€	47%	
Regional insurer	-8 M€	53%	
Coverage V2 Île-de-France		3 % profit margin	
	Result	Market share	
Departmental insurer	11 M€	47%	
Regional insurer	-15 M€	53%	

FIGURE 16 – Results of the competitive market for Île-de-France.

For the Nouvelle-Aquitaine region, it is observed that the regional mutual insurance plan exhibits deficiencies compared to the departmental mutual insurance for both levels of coverage. These shortcomings are attributed to the emergence of an adverse selection issue, resulting in a net loss of -35 million euros for coverage V1 and -45 million euros for coverage V2. Indeed, residents in departments where the rates were lower than those of the region did not opt for the regional offering. Simultaneously, an underpricing of the plan was noted for other departments, leading to the attraction of residents from these areas.

Nonetheless, for the Île-de-France region, it was not foreseen that the regional insurer would also display a negative outcome. Indeed, in reality, the Île-de-France region has successfully established a regional mutual insurance plan that boasts solid financial health, as attested by professionals operating within the insurance entities. This disparity can be accounted for by the circumstance that in the context of this dissertation, policyholders automatically gravitate toward the insurer offering the most competitive rates. However, in practicality, it remains conceivable that certain individuals might not necessarily dedicate the time to meticulously compare rates, whereas others place greater trust in a regional entity as opposed to a departmental one. It's also plausible that some individuals might only be apprised of the existence of the regional mutual insurance, without being cognizant of the municipal/departmental mutual insurance alternatives.

In order fit more to reality, it would be necessary to take into account, for each line of the fictitious base, a group of individuals likely not to opt for the lowest price.

Conclusion

The results of this study have demonstrated the potential to delve into the information extracted from the *Open* DAMIR database at a finer scale. Despite certain limitations, the established rates have shown a relative coherence with the trends observed in the health insurance domain.

To further delve into these investigations, it would be opportune to start from the most recent version of the *Open* DAMIR database. This more recent update could provide a better grasp of the impacts of the 100% Santé policy, which was only considered for costs in this study, but not for frequency.

Regarding the rates, significant disparities were observed among different age brackets, particularly among seniors. The application of smoothing techniques could be relevant, either at the departmental level, based on age, or even both simultaneously.

External variables were not integrated into the pricing models due to their lesser significance compared to other explanatory variables, namely age and location. An alternative approach would involve excluding the location variable from the models and basing the pricing solely on age and external variables. Furthermore, the creation of zoning based on the pure premiums obtained in this study could be considered.

Hence, communal mutual insurance, despite gaining popularity in recent years, continues to represent a niche market. This form of partnership between local communities and insurers remains largely unfamiliar to the general public. Moreover, despite the pessimistic tone of the conclusions in this paper regarding the future of regional mutual insurance, it has been shown that possible disparities exist between the behaviors of real residents and the specific context examined here. Consequently, even though not observed in this paper, considering regional mutual insurance as business opportunities while simultaneously improving access to care for vulnerable populations remains feasible.

The Île-de-France region exemplifies this dynamic with the creation of the first regional mutual insurance in France, sparking increasing interest among young professionals according to testimonies from experts in these organizations. This development suggests the potential for a broader adoption of these offerings, potentially offering an encouraging path for other regions to follow suit.

TABLE DES MATIÈRES

Introduction	1
I L'assurance santé et l'<i>Open Data</i> en France	3
1 L'assurance santé en France	7
1.1 Le système français de protection sociale	7
1.1.1 Les différents régimes de la Sécurité Sociale	7
1.1.2 Les branches de la Sécurité Sociale	7
1.2 L'Assurance Maladie	8
1.3 Les régimes complémentaires	9
1.3.1 Le rôle des complémentaires santé	9
1.3.2 Les types de contrats d'Assurance Maladie complémentaire	9
1.3.3 Les aides sociales pour l'accès aux soins	9
1.3.4 Le remboursement	10
1.3.5 Les grilles de garanties	13
1.3.6 La nomenclature des actes	14
1.4 Des réformes clés	15
1.4.1 Le contrat responsable	15
1.4.2 Le 100% santé	15
1.4.3 L'Accord National Interprofessionnel de 2016 (ANI)	16
2 Les mutuelles communales	17
2.1 Un défi pour la puissance publique	17
2.2 Définition d'une mutuelle communale	18
2.3 Les acteurs du marché	18
2.4 Les étapes de la mise en place d'une mutuelle communale	19
2.5 Les conditions de mise en place – Point de juridique	20

2.6	Les chiffres clés	22
2.7	Un équilibre économique menacé pour les assureurs?	23
3	Les enjeux de l'Open Data	25
3.1	Les principes de l'Open Data	26
3.2	SNIIRAM : la plus grande base de données de santé	26
3.3	L'Open DAMIR parmi les Open Data en santé	27
	Synthèse de la première partie	28
II	Les Open Data dans la création de la base de tarification	29
4	Les retraitements de l'Open DAMIR	33
4.1	Choix de la base de données	34
4.2	L'Open DAMIR	35
4.3	Les traitements déjà réalisés	37
4.3.1	Récapitulatif des retraitements déjà effectués	37
4.3.2	La classification des actes de soins	38
4.3.3	Backtesting des classifications	41
4.3.4	La segmentation pour la tarification	41
4.3.5	Base à l'issue des retraitements	43
5	La création d'une nouvelle base de tarification	45
5.1	Le retraitement des montants remboursés par la Sécurité Sociale	45
5.1.1	La méthode de retraitement	45
5.1.2	Des codes actes manquants	47
5.1.3	Des montants retraités anormaux	48
5.2	La répartition des montants par département	49
5.2.1	Méthode 1 : hypothèse de dépense uniforme	50
5.2.2	Méthode 2 : recours à la base DAM_DPT	51
5.2.3	Combinaison des méthodes	53
5.3	Inflation des dépenses de santé	55
5.4	NEW_DAMIR : la nouvelle base de données	55
5.4.1	Ajout des montants remboursés par la complémentaire santé	56
5.4.2	Ajout de la variable Exposition	57
6	Les variables externes	59
6.1	Le taux de chômage	60
6.2	L'indice de vieillissement	60
6.3	Le niveau médian de vie	61
6.4	La densité d'établissement de santé : FINESS	61
7	Analyse descriptive de la base NEW_DAMIR	65
7.1	Analyse des valeurs extrêmes	66
7.2	Statistiques de la base NEW_DAMIR	67
7.2.1	Analyse par classe d'âge	67

7.2.2	Analyse par département	68
	Synthèse de la deuxième partie	71
	III La tarification des mutuelles communales	73
8	Rappels théoriques sur la tarification	77
8.1	Le contrat d'assurance et les types de primes	77
8.2	Modélisation de la prime pure	78
8.3	Les méthodes de tarifications	80
8.3.1	Modèles linéaires généralisé (GLM)	80
8.3.2	Légitimité des GLM	80
8.3.3	La statistique directe	81
8.3.4	Les méthodes retenues par segment	81
9	Analyse Univariée	83
9.1	Analyse des catégories	84
9.2	Analyse en fréquence	86
9.3	Analyse en coût	88
10	Modélisation de la prime pure	91
10.1	La modélisation par deux modèles	92
10.1.1	Test de loi de fréquence	93
10.1.2	Test de loi pour le coût	94
10.1.3	Sélection des variables	96
10.1.4	Validation des modèles	98
10.2	La modélisation par un modèle : Tweedie	99
	Synthèse de la troisième partie	101
	IV Les tarifs et le marché concurrentiel des mutuelles communales	103
11	Les résultats	107
11.1	Les tarifs de l'assureur régional	107
11.1.1	La garantie entrée de gamme V1	107
11.1.2	La garantie moyenne de gamme V2	108
11.1.3	La comparaison des garanties	109
11.2	Les tarifs de l'assureur départemental	110
11.3	La comparaison des tarifs départementaux et régionaux	111
11.3.1	Garantie V1	111
11.3.2	Garantie V2	114
12	Le marché concurrentiel	117
12.1	Les adhérents potentiels de l'offre	118
12.2	Les assureurs	119
12.3	Application à la région Nouvelle-Aquitaine	120

12.4 Application à la région Ile-de-France	121
12.5 Réflexion sur les résultats obtenus	122
Synthèse de la quatrième partie	123
Conclusion	125
Annexes	129
Annexe A : Grilles de garanties	130
Annexe B : Les variables de l'Open DAMIR	131
Annexe C : Backtesting de la classification des codes actes	134
Annexe D : Retraitement des classes d'âges	136
Annexe E : Modèle linéaire généralisé (GLM)	137
12.5.1 Généralités sur les GLM	137
12.5.2 La famille exponentielle	138
12.5.3 Estimation des paramètres par la méthode du maximum de vraisemblance	139
Annexe F : La distribution Tweedie	140
Annexe G : Mesure de corrélation	141
12.5.4 Coefficient de Pearson	141
12.5.5 Tau de Kendall	141
12.5.6 Rho de Spearman	142
Annexe H : La déviance	143
Annexe I : Les cartes de France	144
12.5.7 Les régions	144
12.5.8 Les départements	145
Annexe J : Les tarifs par département	146
12.5.9 Les tarifs par département - Garantie V1	146
12.5.10 Les tarifs par département - Garantie V2	148
Annexe K : Base fictive Ile-de-France	150
Liste des acronymes	151
Bibliographie	153
Articles	153
Mémoires	153
Sites internet	153
Formations	154

TABLE DES FIGURES

1	Dépense unitaire moyenne des "Consultations/Visites" (2019)	11
2	Offre de soins par département en 2019	11
3	Résultats modélisation par 2 GLM (Actes d'imagerie et de radiologie)	12
4	Primes pures mensuelles Garantie V1	13
5	Tarif par département Garantie V1	14
6	Tarif par département Garantie V2	14
7	Résultats du marché concurrentiel pour la Nouvelle-Aquitaine	15
8	Résultats du marché concurrentiel pour l'Île-de-France	15
9	Average annual expenditure for "Consultations/Visits"	19
10	Number of employees in healthcare establishments	19
11	Results of modeling using 2 GLMs (Imaging and Radiology Procedures)	20
12	Pure monthly premiums Coverage V1	21
13	Premium by department Coverage V1	22
14	Premium by department Coverage V2	22
15	Results of the competitive market for Nouvelle-Aquitaine	23
16	Results of the competitive market for Île-de-France.	23
1.1	Décomposition des remboursements	11
1.2	Principaux taux de remboursement de la Sécurité Sociale	11
1.3	friseRecapRemb	13
1.4	Exemple de calcul du tarif conventionnel pour une séance rééducation du pied	14
2.1	Les étapes de la mise en place d'une mutuelle communale	19
2.2	Synthèse des niveaux juridiques	22
2.3	Les chiffres clés des mutuelles communales	23
4.1	Les variables de l'Open DAMIR	35
4.2	Composition d'un acte de soin majoré	36
4.3	Exemples de modalité et libellé de la variable "PRS_NAT"	38

4.4	Segmentation principale retenue des actes	40
4.5	Backtesting de la classification sur un échantillon de 50 codes actes	41
4.6	Segmentation pour la tarification	42
4.7	Variables de la base OLD_DAMIR à l'issue des retraitements	43
5.1	Fréquence d'apparition des postes de soins pour les codes non répertoriés	47
5.2	Fréquence d'apparition des postes de soins ayant une anomalie après retraitement	48
5.3	Dépenses de santé non réparties de la région IDF	49
5.4	Dépenses de santé réparties au sein des départements de la région IDF	49
5.5	Répartition de la population (60-69 ans) en région IDF	51
5.6	Répartition des frais réels de Consultations/Visites (60-69 ans) en région IDF . .	51
5.7	Dépenses de santé réparties au sein des régions (OLD_DAMIR)	52
5.8	Dépenses de santé réparties au sein des régions (DAM_DPT)	52
5.9	Coefficient de répartition pour les Consultations/Visites	53
5.10	Tableau récapitulatif des méthodes utilisées	54
5.11	Les taux de capitalisation sélectionnés	55
5.12	Les variables de NEW_DAMIR	57
6.1	Interprétation de l'indice de vieillissement	60
6.2	Tranche d'effectifs de salariés	63
7.1	Boîtes à moustaches des montants unitaires remboursés pour "Forfait journalier" par la garantie V2	66
7.2	Dépense moyenne par an par individu Consultations Visites (2019)	67
7.3	Fréquence unitaire par classe d'âge en 2019	68
7.4	Coût unitaire par classe d'âge en 2019	68
7.5	Dépense unitaire moyenne des "Consultations/Visites" (2019)	68
7.6	Offre de soins par département en 2019	68
7.7	Montant des dépenses annuelles en Île-de-France (2019)	69
7.8	Montant des dépenses annuelles en Nouvelle-Aquitaine (2019)	70
8.1	Méthode de tarification par segment	82
9.1	Cartographie des différentes catégories pour les variables externes	84
9.2	Analyse univariée en fréquence pour les actes médicaux techniques	86
9.3	Analyse univariée en coût pour les actes médicaux techniques	88
10.1	Résultats des tests de corrélations (Actes d'imagerie et de radiologie)	92
10.2	Diagramme de Cullen et Frey pour la loi de fréquence (Actes d'imagerie et de radiologie)	93
10.3	QQ Plot - Binomiale négative (Actes d'imagerie et de radiologie)	94
10.4	Diagramme de Cullen et Frey pour la loi de coût (Actes d'imagerie et de radiologie)	94
10.5	QQ Plot - Gamma (Actes d'imagerie et de radiologie)	95
10.6	AIC selon la loi candidate de coût - (Actes d'imagerie et de radiologie)	95
10.7	AIC selon la loi candidate de coût - (Actes d'imagerie et de radiologie)	96
10.8	AIC - Selection de variables (Actes d'imagerie et de radiologie)	97

10.9 Variables explicatives retenues	98
10.10 Validation du modèles par test de déviance	98
10.11 Résultats modélisation par 2 GLM (Actes d'imagerie et de radiologie)	99
10.12 Résultats des tests de corrélations (Actes médicaux techniques)	99
10.13 Histogramme empirique VS densité théorique (Actes médicaux techniques - Garantie V2)	100
10.14 Histogramme empirique VS densité théorique (Actes médicaux techniques - Garantie V2)	100
11.1 Primes pures mensuelles Garantie V1	107
11.2 Primes pures mensuelles Garantie V2	108
11.3 Comparaison des primes pures mensuelles (Ile-de-France)	109
11.4 Comparaison des primes pures mensuelles (Nouvelle-Aquitaine)	109
11.5 Tarif par département Garantie V1	110
11.6 Tarif par département Garantie V2	110
11.7 Comparaison des tarifs de l'assureur départemental et régional	111
11.8 Ecart relatif entre l'assureur départemental et régional pour la région Ile-de-France	112
11.9 Comparaison des tarifs de l'assureur départemental et régional pour la région Nouvelle-Aquitaine	112
11.10 Ecart relatif entre l'assureur départemental et régional pour la région Nouvelle-Aquitaine	113
11.11 Tarifs département vs région - Ile-de-France (V2)	114
11.12 Tarifs département vs région - Nouvelle-Aquitaine (V2)	115
12.1 Composition démographique de la base fictive	118
12.2 Résultats du marché concurrentiel pour la Nouvelle-Aquitaine	120
12.3 Résultats du marché concurrentiel pour l'Ile-de-France	121
12.4 grilles	130
12.5 Les variables de l' <i>Open</i> DAMIR 1	131
12.6 Les variables de l' <i>Open</i> DAMIR 2	132
12.7 Les variables de l' <i>Open</i> DAMIR 3	133
12.8 Backtesting de la classification des codes actes 1	134
12.9 Backtesting de la classification des codes actes 2	135
12.10 Répartition de la population par tranche d'âge INSEE	136
12.11 Résultat outil en ligne pour une tranche d'âge	136
12.12 Résultat après retraitement des classes d'âges	136
12.13 Carte des régions de France	144
12.14 Carte des départements de France	145
12.15 Primes pures mensuelles par département (Garantie V1) page 1/2	146
12.16 Primes pures mensuelles par département (Garantie V1) page 2/2	147
12.17 Primes pures mensuelles par département (Garantie V2) page 1/2	148
12.18 Primes pures mensuelles par département (Garantie V2) page 2/2	149
12.19 Base fictif Ile-de-France	150

INTRODUCTION

En France, l'assurance santé joue un rôle fondamental dans la protection et l'accès aux soins pour l'ensemble des citoyens. La branche maladie de la Sécurité Sociale a pour mission d'apporter un soutien financier aux individus face aux conséquences des risques tels que les maladies, la maternité, l'invalidité et le décès. Néanmoins, les prestations fournies par la Sécurité Sociale ne couvrent souvent pas tous les coûts associés aux soins médicaux. C'est ici que les organismes assureurs santé interviennent, offrant un appui financier supplémentaire aux patients en échange du paiement d'une cotisation. Pourtant, il est à noter que près de 30 % des Français ont déjà reporté ou renoncé à des soins pour des raisons financières, et plus de trois millions de citoyens ne possèdent aucune complémentaire santé. Cette situation met en évidence la nécessité de prêter attention aux politiques de santé et de garantir l'accessibilité financière aux soins afin d'assurer une couverture adéquate pour l'ensemble de la population française.

Le concept de mutuelle communale s'inscrit dans la perspective d'améliorer l'accessibilité aux soins. Contrairement à ce que son nom pourrait laisser penser, il ne s'agit pas d'une entité mutualiste créée directement par la commune, mais plutôt d'une alliance entre celle-ci et un organisme d'assurance. L'objectif de cette synergie entre le secteur public et privé consiste à offrir une couverture individuelle en matière de santé à tous les résidents de la commune, négociée à des tarifs préférentiels. Cependant, depuis la mise en place de la loi ANI, qui impose aux entreprises de fournir une complémentaire santé prise en charge à au moins 50 % par l'employeur à l'ensemble de leurs salariés, une grande partie de la population ne porte que peu d'intérêt aux offres de mutuelles communales, ce qui laisse une population plus vulnérable aux portes de ces offres. En effet, c'est principalement la population senior qui représente la majeure partie des adhérents à ces offres de mutuelle communale, ce qui soulève des interrogations concernant l'équilibre technique de l'organisme assureur.

Ce mémoire adopte la perspective d'un assureur qui opère dans le domaine des mutuelles communales.

Face aux difficultés inhérentes à la mutualisation des risques à l'échelle communale, quelle serait la solution ?

L'intuition de l'assureur le pousse à explorer la possibilité d'une mutualisation au niveau régional, jugée potentiellement plus viable selon le principe de mutualisation. Par conséquent, dans un objectif de démarchage, une analyse approfondie est entreprise pour évaluer la pertinence de maintenir l'échelle communale par rapport à une démarche commerciale ciblant les élus locaux des régions. Cette démarche se matérialise d'abord par la création de tarifs distincts pour chaque département ainsi que pour chaque région, suivis d'une phase comparative pour évaluer leur différence. Ainsi, en raison des contraintes liées à la disponibilité d'information, la notion de commune sera équivalente à celle du département dans le contexte de cette étude.

Dans un premier temps, la base de données utilisée sera construite à partir des bases de données *Open DAMIR* des années 2018 et 2019 qui ont fait l'objet de nombreux retraitements dans le cadre d'un précédent mémoire au sein du cabinet de conseil Actuelia. Ces bases de données se distinguent par la présence de données agrégées, telles que les dépenses de santé regroupées par région. Ainsi, la constitution de la base de données qui sera utilisée dans ce mémoire impliquera un affinement de la granularité des données provenant de l'*Open DAMIR*, accompagné d'une mise à jour des informations pour prendre en compte les impacts de l'inflation et de la réforme 100% Santé. La base sera ensuite enrichie de variables externes pour identifier les tendances de consommation de soins parmi les résidents des divers départements. Si ces tendances sont observables, elles seront prises en compte lors de l'établissement des tarifs.

Dans un second temps, un récapitulatif théorique des méthodes de tarification est effectué, suivi de la sélection des méthodes appropriées. Dans cette optique, les modèles linéaires généralisés seront privilégiés, du fait de leur grande interprétabilité. Deux perspectives distinctes seront ainsi exposées dans cette section en vue d'une comparaison ultérieure : celle d'un assureur qui mutualise son risque au niveau communal/départemental, et celle d'un assureur qui mutualise son risque à l'échelle régionale.

Après avoir examiné les disparités tarifaires entre les départements et les régions, un critère sera établi pour évaluer la pertinence de la mise en place d'une mutuelle régionale. Si la décision est prise d'instaurer une mutuelle régionale pour une région spécifique, il sera nécessaire de prendre en compte la présence préexistante de mutuelles communales sur le territoire. La dynamique concurrentielle sera ensuite modélisée au moyen d'un marché concurrentiel fictif opposant la mutuelle communale/départementale à la mutuelle régionale.

Première partie

L'assurance santé et l'*Open Data* en
France

Cette première partie vient présenter le contexte réglementaire dans lequel s'inscrit ce mémoire. Celui-ci portant sur la tarification des mutuelles communales, cette partie vient dans un premier temps présenter le système français de protection sociale. Ensuite, elle introduit la notion des mutuelles communales et leurs spécificités. En dernier lieu, elle aborde la notion d'*Open Data* en tant que seule source de données utilisée dans le mémoire.

CHAPITRE

1

L'ASSURANCE SANTÉ EN FRANCE

Ce chapitre vient présenter les principaux piliers du système français de protection sociale, mettant en évidence sa structure, ses acteurs clés, les notions importantes du monde de l'assurance santé ainsi que les réformes de santé.

1.1 Le système français de protection sociale

1.1.1 Les différents régimes de la Sécurité Sociale

Le système français de protection sociale aussi appelé Sécurité Sociale est constitué d'un ensemble de régimes légaux regroupés en trois grands blocs :

- le **Régime général** et le **Régime local Alsace Moselle** avec les salariés de l'industrie, du commerce et de services, ainsi que les travailleurs indépendants ;
- le **Régime agricole** avec les salariés et les non-salariés agricoles ;
- les **Régimes spéciaux** qui regroupent les fonctionnaires, les clercs de notaires, les mines, les cultes et certains salariés d'entreprise telles que la SNCF, EDF, RATP, marine marchande.

Le présent mémoire portera exclusivement sur l'étude du régime général de la Sécurité Sociale, sans prendre en compte les autres régimes existants.

1.1.2 Les branches de la Sécurité Sociale

La Sécurité Sociale est organisée en plusieurs branches, chacune ayant des missions spécifiques. Les branches de la Sécurité Sociale sont :

- la branche Maladie (couvrir la maladie, maternité, invalidité, décès),
- la branche Famille (réduire les inégalités de niveau de vie entre les ménages),

- la branche Accidents du travail – Maladies professionnelles (gérer la politique de prévention),
- la branche Retraite (préparer les retraites),
- la branche Autonomie (gérer les dépenses liées à l'autonomie des personnes âgées et des personnes handicapées)
- la branche Recouvrement (gérer de la trésorerie de la Sécurité Sociale).

Chaque branche de la Sécurité Sociale est organisée de manière autonome, avec sa propre gestion financière et ses propres règles et prestations. Cette organisation en branches permet de mieux répondre aux différents besoins de chaque assuré social, garantissant ainsi une gestion efficace et transparente des fonds de la Sécurité Sociale.

La branche Maladie sera celle étudiée dans le cadre du mémoire.

1.2 L'Assurance Maladie

Comme mentionné précédemment, l'Assurance Maladie est une branche de la Sécurité Sociale. Elle a pour objectif d'aider financièrement les personnes à faire face aux conséquences des risques maladies, maternité, invalidité et décès. Elle est gérée par la Caisse Nationale d'Assurance Maladie (CNAM), qui supervise également les différents organismes locaux chargés de gérer les remboursements de soins de santé sur l'ensemble du territoire français.

L'Assurance Maladie verse des prestations à des bénéficiaires. Il en existe deux types :

- **l'assuré social** : qui est personnellement assuré et cotise grâce aux prélèvements sur le salaire par exemple,
- **ayant droits** : ce sont des personnes qui bénéficient de l'Assurance Maladie d'un assuré social car ils ne peuvent pas être assurés (les enfants par exemple).

Chaque assuré social est rattaché à une Caisse Primaire d'Assurance Maladie (CPAM), c'est l'affiliation. L'assuré reçoit alors une carte vitale qui est informatisée justifiant du droit aux prestations et contient les informations nécessaires à la prise en charge de la personne.

Une prestation est une somme versée pour indemniser des dépenses liées aux risques couverts par l'Assurance Maladie. Il existe deux types de prestations :

- **prestations en nature** : remboursement des dépenses engagées par l'assuré et ses ayants droits pour se soigner tels que les consultations médicales, les médicaments, les examens et autres soins nécessaires.
- **prestations en espèce** : revenus de remplacement versés à l'assuré social. Les indemnités journalières remplacent le salaire pendant un arrêt de travail.

Le mémoire se concentrera exclusivement sur l'étude des prestations en nature de l'Assurance Maladie.

1.3 Les régimes complémentaires

1.3.1 Le rôle des complémentaires santé

Les prestations en nature de la Sécurité Sociale, bien qu'essentielles pour la prise en charge des dépenses de santé des assurés, ne sont pas toujours suffisantes pour couvrir l'ensemble des coûts liés aux soins. En effet, les remboursements proposés par l'Assurance Maladie peuvent être limités ou ne pas couvrir l'intégralité des frais médicaux, laissant les patients supporter une partie de la charge financière qui peut parfois s'avérer conséquente.

C'est pour cette raison que de nombreuses personnes souscrivent à des assurances maladie complémentaires, souvent appelées à tort "mutuelles". Ces assurances complémentaires santé offrent la possibilité de prendre en charge la totalité ou une portion de la charge financière restante. Cette couverture permet aux assurés sociaux d'accéder à des soins de santé en réduisant les coûts financiers engendrés.

Par ailleurs, le **système du tiers payant** est un dispositif qui permet à un assuré social de ne pas avancer les frais de santé pris en charge par l'Assurance Maladie obligatoire et/ou sa complémentaire santé.

1.3.2 Les types de contrats d'Assurance Maladie complémentaire

Il existe deux types de contrats d'Assurance Maladie complémentaire :

- le **contrat collectif** signé entre un groupe de personnes et l'assureur. C'est le plus souvent une entreprise qui négocie le contrat pour l'ensemble de ses salariés ou pour une catégorie de salariés tels que les cadres par exemple.
- le **contrat individuel** est souscrit par une personne physique directement auprès de l'organisme assureur. Les garanties proposées par ce type de contrat peuvent être personnalisées en fonction des besoins et des exigences de l'assuré.

Pour rappel, la loi ANI entrée en vigueur en 2016 impose aux entreprises du secteur privé de proposer une couverture complémentaire santé à l'ensemble de leurs salariés.

Les contrats individuels seront examinés dans ce mémoire.

1.3.3 Les aides sociales pour l'accès aux soins

Des dispositifs d'aides sociales sont mis en place pour permettre aux personnes disposant de faibles revenus ou en situation de précarité d'accéder tout de même à une couverture santé de base ou complémentaire.

En 2016, l'entrée en vigueur de la **PUMA (Protection Universelle Maladie)** a permis de remplacer la CMU (Couverture Maladie Universelle de base) tout en assurant une continuité de la prise en charge des frais de santé, même en cas de changement de situation professionnelle (perte d'emploi, etc.), familiale (séparation, etc.) ou de résidence.

Initialement, la **Couverture Maladie Universelle Complémentaire (CMU-C)** et l'**Aide à la Complémentaire Santé (ACS)** étaient des dispositifs français visant à assurer l'accès aux soins médicaux pour les personnes à faibles revenus, mais elles ont été remplacées par la

Complémentaire santé solidaire (CSS) le 1^{er} novembre 2019. Cette nouvelle couverture reste gratuite pour les bénéficiaires de la CMU-C, mais requiert une participation financière de la part des anciens bénéficiaires de l'ACS. Concrètement, elle permet de rembourser la part complémentaire des dépenses de santé. Enfin, l'**Aide Médicale de l'Etat (AME)** permet l'accès aux soins des personnes en situation irrégulière en France et sans ressources suffisantes pour payer leur frais.

1.3.4 Le remboursement

Le remboursement des frais de santé est le mécanisme le plus important du mémoire, il s'agit de l'essence même des assureurs santé.

Un tarif conventionnel également appelé **Base de remboursement de la Sécurité Sociale (BRSS)** est défini pour chaque prestation médicale remboursable par le régime obligatoire (consultation chez un médecin généraliste, séance de kinésithérapie,...).

Un **dépassement d'honoraires** est observé lorsque les **Frais Réels (FR)** sont plus élevés que le tarif conventionnel. C'est le cas lorsque le patient a payé plus cher sa consultation chez le médecin que le prix de base.

Ainsi, le processus de remboursement se déroule en plusieurs étapes : initialement, c'est la Sécurité Sociale qui effectue le premier remboursement (MT_RO), suivi de l'intervention de l'organisme d'assurance pour couvrir la part restante (MT_RC). En dernier lieu, s'il reste des frais à payer après ces deux remboursements, le patient est responsable de les régler.

Pour calculer les montants de remboursements effectués par le régime obligatoire, un **Taux de Remboursement (TR)** est défini pour chaque type de prestation médicale puis appliqué à la BRSS correspondante.

Le montant remboursé par le régime obligatoire (MT_RO) se calcule ainsi :

$$MT_RO = TR \times BRSS \quad (1.1)$$

Avec :

MT_RO : Montant remboursé par la Sécurité Sociale

TR : Taux de remboursement de la Sécurité Sociale

$BRSS$: Base de remboursement de la Sécurité Sociale

Enfin, Le **ticket modérateur (TM)** est la part restante de la Base de remboursement après déduction du montant couvert par la Sécurité Sociale.

1.3. LES RÉGIMES COMPLÉMENTAIRES

Le schéma ci-dessous permet de synthétiser les différentes notions introduites dans cette partie :

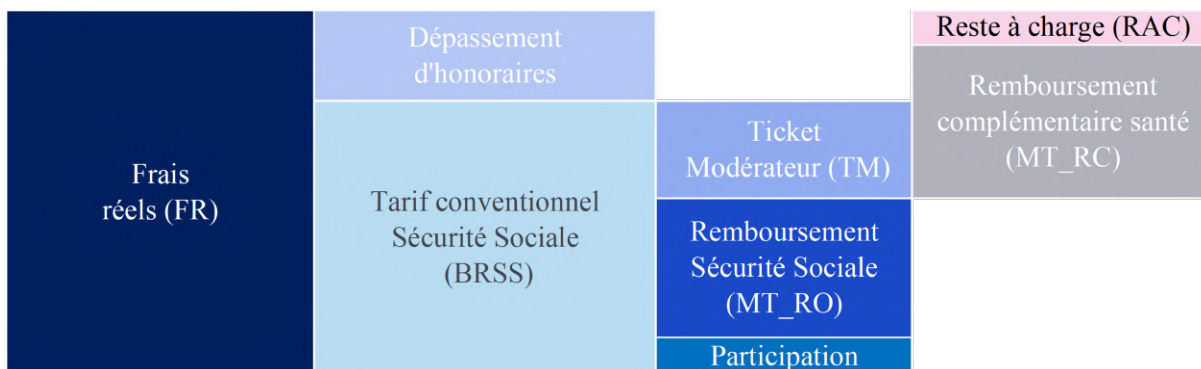


FIGURE 1.1 – Décomposition des remboursements

Le tableau ci-dessous indique les principaux taux de remboursements usuels afin d'avoir une idée générale.

Prestation	Taux de Remboursement
Médecins, praticiens	70%
Auxiliaires médicaux	60%
Médicaments	15%; 30% ; 65%; 100%
Optique	60%
Dentaire	70%
Aides auditives	60%
Hospitalisation	80%
Frais de transport	65%
Cures thermales	65%
Prévention	65%

FIGURE 1.2 – Principaux taux de remboursement de la Sécurité Sociale

Comme l'indique le tableau ci-dessus, les taux de remboursement de la Sécurité Sociale sont inférieurs à 100 %, à l'exception des médicaments reconnus comme indispensables et coûteux, tels que les médicaments antirétroviraux pour le traitement du VIH/sida.

Ainsi, la différence entre la base de remboursement et le remboursement de la Sécurité Sociale est rarement nulle. De plus, les dépassements d'honoraires s'ils existent, vont s'ajouter au reste à charge du patient.

Par ailleurs, des petites sommes peuvent être déduites par l'Assurance Maladie sur les remboursements du régime obligatoire, il s'agit d'une **participation forfaitaire** d'un euro afin de préserver le système de santé [1] ou d'une **franchise médicale** pour contribuer à financer la lutte contre le cancer, la maladie d'Alzheimer ou encore les soins palliatifs [2] . Ces montants ne sont généralement pas pris en charge par les organismes complémentaires.

Exemple de remboursement

Une étudiante de 24 ans souhaite commencer l'escalade en salle. Malheureusement, lors de sa première séance, elle fait une mauvaise chute et se tord le bras. Son médecin traitant lui prescrit une ordonnance chez un kinésithérapeute de trente séances pour une rééducation d'un membre.

Heureusement, la patiente bénéficie d'un contrat d'assurance santé individuel où il est indiqué qu'elle peut bénéficier d'un remboursement s'élevant à « 200% de la BRSS y compris SS ». Elle n'a pas de sur-complémentaire.

Pour faciliter la compréhension, le montant de franchise est supposé dépassé et ne sera pas pris en compte dans le calcul.

Les **frais réels** de la séance de kinésithérapie s'élèvent à **38 €**.

La Sécurité Sociale estime qu'une séance de kinésithérapie pour l'acte « rééducation d'un membre » coûte 16,13€ (BRSS). Le taux de remboursement associé est de 60%, ce qui signifie que la Sécurité Sociale remboursera 60% de **16,13€**.

Ainsi, le montant remboursé par la Sécurité Sociale par séance (Montant RO) vaut :

$$MT_RO = 60\% \times BRSS = 60\% \times 16,13 = 9,68 \text{ €} \quad (1.2)$$

Le **Ticket Modérateur (TM)** vaut alors :

$$TM = BRSS - MT_RO = 16,13 - 9,68 = 6,45 \text{ €} \quad (1.3)$$

Ensuite, il est indiqué que la complémentaire santé prend en charge les dépenses de santé jusqu'à « 200% de la BRSS y compris SS ». Cela signifie que la somme des remboursements provenant des régimes obligatoires et complémentaires ne peut pas excéder deux fois la Base de Remboursement de la Sécurité Sociale (BRSS).

$$Limite_Garantie = 200\% \times BRSS = 200\% \times 16,13 = 32,26 \text{ €} \quad (1.4)$$

Ainsi, le montant que doit rembourser l'assureur santé (MT_RC) vaut :

$$\begin{aligned} MT_RC &= \min(Frais\ reels - MT_RO; Limite_Garantie - MT_RO) \\ MT_RC &= \min(38 - 9,68 ; 32,26 - 9,68) \\ MT_RC &= \min(28,32 ; 22,58) \\ MT_RC &= 22,58 \text{ €} \end{aligned} \quad (1.5)$$

Le reste à charge (RAC) pour le patient vaut :

$$RAC = Frais\ Reels - MT_RO - MT_RC = 38 - 9,68 - 22,58 = 5,74 \text{ €} \quad (1.6)$$

Si la franchise avait été prise en compte, le reste à charge aurait été augmenté de 50 centimes.

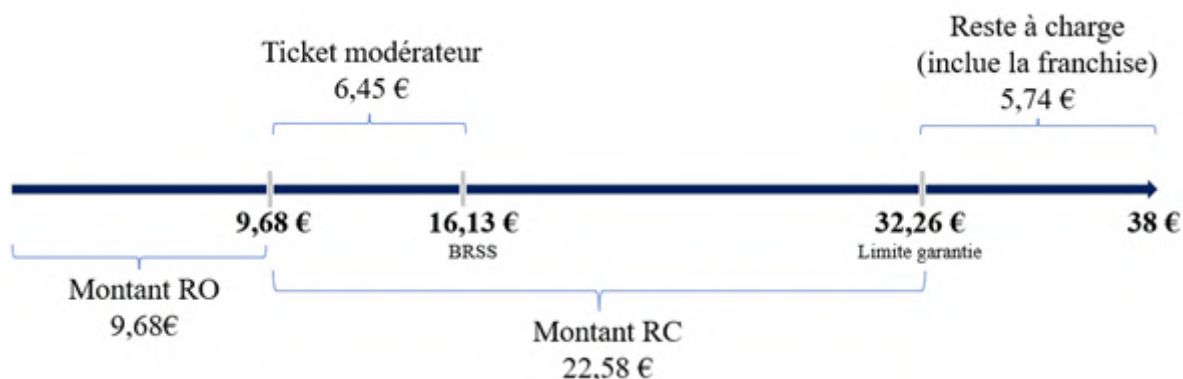


FIGURE 1.3 – Exemple des remboursements

1.3.5 Les grilles de garanties

La grille de garanties d'une complémentaire santé détaille les prestations couvertes par l'assurance complémentaire. En effet, elle indique les niveaux de remboursements associés aux différents actes médicaux, traitements et soins pris en charge par la complémentaire santé. Ainsi, cette grille permet aux assurés de mieux comprendre les conditions de remboursement de leur assurance, d'évaluer les avantages et les limites de leur contrat, puis de faire des choix éclairés en matière de soins de santé leur permettant de bénéficier d'une couverture adaptée à leurs besoins médicaux. Dans un premier temps, les grilles de garanties sont décomposées de différents postes de soins, il s'agit d'une liste non exhaustive :

- les **soins courants** Ce poste comprend les consultations chez le médecin ou spécialiste mais aussi les prestations délivrées par les auxiliaires médicaux ;
- l'**hospitalisation** contient généralement les frais de séjour, les frais de salle d'opération, les chambres particulières ;
- l'**optique** regroupe les montures, les lentilles, les verres, etc ;
- le **dentaire** avec l'orthodontie, les prothèses dentaires et les soins dentaires.

Dans le cadre de ce mémoire, un dernier grand poste intitulé « **Autres** » regroupera les cures thermales, la médecine douce, la prévention et les prestations supplémentaires. Pour chaque poste de soin, une granularité plus fine est souvent présentée dans les grilles tarifaires, comprenant des remboursements spécifiques pour les différentes sous-familles d'un même poste. De plus, les remboursements peuvent être exprimés en pourcentage ou en montant forfaitaire.

Pour les remboursements en pourcentage, ils peuvent être présentés de ces manières :

- en pourcentage de la BRSS y compris la part prise en charge par le régime obligatoire ;
- en pourcentage de la BRSS sans la part prise en charge par le régime obligatoire ;
- en pourcentage des frais réels engagés par le patient ;
- en pourcentage du Plafond Mensuel de la Sécurité Sociale (PMSS¹).

1. Le PMSS 2023 vaut 3 666€

Certains postes de soin, mal remboursés par la Sécurité Sociale ou non pris en charge, bénéficient d'un remboursement en forfait. Par exemple, la chirurgie de l'œil est remboursée à hauteur de 100 € par an. Un exemple de grille simplifiée est inclus en annexe 12.5, présentant les garanties de la mutuelle communale qui sera étudiée dans la suite du mémoire. Tous les remboursements des gilles de la mutuelle sont exprimés en pourcentage de la BRSS y compris la part prise en charge par le régime obligatoire ou en prenant la forme d'un forfait.

1.3.6 La nomenclature des actes

En France, les actes médicaux remboursables par l'Assurance Maladie obligatoire sont répertoriés dans différentes nomenclatures codifiées qui leur permettent d'être identifiés, classés et tarifés. Il existe plusieurs nomenclatures, mais seules les principales seront présentées dans ce mémoire. La première nomenclature est la **Nomenclature Générale des Actes Professionnels (NGAP)** qui définit les honoraires des praticiens du secteur libéral et hospitalier.

Chaque acte médical est identifié par :

- une lettre clé (consultation, soins infirmiers, acte chirurgical, etc..) qui est affectée à une base de remboursement unitaire ;
- un coefficient multiplicateur (complexité de l'acte, qualification du praticien, ..) qui est appliqué à la base de remboursement unitaire.

Exemple

La rééducation d'un pied possède la lettre clé « **AMS** » qui concerne les « *actes de rééducation des affections orthopédiques et rhumatologiques effectués par le masseur-kinésithérapeute* » associé à un tarif conventionnel : **2,15 €**

Le coefficient multiplicateur est de 7,5

Le coût d'une séance estimé par la Sécurité Sociale (BRSS) vaut : $2,15 \times 7,5 = 16,13\text{€}$



Coût de la séance : tarif conventionnel x coefficient

$$2,15 \text{ €} \times 7,5 = \underline{16,13 \text{ €}}$$

FIGURE 1.4 – Exemple calcul du tarif conventionnel pour une séance rééducation du pied

Source de l'image : Hyperassur 2023

La seconde nomenclature présentée est la **Classification Commune des Actes Médicaux (CCAM)**. Il s'agit de la nomenclature adaptée aux actes techniques réalisés par les médecins, plus détaillée que celle de la NGAP. Elle décrit l'appareil anatomique, l'organe, l'action effectuée et la technique utilisée à travers sa codification.

Chaque code médical est associé à un tarif de base, mais il est également possible que ces codes soient complétés par d'autres codes qui viennent majorer le tarif. Par exemple, lorsque l'acte médical est réalisé de nuit en situation d'urgence, des codes supplémentaires sont ajoutés pour refléter la majoration tarifaire correspondante.

Enfin, la dernière nomenclature énoncée dans ce mémoire concerne la **Nomenclature des Actes de Biologie Médicale (NABM)** qui regroupe les informations sur les actes de biologie médicale remboursés par le régime obligatoire.

1.4 Des réformes clés

1.4.1 Le contrat responsable

Le 1^{er} janvier 2006, les contrats dits « responsables » sont mis en place dans l'objectif de réduire le déficit de l'Assurance Maladie. Comme leur nom l'indique, ce type de contrat incite le patient à avoir une attitude raisonnable concernant les dépenses de santé qu'il engendre. Concrètement, ce dispositif permet à un contrat d'Assurance Maladie complémentaire de bénéficier d'avantages fiscaux et sociaux, dans la mesure du respect d'un cahier des charges précis.

Par conséquent, pour bénéficier des avantages de ce contrat, le patient doit respecter le parcours de soin. Par exemple, les consultations des médecins spécialistes doivent être précédées d'une consultation préalable chez le médecin traitant, sauf en cas d'urgence ou de demande spécifique du médecin traitant. Cette démarche vise à limiter la surconsommation des consultations de spécialistes.

Les principales mesures mises en place sont les suivantes :

- l'exclusion de la prise en charge de la participation forfaitaire de 1 €, la franchise médicale de 0,5 € ou toutes autres participations ;
- l'exclusion de la prise en charge des dépassements d'honoraires sur les actes cliniques et techniques hors parcours de soins ;
- la prise en charge obligatoire du ticket modérateur pour certains actes et certains médicaments ;
- la limitation des remboursements en cas de non-respect du parcours de soins.

1.4.2 Le 100% santé

L'objectif premier de la réforme est de lutter contre le renoncement aux soins pour des raisons financières en proposant à tous les Français couverts par un organisme d'assurance (et ayant un contrat dit « responsable ») l'accès à des soins de base sans reste à charge dans les domaines de l'optique, des prothèses dentaires et auditives. L'assuré n'est bien entendu pas contraint à choisir le panier « zéro » (celui qui aboutit à un reste à charge nul) et conserve sa liberté de sélectionner des paniers à tarifs plus flexibles, ce qui limiterait moins ses choix en termes de matériaux et d'esthétique pour ses équipements. De plus, la mise en place de ces paniers entraîne

une révision des nomenclatures CCAM et LPP, impliquant la création de nouveaux codes de prestations et une redéfinition des codifications plus précises. Par exemple, la création de paniers d'actes prothétiques nécessite la subdivision de certains codes CCAM en plusieurs codes distincts. Cette révision des nomenclatures permet également l'établissement de limites distinctes pour les honoraires de facturation.

1.4.3 L'Accord National Interprofessionnel de 2016 (ANI)

En France, la réforme ANI impose aux employeurs de proposer une « mutuelle d'entreprise » à ses salariés, c'est-à-dire une complémentaire santé financée à hauteur de 50% au minimum par l'employeur au profit des salariés. Cette mesure vise à garantir une protection sociale optimale pour les salariés.

Cependant, bien que cette réforme constitue un grand pas en avant pour l'égalité des soins dans le monde salarial, elle laisse certains profils de côté comme par exemple, les demandeurs d'emplois, les auto-entrepreneurs ou encore les retraités, qui ne peuvent plus bénéficier des contrats collectifs d'entreprise et se retrouvent avec des frais de cotisations très élevés en raison de leur âge et de leur état de santé.

Face à ces exclusions et difficultés d'accès à une couverture santé adéquate pour certaines catégories de personnes, les « mutuelles communales » se présentent comme une solution prometteuse. Est-ce réellement le cas ?

CHAPITRE

2

LES MUTUELLES COMMUNALES

Dans cette section, la notion des mutuelles communales est introduite. Les acteurs impliqués, le fonctionnement du système ainsi que certains aspects juridiques seront également abordés.

2.1 Un défi pour la puissance publique

Selon un sondage réalisé par la Mutualité Française en octobre 2018, les chiffres indiquent une tendance préoccupante en ce qui concerne l'accès aux soins en France. En effet, près de **30 %** des Français auraient renoncé ou retardé des soins, principalement en raison de contraintes financières. Ce pourcentage est en augmentation ces dernières années, soulignant ainsi les difficultés financières auxquelles de nombreux français sont confrontés lorsqu'ils ont besoin de se faire soigner.

De plus, selon la même enquête, **51 %** des Français ont mentionné des restes à charge jugés "trop élevés", en particulier parmi les seniors. Ces chiffres mettent en évidence les obstacles financiers qui entravent l'accès aux soins pour une part importante de la population française.

Par ailleurs, il est alarmant de constater que **5 %** des Français, soit environ 3 millions de personnes, n'ont souscrit à aucune complémentaire santé, selon la Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques (DREES) dans son édition 2019 sur la complémentaire santé.

Cette situation souligne l'importance d'une réflexion sur les politiques de santé et l'accessibilité financière des soins pour garantir une couverture adéquate à tous les français. C'est dans cette logique de faciliter l'accès aux soins que s'inscrivent les mutuelles communales.

2.2 Définition d'une mutuelle communale

Le nom de « **mutuelle communale** » peut parfois prêter à confusion, donnant l'impression qu'il s'agit d'un organisme d'assurance propre à une commune qui se substitue aux complémentaires santé traditionnelles. Toutefois, il convient de souligner que les mutuelles communales sont en réalité le fruit d'un partenariat établi entre une commune ou son Centre Communale d'action Sociale (CCAS) et un organisme d'assurance (mutuelles ou compagnie d'assurances), dans le but de proposer une couverture santé¹ aux habitants de ladite commune.

C'est la commune qui initie ce partenariat, consciente des besoins spécifiques de sa population en matière d'accès aux soins et de protection sociale. Ainsi, en travaillant avec un assureur, la commune cherche à proposer une offre de soin à tarif préférentiel à adhésion facultative. Ce type de partenariat est un exemple concret d'initiative locale visant à améliorer l'accès aux soins pour tous les habitants d'une commune donnée.

Deux principaux modèles sont utilisés pour la mise en place d'une mutuelle communale : le premier modèle implique l'utilisation d'un agent intermédiaire chargé d'établir une liaison entre une commune et un assureur santé, tandis que le second modèle concerne les organismes de santé qui développent directement leur propre gamme de produits de santé dans ce cadre.

2.3 Les acteurs du marché

ACTIOM

Le principal acteur sur le marché est ACTIOM (Actions de mutualisation pour l'amélioration du pouvoir d'achat). Il s'agit d'une association d'assurés indépendante qui joue le rôle d'un courtier en mettant en lien les communes et les organismes d'assurance santé.

Comme son nom l'indique, l'association vise à renforcer le pouvoir d'achat des adhérents en leur permettant de bénéficier de tarifs avantageux sur des complémentaires santé. Présent dans 2700 communes sur les 3 000 mutuelles communales mises en place. Il existe également d'autres associations d'assurés telles que Mandarine, Armoria Santé Prévoyance.

Dans le cas où la commune fait le choix de négocier directement auprès de l'association ACTIOM, elle n'aura pas à rédiger un cahier des charges précis et bénéficiera d'une offre standard. Cependant, elle aura un choix vaste parmi les complémentaires santé partenaires de cette association. Toutefois, il est possible qu'aucune de ces offres ne soit adaptée à la situation locale. Par ailleurs, même si le tarif est préférentiel car proposé à un groupe, des frais de courtage sont appliqués.

1. Certains assureurs proposent également une offre de prévoyance dans le cadre de ce partenariat mais puisqu'il s'agit d'une minorité et que l'objet du mémoire porte uniquement sur la tarification d'une offre de santé. La prévoyance ne sera pas abordée.

Les autres acteurs

Si ACTIOM est le moteur principal des mutuelles communales, il se repose essentiellement sur ses partenariats avec d'autres acteurs qui se sont positionnés initialement sur le marché tels que les associations, les petites et moyennes mutuelles (Miel Mutuelle, Smatis, Pavillon Prévoyance, etc.) ou des plus grandes tels que Mutualia. Par la suite, des compagnies d'assurances telles que AXA ont également développé ce modèle d'offre intermédiaire en collaboration avec ACTIOM.

La segmentation des publics assurés, due à la généralisation de la complémentaire d'entreprise et à la sélection des "bons risques" par les assureurs, "compromet la dimension redistributive de l'Assurance Maladie complémentaire" selon l'article "*Liberté ou solidarité : le dilemme des complémentaires*, 2010 [3]". Autrement dit, les principes de solidarité sont dégradés entre les malades et les bien portants, ainsi qu'entre les actifs et les retraités.

Certaines mutuelles se positionnent alors directement sur le marché sans avoir recours à des agents intermédiaires. Ce sont souvent des mutuelles territorialisées membres de la Fédération Diversité Proximité (FDPM) telles que la Mutuelle catalane et la Mutuelle générale d'Avignon. Elles s'opposent à la concentration du secteur assurantiel et prônent le retour aux principes d'origines de la mutualité en développant leur propre gamme de complémentaire santé communale. Ainsi, si la commune choisit de ne pas négocier directement avec une association d'assurés (ACTIOM) et préfère rédiger un cahier des charges répondant aux besoins spécifiques de sa population, elle devra suivre un processus d'appel d'offres en sollicitant les organismes d'assurance santé. Cependant, il peut se révéler difficile de mener à bien ces étapes sans une expertise dans le domaine.

2.4 Les étapes de la mise en place d'une mutuelle communale

Bien qu'il n'existe pas de processus à respecter impérativement pour la mise en place d'une mutuelle communale, les communes, ou plus souvent les CCAS suivent généralement les étapes suivantes :

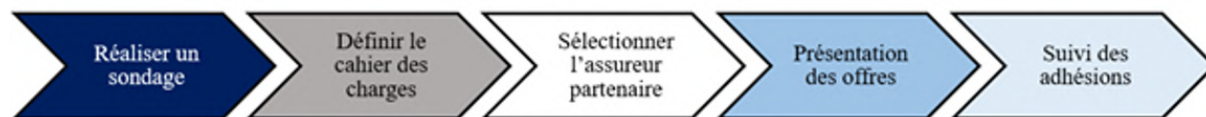


FIGURE 2.1 – Les étapes de la mise en place d'une mutuelle communale

Étape 1 : Réaliser un sondage ou une enquête auprès des habitants de la commune afin de déterminer le nombre d'individus intéressés par l'adhésion à une mutuelle communale. Cerner les attentes des habitants en matière de couverture santé (garanties souhaitées et tarifs acceptables).

Étape 2 : C'est le plus souvent le Centre communal d'action sociale (CCAS) qui se charge de définir le cahier des charges et/ou de soumettre une requête officielle auprès des agents intermédiaires (ACTIOM) ou directement aux organismes assureurs de santé. Cette demande exprime

la volonté de la commune de mettre en place un tel dispositif.

Etape 3 : Une fois les attentes des adhérents identifiées et la demande formulée, la commune peut procéder au choix d'un assureur partenaire. Il est possible d'avoir plusieurs assureurs partenaires, offrant ainsi une diversité de choix aux habitants de la commune.

Etape 4 : Lorsque le ou les organismes partenaires ont été sélectionnés, ils organisent une réunion publique au cours de laquelle ils se présentent ainsi que leur offre de mutuelle communale. Ensuite, des permanences en présence physique sont mises en place afin de permettre aux habitants de poser des questions, d'obtenir des informations supplémentaires et de procéder à leur adhésion à la mutuelle communale.

Etape 5 : Une fois la mutuelle communale établie, un suivi régulier des niveaux de cotisations est effectué. L'objectif est de maintenir des cotisations abordables pour les adhérents. Des *reporting* sont générés pour suivre le nombre de nouveaux adhérents, des résiliations, les offres les plus sélectionnées et les raisons de souscription. Enfin, les communications sont principalement réalisées à travers le bulletin municipal, des affiches de bus, panneaux ou encore sur la page Facebook de la commune.

2.5 Les conditions de mise en place – Point de juridique

Avant de mettre en place une mutuelle communale, les élus locaux s'interrogent sur la nature juridique du système. L'objectif de cette sous-partie consiste à éclairer ces interrogations. En effet, l'intervention de la mairie est une intervention sociale dans un secteur économique, il y a donc une première question juridique. Quelles sont les conditions de mise en place d'une mutuelle communale ?

Ainsi, Hugo-Bernard Pouillaude, Avocat Associé chez BRL Avocats, responsable du pôle Droit public et Fonction publique décrit les mutuelles communales comme « une forme de socialisme municipale » car au début du 20^e siècle il y avait une jurisprudence très restrictive sur l'articulation entre l'action sociale des mairies et la liberté d'entreprendre. En effet, elle permettait aux communes de n'intervenir qu'en cas de carence de l'initiative privée. Aujourd'hui, les communes peuvent initier certains projets telles que la mise en place d'une mutuelle communale mais une réglementation stricte reste en vigueur concernant les acteurs privés.

Bien que les mutuelles mènent des actions de prévoyance, de solidarité et d'entraide, elles sont qualifiées par la Cour de justice de l'Union européenne d'activités économiques car elles impliquent des opérateurs économiques intervenant sur un marché. Par conséquent, l'intervention de la commune dans le domaine de complémentaire santé doit respecter le droit de la concurrence et la liberté d'entreprendre. Cela signifie que la commune ne peut en aucun cas fausser la concurrence et ne peut favoriser aucun acteur, y compris elle-même. Par exemple, elle ne peut pas créer un organisme ou une association proposant des prix anormalement bas, compensés par des subventions publiques, ce qui fausserait la concurrence au détriment des autres acteurs du marché. Enfin, elle doit également respecter les obligations du code de la communication publique notamment des règles de publicités.

L'Association des Maires de France (AMF) vient compléter les propos de Maître Pouillaude en rédigeant une note sur « *Les aspects juridiques liés à la mise en place de mutuelles communales* » [4] dans laquelle il est détaillé quatre cas de figure permettant de déterminer le lien juridique entre la commune et l'organisme assureur selon le degré d'implication de la commune dans la mise en place d'une mutuelle communale.

Le **premier niveau** d'implication consiste à positionner la commune en tant que "facilitateur". Cela implique que les associations d'assurés, telles que ACTIOM, approchent la commune, qui à son tour envoie une lettre d'accréditation². Cette lettre a pour but d'informer les habitants que la commune soutient cette action et de leur indiquer les démarches à suivre. Ainsi, il n'y a aucun lien contractuel entre la commune et les associations, ni participation financière de sa part mais simplement le relai d'informations auprès des habitants.

Il arrive parfois que la commune mette à disposition un bureau ou un local au sein de la mairie/CCAS afin de permettre aux habitants de discuter avec les professionnels pour répondre à des points de questionnement. Cependant, cette mise à disposition d'une permanence doit respecter l'article L. 2125-1 du code général de la propriété des personnes publiques qui indique que **toute occupation ou utilisation du domaine public d'une personne publique donne lieu au paiement d'une redevance**. Ceci est également valable pour les prochains niveaux qui seront présentés ci-dessous.

Dans le **deuxième niveau** d'implication, la commune (mairie) négocie et sélectionne un contrat d'assurance complémentaire santé à proposer aux habitants avec des tarifs attractifs. Ainsi cette mise en concurrence se fait soit par une procédure d'appel à partenariat, soit par un démarchage direct des mutuelles³.

Le **troisième niveau** est atteint lorsque la commune décide de créer via son CCAS, une association d'assurés dont l'objectif est de mutualiser les risques à l'échelle de la commune. Dans ce cas de figure, les habitants payent une cotisation symbolique pour permettre à l'association de négocier des tarifs, de rédiger un cahier des charges puis sélectionner les organismes assureurs partenaires. Cependant, l'objet de l'association n'est pas de souscrire les contrats au nom des adhérents. **Les contrats proposés sont de caractère individuel à adhésion facultative**.

Le **quatrième** et dernier niveau d'implication est le cas où un CCAS souscrit un contrat collectif pour la population de la commune. Dans ce schéma, le CCAS règle l'intégralité des cotisations à l'organisme assureur puis se tourne vers les adhérents pour recouvrer les cotisations.

Par ailleurs, les trois premiers niveaux ne sont pas soumis au Droit des marchés publics car il n'y a pas de transaction financière entre les entités publiques et privées. Cependant, c'est le cas dans le dernier niveau.

2. La lettre d'accréditation n'a pas, a priori, de valeur juridique, il s'agit d'un simple document de communication de l'offre proposée par les associations.

3. Dans la pratique, certaines communes sollicitaient plusieurs assureurs, mais ne mettaient en avant que l'assureur local lorsqu'il s'agissait de communiquer. Cela soulève un problème en ce qui concerne la préférence accordée aux acteurs économiques par les pouvoirs publics.

Enfin, un débat existe dans ce dernier cas de figure sur la notion de « contrat collectif ». En effet, l'article L. 221-2 du code de la mutualité énonce qu'une opération de mutuelle collective facultative est celle où, sur la base d'un contrat souscrit par une personne morale, les membres de cette personne morale adhèrent librement à une mutuelle. Dans le cas d'un CCAS, qui est effectivement une personne morale, il est contestable que les habitants de la commune soient considérés comme des "membres" au sens de cette disposition. L'article ne conclut pas à ce sujet.

Cependant, après avoir interrogé différents acteurs présents dans la mise en place des mutuelles communales, il semble que les troisième et quatrième niveaux d'implication ne sont quasiment jamais rencontrés. La grande majorité des communes ayant mis en place une offre de ce type s'arrêtent au deuxième niveau d'implication.

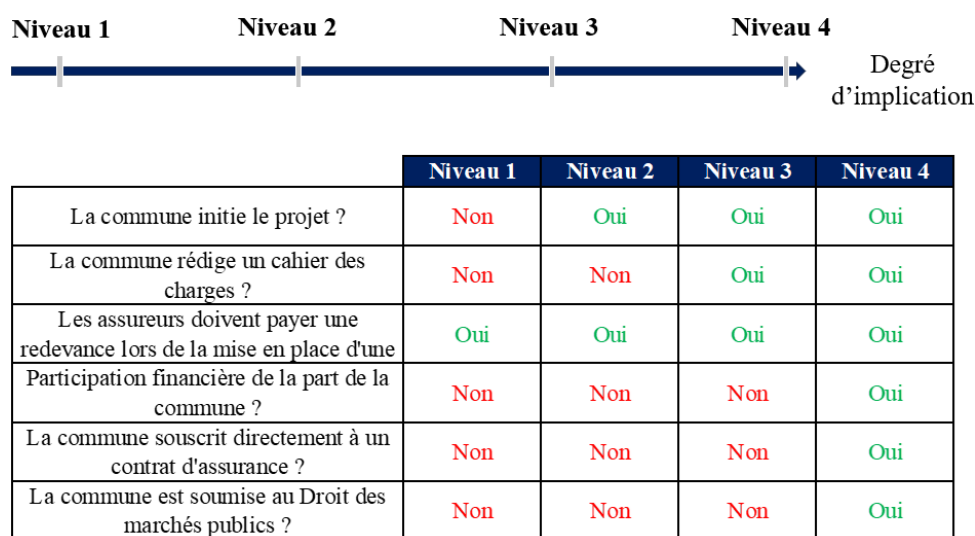


FIGURE 2.2 – Synthèse des niveaux juridiques

2.6 Les chiffres clés

Depuis l'initiative en **2013** avec la commune de Caumont-sur-Durance dans le Vaucluse [5], les mutuelles communales ont connu une expansion significative en France. En **2018**, Montreuil (93) est devenue la première ville de plus de 100 000 habitants à mettre en place une mutuelle communale [6], marquant ainsi une étape importante dans leur développement.

En **2022**, le nombre de communes proposant une mutuelle communale a dépassé les 3 000, ce qui représente une couverture pour environ 30 000 personnes. Cette croissance rapide témoigne de l'engouement et de la reconnaissance de l'importance de ces partenariats entre les communes et les mutuelles pour améliorer l'accès aux soins. En **2023**, une nouvelle étape est franchie avec le lancement de la première mutuelle communale à l'échelle d'une région en Île-de-France [7]. Ce progrès montre l'évolution continue du concept de mutuelle communale, démontrant son potentiel à répondre aux besoins de santé des résidents à différentes échelles territoriales.

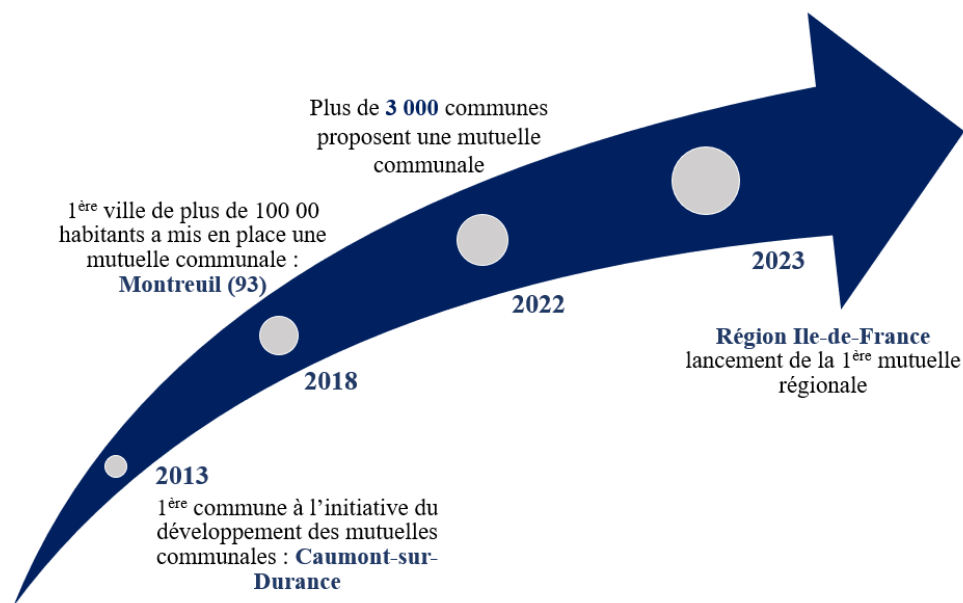


FIGURE 2.3 – Les chiffres clés des mutuelles communales

Ainsi, selon Thomas Frinault et Rémy Le Saout, respectivement enseignant-chercheur et sociologue [8], les maires apprécient la mise en place d'une mutuelle communale car elle leur offre des bénéfices symboliques immédiats, renforçant ainsi l'image sociale nécessaire à leur carrière politique, en particulier en milieu rural. Par exemple, le maire de Mornant a reçu en 2015 le trophée "de la Solidarité" de l'Association des Maires du Rhône pour avoir mis en place une mutuelle dans sa commune, ce qui souligne l'importance de l'attachement au territoire pour la légitimité politique des élus ruraux. C'est également pour cette raison que ces dernières années, les mutuelles communales connaissent un gain en popularité et qu'il est possible de croire à une démocratisation de ce dispositif car au-delà des milieux ruraux, les zones urbaines connaissent également un gain en popularité. En effet, la région Ile-de-France a lancé son offre de mutuelle régionale depuis le 1^{er} février 2023.

2.7 Un équilibre économique menacé pour les assureurs ?

Dans son article intitulé "Légitimité de l'intervention privée dans un service public" [9], l'économiste Christophe Defeuilley interroge la validité de l'intervention du secteur privé dans le domaine du service public. Cette question résulte de l'existence supposée d'un contrat équilibré où les deux parties trouvent un accord permettant de concilier leurs objectifs, à savoir la recherche de rentabilité pour l'une et la capacité de promouvoir et défendre l'intérêt général pour l'autre. En ce sens, il peut alors exister un conflit d'intérêts.

Il est difficile pour les organismes assureurs de maintenir une rentabilité économique car le portefeuille d'assurés est composé à plus de 70% de seniors. En effet, selon de nombreux retours d'expérience parmi les communes ayant mis en place une mutuelle communale, toutes affirment que ce sont majoritairement les seniors qui adhèrent à ce type de contrat. Par exemple, le directeur du CCAS de la ville de Saintes, Xavier Jouzel, indique que 75% des adhérents ont plus de 60 ans.

Il est légitime de se demander pourquoi les assureurs partenaires se positionnent alors sur ce marché de mutuelle communale s'il n'est à priori pas rentable. Malgré le risque d'anti-sélection, plusieurs raisons incitent les assureurs à investir sur le marché des mutuelles communales. D'une part, certaines mutuelles cherchent à renouer avec les principes de solidarité et d'altruisme. D'autre part, des mutuelles de petite taille, ayant peu d'adhérents, voient dans cette offre l'opportunité d'accroître leur part de marché.

Par ailleurs, pour certaines entités, le marché des mutuelles communales constitue une porte d'entrée stratégique leur permettant de vendre d'autres produits d'assurance simultanément. Bien que des critiques émanent des communes concernant ces pratiques, des retours d'expérience confirment cette réalité. De plus, en cas de difficultés financières, certains assureurs n'hésitent pas à ajuster les montants de cotisations pour rétablir leur équilibre.

Selon certains responsables de mutuelles, la pérennité des mutuelles communales repose sur un modèle économique axé sur la réduction maximale des frais de gestion. Ils estiment qu'il est possible d'assurer la couverture de la population fragile et à faibles revenus tout en maintenant l'équilibre financier des contrats en se retirant de l'environnement concurrentiel qui pousse à des dépenses publicitaires exponentielles.

Ainsi, la mutualisation au sein d'une commune semble être difficile s'il n'y a pas assez d'adhérents dans le portefeuille mais peut-être qu'à une échelle plus grande par exemple celle d'un département ou d'une région, il serait possible d'espérer une mutualisation des risques.

Il est donc plausible qu'un assureur opérant sur le marché des mutuelles communales souhaiterait établir des partenariats à une échelle plus vaste. Et c'est effectivement le cas. Après avoir échangé avec des professionnels travaillant dans le domaine des mutuelles communales, il est fréquent que ces dernières cherchent à établir des partenariats à une échelle supérieure à celle de la commune.

CHAPITRE

3

LES ENJEUX DE L'*OPEN DATA*

Ce segment du mémoire examine l'univers de l'*Open Data* dans le domaine de la santé, en mettant l'accent sur deux bases de données cruciales : la SNIIRAM et l'*Open DAMIR*.

3.1 Les principes de l'*Open Data*

L'idée fondamentale des *Open Data* est de rendre accessible au public des ensembles de données variées, et détenues par des institutions publiques, des organisations ou des entreprises. Afin d'être qualifiées de données ouvertes, ces dernières doivent respecter ces trois critères :

- **disponibilité** : les données doivent être facilement accessibles, de préférence via des téléchargements en ligne, et être disponibles à un coût de reproduction raisonnable ;
- **réutilisation et redistribution** : les données doivent être fournies sous des conditions qui permettent leur réutilisation et leur redistribution, y compris la possibilité de les combiner avec d'autres ensembles de données ;
- **participation universelle** : tout le monde doit être en mesure de les utiliser, de les réutiliser, favorisant ainsi une participation équitable et inclusive de la société dans son ensemble.

De plus, le Règlement Général sur la Protection des Données (RGPD), en vigueur depuis le 25 mai 2018, vise à responsabiliser la gestion des données personnelles par les organismes publics et privés, avec des principes clés incluant la transparence, l'obligation de signaler les fuites de données sous 72 heures, la création d'un Comité européen de la protection des données, et la prévision de sanctions progressives en cas de non-respect.

Ainsi, les données de santé en *Open Data* constituent des éléments essentiels car presque aucune donnée privée n'a été utilisée dans le cadre de ce mémoire.

3.2 SNIIRAM : la plus grande base de données de santé

Le Système National Inter Régimes d'Assurance Maladie (SNIIRAM) est la base de données considérée comme la plus importante en termes de volume en France. Elle a été créée en 1999 dans une Loi de Financement de la Sécurité Sociale (LFSS) et renseigne sur le parcours des patients et l'organisation du système de soins.

Cette initiative, inscrite dans l'article L161-28-1 du Code de la Sécurité Sociale, vise à atteindre plusieurs objectifs. Tout d'abord, elle a pour but de contribuer à une meilleure gestion de l'Assurance Maladie. De plus, elle cherche à améliorer la gestion des politiques de santé. Un autre objectif important est d'améliorer la qualité des soins en permettant une évaluation basée sur des données fiables. Enfin, cette initiative cherche à transmettre aux professionnels de santé des informations pertinentes sur leur activité, afin de faciliter leur prise de décision et leur pratique médicale. En revanche, l'accès à cette base de données est strictement réglementé. En effet, les utilisateurs doivent être identifiés de manière formelle et nominative, et leur activité est tracée.

Seuls les organismes autorisés, tels que les régimes d'Assurance Maladie, les services ministériels, les agences sanitaires et les organismes publics de recherche, peuvent y accéder contrairement aux organismes à but lucratif comme les organismes assureurs qui ne sont pas inclus dans la liste des entités autorisées.

3.3 L'Open DAMIR parmi les Open Data en santé

L'Open DAMIR, Dépense D'Assurance Maladie Inter-Régimes est une base gérée par la CNAM qui recense l'ensemble des dépenses et remboursements liés à l'Assurance Maladie excepté une partie des prestations hospitalières du secteur public. Cette base est en réalité une extraction du Système National Inter Régimes d'Assurance Maladie (SNIIRAM) mais l'Open DAMIR présente l'avantage d'être accessible gratuitement depuis le 25 janvier 2015.

Cette base est disponible en ligne [10] sous la forme de bases mensuelles, chacune valant environ 5 Go avec environ une trentaine de millions de lignes et chaque prestation est décrite par 55 variables.

Par ailleurs, afin de respecter la réglementation concernant l'anonymat des patients ainsi que des professionnels de santé, les données présentes dans la base sont affichées sous forme agrégées. Autrement dit, une ligne rassemble un groupe d'individus présentant des caractéristiques similaires (tranche d'âge, sexe, localisation, ...). Les montants affichés sont la résultante de la somme de tous les individus regroupés (dépense réelle engagée, montant remboursé par le régime obligatoire, quantité d'actes, etc). Concernant les localisations, elles sont limitées à 13 zones géographiques regroupées similairement aux régions administratives nouvellement crée en 2015.

Seule la base Open DAMIR a été présentée dans ce premier chapitre car il s'agit de la base principale à l'origine des travaux réalisés. Toutefois, les chapitres suivants présenteront toutes les Open Data qui ont contribué à ce mémoire.

Ce mémoire conduit ses travaux exclusivement au moyen de données ouvertes. Toutefois, l'incorporation des Open Data dans le processus de tarification des mutuelles communales en tant que complément aux données existantes pourrait susciter un intérêt grandissant en raison de son potentiel à accroître la transparence, à optimiser la pertinence des modèles tarifaires et à renforcer la confiance des adhérents.

Synthèse de la première partie

Le système d'assurance santé en France se compose de deux volets : le régime obligatoire et le régime complémentaire. D'un côté, la Sécurité Sociale est organisée en plusieurs branches, chacune ayant des missions spécifiques. Dans le cadre de ce mémoire seule l'Assurance Maladie sera étudiée. Elle a pour objectif d'aider financièrement les personnes à faire face aux conséquences des risques maladie, maternité, invalidité et décès en versant des prestations à ses bénéficiaires.

Ainsi, une prestation est une somme versée pour indemniser des dépenses liées aux risques couverts par l'Assurance Maladie. Ce sont les prestations en nature qui seront étudiées dans ce mémoire, c'est-à-dire le remboursement des dépenses engagées par l'assuré et ses ayants droit pour se soigner telles que la consultation du médecin, les médicaments, les examens et autres soins nécessaires.

Parfois ces prestations en nature ne sont pas suffisantes, d'où la nécessité d'avoir des acteurs complémentaires qui viennent alléger le reste à charge des patients. Ces derniers proposent principalement deux types de contrat : collectif ou individuel. Ce sont les contrats individuels qui seront étudiés dans le cadre de ce mémoire.

Par ailleurs, les actes de soins sont répertoriés dans différentes nomenclatures codifiées, ce qui permet à la Sécurité Sociale d'établir un tarif conventionnel pour chaque acte. Les assureurs santé présentent leurs garanties sous forme de grille décomposée par poste de soins, en se basant le plus souvent sur ce tarif de base.

Certaines communes, dans le cadre de leur action sociale souhaitent proposer une couverture santé à tarif préférentiel à leurs habitants. Elles mettent alors en place un partenariat avec un organisme d'assurance santé qui porte le nom de "mutuelle communale". Cependant, la grande majorité des adhérents de ce type d'offre sont en réalité des seniors de plus de 65 ans faisant courir un risque de sélection adverse et pose ainsi la question sur l'équilibre technique de l'organisme assureur.

Le mémoire établira ses travaux sur la base d'*Open Data* et plus précisément sur la base de l'*Open DAMIR* qui est une base gérée par la CNAM. La base recense l'ensemble des dépenses et remboursements liés à l'Assurance Maladie excepté une partie des prestations hospitalières du secteur public.

Deuxième partie

Les *Open Data* dans la création de la base de tarification

Le contenu de cette seconde partie porte dans un premier temps sur les travaux réalisés par Sandrine HUYNH dans le cadre de son mémoire intitulé "*Open Data* et assurance santé : l'union fait la force?" (2021) [11], car les principales données de ce mémoire sont issues de ses travaux. Ensuite, sont présentés de manière plus détaillée quelques retraitements supplémentaires conduisant à la création d'une nouvelle base de tarification. De plus, des variables externes ont été ajoutées à la base de données et proviennent de différentes sources d'*Open Data*. Elles sont également mises en avant dans cette partie. Enfin, une analyse des valeurs extrêmes et quelques statistiques sur la nouvelle base qui a été créée sont présentées afin de délimiter le périmètre d'étude du mémoire.

CHAPITRE

4

LES RETRAITEMENTS DE L'OPEN DAMIR

Ce chapitre présente la base de travail initial de ce mémoire ainsi que les retraitements effectués par Sandrine HUYNH. Des vérifications de cohérence seront effectuées afin de s'appropriier les données.

Sandrine HUYNH, dans son mémoire intitulé "*Open Data* et Assurance santé : l'union fait la force?" (2021) a traité les bases *Open* DAMIR 2018 et 2019 à l'aide des machines virtuelles. En effet, ses travaux rigoureux ont permis d'aboutir à la naissance d'une première base de données déjà retraitée de l'*Open* DAMIR au sein du cabinet de conseil Actuelia. Ses travaux peuvent alors servir de base initiale pour débiter les travaux de ce mémoire, mais est-ce vraiment la meilleure option ? Ne vaut-il pas mieux utiliser les données les plus récentes de l'*Open* DAMIR pour une tarification optimale ?

4.1 Choix de la base de données

Dans cette section, il sera discuté du choix de la base de données pour commencer les travaux de tarification.

Il faut choisir entre :

- les bases de données les plus récentes disponibles de l'*Open* DAMIR non retraitées datant de 2020 et 2021 ;
- une base *Open* DAMIR post retraitement datant de 2018 et 2019.

Le choix s'est naturellement porté pour la deuxième option, c'est-à-dire pour la base de données de l'*Open* DAMIR où de nombreux retraitements ont déjà été effectués. La raison principale soutenant le choix de la base post retraitement est la difficulté logistique à ouvrir les bases de données mensuelles de l'*Open* DAMIR sur les ordinateurs de l'entreprise.

D'autres raisons ont appuyé le choix de la seconde option comme la présence des données COVID en 2020 et 2021. En effet, en raison du confinement en 2020, le nombre de consultations a considérablement diminué, tandis qu'en 2021, à la suite du déconfinement, un rattrapage des prestations s'est produit, entraînant une surconsommation de prestations médicales. Il semble difficile d'établir un tarif en se reposant sur ces années exceptionnelles. Quant à la base de l'année 2022, elle n'était pas publiée à la date de l'étude.

4.2 L'Open DAMIR

Dans cette section, l'Open DAMIR originale est présentée. Il s'agit de la base avec laquelle Sandrine HUYNH a travaillé. Parmi les 55 variables présentes disponibles en annexe 12.5, seules les principales seront présentées ci-dessous.

BENEFICIAIRE	
AGE_BEN_SNDS	Tranche d'Age du Bénéficiaire au moment des soins, segmentée par classe de 10 ans, excepté pour les 0-19 ans et 80+ qui sont regroupés ensemble
BEN_CMU_TOP	Identifie les bénéficiaires de la CMU-C
BEN_QLT_COD	Qualité du Bénéficiaire (souscripteur ou type d'ayant droit)
BEN_RES_REG	Région administrative de résidence du Bénéficiaire
BEN_SEX_COD	Sexe du Bénéficiaire
PERIODE	
SOI_ANN	Année de Soins
SOI_MOI	Mois de Soins
PRESTATION	
PRS_NAT	Nature de Prestation, il s'agit d'un code acte par exemple le code 1192 correspond à une "téléconsultation généraliste"
PRS_REM_TAU	Taux de Remboursement de la Sécurité Sociale
PRS_REM_TYP	Type de Remboursement
INDICATEURS BRUTS	
PRS_ACT_COG	Coefficient Global correspondant au coefficient intervenant dans la nomenclature des actes NGAP
PRS_ACT_NBR	Dénombrement
PRS_ACT_QTE	Quantité d'acte
PRS_DEP_MNT	Montant du dépassement d'honoraires par rapport au tarif de base de la Sécurité Sociale
PRS_PAI_MNT	Montant de la Dépense, c'est-à-dire les frais réellement engagés par les patients
PRS_REM_BSE	Base de Remboursement soit le tarif de base de la Sécurité Sociale
PRS_REM_MNT	Montant Versé/Remboursé par la Sécurité Sociale
INDICATEURS PREFILTRES	
FLT_ACT_COG	Coefficient Global de la Prestation Préfiltré
FLT_ACT_NBR	Dénombrement de la Prestation Préfiltré
FLT_ACT_QTE	Quantité de la Prestation Préfiltrée
FLT_DEP_MNT	Montant du Dépassement de la Prestation Préfiltré
FLT_PAI_MNT	Montant de la Dépense de la Prestation Préfiltré
FLT_REM_MNT	Montant Versé/Remboursé Préfiltré

FIGURE 4.1 – Les variables de l'Open DAMIR

Les indicateurs préfiltrés

Deux types d'indicateurs existent dans la base de données :

- les indicateurs bruts avec "PRS_" en préfixe ;
- les indicateurs préfiltrés avec "FLT" en préfixe.

Par exemple, un acte de soin réalisé de nuit en situation d'urgence dans un hôpital fait l'objet de plusieurs majorations sur le montant des dépenses de santé :

- une majoration d'un acte réalisé de nuit ;
- une majoration d'un acte réalisé en urgence.

Ces montants de majoration sont alors affectés dans les indicateurs préfiltrés. Cette approche permet d'éviter le double comptage.

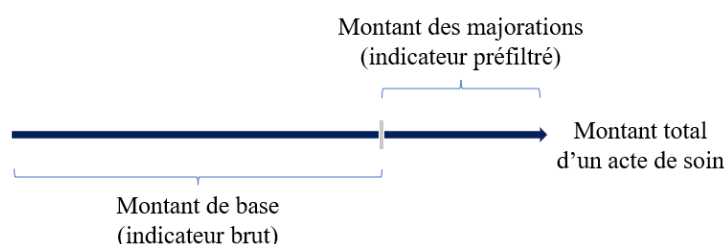


FIGURE 4.2 – Composition d'un acte de soin majoré

Des montants de dépense négatifs

Dans la base *Open* DAMIR, les montants de remboursement par le régime obligatoire incluent les montants des participations forfaitaires et franchises si elles existent. En contrepartie, ces montants qui restent à la charge des patients se voient attribuer une ligne supplémentaire avec un montant négatif. Concernant les autres montants négatifs, ce sont souvent des opérations de régularisation qui en sont responsables et sont alors considérées comme des anomalies puis exclues de la base d'étude.

La différence entre dénombrement et quantité d'actes

Les deux variables permettent de compter le nombre d'actes par lignes agrégées. Cependant, malgré leur concordance générale, des différences apparaissent dans des cas spécifiques :

- Pour les transports, la quantité d'actes indique le nombre de factures, tandis que le dénombrement compte le nombre de courses effectuées. ;
- Pour les frais de séjour, la quantité d'actes indique encore le nombre de factures, tandis que le dénombrement compte le nombre de jours hospitalisés.

Ainsi, il sera nécessaire de choisir entre la variable de quantité ou de dénombrement en fonction de l'acte étudié. De plus, un certain nombre de régimes d'Assurance Maladie ne fournissent pas le dénombrement des actes, mais seulement la quantité, entraînant la présence de valeurs manquantes pour la variable de dénombrement. C'est pour cette raison que, pour la suite du mémoire, la variable "quantité d'actes" sera privilégiée pour les travaux.

4.3 Les traitements déjà réalisés

Un nombre conséquent de retraitements a été effectué par Sandrine HUYNH, il s'agira d'en présenter quelques-uns à titre indicatif. Par la suite, d'autres retraitements complémentaires réalisés cette fois-ci dans le cadre de ce mémoire seront mis en lumière.

4.3.1 Récapitulatif des retraitements déjà effectués

Sandrine HUYNH dans son mémoire choisit de travailler sur une profondeur d'historique de données de deux ans, car la santé est une branche courte. Elle parvient à réduire le nombre de variables initiales de 55 à 17 en entreprenant une série d'étapes.

Tout d'abord, les lignes et variables hors périmètre d'études ont été supprimées, c'est-à-dire que seules les variables concernant le régime général de la Sécurité Sociale et sa branche maladie sont pertinentes dans le cadre de l'étude.

Par exemple, la base de données est filtrée en fonction de la nature des prestations (**ASU_NAT**) pour ne conserver que les lignes correspondant à la branche maladie, en excluant toutes les autres modalités.

De la même manière, les lignes indiquant des bénéficiaires de la CMU-C (**BEN_CMU_TOP**) et toutes les enveloppes différentes du régime général (**CPT_ENV_TYP**) ont été supprimées.

Par ailleurs, Sandrine HUYNH travaillait à l'époque avec une base de données privée d'une mutuelle appelée VirtuaMut'. Par conséquent, toutes les variables de l'*Open* DAMIR qui n'étaient pas présentes dans la base de VirtuaMut' ont été éliminées, à condition qu'elles ne servaient pas à construire une nouvelle variable pertinente. Par exemple, les variables de dates de soins n'étaient pas au même format, elles ont été retraitées.

Après ces étapes de suppression, quelques vérifications traditionnelles ont été effectuées pour assurer la qualité des données, telles que la suppression des lignes présentant un montant négatif ou des quantités négatives. Ces lignes sont probablement liées à des opérations de régularisation de la Sécurité Sociale ou encore à des lignes affichant un taux de remboursement de la Sécurité Sociale supérieur à 100%, ce qui n'est pas possible.

Cependant parmi ces vérifications, une attention particulière a été attribuée à la variable concernant le type de remboursement (**PRS_REM_TYP**), car il est important que les quantités des compléments, majorations et suppléments d'acte soient bien nulles. Si ce n'est pas le cas, cela pourrait biaiser les lois de fréquence par la suite lors de la tarification. Par exemple, la consultation chez un médecin le 14 juillet, qui est un jour férié peut entraîner une majoration des dépenses engagées. Cela se traduit dans la base par l'apparition d'une ligne supplémentaire pour l'acte associé avec le montant de majoration affiché. Le problème a été géré en implémentant une fonction indicatrice pour toutes les lignes de majorations afin qu'elles ne soient pas comptées en double dans le calcul de la fréquence. Ensuite les montants supplémentaires ont été fusionnés avec leurs actes de référence dans la mesure du possible.

Dans la suite, bien que Sandrine HUYNH soit consciente qu'elle aurait pu effectuer davantage de retraitements concernant les données aberrantes ou manquantes telles que le genre, l'âge ou une région non renseignée dans la base, elle a décidé de les supprimer par manque de temps. Cela correspond à "l'analyse des cas complets".

Par ailleurs, pour chaque ligne, les variables de sous-familles d'actes (SS_FAM), de familles d'actes (FAM) et de grands postes de soin (POSTE) ont été une créée. L'intérêt de ce retraitement fera l'objet d'une présentation plus approfondie dans la section suivante.

En fonction de la classification, un montant de remboursement de la complémentaire santé VirtuaMut¹ a été affecté. Cependant, ces montants ne trouveront pas d'application au sein de ce mémoire. En effet, les grilles de garanties diffèrent et nécessiteront l'ajout de nouveaux montants de remboursement. Ainsi, ce montant constituera par la suite le "coût" dans le cadre d'une tarification par fréquence x coût. Le modèle sera introduit dans la partie 3 du mémoire.

Enfin, tous les retraitements évoqués ci-dessus ont été répétés sur chacune des 24 bases mensuelles utilisées afin d'obtenir une base globale post retraitement.

4.3.2 La classification des actes de soins

Un des objectifs du mémoire de Sandrine HUYNH [11] était d'établir un tarif global pour une mutuelle privée du nom de VirtuaMut². Par ailleurs, un portefeuille d'assurance couvre un risque spécifique, ce qui signifie que les pertes sont supposées suivre la même distribution de probabilité, impliquant ainsi une tarification par type de garantie. C'est donc pour cette raison que Sandrine HUYNH a d'abord procédé à l'établissement de plusieurs "sous-tarifs" correspondant à chaque type de soin. Le tarif final de la mutuelle serait alors calculé en additionnant tous ces sous-tarifs.

Dans la base *Open* DAMIR, les données ne sont pas classées par poste de soins rendant difficile la segmentation de la base pour une tarification. En revanche, la variable "PRS_NAT" renseigne un code de 4 chiffres permettant de retrouver la nature des prestations. Pour information, une modalité se présente de cette manière :

PRS_NAT	Libellé Nature de Prestation en B2	Libellé Nature de Prestation
0		SANS OBJET
1096	TTE	TELECONSULTATION MEDECIN TRAITANT AVEC EHPAD
1097	TDT	TELE EXPERTISE DOSSIER TRAITANT
1098	U03	CONSULTATION CCMU 3
1099	U45	CONSULTATION CCMU 4 ET 5
1100	RNM	PROTOCOLE MURAIN - BILAN VISUEL

FIGURE 4.3 – Exemples de modalité et libellé de la variable "PRS_NAT"

Le code à 4 chiffres dans la colonne "PRS_NAT" sera appelé "code acte" dans la suite du mémoire, il permet d'identifier la nature de la prestation.

1. Nom fictif de la mutuelle
2. Nom fictif de la mutuelle

Il faut alors classer les 1 080 modalités de la variable PRS_NAT pour effectuer une tarification. Ainsi, les variables suivantes ont été créées :

- les **sous-familles** qui correspondent à la maille la plus fine utilisée et priorisée pour la tarification, cette classe contient par exemple la distinction entre les consultations/visites et actes médicaux techniques ;
- les **familles** sont des regroupements de sous-familles, qui seront utilisées dans la tarification si l'information par sous-famille prise individuellement n'est pas assez précise ou complète ;
- les **grands postes** tels que les soins courants, le dentaire, l'optique, ... permettent d'effectuer des analyses à une plus grande échelle et sont à leur tour des regroupements de familles.

Méthode de classification des codes actes

Pour rappel, Sandrine HUYNH disposait d'une base de données de la mutuelle VirtuaMut³ dans le cadre de son mémoire. Cette base possédait également des codes d'actes qui étaient déjà classés par la mutuelle partenaire, ce qui a permis d'entamer la classification en effectuant des correspondances. Ensuite, un rapprochement a été fait avec un fichier des *Données Nationales : dépenses d'Assurance Maladie (hors prestations hospitalières) du régime général*. Ce dernier fichier est une extraction de l'*Open DAMIR* et possède une variable "l_serie" qui propose, pour les codes actes, un libellé plus compréhensible que l'*Open DAMIR*. Grâce à cette opération, plus d'un tiers des 1 080 actes ont pu être classés. Par la suite, la consultation de la section "Q&A" du site *data.gouv.fr* suivie de nombreuses recherches internet ont permis d'avancer dans le processus de classification. Enfin, lorsque les bases disponibles et données d'internet ne suffisaient plus, un avis d'expert a été sollicité.

3. Nom fictif donné à la mutuelle partenaire

Ainsi, la classification retenue des actes dans le cadre du mémoire de Sandrine HUYNH "*Open Data et Assurance santé : l'union fait la force ? (2021)*" est gardée. Elle est présentée dans la figure ci-dessous :

Grands postes	Familles d'actes	Sous-familles d'actes
Soins courants	Honoraires médicaux	Consultations/visites Actes médicaux (techniques) Autres honoraires médicaux
	Honoraires paramédicaux	Auxiliaires médicaux (kiné, infirmiers, ...)
	Analyse et examen de laboratoire	Analyse médicale et examens labo
	Imagerie médicale	Actes d'imagerie, de radiologie et ostéodensitométrie
	Transport, ambulances	Transport
	Pharmacie	Petit appareillage Grand appareillage Pharmacie Vaccin anti-grippe
Hospitalisation	Hospitalisation	Frais de séjour Honoraires et actes chirurgicaux Forfait journalier Chambre particulière Lit accompagnant Autres - hospitalisation
Dentaire	Dentaire	Prothèse Soins dentaires Parodontologie Implantologie Orthodontie Autres - dentaire
Optique	Optique	Monture Verres Chirurgie œil Lentilles Autres - Optique
Aides auditives	Audio	Audioprothèse Pile, accessoire Autres - audio
Autres	Cure thermique	Cure thermique
	Médecine douce	Osthéopathie, diététique, chiropractie, ...
	Prévention	Prévention
	Prestations supplémentaires	Maternité Autres (aides, inclassable, ...)

FIGURE 4.4 – Segmentation principale retenue des actes

4.3.3 Backtesting des classifications

La classification présentée ci-dessus a un impact fort sur la tarification. Afin de s'approprier les données, des vérifications ponctuelles ont été effectuées. Idéalement, il faudrait vérifier chaque code acte individuellement, mais par manque de temps, un échantillon de 50 codes actes a été sélectionné aléatoirement pour vérification.

Concrètement, un classement est réalisé sans tenir compte de ceux déjà effectués par Sandrine HUYNH. Ensuite, une comparaison est effectuée entre les deux résultats. Ainsi, dans la figure ci-dessous, la classification affichée pour chaque code correspond à celle qui avait déjà été réalisée, et la dernière colonne intitulée "Check" indique si l'affectation de la "sous-famille d'acte" est la même pour la nouvelle version.

PRS_NAT	Libellé Nature de Prestation	Sous-famille d'acte	Poste	CHECK
1096	TELECONSULTATION MEDECIN TRAITANT AVEC EHPAD	Consultations/visites	Soins courants	OK
1164	TÉLÉ CONSULTATION - ALD ET / OU EHPAD	Consultations/visites	Soins courants	OK
1312	ACTES DE SPECIALITE EN K	Actes médicaux (techniques)	Soins courants	OK
1975	FRANCHISE HORS TIERS PAYANT SUR MEDICAMENT	Autres (aides, inclassable, ...)	Autres	OK
2112	FRAIS DE SEJOUR SUPPLEMENTAIRE AU GHS	Frais de séjour	Hospitalisation	OK
1171	RÉMUNÉRATION MÉDECIN TRAITANT CENTRES DE SANTÉ	Autres (aides, inclassable, ...)	Autres	OK
1408	COURONNE ZIRCONNE RAC MODERE	Prothèse	Dentaire	OK
9211	VACCIN (MILITAIRES) / VACCIN GRIPPE (CRPCEN)	Autres (aides, inclassable, ...)	Autres	NOK
1612	CONTRAT DE BONNES PRATIQUES ZONE URBAINE	Autres (aides, inclassable, ...)	Autres	OK

FIGURE 4.5 – Backtesting de la classification sur un échantillon de 50 codes actes

Sur la figure, seules les premières lignes du backtesting sont présentées pour une meilleure visibilité. La suite des résultats peut être retrouvée en annexe 12.5. Dans l'échantillon sélectionné, une seule ligne ne correspond pas entre les classifications. D'un côté, le code VCC/VAC correspondant à "VACCIN (MILITAIRES) / VACCIN GRIPPE (CRPCEN)" a été classé dans la sous-famille "Autres", tandis que de l'autre côté, il a été classé dans la sous-famille "Vaccin antigrippe". Le changement de classe a été effectué, cependant son impact est négligeable, car les montants associés sont très faibles dans la base de données par rapport au montant global de chaque poste.

4.3.4 La segmentation pour la tarification

Pour la mise en place des tarifs, l'approche envisagée repose sur l'utilisation de modèles linéaires généralisés, qui seront détaillés dans la partie 3 du mémoire. Cette démarche requiert toutefois une vérification préalable de la disponibilité de données pour exécuter ces tarifications. Cette validation a déjà été effectuée par Sandrine HUYNH, et il a été constaté que toutes les sous-familles de la base *Open* DAMIR possèdent une quantité d'actes suffisante pour permettre la mise en œuvre de ces modèles.

Néanmoins, même si l'approche idéale consisterait à établir des tarifs pour chacune des 36 sous-familles d'actes identifiées dans la figure 4.4, ce processus serait excessivement chronophage et complexe. En effet, il est possible que les données disponibles dans l'*Open* DAMIR ne soient pas adéquates pour chaque sous-famille d'actes. En d'autres termes, pour certaines sous-familles, la classification comporte de nombreuses incertitudes, ce qui risquerait de mettre en péril la fiabilité des tarifs obtenus. Dans cette perspective, il est préférable d'opter pour une tarification plus globale, même si cela peut introduire un certain biais.

Ainsi, la segmentation tarifaire permet de déterminer quelles sous-familles seront soumises aux modèles linéaires généralisés (GLM). La segmentation retenue coïncide en grande partie avec celle présentée dans le mémoire de Sandrine HUYNH, à l'exception de quelques sous-familles surlignées en rose. Initialement, ces sous-familles avaient été exclues en raison des quantités d'actes jugées insuffisantes dans le portefeuille de la mutuelle partenaire, VirtuaMut'. Cela visait à aligner les données sur la base privée de VirtuaMut'. Cependant, pour les besoins de ce mémoire, aucune donnée privée n'est utilisée, éliminant ainsi la nécessité d'exclure ces sous-familles.

Grands postes	Familles d'actes	Sous-familles d'actes	Choix de segmentation	
Soins courants	Honoraires médicaux	Consultations/visites	1 segment de tarification	
		Actes médicaux (techniques)	1 segment de tarification	
		Autres honoraires médicaux	Non considéré	
	Honoraires paramédicaux	Auxiliaires médicaux (kiné, infirmiers, ...)	1 segment de tarification	
	Analyse et examen de laboratoire	Analyse médicale et examens labo	1 segment de tarification	
	Imagerie médicale	Actes d'imagerie, de radiologie et ostéodensitométrie	1 segment de tarification	
	Transport, ambulances	Transport	Non considéré	
Hospitalisation	Hospitalisation	Petit appareillage	1 segment de tarification	
		Pharmacie		
		Vaccin anti-grippe		
		Frais de séjour		
		Honoraires et actes chirurgicaux	1 segment de tarification	
Dentaire	Dentaire	Grand appareillage	1 segment de tarification	
		Forfait journalier		
		Chambre particulière		
		Lit accompagnant		
		Autres - hospitalisation	Non considéré	
Dentaire	Dentaire	Prothèse	1 segment de tarification	
		Soins dentaires	1 segment de tarification	
		Parodontologie	1 segment de tarification	
		Implantologie	1 segment de tarification	
		Orthodontie	1 segment de tarification	
Optique	Optique	Autres - dentaire	Non considéré	
		Monture	1 segment de tarification	
		Verres	1 segment de tarification	
		Lentilles	1 segment de tarification	
		Chirurgie œil	Non considéré	
Aides auditives	Audio	Autres - Optique	Non considéré	
		Audioprothèse	1 segment de tarification	
		Pile, accessoire	1 segment de tarification	
Autres	Prestations supplémentaires	Autres - audio	Non considéré	
		Cure thermique	Non considéré	
		Médecine douce	Ostéopathie, diététique, chiropractie, ...	Non considéré
		Prévention	Prévention	Non considéré
		Maternité	Maternité	Non considéré
Autres (aides, inclassable, ...)	Autres (aides, inclassable, ...)	Non considéré		

FIGURE 4.6 – Segmentation pour la tarification

En ce qui concerne les sous-familles non prises en compte, ces segments ont généralement été écartés en raison des incertitudes entourant la classification des prestations. Par exemple, pour la sous-famille "Maternité", il apparaît que la plupart des prestations associées sont dispersées et incluses dans d'autres sous-familles.

En définitive, 18 segments tarifaires ont été sélectionnés pour l'application des GLM. Les sous-familles restantes seront néanmoins incorporées dans le processus tarifaire par le biais d'une méthode alternative, à savoir la "statistique directe", qui sera également exposée dans la troisième section du mémoire.

4.3.5 Base à l'issue des retraitements

Voici la base de données obtenue à l'issue des retraitements qui servira pour le début des travaux de ce mémoire. Le nom de **OLD_DAMIR**⁴ lui est attribué pour simplifier son appellation par la suite.

Nom variable	Libellé
CODE ACTE	Code identifiant la nature de prestation
ANNEE_SOINS	L'année de soin
ANNEE_REGLT	L'année de règlement
SS_FAM	La sous-famille de l'acte
FAM	La famille de l'acte
POSTE	Le poste de soin de l'acte
QLT	La qualité des adhérents (souscripteur/ conjoint ou ayant droit)
SEXE	Le sexe des adhérents
CLASSE_AGE	La classe d'âge des adhérents par tranche de 10 ans sauf pour les moins de 20 ans et plus de 80 ans (0-19 ; 20-29 ans; .. ; 80+)
REGION	La région de résidence des adhérents de la ligne
DEPENSE	Le montant agrégé des dépenses engagées des adhérents de la ligne
MT_RO	Le montant RO agrégé remboursé aux individus de la ligne
MT_RC_RN	Le montant RC de la garantie A de VirtuaMut' agrégé remboursé aux individus de la lignes
MT_RC_RS	Le montant RC de la garantie B de VirtuaMut' agrégé remboursé aux individus de la lignes
QTE_ACTE	La quantité d'actes agrégés
EXPO TOTALE	L'exposition totale de la ligne (nombre de bénéficiaires exposés)
DENOMBREMENT	Le denombrement des actes

FIGURE 4.7 – Variables de la base OLD_DAMIR à l'issue des retraitements

Pour rappel, les montants RO et RC correspondent respectivement aux montants remboursés par le régime obligatoire et le régime complémentaire. De plus, parmi ces 17 variables, certaines ont été supprimées telles que les variables MT_RC_RN et MT_RC_RS qui correspondent respectivement aux montants remboursés par les garanties A et B de la complémentaire santé VirtuaMut'. En effet, dans la suite du mémoire, la tarification portera sur des garanties de mutuelles communales qui diffèrent donc des garanties utilisées dans le mémoire de Sandrine HUYNH.

Ainsi, à l'issue de ces retraitements, Cette base OLD_DAMIR constituera le point de départ des travaux de ce mémoire. Elle présente cependant des forces et des faiblesses pour la suite de l'étude.

4. L'appellation "OLD_DAMIR" a été attribuée car cette base de données sera utilisée pour créer une nouvelle base "NEW_DAMIR" qui servira à la tarification.

Les forces de la base OLD_DAMIR

- les retraitements déjà effectués sont assez généraux, ce qui permet la réutilisation de la base ;
- évite d'avoir recours à des machines virtuelles pour traiter les bases de l'*Open* DAMIR ;
- contient les dépenses de santé réelles par classe d'âge et par région. Cela permettra dans la suite de recalculer les montants de l'organisme complémentaire ;
- possède 304 393 lignes donc traitables sur le logiciel R.

Les faiblesses de la base OLD_DAMIR

Bien qu'un travail rigoureux ait été effectué de la part de Sandrine HUYNH, elle n'avait pas prévu que la base de données serait reprise 3 années plus tard par une autre personne. En effet, la base fournie est une version de sauvegarde intermédiaire extraite de l'*Open* DAMIR sur SAS. Elle n'avait pas pour objectif d'obtenir une base entièrement retraitée sur SAS, mais plutôt de générer une base importable sur R.

- les montants de dépenses datent de 2018 et 2019, il faudra les mettre à jour en tenant compte de l'inflation ;
- les variables base et taux de remboursement de la Sécurité Sociale ont été supprimées. Elles sont donc absentes ;
- les montants de la variable MT_RO ne sont pas complets.

Il est à noter que les montants remboursés par la Sécurité Sociale ne sont pas complets dans la version fournie en raison d'un problème d'output ou de réagrégation. Cependant, dans le mémoire de Sandrine HUYNH, les montants étaient corrects.

Ainsi, il a été décidé d'utiliser la base OLD_DAMIR pour entamer les travaux de ce mémoire. Cependant, l'information contenue dans cette base est disponible à une granularité régionale. Or, l'un des objectifs du mémoire consiste à établir une tarification pour les mutuelles communales, ce qui requiert une information à une maille plus fine que celle de la région. Par conséquent, des ajustements seront nécessaires pour combler les faiblesses de cette base.

CHAPITRE

5

LA CRÉATION D'UNE NOUVELLE BASE DE TARIFICATION

Dans le chapitre précédent, plusieurs faiblesses de la base OLD_DAMIR ont été identifiées. Parmi celles-ci, les montants remboursés par le Régime Obligatoire ne sont pas complets, et la granularité de l'information n'est pas suffisamment fine pour répondre aux objectifs du mémoire. Ainsi, ce chapitre vise à présenter, dans un premier temps, les retraitements effectués concernant les montants remboursés par la Sécurité Sociale. Ensuite, il mettra en avant une solution pour obtenir une vision de l'information à une échelle plus fine.

5.1 Le retraitement des montants remboursés par la Sécurité Sociale

En effectuant quelques vérifications traditionnelles de valeur manquantes ou aberrantes, il s'est avéré que la base de données présentait un montant total remboursé par le régime obligatoire de 187 millions d'euros sur les années 2018 et 2019. Le montant est manifestement faible par rapport aux ordres de grandeur trouvables sur internet qui sont de l'ordre d'une centaine de milliards d'euros par an. Une attention particulière doit être accordée à cette variable, car elle revêt une importance fondamentale pour le calcul des montants de remboursements de la complémentaire santé dans la suite du mémoire. Ces montants joueront également un rôle crucial dans la mise en place de la tarification.

5.1.1 La méthode de retraitement

L'idée retenue est de recalculer les valeurs de la variable **MT_RO** qui correspondent aux valeurs de remboursements effectués par la Sécurité Sociale en appliquant l'équation 1.1. Pour cela, il faut deux variables qui sont absentes de la base de données : **une base et un taux de remboursement**.

Il faut donc réussir à retrouver pour chaque code acte, c'est-à-dire pour chaque ligne de la base de données, son taux de remboursement ainsi que sa base de remboursement correspondante. Idéalement, il faudrait disposer d'une liste complète de tous les codes actes existants avec leurs taux de remboursement et bases de remboursement associées. Cependant, aucune liste publique de cette nature n'a été trouvée.

La solution retenue fut d'effectuer une correspondance entre les codes actes de la variable "CODE_ACTE" et ceux présents dans une autre base de données nationale gérée par la CNAM qui contient les deux indicateurs énoncés précédemment. Cette base externe intitulée «*Dépenses d'Assurance Maladie hors prestations hospitalières par caisse primaire/département*» sera renommée **DAM_DPT** pour la suite du mémoire.

Introduction de la base DAM_DPT

Comme l'*Open* DAMIR, cette base de données disponible sur le site de *data.gouv.fr*[12] est également extraite de la SNIIRAM. Elle contient tous les remboursements mensuels effectués par le régime général de l'Assurance Maladie, à l'exception des prestations hospitalières. Ces remboursements sont répertoriés par caisse primaire/département, par type de prestations (soins et prestations en espèces), type d'exécutants (médecins par spécialité, chirurgiens dentistes, auxiliaires médicaux, laboratoires d'analyse, pharmaciens, etc.) et par type de prescripteurs.

A l'instar de la base *Open* DAMIR, une année est décomposée en 12 fichiers mensuels, comptant entre 3 et 4 millions de lignes pour chaque fichier. Par conséquent, la taille des fichiers est plus réduite et traitable sur R. La base de données présente également une particularité avec des lignes agrégées, mais elle ne peut pas remplacer complètement l'*Open* DAMIR, car elle ne contient pas la variable d'âge, qui sera une information indispensable durant l'étape de tarification.

Cette base présente tout de même une double utilité, car elle permet en premier lieu de recalculer les montants remboursés par la Sécurité Sociale. De plus, elle offre l'avantage de fournir l'information à l'échelle départementale, d'où le nom "_DPT". Ainsi, elle sera d'une grande utilité dans la suite du mémoire pour obtenir une granularité plus fine.

Parmi les 37 variables disponibles dans DAM_DPT, celles qui sont intéressantes pour le retraitement des montants RO sont :

- **PRS_NAT** : l'équivalent de CODE_ACTE ;
- **REM_TAU** : le taux de remboursement **unitaire** de la Sécurité Sociale ;
- **REC_MON** : les bases de remboursement **agrégées** de la ligne ;
- **DENOMBREMENT** : le nombre d'actes de la ligne.

Les bases de remboursements sont agrégées, ce qui signifie qu'il s'agit d'une somme de bases de remboursement. Or, il faut revenir à une base de remboursement unitaire avant de pouvoir effectuer la correspondance avec la base OLD_DAMIR. Par conséquent, les bases de remboursements unitaires sont stockées dans une nouvelle variable "BR_UNITAIRE" en divisant REC_MON par DENOMBREMENT. Ensuite, par correspondance en utilisant la clé CODE_ACTE et

5.1. LE RETRAITEMENT DES MONTANTS REMBOURSÉS PAR LA SÉCURITÉ SOCIALE

PRS_NAT, les taux et bases de remboursements du régime obligatoire sont affectés dans la base de données OLD_DAMIR. Il est à noter que la distinction entre les variables DENOMBREMENT et QTE_ACTE n'existe pas dans DAM_DPT.

Enfin, dans la base OLD_DAMIR, il ne reste plus qu'à appliquer la formule suivante pour obtenir les montants remboursés par la Sécurité Sociale corrigée :

$$MT_RO_Corrige = TR \times BR_UNITAIRE \times QTE_ACTE \quad (5.1)$$

Avec :

MT_RO_Corrige : Montant corrigé remboursé par la Sécurité Sociale

TR : Taux de remboursement de la Sécurité Sociale

BR_UNITAIRE : Base de remboursement unitaire de la Sécurité Sociale

QTE_ACTE : La quantité d'actes de la ligne permettant d'obtenir un montant agrégée.

5.1.2 Des codes actes manquants

Cette correspondance a permis de recalculer la grande majorité des montants RO. Cependant, 38 modalités parmi les 482 de la variable "CODE_ACTE" dans OLD_DAMIR ne trouvaient pas de correspondance dans la variable "PRS_NAT" de DAM_DPT.

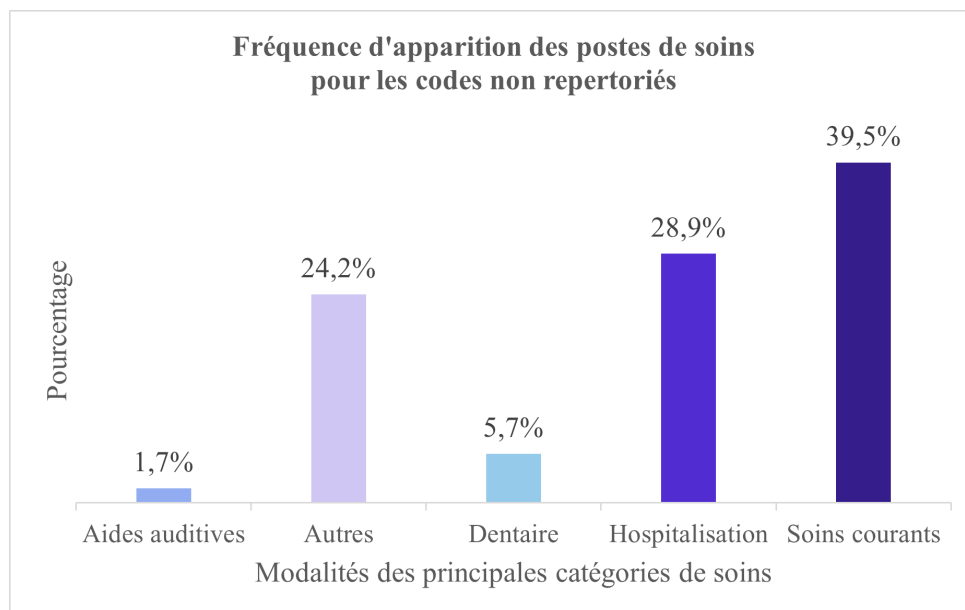


FIGURE 5.1 – Fréquence d'apparition des postes de soins pour les codes non répertoriés

Il est observé que plus de 50% des codes actes n'ayant pas trouvé de correspondance sont classés dans les postes d'hospitalisation et autres. Cette situation est plausible car la base DAM_DPT ne contient pas de code liés aux prestations hospitalières. De plus, il est également possible que parmi les soins courants, certaines consultations chez le médecin ou autres actes médicaux techniques aient été effectuées dans un hôpital, auquel cas, le code n'est pas répertorié dans la base DAM_DPT.

La part des montants non corrigés est égale à 0,47% du montant total remboursé par la Sécurité Sociale. Il a donc été décidé que l'impact serait négligeable. Ainsi, les valeurs ont été laissées telles quelles.

5.1.3 Des montants retraités anormaux

Une fois les valeurs de la variable "MT_RO" recalculée, il est parfois observé que les montants remboursés par la Sécurité Sociale sont supérieurs aux dépenses réelles engagées par les patients. Autrement dit, les patients s'enrichissent grâce au régime obligatoire, ce qui n'est pas possible par principe indemnitaire¹. Cette anomalie a été constatée pour 1,2 % de la base, ce qui équivaut à 3 652 lignes sur un total de 304 393.

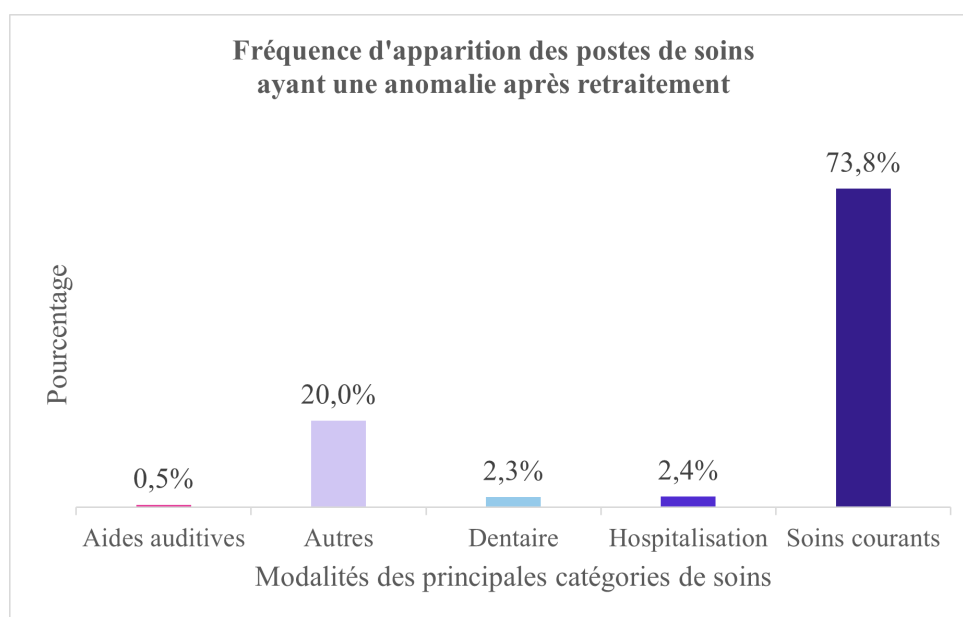


FIGURE 5.2 – Fréquence d'apparition des postes de soins ayant une anomalie après retraitement

Parmi les postes présentant cette anomalie, 74% proviennent des "Soins courants", car ce poste contient le plus grand nombre de modalités différentes. Pour l'ensemble des codes actes touchés par cette anomalie, il a été remarqué que la plupart provenaient de la sous-famille "Pharmacie". Une hypothèse a été formulée en supposant qu'il n'y avait pas de dépassement d'honoraires pour ces codes actes anormaux, car les dépassements d'honoraires étaient rares pour les médicaments. Par conséquent, les frais réels ont été ajustés de manière à correspondre à la base de remboursement.

Ainsi, les montants remboursés par la Sécurité Sociale ont été considérés comme égaux aux frais réels multipliés par le taux de remboursement.
Le poids des dépenses pour les codes "anormales" était de l'ordre de moins de 0,3 % par rapport aux dépenses totales de la base.

1. L'indemnité versée par l'assureur ne doit pas améliorer la situation financière de l'assuré avant le soin.

5.2 La répartition des montants par département

Pour rappel, un des objectifs du mémoire consiste à mettre en place un tarif pour des mutuelles communales, et idéalement, il serait souhaitable d'établir un tarif pour chacune des 34 970 communes. Cependant, étant donné que les informations nécessaires ne sont pas disponibles à la maille de la commune, une granularité supérieure a été choisie, celle des départements du territoire français.

Dans la suite du mémoire, les mutuelles communales seront confondues avec les mutuelles départementales pour combler le manque d'information à l'échelle communale.

Pour atteindre cet objectif de tarification, il faut réussir à obtenir une base de données qui recense les **indicateurs** suivants à la maille départementale :

- dépense de santé réelle ;
- montant remboursé par la Sécurité Sociale ;
- quantité d'actes ;
- base de remboursement ;
- nombre de bénéficiaires exposés.

Il est primordial de réussir à ventiler les montants des différents indicateurs par région au sein des différents départements qui la composent.

Un exemple est présenté dans la suite de la section pour les montants de dépenses de santé de la région Île-de-France. Les mêmes méthodes seront utilisées pour les autres indicateurs.

Ainsi, le montant total de dépenses de la région peut être considéré comme la somme des dépenses individuelles X_k des n départements qui composent la région.

Un schéma est présenté ci-dessous pour illustrer l'objectif à atteindre :

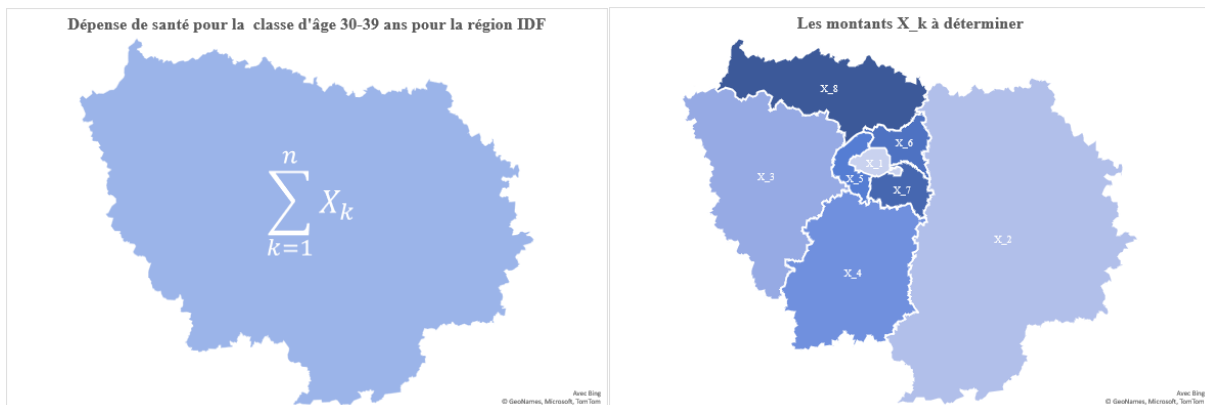


FIGURE 5.3 – Dépenses de santé non réparties de la région IDF

FIGURE 5.4 – Dépenses de santé réparties au sein des départements de la région IDF

L'opération sera effectuée pour tous les indicateurs, régions, classes d'âges et sous-famille de soins. Par ailleurs, deux méthodes ont été retenues pour ventiler l'ensemble des indicateurs (Dépense de santé, montants RO, quantité d'actes). Les résultats présentés concerneront uniquement les dépenses de santé réelles, car la méthode retenue pour les autres indicateurs est la même.

5.2.1 Méthode 1 : hypothèse de dépense uniforme

Une première idée consiste à répartir les montants de dépense de santé en fonction de la population de chaque département.

Ainsi, la part des dépenses de santé attribuée à un département serait proportionnelle à sa population. Ainsi, plus un département est peuplé, plus son poids sera important dans l'attribution des dépenses de santé de la région.

Les " X_k " à déterminer dans la section précédente dépendent alors du département, de la sous-famille de soin et de la classe d'âge. Ensuite, afin de faciliter la compréhension de la formule, qui est en réalité une règle de 3, les notations seront le plus explicites possible.

$$\text{Dépense_réelle}_{\text{dpt,SS_FAM,age}} = \frac{\text{Nb habitants}_{\text{dpt,age}} \times \text{Dépense_réelle}_{\text{region,SS_FAM,age}}}{\text{Nb habitants}_{\text{region,age}}}$$

(5.2)

Avec :

- Dpt : Le département associé à la dépense
- SS_FAM : La sous-famille associée à la dépense
- $Region$: La région associée à la dépense

Voici un exemple illustratif pour la région Ile-de-France. Si 50% de la population de l'Ile-de-France habite à Paris, alors Paris est "responsable" de 50% des dépenses de la région Ile-de-France.

Chaque année, l'INSEE estime la population des régions et des départements (France métropolitaine et DOM) à la date du 1er janvier. Ainsi, une base externe de l'INSEE a permis d'obtenir le nombre d'habitants par classe d'âge et par département.

En revanche, les classes d'âges ne sont pas les mêmes entre la base de l'INSEE qui propose des tranches de vingt ans (20-39 ans par exemple) et la base OLD_DAMIR qui propose des tranches d'âges de 10 ans (20-29 ans par exemple).

Ainsi, les classes d'âges de la base INSEE ont été adaptées à l'aide d'un outil en ligne [13] de l'INSEE pour correspondre à la base OLD_DAMIR. La méthode est détaillée en annexe 12.5.

Les figures ci-dessous présentent la répartition de la population d'un côté et la répartition des dépenses de santé de la sous-famille "Consultations/Visites" de l'autre :

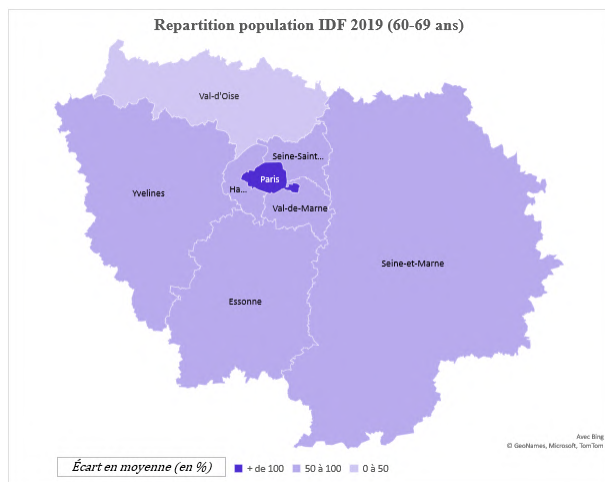


FIGURE 5.5 – Répartition de la population (60-69 ans) en région IDF

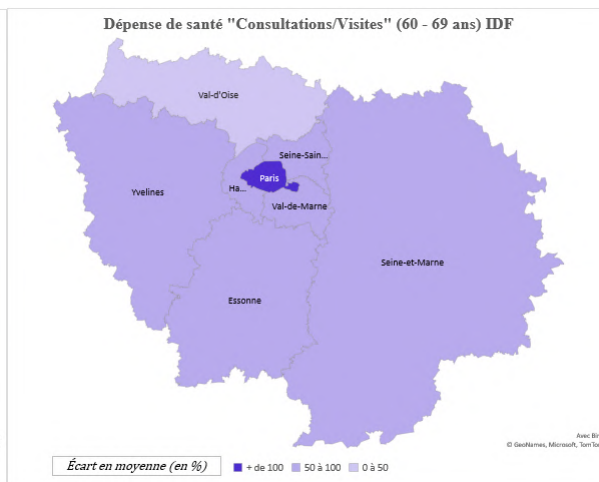


FIGURE 5.6 – Répartition des frais réels de Consultations/Visites (60-69 ans) en région IDF

Paris, qui est la ville la plus peuplée de l'Ile-de-France, consomme en moyenne plus que les autres départements moins peuplés.

Limites

Cette hypothèse ne prend pas en compte certaines spécificités comme les dépassements d'honoraires et les tendances de consommation des habitants, dont la variabilité territoriale ne dépend pas de la démographie.

C'est pourquoi une seconde méthode est introduite dans la prochaine section.

5.2.2 Méthode 2 : recours à la base DAM_DPT

Pour rappel, DAM_DPT offre l'avantage d'avoir des données au niveau départemental et provient de la SNIIRAM, tout comme l'Open DAMIR. Cela signifie qu'il est possible de déduire la répartition des dépenses de santé au sein des différents départements pour une sous-famille d'actes de soins donnée.

Ainsi, l'objectif de la méthode 2 consiste à étudier la répartition des dépenses de santé au niveau départemental dans la base DAM_DPT pour en déduire des coefficients de répartition. Ces coefficients seront ensuite utilisés pour ventiler les montants de dépenses "régionaux" de la base OLD_DAMIR.

De plus, afin d'établir une certaine légitimité à cette seconde méthode, il a été décidé de vérifier si la répartition des montants de dépenses réels à l'échelle régionale était identique entre les bases OLD_DAMIR et DAM_DPT. Si tel est le cas pour les régions, alors une hypothèse sera formulée pour les départements, justifiant ainsi l'application des coefficients.

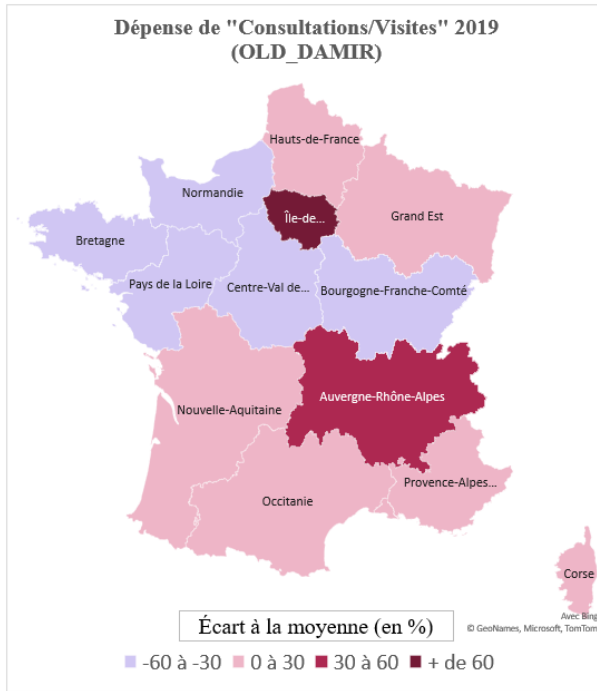


FIGURE 5.7 – Dépenses de santé réparties au sein des régions (OLD_DAMIR)

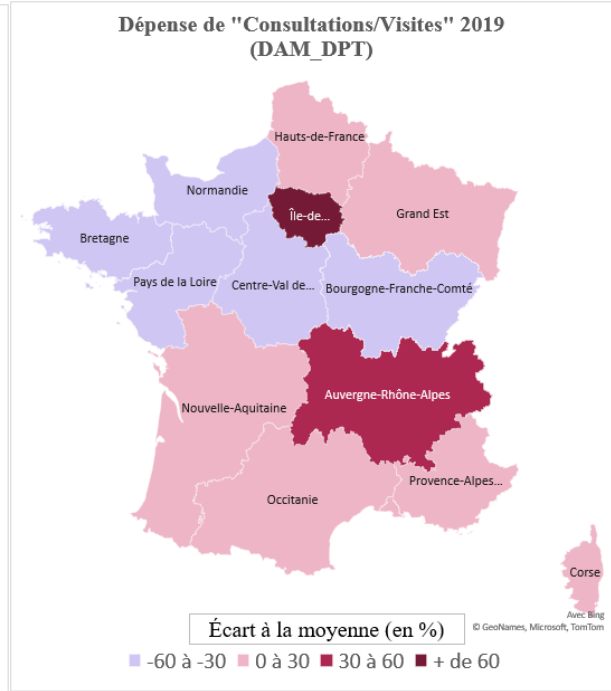


FIGURE 5.8 – Dépenses de santé réparties au sein des régions (DAM_DPT)

Il est observé que les répartitions sont équivalentes entre les deux bases pour l'année 2019 et la sous-famille "Consultations/Visites".

Si les répartitions des dépenses de santé sont identiques à la maille région alors elle est supposée être la même pour la maille département.

Les étapes de la seconde méthode se présentent comme suivent :

Étape 1 : Attribuer des sous-familles aux codes actes présents dans DAM_DPT

Chaque code acte dans la base DAM_DPT se voit attribuer la sous-famille correspondante selon la table de classification présentée dans la figure 4.4.

Étape 2 : Sélectionner une sous-famille ainsi qu'une région afin de calculer les coefficients de ventilation

L'objectif de cette étape permet de répondre à la question : "Comment se répartissent les dépenses d'une région au sein de ses départements?" L'exemple est pris pour la région "Ile-de-France" et la sous-famille "Consultations/Visites".

Les coefficients ont été calculés de la manière suivante et sont basés sur DAM_DPT :

$$Repartition_{Dpt,SS_FAM,Region} = \frac{Depense_{Dpt,SS_FAM,Region}}{Dpense_{SS_FAM,Region}} \quad (5.3)$$

Avec :

- *Dpt* : Le département associé à la dépense ;
- *SS_FAM* : La sous-famille associée à la dépense ;
- *Region* : La région associée à la dépense.

Le terme au numérateur correspond alors au montant total de dépense d'un département tandis que le dénominateur correspond au montant total des dépenses de la région à laquelle ce département appartient.

Les résultats sont présentés ci-dessous :

Départements	SS_FAM	Répartition
Paris	Consultations/visites	22%
Seine-et-Marne	Consultations/visites	10%
Yvelines	Consultations/visites	13%
Essonne	Consultations/visites	9%
Hauts-de-Seine	Consultations/visites	14%
Seine-Saint-Denis	Consultations/visites	12%
Val-de-Marne	Consultations/visites	11%
Val-d'Oise	Consultations/visites	9%

FIGURE 5.9 – Coefficient de répartition pour les Consultations/Visites de la région Ile-de-France

Ainsi, 22% des dépenses de Consultations/Visites de la région Ile-de-France sont concentrés à Paris.

Etape 3 : Appliquer la ventilation aux montants de dépenses de santé de la région sélectionnée dans la base OLD_DAMIR.

Etape 4 : Répéter le processus pour toutes les régions et toutes les sous-familles existantes

Limites

Une hypothèse particulière est accordée à la classe d'âge. En effet, étant donné qu'il n'y a pas de variable "CLASSE_AGE" dans DAM_DPT, les coefficients seront les mêmes pour toutes les classes d'âge de la base OLD_DAMIR.

5.2.3 Combinaison des méthodes

Ainsi, deux méthodes ont été introduites, mais il a été décidé que la seconde méthode était plus adaptée à la problématique. En effet, elle permet de mieux représenter les disparités territoriales. Cependant, dans la base DAM_DPT, certaines sous-familles ne sont pas présentes, ce qui rend impossible le calcul de coefficients de ventilation.

Dans ces cas, la méthode 1 est utilisée pour les sous-familles absentes de la base DAM_DPT.

Un tableau récapitulatif des méthodes utilisées en fonction de la sous-famille est présenté ci-dessous :

POSTE	SS FAM	Méthode
Soins courants	Analyse médicale et examens labo	Méthode 2
Soins courants	Auxiliaires médicaux (kiné, infi)	Méthode 2
Soins courants	Actes d'imagerie, de radiologie	Méthode 2
Soins courants	Actes médicaux (techniques)	Méthode 2
Soins courants	Consultations/visites	Méthode 2
Soins courants	Grand appareillage	Méthode 2
Soins courants	Petit appareillage	Méthode 2
Soins courants	Pharmacie	Méthode 2
Soins courants	Transport	Méthode 2
Soins courants	Vaccins anti-grippe	Méthode 2
Soins courants	Autres honoraires médicaux	Méthode 1
Optique	Autres - Optique	Méthode 2
Optique	Lentilles	Méthode 2
Optique	Monture	Méthode 2
Optique	Verres	Méthode 2
Inconnu	Inconnu	Méthode 1
Hospitalisation	Frais de séjour	Méthode 1
Hospitalisation	Autres - hospitalisation	Méthode 1
Hospitalisation	Forfait journalier	Méthode 2
Hospitalisation	Honoraires et actes chirurgicaux	Méthode 2
Hospitalisation	Hospitalisation	Méthode 2
Dentaire	Prothèse	Méthode 2
Dentaire	Autres - dentaire	Méthode 2
Dentaire	Implantologie	Méthode 2
Dentaire	Orthodontie	Méthode 2
Dentaire	Parodontologie	Méthode 2
Dentaire	Soins dentaires	Méthode 2
Autres	Ostéopathie, diététique, chirop	Méthode 2
Autres	Autres (aides, inclassable, ...)	Méthode 2
Autres	Cure thermale	Méthode 2
Aides auditives	Audioprothèse	Méthode 2

FIGURE 5.10 – Tableau récapitulatif des méthodes utilisées

Le poste "Inconnu" désigne les lignes présentes dans la base *Open DAMIR* qui ne sont pas associées à un code acte. Chaque année, la base est enrichie de nouvelles données et de nouveaux codes actes sont ajoutés. Ainsi, dans le cadre de ce mémoire, les postes inconnus ne seront pas traités.

Enfin, lorsque la méthode est définie pour chaque sous-famille, les montants des différents indicateurs sont ventilés au sein des différents départements de la France. Cela permet d'aboutir à la création d'une nouvelle base de données qui se nommera **NEW DAMIR**.

5.3 Inflation des dépenses de santé

Pour rappel, les dépenses de santé de la base OLD_DAMIR datent de 2018 et 2019. Par conséquent, afin de tarifier pour l'année 2023, il faut ramener les montants au moment de la tarification en mettant ces montants en "as-if" : cela implique de capitaliser les montants sur une ou plusieurs périodes pour les ramener à la valeur d'aujourd'hui. Ainsi, les montants de 2018 et 2019 sont respectivement ajustés à l'inflation jusqu'en 2022 et 2023 afin de conserver une profondeur historique de deux ans.

Une pratique courante dans le domaine de la santé consiste à utiliser un taux de capitalisation égal à l'**Objectif National de Dépenses d'Assurance Maladie (ONDAM)**. L'ONDAM représente un objectif de dépenses à ne pas dépasser pour les soins de ville et d'hospitalisation dispensée dans les établissements privés ou publics, ainsi que les centres médico-sociaux. Chaque année, le taux prévisionnel de l'année en cours et le taux réel de l'année précédente sont publiés dans le **Projet de loi de financement de la Sécurité Sociale (PLFSS)**.

Année	Taux sélectionnés	Réel/Prévisionnel
2019	2,60%	Taux réel
2020	7,60%	Taux réel
2021	8,70%	Taux réel
2022	2,20%	Taux réel
2023	-0,80%	Taux prévisionnel

FIGURE 5.11 – Les taux de capitalisation sélectionnés

Ainsi, pour les années de 2018 à 2022, les taux sélectionnés sont les taux réels d'augmentation des dépenses de santé publiées par le PLFSS. En revanche, pour l'année 2023, le taux réel n'est pas encore disponible, donc le taux prévisionnel a été choisi.

Exemple : Pour un montant de 100 € en 2019, le nouveau montant mis en "as-if" vaut :

$$100 \times (1 + 7,6\%) \times (1 + 8,7\%) \times (1 - 0,8\%) = 118,6 \text{ €}$$

Limites

Idéalement, il aurait fallu mettre à jour les données par sous-famille, mais par manque d'information, un coefficient global a été appliqué à l'ensemble des indicateurs. De plus, l'impact du 100% Santé n'a pas pu être pris en compte pour la fréquence de consommation des soins. En effet, la part supplémentaire prise en charge par les organismes complémentaires de santé est difficilement quantifiable et l'Ondam ne prend pas en compte le changement de comportement des assurés. Il faudra donc garder à l'esprit que les quantités d'actes n'ont pas été mis en "as-if".

5.4 NEW_DAMIR : la nouvelle base de données

La base NEW_DAMIR est une version partiellement désagrégée de OLD_DAMIR, les dépenses de santé ont été réparties au sein des départements. C'est également le cas pour les autres

indicateurs tels que les montants remboursés par la Sécurité Sociale, les bases de remboursements agrégées et les quantités d'actes.

5.4.1 Ajout des montants remboursés par la complémentaire santé

Les mutuelles communales proposent à leurs adhérents la possibilité de choisir entre plusieurs niveaux de garantie. Dans cette optique, deux garanties ont été sélectionnées pour tarification dans le cadre de ce mémoire :

- une garantie entrée de gamme V1 ;
- une garantie milieu de gamme V2.

Les grilles de garanties simplifiées² sont disponibles en annexe 12.5.

Afin d'effectuer une tarification pour ces deux garanties, il faut dans un premier temps ajouter deux variables "MT_RC_V1" et "MT_RC_V2" qui correspondent respectivement aux montants remboursés par la complémentaire santé de la garantie entrée et moyenne gamme.

À l'instar de Sandrine HUYNH dans son mémoire [11], le montant des deux nouvelles variables a été calculé de cette manière pour chaque ligne dans la base NEW_DAMIR :

$$MT_RC_k = \min \left(\begin{array}{l} DEPENSE - MT_RO ; \\ BR \times TR_RC_k + FORFAIT_k \times QTE_ACTE ; \\ LIMIT_k \times QTE_ACTE - MT_RO \end{array} \right) \quad (5.4)$$

Avec :

$k \in \{V1, V2\}$: L'indice de la grille de garantie.

DEPENSE : Le montant agrégé des frais réels d'un soin ou d'un équipement pour une ligne donnée.

MT_RO : Le montant agrégé du remboursement de la Sécurité Sociale pour une ligne donnée.

BR : La base de remboursement agrégée pour une ligne donnée.

QTE_ACTE : La quantité totale d'actes pour une ligne donnée.

TR_RC_k : Le taux de remboursement de la mutuelle selon la grille de garantie "k" après remboursement de la Sécurité Sociale.

FORFAIT_k : Le forfait offert par la mutuelle pour la garantie "k".

LIMIT_k : La limite de remboursement de la mutuelle pour l'acte concerné selon la garantie "k".

Le premier terme du minimum représente la part restante après le remboursement de la Sécurité Sociale et indique que le montant remboursé par la mutuelle ne peut pas dépasser cette part restante tandis que le deuxième terme est égal au montant de remboursement de la mutuelle selon ses grilles de garanties, sans prendre en compte les éventuelles limites de remboursement fixées par la mutuelle. Enfin, le dernier terme indique que la mutuelle ne peut pas rembourser au-delà de la limite de remboursement qu'elle impose, si une telle limite existe.

2. Les grilles sont simplifiées afin de préserver l'anonymat.

Ainsi, le calcul du remboursement de la mutuelle est basé sur un minimum entre ces trois termes, en tenant compte à la fois du remboursement de la Sécurité Sociale, des garanties de la mutuelle et de ses éventuelles limites de remboursement.

5.4.2 Ajout de la variable Exposition

L'exposition représente la période normalisée durant laquelle un assuré est exposé aux risques au sein du portefeuille de la mutuelle pendant l'année de souscription d'un contrat de santé. Ce dernier est généralement annuel et à tacite reconduction. Ainsi, un assuré couvert pendant :

- une année entière aura une exposition évaluée à 1.
- la moitié d'une année aura une exposition évaluée à 0,5.

Dans le cadre de ce mémoire, la variable exposition qui sera créée représente le nombre de bénéficiaires exposés par ligne. Il est supposé que l'assureur fournit une couverture santé à l'ensemble de la population française. En effet, les dépenses présentes de la base de données sont issues du régime de la Sécurité Sociale dont 92 % de la population française est affiliée.

Les populations de la variable exposition "EXPO" ainsi introduites sont dépendantes de la classe d'âge et du département, et elles ont été extraites d'une base de données de l'INSEE. Il convient de rappeler que cette base de données a été modifiée pour obtenir la correspondance des tranches d'âges entre la base OLD_DAMIR et la base de l'INSEE.

A l'issue de cette étape, la base de données contient les variables suivantes :

Nom variable	Libellé
DPT	Le département de résidence des adhérents
L_DPT	Le libellé du département
REGION	La région de résidence des adhérents
POSTE	Le poste de soin de l'acte
FAM	La famille de soin de l'acte
SS_FAM	La sous-famille de l'acte
ANNEE_REGLT	L'année de règlement
CLASSE_AGE	La classe d'âge des adhérents
DEPENSE	Le montant agrégée des dépenses réels engagés par les patients
MT_RO	Le montant agrégée remboursé par le régime obligatoire
QTE_ACTE	La quantité d'acte agrégée
EXPO	Le nombre de bénéficiaires exposés
BR	La base de remboursement agrégée de l'acte de soin
MT_RC_V1	Le montant agrégé remboursé par la complémentaire santé pour la garantie V1
MT_RC_V2	Le montant agrégé remboursé par la complémentaire santé pour la garantie V2

FIGURE 5.12 – Les variables de NEW_DAMIR

Dans la base de données NEW_DAMIR, certaines variables présentes dans OLD_DAMIR sont manquantes. En effet, les variables "QLT" et "SEXE", qui indiquent respectivement la qualité des adhérents (souscripteur/conjoint ou enfant) et le sexe de l'adhérent, ne sont pas incluses dans la nouvelle base.

La décision de ne pas inclure ces deux variables découle de l'étape de tarification qui sera abordée ultérieurement dans ce mémoire. En effet, comme précisé dans le chapitre précédent, l'implémentation de modèles linéaires généralisés est prévue. Au sein de ce modèle, la valeur moyenne d'une variable cible est prédite et expliquée à partir de variables explicatives, avec l'application d'une fonction de lien à la combinaison linéaire de ces dernières. À titre d'illustration, le tarif pourrait être estimé à l'aide de diverses variables telles que l'âge, la localisation géographique, la catégorie socio-professionnelle, etc.

En revanche, il est nécessaire pour l'application de ces modèles de ne pas avoir de corrélation entre les variables explicatives, sinon des biais dans l'estimation de la variable cible pourraient être introduits.

Une des hypothèses de ce mémoire a été de reprendre les résultats des analyses effectuées par Sandrine HUYNH [11]. À l'issue de ses travaux d'analyse, elle avait conclu que la variable "QLT" représentant la qualité des adhérents (souscripteur/conjoint ou enfant) était fortement corrélée à la classe d'âge de ces derniers. En effet, un enfant ne peut pas être souscripteur d'un contrat d'assurance. Par conséquent, elle n'avait pas retenu cette variable dans la tarification.

En ce qui concerne la variable "SEXE", elle n'a pas été incluse dans la base NEW_DAMIR, car elle ne sera pas considérée comme une variable explicative dans les modèles de tarification, conformément à la réforme de la "Gender Directive". Depuis fin 2012, cette réforme interdit l'utilisation du genre d'une personne dans la détermination des tarifs d'assurance. Toutefois, en pratique, la proportion d'hommes et de femmes est prise en compte pour établir un tarif commun aux deux sexes. Dans le cadre de ce mémoire, cependant, il était chronophage de ventiler les indicateurs en fonction du département, du sexe et de la classe d'âge. C'est pourquoi le choix a été fait de ne pas différencier les montants de dépenses de santé, de remboursements, etc., en fonction du sexe.

Ce chapitre a permis de pallier les lacunes de la base OLD_DAMIR. Les données ont été mises à jour, et la granularité de la base a été ajustée. À ce stade, l'état actuel de la base NEW_DAMIR est suffisant pour établir une tarification par département. Cependant, il est possible d'ajouter des variables externes afin d'expliquer les éventuelles variations de tarifs entre les différents départements.

CHAPITRE

6

LES VARIABLES EXTERNES

Du point de vue d'un assureur cherchant à mieux appréhender le marché, il peut être pertinent d'analyser les disparités entre les départements. En comprenant les particularités socio-économiques des résidents, il pourrait anticiper et saisir les demandes en matière d'assurance santé de manière plus précise.

Ainsi, ce chapitre a pour objectif d'introduire des variables externes à la base NEW_DAMIR ainsi que les retraitements effectués pour ces derniers. Les variables concernées sont :

- le taux de chômage du département ;
- l'indice de vieillissement par département ;
- le niveau de vie médian par département ;
- la disponibilité des offres de soins par département.

Il serait bénéfique pour l'assureur de déceler des schémas de consommation à travers ces variables externes, et ainsi identifier les populations qui seraient avantageuses à inclure dans son portefeuille d'assurés.

6.1 Le taux de chômage

Les mutuelles communales sont ouvertes à tous, mais elles n'arrivent généralement pas à attirer les salariés, étant donné qu'ils bénéficient souvent d'une couverture santé collective fournie par leur entreprise. En revanche, les demandeurs d'emploi pourraient constituer une population intéressée par ces mutuelles communales.

Les taux de chômage de l'année 2022 et 2023 ont alors été récupérés pour l'ensemble des départements sur le site de l'[INSEE](#). Ensuite, la variable externe "**TX_CHOMAGE**" a été ajoutée dans la base NEW_DAMIR. L'attribution des valeurs pour chaque ligne s'est effectuée par correspondance entre l'année de règlement et le département.

6.2 L'indice de vieillissement

À travers les retours d'expérience des différents dispositifs de mutuelles communales, il est maintenant évident que la grande majorité des adhérents à ce type d'offre sont des seniors. D'une part, parce qu'ils sont à la retraite et ne sont donc plus éligibles à leur complémentaire santé d'entreprise. D'autre part, le coût des complémentaires santé individuels traditionnels est souvent très élevé pour les seniors.

Ainsi, l'ajout de la variable "**INDICE_V**" dans la base NEW_DAMIR pourrait mettre en évidence certaines tendances concernant les montants remboursés par les mutuelles partenaires dans le cadre du dispositif de mutuelle communale.

L'indice de vieillissement est le rapport de la population des 65 ans et plus sur celle des moins de 20 ans. Un indice autour de 100 indique que les 65 ans et plus et les moins de 20 ans sont présents dans à peu près les mêmes proportions sur le territoire.



FIGURE 6.1 – Interprétation de l'indice de vieillissement

Plus l'indice est faible plus le rapport est favorable aux jeunes, plus il est élevé plus il est favorable aux personnes âgées. x représente l'indice de vieillissement.

Les valeurs de la nouvelle variable ont été récupérées sur le site de l'[Agence Nationale de la Cohésion des Territoires \(ANCT\)](#). Les indices de vieillissement des années 2022 et 2023 ont été attribués par la même méthode que le taux de chômage, c'est-à-dire par correspondance avec les départements et les années de règlement.

6.3 Le niveau médian de vie

Bien que ces dernières années les mutuelles communales connaissent un gain en popularité, elles restent très peu développées aujourd'hui. Si d'un côté la méconnaissance de ce dispositif empêche ce dernier de se développer, certaines communes refusent d'y adhérer, car le niveau de vie est relativement élevé. Les habitants de la commune n'estiment pas avoir besoin de ce genre de dispositif.

Également, cette variable pourrait permettre d'observer si le niveau vie médian d'un département a un impact sur la fréquence de consommation des soins, car il est estimé qu'aujourd'hui 30% des Français auraient déjà renoncé ou retardé des soins pour des raisons financières.

Ainsi, la variable "**NIV_VIE_MEDIAN**" a été ajoutée à la base NEW_DAMIR. Les données ont été récupérées sur le site de l'[INSEE](#). Cependant, il convient de noter que les données les plus récentes disponibles sont pour l'année 2019. Par conséquent, il a été supposé que le niveau de vie n'a pas connu de variation significative entre 2019 et 2023. Ainsi, les données de 2019 ont été utilisées pour les deux années de règlement, étant donné que la dernière publication des données avant 2019 remontait à 2014.

6.4 La densité d'établissement de santé : FINESS

La dernière variable externe abordée dans ce mémoire concerne la problématique de la répartition inégale des établissements de santé en France. En effet, certaines zones géographiques rencontrent des difficultés à accéder à des soins de santé en raison d'un manque d'établissements ou de personnel sur le territoire. Ces zones sont appelées des "déserts médicaux".

L'étude de cette variable pourrait déterminer d'éventuels impacts sur la consommation des soins. En effet, une accessibilité limitée peut entraîner des distances à parcourir plus longues pour consulter un médecin, ce qui pourrait potentiellement décourager certains patients et impacter la fréquence de leur consommation. Des délais d'attente prolongés peuvent également entraîner une possible aggravation de l'état de santé d'un patient, ce qui pourrait augmenter le coût des soins nécessaires.

L'objectif de cette sous-partie est d'ajouter une variable qui indique le nombre de salariés dans des établissements de santé pour chaque département. Autrement dit, un département qui possède peu d'employés dans le domaine médical sera considéré comme un "désert médical".

Dans un premier temps, il est essentiel de réussir à cartographier tous les établissements de santé en France. Pour cela, le **FINESS (Fichier National des Établissements Sanitaires et Sociaux)** apporte une aide conséquente puisque ce dernier recense tous les établissements, structures et équipements dans les domaines sanitaires, médico-sociaux et sociaux. Cependant, il ne contient pas d'enregistrement des fournisseurs de biens médicaux tels que les opticiens et les audioprothésistes. La base est mise à jour très régulièrement et la version téléchargée sur le site data.gouv.fr est celle du 3 juillet 2023. Le fichier contient la liste des établissements de soins tels que les centres hospitaliers, les EHPAD, les laboratoires, etc... De plus, les coordonnées de ces derniers sont également répertoriées.

La base FINESS présente 32 variables, mais les noms des colonnes ne sont pas accessibles directement dans la base. Il faut se référer à un fichier PDF disponible sur le même site afin de connaître les noms des colonnes. Cependant, dans les ressources communautaires du site, un code R mis en ligne par Philéas Condemine permet de nettoyer automatiquement la base de données et d’y associer les noms des colonnes correspondantes.

Parmi les 32 variables disponibles, les variables intéressantes dans le cadre de l’étude sont :

- siret : le numéro SIRET de l’établissement permettant d’identifier les établissements ;
- departement : le département de l’établissement ;
- categetab : la catégorie de l’établissement (EHPAD, hôpital,..) ;

Avec ces variables, il est possible de compter le nombre d’établissements par département. Cependant, il est souhaitable de prendre en compte la différence entre un centre hospitalier et un cabinet indépendant. En effet, le nombre de patients pris en charge dans les centres hospitaliers diffère significativement du nombre pris en charge par des cabinets indépendants. Le simple fait de rapporter le nombre d’établissements médicaux ne suffit pas à caractériser complètement les services de soins disponibles dans le département.

Ainsi, il est préférable de raisonner en termes de nombre de salariés travaillant dans chaque établissement. Cette approche permet de mieux refléter l’étendue des activités de chaque établissement de santé et de considérer sa contribution globale aux soins de santé au sein du département. En revanche, il n’existe pas de variable indiquant le nombre de salariés par établissement dans le FINESS.

Pour obtenir le nombre de salariés pour chaque établissement de santé, il a été décidé d’avoir recours à la base SIRENE [14], qui met à disposition l’ensemble des SIRENE/SIRET créés historiquement depuis 1950, ainsi que les tranches d’effectifs associées. Par conséquent, pour tous les SIRET répertoriés dans le FINESS, les informations concernant le nombre de salariés ont été récupérées à partir de la base SIRENE.

En revanche, la base SIRENE contient 36 millions de lignes pour 6,4 Go. Il est difficile de traiter la base sur les ordinateurs de l’entreprise. Ainsi, par insuffisance de RAM (mémoire vive de l’ordinateur), la récupération du nombre de salariés a été effectuée en Python sur un notebook Jupyter hébergé par Google Cloud Platform¹.

Enfin, la variable de la base SIRENE contenant l’information sur le nombre de salariés est exprimée sous forme de tranches d’effectifs, ce qui signifie que le nombre exact de salariés par numéro SIRET n’est pas connu.

1. La plateforme permet de mettre en place des machines virtuelles avec des pré-installations comme Python, pandas, numpy, etc.

Pour effectuer le comptage, la moyenne de chaque tranche a été sélectionnée. Les résultats sont présentés dans la figure ci-dessous :

CLASSE	Nombres salariés	Effectifs retenus
NN	Unités non employeuses ou effectifs inconnus	0
00	0	0
01	1 ou 2	2
02	3 à 5	4
03	6 à 9	8
11	10 à 19	15
12	20 à 49	35
21	50 à 99	75
22	100 à 199	150
31	200 à 249	225
32	250 à 499	375
41	500 à 999	750
42	1000 à 1 999	1 500
51	2000 à 4 990	3 495
52	5000 à 9 999	7 500
53	10 000 et plus	10 000

FIGURE 6.2 – Tranche d'effectifs de salariés

Parmi l'ensemble des établissements répertoriés dans la base de données, un seul a été classé dans la catégorie "53" correspondant à plus de 10 000 salariés. Par conséquent, la moyenne attribuée à la tranche de 10 000 salariés et plus est de 10 000, en considérant que cette approximation n'introduirait pas de biais significatif dans les effectifs.

Après avoir effectué la jointure des numéros SIRET avec leurs tranches d'effectifs, il a été constaté que 11 935 lignes sur un total de 97 091 contenaient des valeurs manquantes. Ces valeurs manquantes indiquent que la catégorie d'effectifs n'était pas renseignée pour ces établissements. Pour pallier cette absence d'information, la valeur moyenne du nombre de salariés a été attribuée en fonction de la catégorie et du département.

Exemple : Le nombre moyen d'employés dans un centre hospitalier à Paris (75) est de 2124 personnes.

Par conséquent, si une ligne de la base de données ne contient pas d'information sur la tranche d'effectifs, mais est identifiée comme étant un centre hospitalier situé à Paris, alors le nombre de salariés attribué serait de 2124.

CHAPITRE

7

ANALYSE DESCRIPTIVE DE LA BASE NEW_DAMIR

Le but de ce chapitre est de procéder à une dernière vérification des données. Bien qu'une première vérification ait déjà été effectuée dans la base OLD_DAMIR, la base NEW_DAMIR a été obtenue en ventilant les données de OLD_DAMIR à l'aide de méthodes dont les limites ont été introduites dans les chapitres précédents. Par conséquent, il est possible qu'après cette ventilation des indicateurs, des montants anormaux puissent être identifiés.

Ainsi, après avoir étudié la présence éventuelle de valeurs considérées comme extrêmes, ce chapitre présentera des statistiques générales permettant une meilleure compréhension du contexte de la santé.

7.1 Analyse des valeurs extrêmes

Cette étape permet d'exclure les lignes associées à des données aberrantes afin de limiter les biais dans la tarification qui sera ensuite réalisée. Pour cela, des boîtes à moustaches ont été utilisées pour identifier les montants minimum et maximum de remboursement de la mutuelle ainsi que les dépenses engagées pour chaque sous-famille d'actes.

Pour effectuer une comparaison équitable, il est nécessaire de rendre unitaire les montants associés à chaque ligne. Ces montants seront alors comparés avec les prix moyens du marché par équipement ou acte de soin.

La mise en place de cette analyse peut être complexe, car pour une sous-famille donnée, il peut exister plusieurs actes différents, chacun avec des montants de remboursement pouvant varier considérablement. Cela rend le processus d'analyse plus délicat, car l'information est agrégée et il n'est pas possible d'identifier quel acte en particulier serait à l'origine d'une valeur extrême. Par exemple, dans la sous-famille "Forfait Journalier", il y a les forfaits journaliers hospitaliers qui correspondent à la somme qu'un patient doit verser en cas d'hospitalisation de plus de 24 heures, mais aussi les forfaits journaliers de sorties précoces.

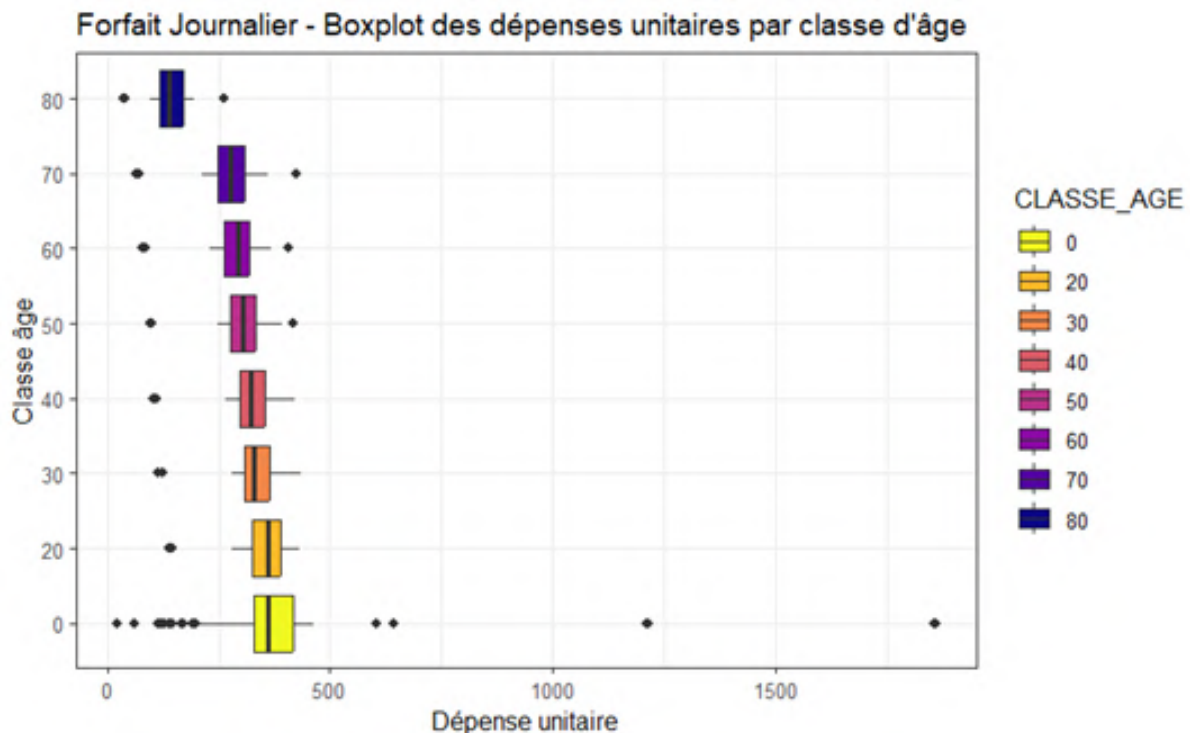


FIGURE 7.1 – Boîtes à moustaches des montants unitaires remboursés pour "Forfait journalier" par la garantie V2

Des montants de dépenses allant jusqu'à 1800 euros unitaire ont été recensés en Haute-Corse et 1200 euros en Corse-du-Sud. Étant donné que le forfait journalier hospitalier est de 20 euros par jour d'hospitalisation, cela implique que les habitants de la Haute-Corse âgés de 0 à 19 ans sont hospitalisés en moyenne 90 jours, soit environ 3 mois par an. Il semble peu probable que cela soit le cas, même pour les nourrissons qui rencontrent des difficultés. En effet, ce scénario ne semble pas être récurrent et n'aurait probablement pas un impact significatif sur une tranche

d'âge aussi large.

Dans le doute, étant donné que dans la majorité des autres sous-familles, la Haute-Corse et la Corse-du-Sud présentaient également des montants très élevés par rapport au reste des départements, ces deux départements seront exclus de l'étude et les tarifs seront établis pour la France continentale.

7.2 Statistiques de la base NEW_DAMIR

Cette partie vise à explorer les données, à mettre en évidence des différences de comportements entre les départements et les classes d'âges. Il s'agit également de discuter de la qualité des données. Ainsi, les montants de dépenses affichés dans cette section ont été normalisés par les quantités d'actes pour être rendus unitaires tandis que les fréquences sont normalisées par le nombre de bénéficiaires. Il est important de préciser que les montants dépendent des retraitements effectués sur les bases de données, et qu'il convient de conserver un regard critique quant à la véracité des statistiques obtenues. De plus, par souci de longueur, seuls les résultats de la sous-famille "Consultations/Visites" seront présentés.

7.2.1 Analyse par classe d'âge

La figure ci-dessous illustre les dépenses de santé moyenne pour un individu en France selon sa tranche d'âge :

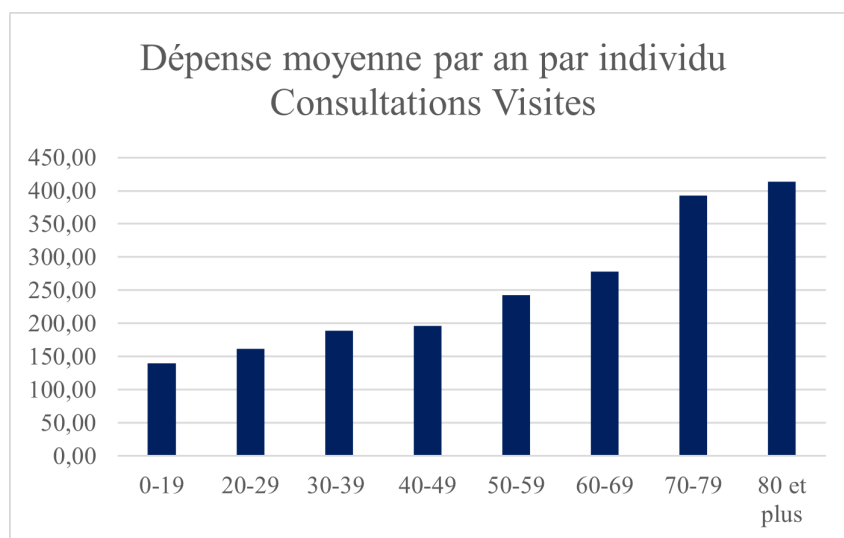


FIGURE 7.2 – Dépense moyenne par an par individu Consultations Visites (2019)

Un français dépense en moyenne 251 € par an dans des consultations chez le médecin. Il s'agit des frais réels sans compter le montant remboursé par la Sécurité Sociale ni par les complémentaires santé. Il est constaté que plus l'individu vieillit plus il dépensera dans les soins et donc qu'un effet lié à l'âge se fera sans doute ressentir dans les tarifs établis.

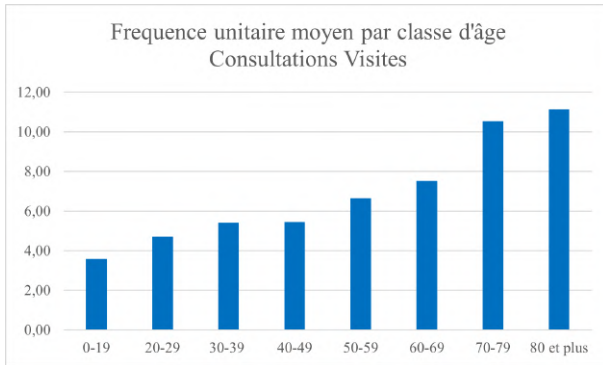


FIGURE 7.3 – Fréquence unitaire par classe d'âge en 2019

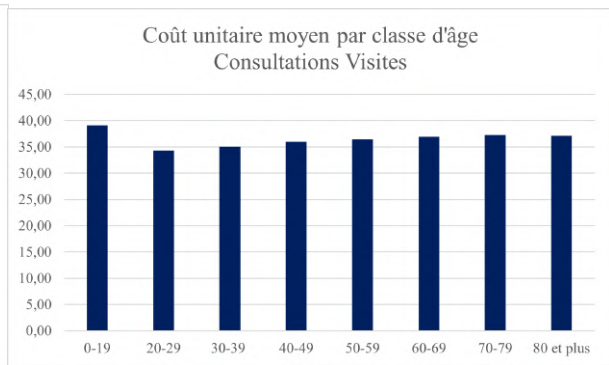


FIGURE 7.4 – Coût unitaire par classe d'âge en 2019

La dépense de santé par an augmente globalement avec l'âge, mais dans la majorité des cas, cela est davantage attribué à la fréquence de consommation. Pour la classe d'âge 0-19 ans, la fréquence de consommation des soins est faible, mais le montant du coût moyen unitaire est le plus élevé parmi les tranches d'âges. À titre indicatif, le tarif d'une consultation chez le généraliste (secteur 1) est fixé à 25 € depuis le 1er mai 2017. Quant aux consultations chez les spécialistes (secteur 1), leurs prix varient généralement de 28 € à 49 €. Les montants étudiés paraissent cohérents et c'est également le cas pour les autres sous-familles d'actes de soins.

7.2.2 Analyse par département

La dépense moyenne par an pour un français a été calculée, ce qui a permis de classer l'ensemble des départements dans des catégories afin d'observer les tendances de consommation de chaque département. Une cartographie des catégories est comparée à celle de la disponibilité de l'offre de soins¹, basée sur la variable externe introduite dans le chapitre précédent.

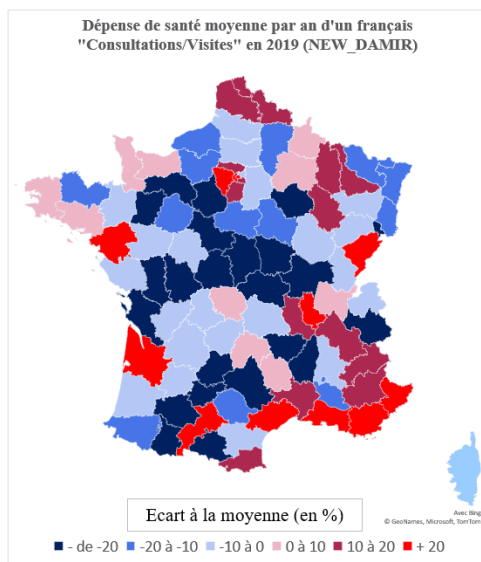


FIGURE 7.5 – Dépense unitaire moyenne des "Consultations/Visites"

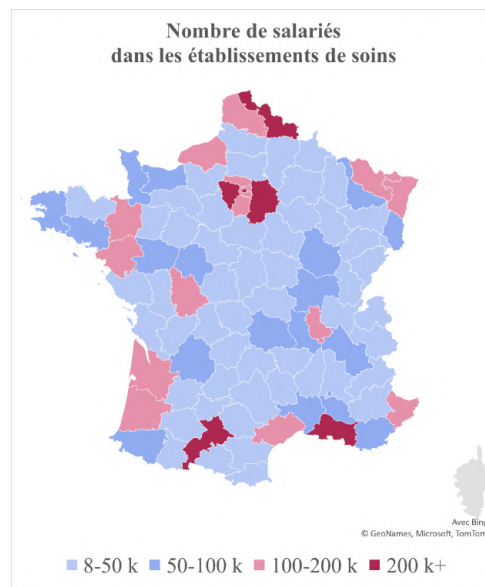


FIGURE 7.6 – Offre de soins par département

1. La cartographie de l'offre de soins a été réalisée dans une étape ultérieure du mémoire, précisément lors de l'analyse univariée, dont le principe sera détaillé dans le chapitre 9.

En 2019, les départements du Sud ainsi que ceux de la région Île-de-France présentent des dépenses annuelles moyennes plus élevées que les départements situés au cœur de la France. Cette disparité s'explique par la présence répandue de dépassements d'honoraires dans ces zones géographiques. Par ailleurs, il est également notable qu'en général, ce sont les grandes agglomérations qui enregistrent les niveaux de consommation les plus élevés, ce phénomène pouvant être attribué à une disponibilité accrue de l'offre de soins dans ces zones.

Le noms des départements est disponible en annexe 12.14.

Pour la suite de l'analyse, la région Ile-de-France qui comprend des zones urbaines principalement est comparée à la région Nouvelle-Aquitaine qui présente une diversité de zones, allant des zones rurales aux zones urbaines.

Ainsi, les montants des dépenses moyennes annuelles de la région Île-de-France sont affichés sur la figure ci-dessous :

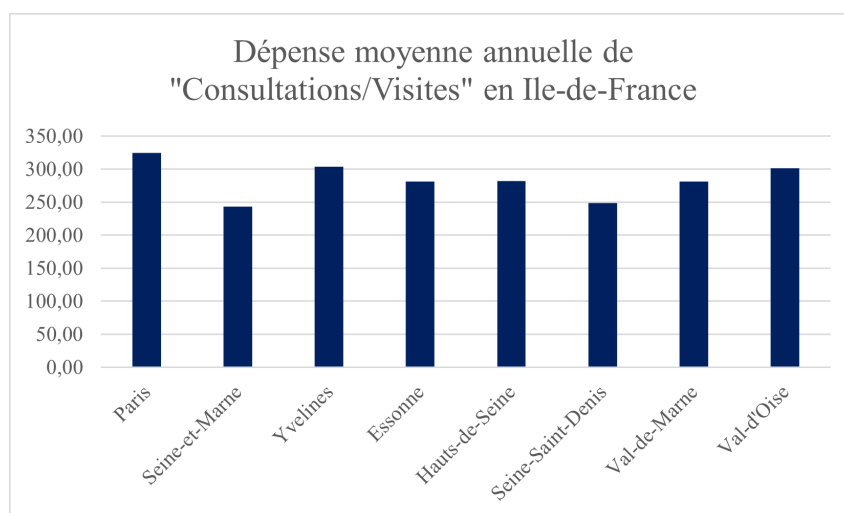


FIGURE 7.7 – Montant des dépenses annuelles en Île-de-France (2019)

Il est observé que Paris est le département qui enregistre la plus haute consommation dans la région, mais qu'il n'existe pas une disparité significative au sein de la région Île-de-France. En effet, avec une dépense moyenne de 283 € par an pour la région, seuls les départements de la Seine-et-Marne et Seine-Saint-Denis sont légèrement en dessous de la moyenne nationale estimée à 251 €.

Les montants des dépenses moyennes annuelles pour les départements de la Nouvelle-Aquitaine sont affichés sur la figure ci-dessous :

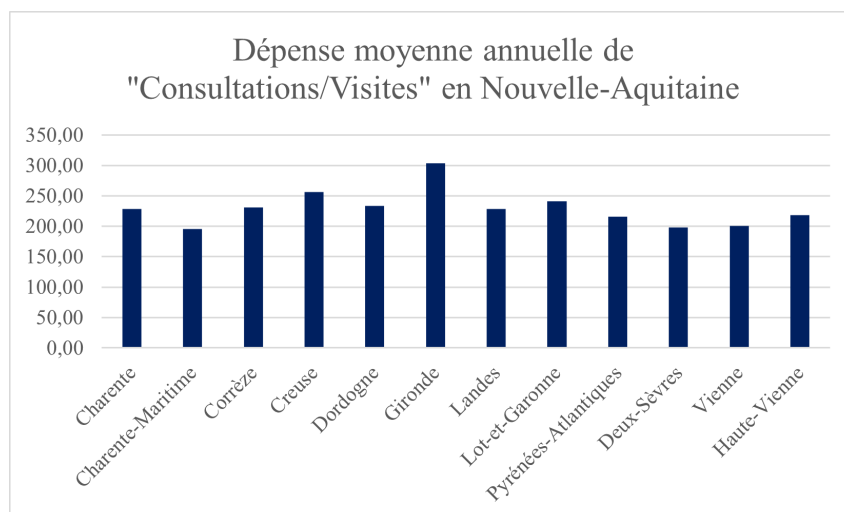


FIGURE 7.8 – Montant des dépenses annuelles en Nouvelle-Aquitaine (2019)

Un pic à 303 € est observé pour le département de la Gironde qui est le seul département au-dessus de la moyenne nationale en termes de dépenses. En effet, les zones rurales de la région présentent en général une consommation moindre par rapport aux zones urbaines. Cette situation peut être attribuée à la fréquence moins élevée des dépassements d'honoraires et à la limitation de l'offre de soins dans ces zones.

Cette analyse a initialement permis d'identifier les tendances de consommation au sein des différents départements et des différentes classes d'âge. De plus, elle a mis en évidence l'existence de deux facteurs qui influenceront ultérieurement le tarif : le premier étant lié à l'âge, et le second à l'effet du département ou de la région, engendrant des variations dans le tarif. Il est également remarqué que les grandes villes, caractérisées par une disponibilité de soins plus élevée, enregistrent une consommation supérieure. Par conséquent, la suite du mémoire prévoit la réalisation d'analyses complémentaires afin de confirmer cette hypothèse.

Synthèse de la deuxième partie

La base de données utilisée dans ce mémoire est appelée OLD_DAMIR. Il s'agit de la fusion des bases de l'*Open* DAMIR publiées en 2018 et 2019, qui avait déjà été retraitée dans le cadre d'un précédent mémoire réalisé au sein de l'entreprise. Cependant, cette version était une sauvegarde intermédiaire qui présentait une anomalie pour les montants remboursés par la Sécurité Sociale. Par conséquent, un retraitement a été effectué afin de les corriger.

Par ailleurs, OLD_DAMIR possède la particularité d'avoir les informations agrégées au niveau régional. Cependant, un des objectifs du mémoire est d'établir un tarif par département. Il faut donc réussir à passer de l'information des différents indicateurs (dépense de santé, montants remboursés par le régime obligatoire, quantité d'actes) de l'échelle régionale à l'échelle départementale.

Pour ce faire, une première méthode a été envisagée, celle de l'hypothèse de dépense uniforme. La part des montants des indicateurs attribués à un département serait proportionnelle à sa population. Bien que cette première méthode fournisse une estimation rapide, elle ne tient pas compte des spécificités des départements, telles que les dépassements d'honoraires ou les tendances de consommation. C'est pourquoi une deuxième méthode a été introduite.

Ainsi, la seconde méthode consiste à utiliser une base externe qui est nommée DAM_DPT, il s'agit d'une base également extraite de la SNIIRAM ressemblant à la base *Open* DAMIR, mais dont les prestations hospitalières et les âges sont manquants. Cette nouvelle base présente l'avantage d'avoir des informations sur les indicateurs au niveau départemental. Des coefficients de répartition sont alors déduits à partir de cette nouvelle base. Ces coefficients obtenus sont ensuite appliqués aux montants de la base OLD_DAMIR pour effectuer la ventilation au niveau départemental et créer par la même occasion une nouvelle base de données appelée NEW_DAMIR.

Afin de tarifier pour l'année 2023, les indicateurs de coûts sont mis à jour en utilisant l'indicateur de l'ONDAM. Il est à noter que cette méthode n'inflète que les coûts et non pas les fréquences de consommation.

Ensuite, pour chaque département, la base de données NEW_DAMIR a été enrichie en incluant des variables externes telles que l'indice de vieillissement, le niveau de vie médian, le taux de chômage et le nombre de salariés dans les établissements de soins. Ces variables auraient pour objectif d'expliquer davantage les différences de tarifs qui seraient obtenus dans la suite du mémoire.

Enfin, en réalisant une analyse des valeurs extrêmes et des statistiques globales sur la base NEW_DAMIR, il a été décidé d'exclure les départements de la Corse. De plus, cette démarche a révélé la nécessité de prendre certaines précautions pendant la mise en place de la tarification pour certaines sous-familles d'actes de soin en raison de possibles biais liés à la qualité des données et aux nombreuses hypothèses formulées dans ce mémoire ainsi que dans celui de Sandrine HUYNH.

Troisième partie

La tarification des mutuelles
communales

La base de données est prête à être utilisée pour établir un tarif, il convient alors de choisir les méthodes de tarification. Les modèles linéaires généralisés seront privilégiés pour cette tâche, car ils présentent l'avantage d'être très interprétables.

Un tarif sera défini pour chacune des deux garanties mentionnées dans l'Annexe 12.5.

La troisième partie est consacrée à la mise en place de la tarification. Elle expose les méthodes de tarification utilisées ainsi que la démarche entreprise pour les mettre en place. Deux visions seront ainsi présentées dans cette partie afin de les comparer par la suite :

- celle d'un assureur qui mutualise son risque à l'échelle départementale ;
- celle d'un assureur qui mutualise son risque à l'échelle régionale.

Ils seront respectivement appelés "Assureur départemental" et "Assureur régional".

Cette comparaison vise à déterminer si une expansion de l'offre à une échelle régionale pourrait être bénéfique pour un assureur qui chercherait à démarcher les élus locaux dans le cadre d'établir un partenariat.

CHAPITRE

8

RAPPELS THÉORIQUES SUR LA TARIFICATION

Ce chapitre effectue un rappel théorique sur la nature des contrats d'assurance ainsi que la modélisation de la prime pure. Il décrit ensuite la théorie des modèles linéaires généralisés.

8.1 Le contrat d'assurance et les types de primes

Une police d'assurance est un accord contractuel impliquant d'un côté le souscripteur, l'assuré et le bénéficiaire, et de l'autre l'assureur. Il convient de noter que le souscripteur est celui qui signe et paie les primes, l'assuré est celui qui est exposé au risque, et le bénéficiaire est celui qui reçoit les indemnités. Les trois rôles peuvent être assumés par la même personne ou par des individus différents.

En cas de sinistre, le bénéficiaire de la police (qui n'est pas obligatoirement le souscripteur du contrat) reçoit le montant prévu dans le contrat en cas de réalisation du sinistre. Ainsi, le risque économique initialement supporté par l'assuré est **transféré à l'assureur**.

En théorie, la mutualisation réalisée par la souscription de nombreux contrats au sein d'une compagnie d'assurance permet une utilisation approximative de la **loi des grands nombres**.

En effet, un portefeuille d'assurance couvre un risque spécifique, ce qui signifie que les pertes sont supposées suivre la même distribution de probabilité, impliquant ainsi une tarification par type de garantie. De plus, les contrats sont a priori considérés comme indépendants les uns des autres.

Ainsi, il existe différents types de prime dont l'assuré n'a pas forcément connaissance :

La **prime pure** qui correspond au montant moyen auquel devra faire face l'assureur. Cela représente les coûts attendus des sinistres. Dans le cadre de ce mémoire, différentes primes pures seront estimées pour chaque type de risque (dentaire, optique, etc..) puis la somme de ces montants constituera la **prime pure globale**. Cependant, il est important de noter que l'opération annoncée précédemment est uniquement possible sous réserve d'indépendance entre les différentes primes pures. Or ce n'est pas obligatoirement le cas. En effet, la consultation d'un médecin pour un vaccin spécifique entraîne généralement l'achat du vaccin en question.

La **prime commerciale** est le montant affiché sur les devis, c'est ce que l'assuré doit payer afin d'être couvert par l'assurance. Elle est composée de la prime pure, à laquelle sont ajoutés des frais, des charges et des taxes supplémentaires et s'écrit :

$$Prime\ commerciale = Prime\ pure + Chargement\ de\ securite + Frais + Taxes \quad (8.1)$$

Avec :

- **Prime pure** : Le coût du risque proprement dit.
- **Chargement de sécurité** : Un montant supplémentaire ajouté pour absorber la volatilité naturelle des sinistres. Il s'agit d'une marge de sécurité.
- **Frais** : D'une part, les frais de gestion du contrat, incluant les salaires et les frais de location des locaux. D'autre part, les frais d'acquisition, rémunérant les apporteurs d'affaires.
- **Taxes** : Imposées par le gouvernement.

Ainsi, la stratégie de l'organisme assureur joue sur le montant de ces chargements.

8.2 Modélisation de la prime pure

Lors de l'établissement des tarifs, l'assureur cherche à estimer les coûts potentiels des prestations sur une période donnée. Pour ce faire, la modélisation la plus couramment utilisée en assurance santé consiste à considérer un portefeuille de risques homogènes et indépendants puis d'appliquer la méthode "fréquence x cout"

Ainsi pour chaque type risques i (distinction des sous-familles de soin), la charge totale des soins risque k noté X_k représente le montant total remboursé par la complémentaire santé :

$$X_k = \sum_{i=1}^N S_i \quad (8.2)$$

Avec :

- N : une variable aléatoire représentant la quantité d'actes du portefeuille considéré dans la période fixée ;

- $(S_i)_{i \geq 1}$: une suite de variables aléatoires représentant les montants remboursés par la complémentaire santé, également appelée coût pour simplifier ;
- Si N est égal à 0, alors X_k est également nul.

Lorsqu'il s'agit de données agrégées, la formule demeure valide, mais il est nécessaire de concevoir que la variable N représente en réalité la somme de variables n_p , où p fait référence au numéro d'un groupe agrégé d'individus. De manière similaire, il conviendrait d'introduire une variable S_{n_p} pour représenter le montant total des soins du p -ème groupe agrégé d'individus. En conséquence, S_{n_p} correspond en réalité à une somme de variables S_i (avec p fixé et i variant). Cette situation complique cependant les équations ainsi que leur notation : contrairement à la sommation précédente qui implique un incrément de 1, dans ce contexte, les incréments seraient les n_p , et les valeurs de S_i ne sont pas connues, uniquement celles de S_{n_p} . Ainsi, l'expression de X_k devient :

$$X_k = \sum_{n_p \in \{n_1, \dots, n_{P_{total}}\}} S_{n_p} \quad (8.3)$$

Pour la suite, les notations usuelles seront gardées pour faciliter la compréhension. Ainsi, les montants remboursés par la complémentaire santé $(S_i)_{i \geq 1}$ sont supposés être indépendants et identiquement distribués (i.i.d). Il en est de même pour les quantité d'actes. La charge totale moyenne des coûts s'écrit alors :

$$\text{Prime pure} = E(X_k) = E(N) \times E(S) \quad (8.4)$$

Avec :

- $E(N)$: estimation de la fréquence de consommation de soins pour le risque k donnée ;
- $E(S)$: estimation du coût moyen remboursé par la complémentaire santé pour le risque k donnée.

La prime pure s'écrit finalement :

$$\text{Prime pure} = \text{Fréquence} \times \text{Coût} \quad (8.5)$$

Avec :

$$\text{Fréquence} = \frac{\text{Quantité d'actes totale}}{\text{Nombre total de bénéficiaires exposés}} \quad (8.6)$$

$$\text{Coût} = \frac{\text{Montant remboursé par la complémentaire santé}}{\text{Quantité d'actes totale}} \quad (8.7)$$

La prime pure s'écrit finalement :

$$\text{Prime pure} = \frac{\text{Montant remboursé par la complémentaire santé}}{\text{Nombre total de bénéficiaires exposés}} \quad (8.8)$$

La prime pure sera également appelée "consommation" dans la suite de l'écrit, car le terme est souvent employé dans le domaine de la santé.

8.3 Les méthodes de tarifications

8.3.1 Modèles linéaires généralisé (GLM)

Bien qu'il existe des modèles plus avancés et performants que les modèles linéaires généralisés, les actuaires et assureurs apprécient ces derniers en raison de leur interprétabilité. En effet, ils permettent d'expliquer de manière précise et compréhensible les variations de tarification entre différents profils de clients.

Ainsi, la méthode préconisée¹ consiste à modéliser séparément la fréquence et le coût en utilisant des modèles linéaires généralisés. Cette approche est décrite dans l'annexe 12.5. Un point d'attention est porté à la légitimité de cette méthode dans le cadre de ce mémoire. En effet, il faut vérifier l'indépendance entre la fréquence et le coût, postulat qui n'est pas toujours vérifié pour tous les segments de tarifications retenus.

Dans le cas où cette indépendance n'est pas vérifiée, un GLM sur la consommation (prime pure) est mis en place. Ceci permet d'atténuer le risque de propagation d'erreur lorsque deux modèles sont mis en place. Cette approche s'intéresse au coût total par contrat et l'information sur la quantité d'actes est perdue.

8.3.2 Légitimité des GLM

Avant de pouvoir appliquer les modèles linéaires généralisés (GLM), il est effectivement important de s'assurer de la légitimité de leur utilisation. Théoriquement, la corrélation entre la variable de fréquence et la variable coût doit être nulle, mais dans le domaine de la santé il est possible que selon le niveau de garantie, la fréquence ne soit pas totalement indépendante du coût. En effet, il s'agit de la problématique de l'aléa moral, un patient s'il sait qu'il est bien remboursé par sa complémentaire santé peut avoir plus tendance à se rendre chez le médecin. C'est la raison pour laquelle ce mémoire adopte une approche de tarification basée sur deux niveaux de garantie, afin de légitimer l'utilisation de l'approche fréquence \times coût.

Dans le cadre de ce mémoire, il est admis qu'il puisse exister une légère corrélation entre la fréquence et le coût, toutefois, celle-ci ne doit pas dépasser un certain seuil d'acceptabilité. Pour effectuer cette vérification, trois tests de corrélation sont réalisés afin de réduire le risque d'utiliser deux modèles de tarification lorsque les hypothèses ne sont pas confirmées.

Les tests sont les suivants :

- Kendall ;
- Spearman ;
- Pearson.

Si les coefficients obtenus sont inférieurs à un certain seuil alors la méthode de modélisation par deux GLM sera privilégiée. Cependant, il n'existe pas de seuil universel à respecter, mais Fatemeh ABDOLLAHI (Tarification d'une complémentaire santé à destination des séniors, modulaire par poste de garanties et l'impact sur la solvabilité, 2017, p.57) [15] propose de prendre une limite

1. En assurance santé, c'est surtout d'un point de vue académique que les modélisations par deux GLM sont préconisées. En réalité, l'approche fréquence \times coût n'est pas toujours celle qui est retenue.

à 0,2 tandis que Sandrine HUYNH (*Open data et assurance santé : l'union fait la force ?* 2021, p73) [11] retient un seuil de 0,25.

Le choix du seuil dépend du contexte spécifique de l'étude, et des objectifs de modélisation. Ainsi, un premier seuil de 0,25 a été sélectionné, mais au vu des résultats des tests sur les différentes sous-familles de soins. Il a été observé que la plupart des résultats lorsqu'ils étaient inférieur à 0,30 en valeur absolue se situaient entre 0,25 et 0,30 en valeur absolue. Il a finalement été retenu le seuil de 0,30 pour l'ensemble des trois tests. Les détails de ces tests peuvent être retrouvés en annexe 12.5.3.

8.3.3 La statistique directe

À l'instar de Sandrine HUYNH dans son mémoire "*Open Data et Assurance santé : l'union fait la force*" (2021) [11], lorsque les méthodes présentées dans le paragraphe précédent ne sont pas adaptées aux données de la base NEW_DAMIR, alors une solution plus simple est sollicitée. Elle sera dénommée dans les travaux comme étant la méthode "statistique directe". Elle consiste à calculer la consommation en appliquant l'équation 8.8

8.3.4 Les méthodes retenues par segment

Si des modèles linéaires sont utilisés, des tests de lois pour la fréquence et la sévérité seront effectués en cas de deux GLM distincts. Si un seul GLM est utilisé pour la consommation, la loi de Tweedie sera systématiquement testée pour ce type de modélisation.

Par exemple, pour la sous-famille "Actes d'imagerie, de radiologie et ostéodensitométrie" (Garantie V1), la méthode de modélisation retenue est celle utilisant deux GLM. En revanche, pour la sous-famille "Actes médicaux techniques" (Garantie V2), la méthode "GLM Prime pure" est préconisée, avec l'utilisation d'un seul GLM. Ces deux exemples seront présentés dans la suite du mémoire.

Familles d'actes	Sous-familles d'actes	Méthode Garantie V1	Méthode Garantie V2	Lois utilisées
Honoraires médicaux	Consultations/visites	GLM : Fréquence x Coût	GLM : Fréquence x Coût	Binomiale Négative x Gamma
	Actes médicaux (techniques)	GLM : Fréquence x Coût	GLM Prime pure	Binomiale Négative x Gamma
	Autres honoraires médicaux	Absent de l'offre	Absent de l'offre	
Honoraires paramédicaux	Auxiliaires médicaux (kiné, infirmiers, ...)	Statistiques directe	Statistique directe	
Analyse et examen de laboratoire	Analyse médicale et examens labo	GLM : Fréquence x Coût	GLM Prime pure	Binomiale Négative x Gamma
Imagerie médicale	Actes d'imagerie, de radiologie et ostéodensitométrie	GLM : Fréquence x Coût	GLM Prime pure	Binomiale Négative x Gamma
Transport, ambulances	Transport	Statistiques directe	Statistique directe	
Pharmacie	Petit appareillage	Statistiques directe	Statistiques directe	
	Pharmacie	Statistiques directe	Statistiques directe	
	Vaccin anti-grippe	Statistiques directe	Statistiques directe	
Hospitalisation	Grand appareillage	Statistiques directe	Statistiques directe	
	Frais de séjour	Statistiques directe	Statistiques directe	
	Honoraires et actes chirurgicaux	GLM : Fréquence x Coût	GLM : Fréquence x Coût	Binomiale Négative x Gamma
	Forfait journalier	Statistiques directe	Statistiques directe	
	Chambre particulière	Absent de l'offre	Statistiques directe	
	Lit accompagnant	Absent de l'offre	Statistiques directe	
Dentaire	Autres - hospitalisation	Absent de l'offre	Absent de l'offre	
	Prothèse	GLM Prime pure	GLM Prime pure	Tweedie
	Soins dentaires	Statistiques directe	Statistiques directe	
	Parodontologie	Statistiques directe	Statistiques directe	
	Implantologie	Statistiques directe	Statistiques directe	
	Orthodontie	Statistiques directe	Statistiques directe	
Optique	Autres - dentaire	Absent de l'offre	Absent de l'offre	
	Monture	Statistiques directe	Statistiques directe	
	Verres	Statistiques directe	Statistiques directe	
	Chirurgie œil	Absent de l'offre	Absent de l'offre	
	Lentilles	Statistiques directe	Statistiques directe	
Audio	Autres - Optique	Absent de l'offre	Absent de l'offre	
	Audioprothèse	GLM : Fréquence x Coût	GLM : Fréquence x Coût	Binomiale Négative x Gamma
	Pile, accessoire	Statistiques directe	Statistiques directe	
Cure thermale	Cure thermale	Statistiques directe	Statistiques directe	
Médecine douce	Osthéopathie, diététique, chiropractie, ...	Statistiques directe	Statistiques directe	
Prévention	Prévention	Statistiques directe	Statistiques directe	
Prestations supplémentaires	Maternité	Absent de l'offre	Absent de l'offre	
	Autres (aides, inclassable, ...)	Absent de l'offre	Absent de l'offre	

FIGURE 8.1 – Méthode de tarification par segment

Finalement, parmi les 18 segments tarifaires définies initialement, les GLM seront utilisés pour seulement 7 sous-familles. Les lois mathématiques retenues seront donc appliquées à la fois pour l'assureur départemental et pour l'assureur régional.

CHAPITRE

9

ANALYSE UNIVARIÉE

L'objectif de ce chapitre consiste à exposer, le cas échéant, l'impact des facteurs externes sur la fréquence des remboursements et les montants remboursés par les organismes complémentaires de santé. Les variables externes incluses sont de nature quantitative, mais elles seront regroupées en catégories pour simplifier la structure des modèles employés et rendre leur interprétation plus accessible.

Idéalement, il faudrait procéder à un regroupement spécifique pour chaque segment tarifaire et pour chaque variable externe. Cependant, une telle approche demanderait un investissement en temps considérable pour des avantages qui ne sont pas encore complètement établis. Par conséquent, l'étude a été réalisée sur trois segments tarifaires issus de grands postes différents :

- les **soins dentaires** (Dentaire) ;
- les **honoraires et actes chirurgicaux** (Hospitalisation) ;
- les **actes médicaux techniques** (Soins courants).

Les regroupements sont identiques pour les trois sous-familles examinées, et il est présumé que ces mêmes regroupements seront applicables aux autres sous-familles.

Pour des raisons de concision, seules les analyses univariées pour la sous-famille "Actes médicaux techniques" seront exposées dans ce mémoire. Les analyses bivariées ont révélé que les modalités de chaque variable externe ne présentaient pas d'interactions significatives entre elles.

Dans cette optique, une analyse univariée a été menée en examinant d'abord la fréquence, puis les coûts. Malgré cela, il n'a pas été possible de déceler un impact significatif des variables externes sur la fréquence de la consommation ni sur le montant des remboursements octroyés par la complémentaire santé partenaire. Par conséquent, cette section se concentrera sur l'exploration des raisons potentielles justifiant l'absence d'un impact notable observé.

9.1 Analyse des catégories

L'objectif des cartographies affichées ci-dessous est de fournir une meilleure vision géographique des différentes modalités qui seront examinées dans la section suivante.

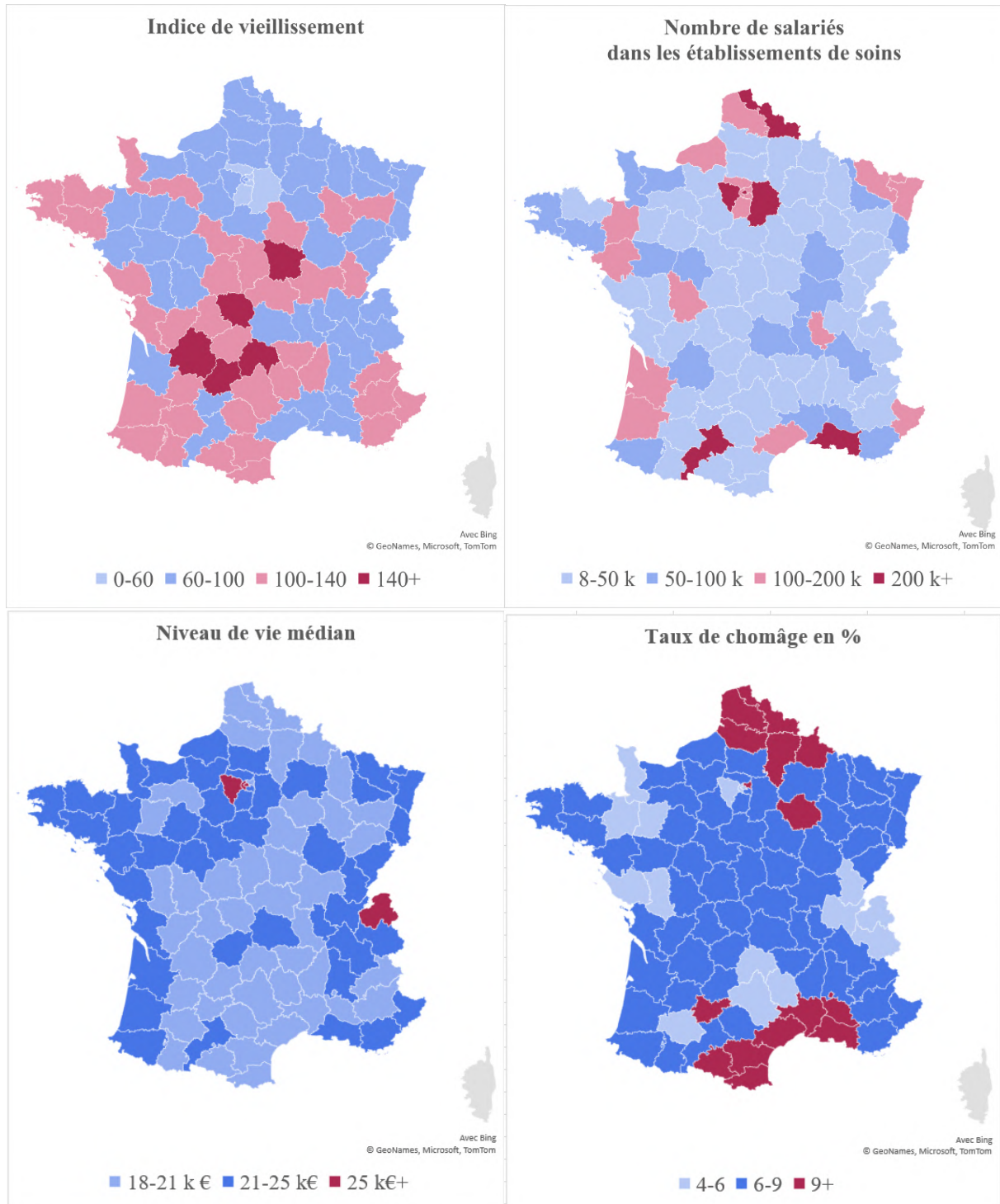


FIGURE 9.1 – Cartographie des différentes catégories pour les variables externes

Il convient également de rappeler que l'exposition aux différentes modalités peut influencer l'interprétation de l'analyse. En effet, une sous-représentation d'une modalité pourrait biaiser les résultats et conduire à une interprétation erronée de l'analyse. Par exemple, il est observé que pour le niveau de vie médian, seuls deux départements sont classés dans la catégorie "25 k€+" et que de manière générale, la répartition des modalités est très dispersée. Il conviendra alors d'être vigilant dans l'interprétation des résultats.

Ainsi, cette section commence par une brève analyse de la répartition des catégories, avant d'exposer leur impact potentiel sur les montants de remboursements ou leur fréquence.

Pour l'indice de vieillissement, il semblerait que les départements les plus jeunes se situent plutôt dans le nord-est de la France, tandis que les départements plus âgés se situent plutôt vers le centre et le sud-est de la France. Cela pourrait être expliqué par le fait que les régions du nord-est, abritant des centres universitaires et des opportunités professionnelles, attirent potentiellement une population plus jeune. En revanche, les climats doux et la qualité de vie dans le centre et le sud-est de la France pourraient favoriser l'installation de retraités.

La disponibilité des offres de soins montre une tendance générale : elle est plus répandue dans les grandes villes. Cette situation peut s'expliquer par la densité de population plus élevée et la demande croissante de services médicaux dans ces zones urbaines. De plus, ces départements ont généralement un accès plus facile aux ressources financières et technologiques nécessaires pour investir dans des services médicaux avancés.

En ce qui concerne le niveau de vie médian, seuls quelques départements de la région Île-de-France ainsi que le département de la Haute-Savoie sont classés dans la catégorie de revenus supérieurs à 25 000 €. Si la présence de la région Île-de-France dans cette catégorie n'est pas surprenante en raison de son attrait professionnel, la situation de la Haute-Savoie est plus inattendue. Cependant, des recherches en ligne révèlent qu'il s'agit en fait d'une région touristique très prisée, située dans les Alpes au pied du Mont-Blanc. Cela pourrait donc expliquer en partie son niveau de vie élevé.

Globalement, les taux de chômage les plus élevés sont observés dans les régions du nord et du sud de la France. Cependant, il a été difficile de mettre en lumière les raisons de ces taux de chômage élevés. À titre indicatif, certains articles de presse suggèrent que ces régions ont été durement touchées par des phénomènes tels que la désindustrialisation et la fermeture d'entreprises.

9.2 Analyse en fréquence

Les analyses qui suivront dépendent de la sous-famille "Actes médicaux techniques", il est possible que ces analyses ne soient pas correctes pour les autres sous-familles. Il est donc recommandé de les interpréter avec prudence.

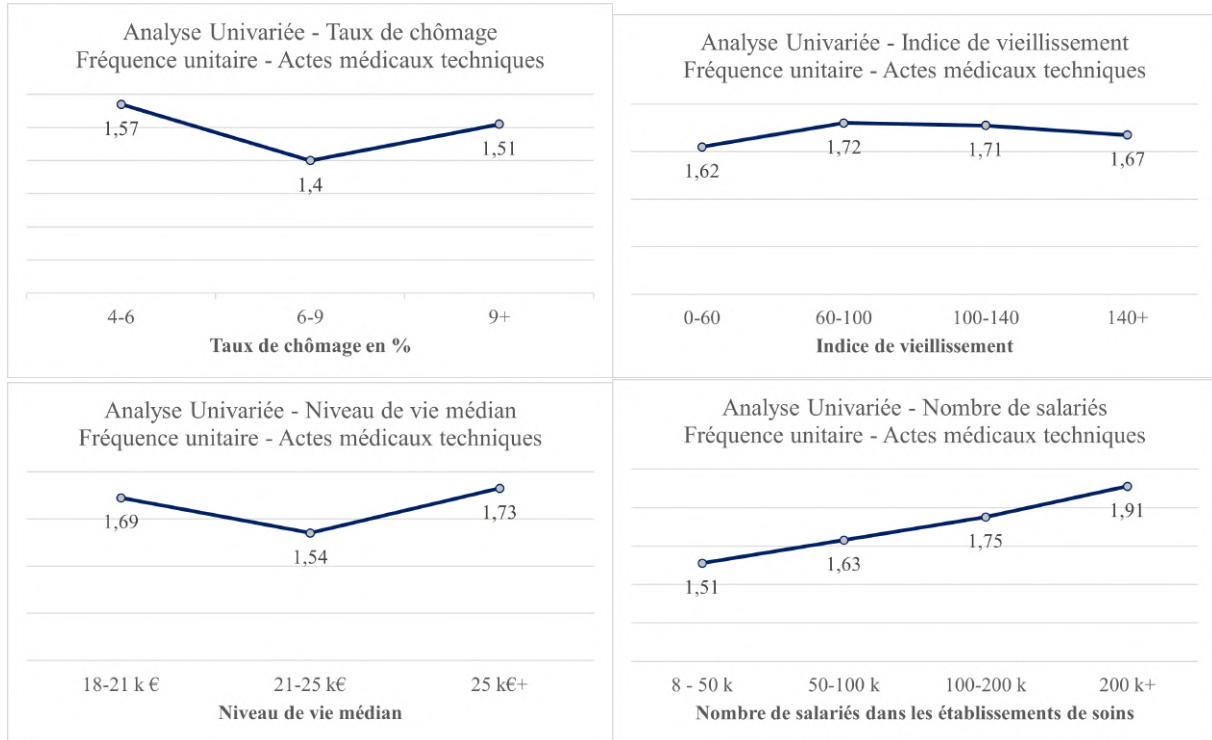


FIGURE 9.2 – Analyse univariée en fréquence pour les actes médicaux techniques

Le taux de chômage

Les départements présentant un taux de chômage modéré (6-9 %) affichent une fréquence de consommation de soins inférieure à celle observée dans les départements ayant un taux de chômage réduit (4-6 %) ou élevé (9%+). Cette disparité pourrait s'expliquer par le fait que les résidents des départements à taux de chômage modéré disposent potentiellement de moins de temps libre que ceux des départements à taux de chômage élevé. En effet, les individus au chômage ont généralement plus de disponibilité que les personnes employées, et peuvent également nécessiter plus de soins en raison de leur instabilité financière et de leurs préoccupations concernant l'avenir. Ce stress financier accru peut avoir un impact négatif sur leur santé, augmentant ainsi la fréquence de consommation de soins par rapport aux départements à taux de chômage modéré. Enfin, les départements à taux de chômage réduit bénéficient généralement de revenus plus élevés, ce qui pourrait leur permettre de consommer davantage de services de santé malgré une disponibilité plus faible due à leurs emplois.

La disponibilité des offres de soins

De manière intuitive, il pourrait être anticipé qu'une disponibilité accrue d'établissements de santé à proximité conduise à une augmentation de la fréquence des consultations médicales. Bien que cette tendance soit observée, l'impact ne semble pas non plus se manifester de manière flagrante. En effet, dans les départements considérés comme ayant une offre de soins limitée (8-50 k), la fréquence des consultations s'élève en moyenne à 1,51 consultations, contre 1,91 dans les départements disposant d'une offre de soins très abondante (200 k+). Les individus résidant dans des départements où les services de soins sont moins accessibles doivent alors parcourir des distances plus importantes pour recevoir des soins, ce qui pourrait en décourager certains.

Le niveau de vie médian

La fréquence de consommation des actes médicaux techniques varie selon les départements ayant un niveau de vie médian faible (18-21 k€), situé entre les départements au niveau de vie médian modéré (21-25 k€) et ceux au niveau de vie médian élevé (25 k€+). Il est envisageable que les départements à revenu plus bas puissent être davantage caractérisés par une proportion plus élevée de métiers exigeants physiquement, tels que des postes ouvriers, ce qui pourrait entraîner une fatigue accrue et des habitudes de vie moins saines. Par conséquent, il se pourrait que la fréquence des actes médicaux techniques soit plus élevée, car ces individus pourraient nécessiter des soins par obligation et ne pas être en mesure de différer leur traitement, même en présence de contraintes financières.

Dans le cas des "revenus moyens", qui sont potentiellement associés à des professions moins physiquement exigeantes que les emplois ouvriers, il est envisageable que des contraintes financières puissent également restreindre leur fréquence de consultations médicales. De plus, leur besoin de recourir aux soins pourrait être moindre. Enfin, les départements ayant des revenus plus élevés pourraient être associés le plus souvent à une majorité de cadres qui seraient alors installés dans des bureaux, avec une contrainte financière moindre. De plus, ces individus pourraient être en mesure de se libérer plus facilement pour des raisons médicales, ce qui pourrait expliquer une augmentation de la fréquence de consommation de soins.

L'indice de vieillissement

Intuitivement, la fréquence de consommation des soins pratiqués par des spécialistes devrait augmenter en lien avec l'indice de vieillissement. Cependant, cela ne se vérifie pas exactement. En effet, cela pourrait s'expliquer par le fait que la plupart des départements classés dans les catégories présentant un indice supérieur à 100 sont situés dans des milieux ruraux. Il est légitime de penser qu'il existerait une interaction entre les deux variables, mais une telle corrélation n'a pas été détectée. Peut-être est-ce dû au fait que les départements d'Île-de-France, bien que situés dans une zone avec une abondance d'offres de soins, figurent parmi les départements les plus jeunes de France. De ce fait, l'analyse de la variable indice de vieillissement semble complexe lorsqu'elle est examinée de manière isolée. Les variations sont encore plus minimes d'une modalité à l'autre.

9.3 Analyse en coût

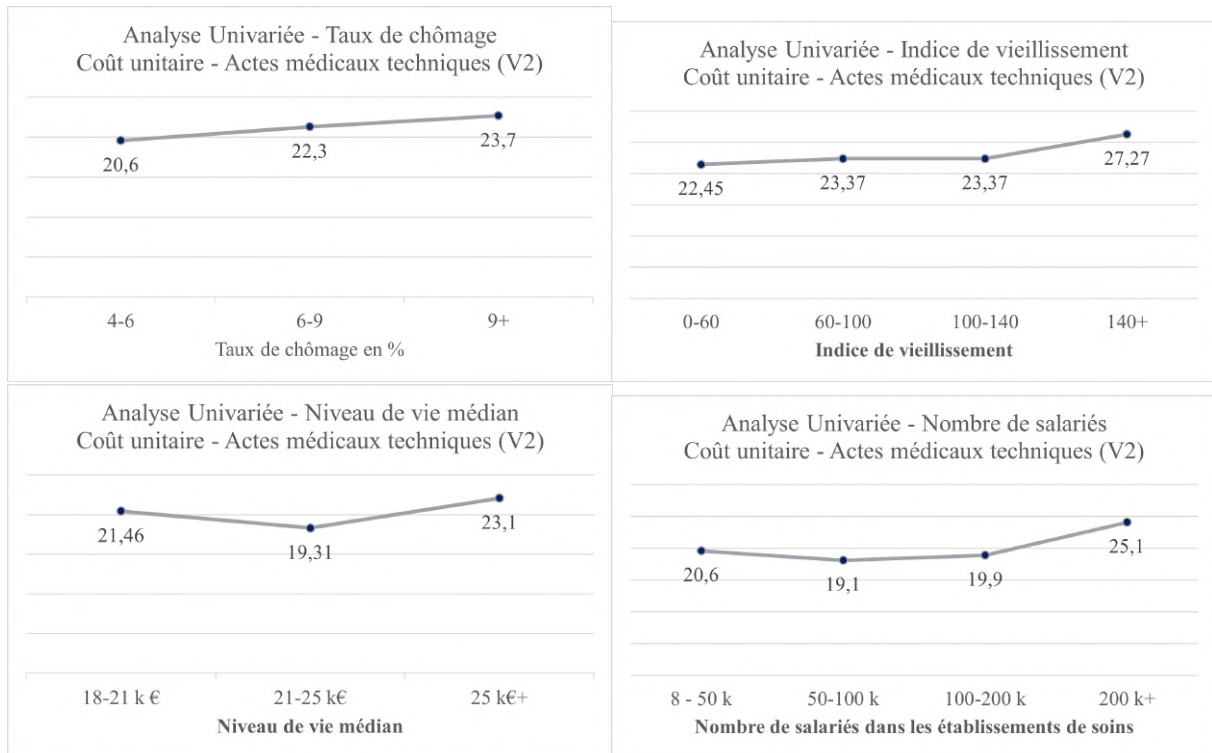


FIGURE 9.3 – Analyse univariée en coût pour les actes médicaux techniques

Le taux de chômage

Il semble que les coûts unitaires des actes médicaux techniques augmentent avec le taux de chômage au niveau départemental. En effet, il est plausible d'envisager que les départements où le taux de chômage est plus bas connaissent une utilisation plus fréquente des services de santé préventive, sans attendre une détérioration de leur état de santé. En revanche, dans les départements où le taux de chômage est modéré, il est possible qu'il y ait un délai dans la recherche de soins, ce qui pourrait potentiellement entraîner une dégradation de l'état de santé. De plus, dans les régions où le chômage est élevé, les comportements à risque pourraient être plus répandus, ce qui pourrait également avoir des conséquences néfastes sur la santé.

La disponibilité des offres de soins

Il est possible d'observer que les coûts unitaires sont légèrement plus élevés dans les zones présentant une offre de soins limitée (8-50 k) par rapport aux départements où l'offre est modérée (50-100 k) voire élevée (100-200 k). Cette situation pourrait être attribuée à des délais d'attente prolongés, qui ont également le potentiel d'aggraver l'état de santé d'un patient, entraînant ainsi une augmentation des coûts des soins nécessaires. En revanche, les grandes agglomérations dotées d'une offre de soins abondante (200 k+) connaissent généralement des situations de dépassement d'honoraires fréquentes, contribuant ainsi à une augmentation du coût unitaire des soins.

Le niveau de vie médian

Il est constaté que les départements caractérisés par un niveau de vie médian faible (18-21 k€) présentent des coûts unitaires plus élevés que ceux ayant un niveau de vie médian situé entre 21 et 25 k€. Cette différence pourrait potentiellement résulter de besoins de soins plus importants dans cette catégorie, en raison des contraintes physiques associées aux emplois. En contraste, il est observé que les départements avec un niveau de vie légèrement plus élevé pourraient également présenter des coûts unitaires accrus, ce qui pourrait également être attribué à la possibilité de dépassements d'honoraires.

L'indice de vieillissement

L'analyse en fréquence de cette variable s'est avérée difficile. En ce qui concerne les coûts, il est tout de même observé que les départements présentant un indice de vieillissement élevé affichent des coûts supérieurs aux autres modalités. Cette observation pourrait être attribuée nécessité accrue de soins spécialisés pour les personnes âgées.

Il est important de souligner que toutes ces analyses peuvent varier en fonction de l'individu. Chaque personne possède ses propres visions et réagit de manière différente. De plus, la culture de la santé peut différer d'un département à un autre. Par conséquent, il n'a pas été possible de dégager des tendances nettes à travers les données de la base NEW_DAMIR. En d'autres termes, il est difficile de produire des analyses unanimement acceptées par tous, car il existe de nombreuses variables et informations potentiellement inconnues au niveau départemental qui pourraient influencer les résultats des analyses.

CHAPITRE

10

MODÉLISATION DE LA PRIME PURE

Un des objectifs opérationnels de ce mémoire consiste à établir la mise en place de deux tarifs distincts. Le premier tarif s'appliquerait à un assureur qui mutualise son risque au niveau départemental, dénommé "l'assureur départemental". Le second tarif serait conçu pour un assureur réalisant une mutualisation de son risque au niveau régional, dénommé "l'assureur régional".

Ce chapitre présente la démarche entreprise pour établir les tarifs. Un exemple de modélisation est donné en utilisant deux modèles GLM, suivi d'un modèle "consommation".

10.1 La modélisation par deux modèles

Dans le cas où, les tests de corrélation sont favorables à une modélisation par deux modèles, il est nécessaire de déterminer les lois correspondantes.

Certaines règles générales peuvent être suivies pour avoir une première idée des lois à tester. Par exemple, la loi de Poisson est utilisée pour compter des données discrètes tandis que la loi Gamma est utilisée lorsque les variables sont continues et positives.

Si aucune intuition ne se dégage quant aux lois appropriées ou si besoin d'une confirmation sur l'intuition, un diagramme de Cullen et Frey peut être utilisé pour obtenir des idées sur les lois potentielles à tester.

Un exemple est présenté pour la sous-famille "Actes d'imagerie et de radiologie", les tests de corrélations donnent les résultats suivants :

		Coefficient V1
Actes d'imagerie et de radiologie	Kendall	-0,07
	Pearson	-0,12
	Spearman	-0,1

FIGURE 10.1 – Résultats des tests de corrélations (Actes d'imagerie et de radiologie)

L'ensemble des trois coefficients en valeur absolue est inférieur au seuil de 0,30 laissant penser qu'une modélisation par deux GLM est envisageable.

10.1.1 Test de loi de fréquence

Une première intuition pousse à tester une loi de Poisson, car il s'agit d'une loi usuelle pour la fréquence. Cependant, la loi de Poisson nécessite que l'espérance et la variance soient égales, ce qui n'est pas le cas dans les données de NEW_DAMIR. En effet, la variance est supérieure à l'espérance dans la majorité des sous-familles. Il s'agit d'un phénomène de sur-dispersion.

Dans ce cas, les lois candidates se tournent alors vers la loi de Poisson sur-dispersé et la binomiale négative qui sont également des lois usuelles dans la modélisation de la fréquence. Un diagramme de Cullen et Frey vient confirmer les éventuelles adéquations des lois énoncées.

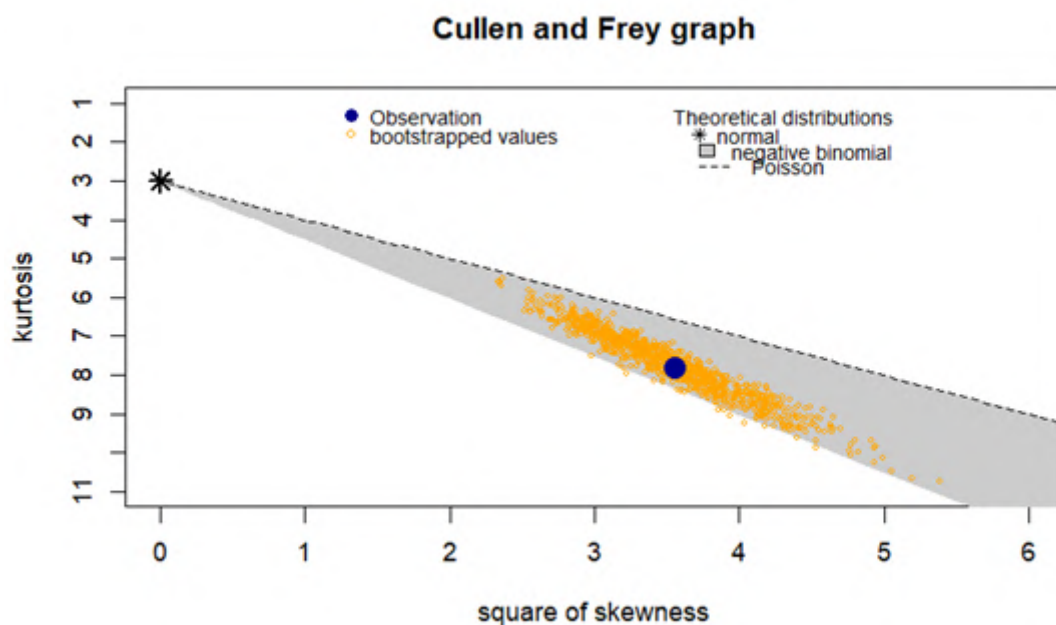


FIGURE 10.2 – Diagramme de Cullen et Frey pour la loi de fréquence (Actes d'imagerie et de radiologie)

Le diagramme indique que les données de fréquence de la sous-famille "Actes d'imagerie et de radiologie" tendent à suivre une loi Binomiale négative. Il est à noter que la variable cible dans cette modélisation est la quantité d'actes agrégée, définie dans le domaine des entiers naturels. L'exposition est prise en compte en tant qu'offset.

Le graphique 10.3 trace le QQ Plot de la loi Binomiale Négative pour la sous-famille "Actes d'imagerie et de radiologie". Le QQ plot, appelée diagramme quantile-quantile est un outil graphique qui compare les quantiles d'une distribution empirique (les observations en gris) aux quantiles théoriques d'une distribution de référence (le modèle théorique en bleu).

Ainsi, il est observé que les données semblent bien correspondre à une loi Binomiale Négative hormis la queue de distribution.

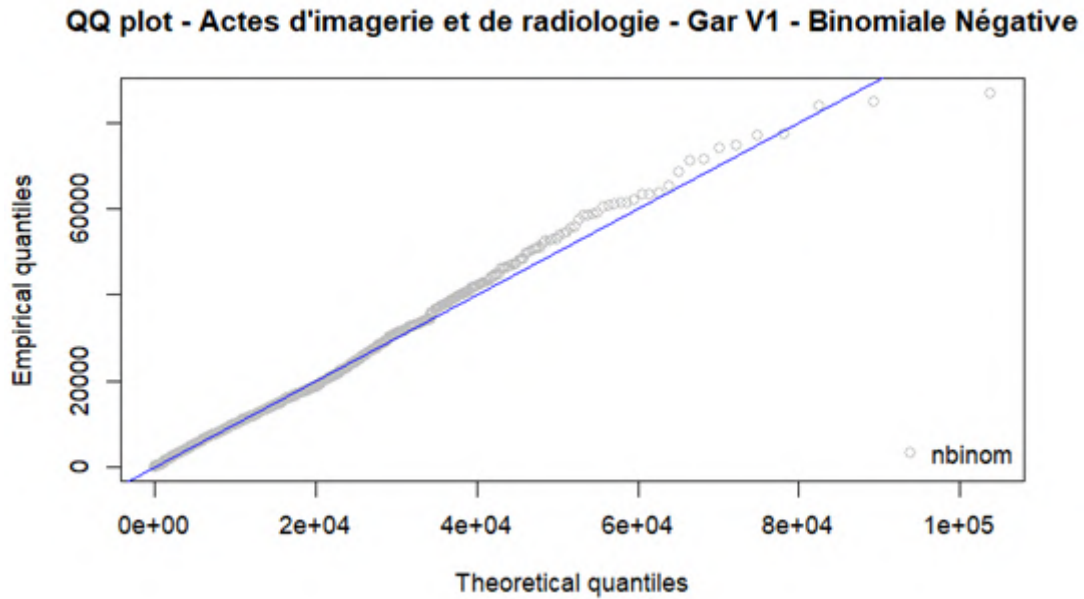


FIGURE 10.3 – QQ Plot - Binomiale négative (Actes d'imagerie et de radiologie)

10.1.2 Test de loi pour le coût

De la même manière, il existe des lois usuelles de coût telles que la loi gamma, normale, log-normale, etc.. Ainsi, un diagramme de Cullen et Frey est également tracé pour les montants remboursés par la complémentaire santé.

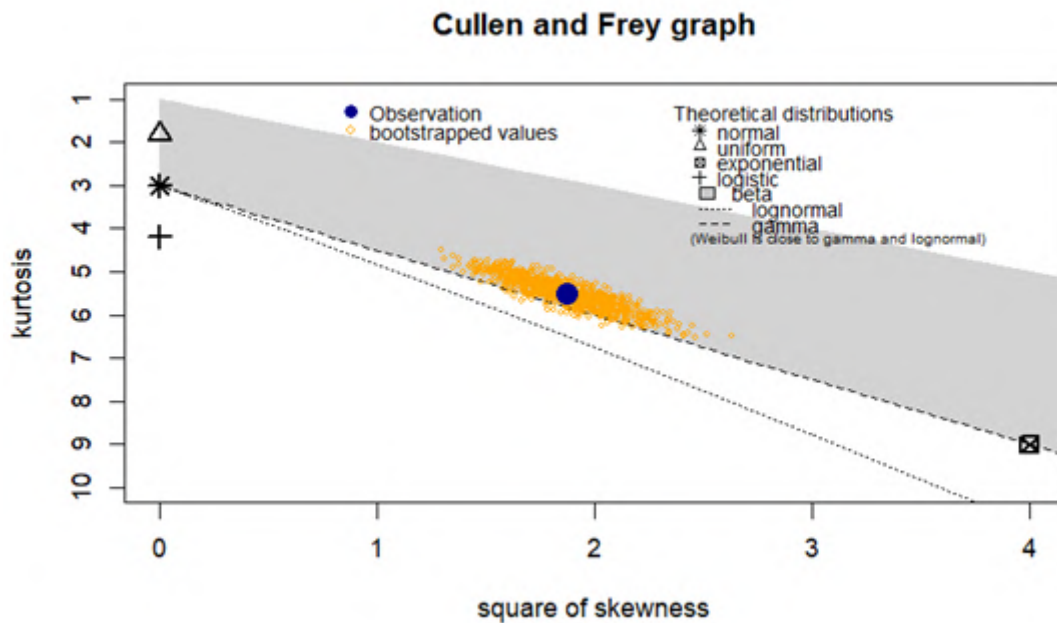


FIGURE 10.4 – Diagramme de Cullen et Frey pour la loi de coût (Actes d'imagerie et de radiologie)

Les montants remboursés par la complémentaire santé semblent suivre une distribution qui pourrait être proche d'une loi gamma ou beta mais un test entre les lois suivantes sera effectué :

- Gamma ;
- Beta ;
- Log Normale ;
- Weibull

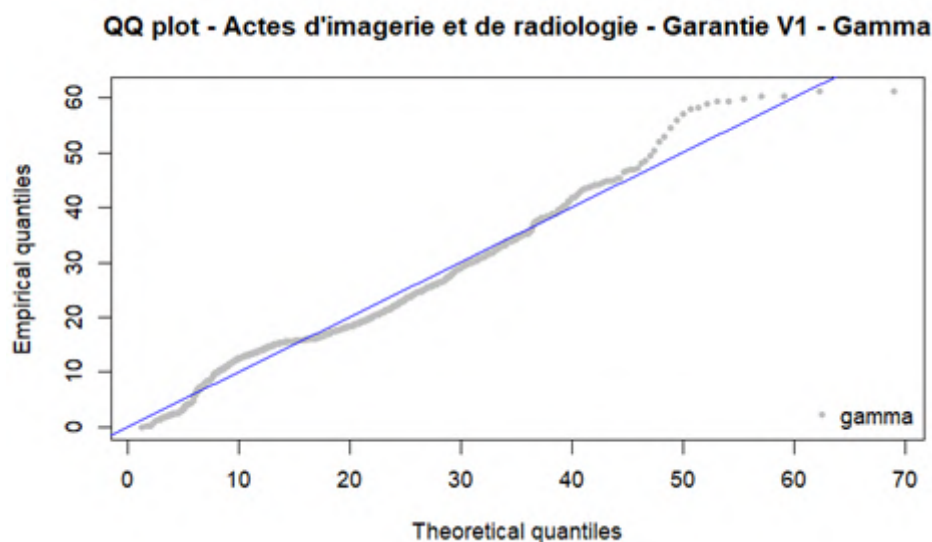


FIGURE 10.5 – QQ Plot - Gamma (Actes d'imagerie et de radiologie)

Il est à noter que la loi Beta est définie sur l'intervalle $[0, 1]$, donc l'ensemble des valeurs a été divisé par la valeur maximale afin d'effectuer le test tout de même.

Ainsi, afin de sélectionner la loi qui serait la plus adaptée aux données, la loi qui présentait le plus petit AIC (Critère d'Information Akaike) était retenue. Ce critère cherche le meilleur compromis entre la vraisemblance et la qualité du modèle.

$$\text{AIC} = 2k - 2 \ln(L) \quad (10.1)$$

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

	AIC
Gamma	12 075
Log normale	12 198
Weibull	29 110

FIGURE 10.6 – AIC selon la loi candidate de coût - (Actes d'imagerie et de radiologie)

Dans toutes les sous-familles où la modélisation par deux GLM a été préconisée, la loi Gamma a été choisie par rapport aux autres lois candidates.

10.1.3 Sélection des variables

Pour la sous-famille "Actes d'imagerie et de radiologie", la loi Binomiale Négative et la loi Gamma ont été sélectionnées respectivement pour la fréquence et la sévérité. Il est nécessaire de réaliser une sélection de variable afin d'obtenir le modèle le plus performant. Les résultats sont présentés pour les deux types d'assureurs du mémoire.

Pour l'assureur qui établit un tarif au niveau départemental

Les variables externes ne peuvent pas être considérées comme explicatives dans le modèle, car un problème de multi-colinéarité est observé entre la variable **DPT** et toutes les variables externes, ce qui est normal étant donné qu'elles ont été ajoutées par correspondance à la variable **DPT**.

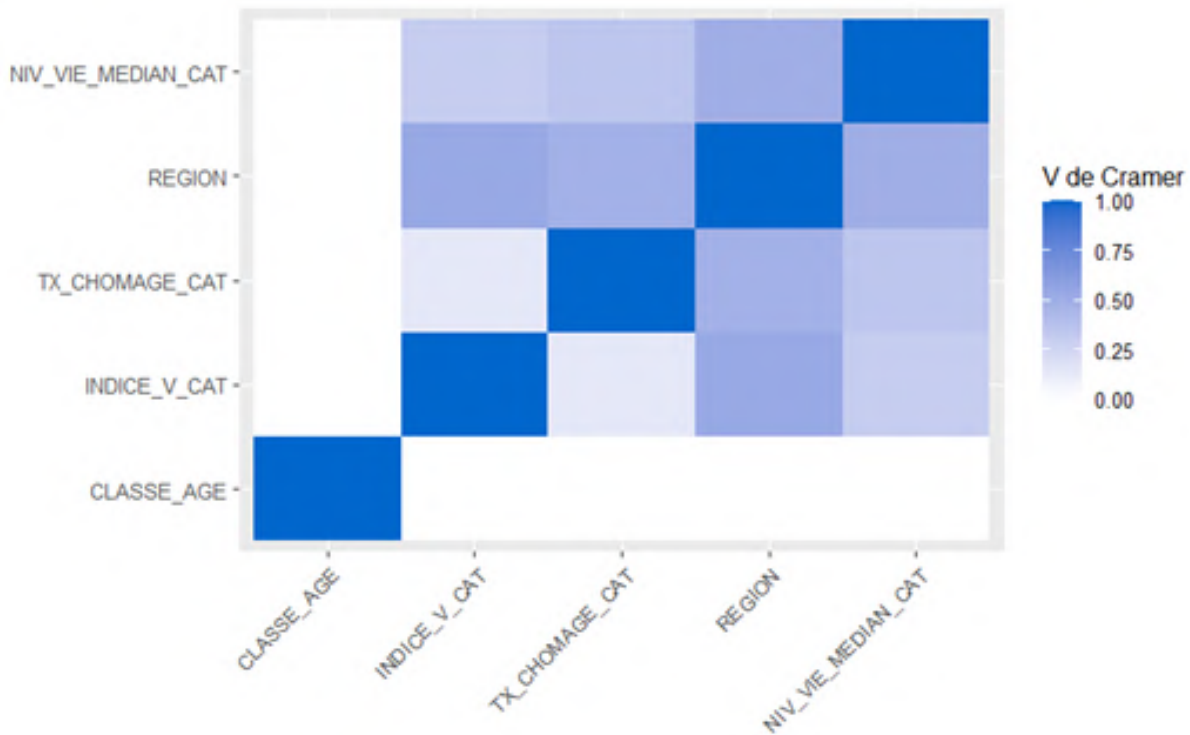


FIGURE 10.7 – Matrice de corrélation par V de Cramer - (Actes d'imagerie et de radiologie)

Le V de Cramer vient confirmer l'impossibilité d'ajouter les variables externes en tant que variables explicatives, car elles sont corrélées avec la variable **DPT**. Il aurait été possible de prendre en compte ces éléments, cependant, cela aurait nécessité de retirer la variable de département. Étant donné que la localisation au sein de la mémoire est un point crucial du sujet, il a été décidé de conserver la variable département par rapport aux autres variables.

Cependant, ces variables externes ne sont que faiblement corrélées entre elles et seront testées pour l'assureur "régional".

La classe d'âge et le département sont donc les seules variables explicatives sélectionnées pour l'assureur "départemental".

Pour l'assureur qui établit un tarif au niveau régional

Dans ce mémoire, l'ajout des variables externes dans le modèle de tarification était initialement prévu. Cependant, les tests de performance n'ont pas été concluants, ce qui signifie que les variables externes ne permettaient pas d'améliorer de manière significative le modèle. En effet, les tarifs obtenus en fin de partie étaient sensiblement égaux à ceux du modèle qui ne prenait pas en compte les variables externes.

La méthode de sélection choisie est la méthode Forward. Cette approche débute avec un modèle ne contenant aucune variable (Null Model) puis ajoute de manière itérative la variable la plus significative. La significativité de la variable a été mesurée par la quantité d'AIC qu'elle retirait au modèle, c'est-à-dire que plus la variable faisait diminuer l'AIC du modèle, plus elle était considérée comme importante.

À titre indicatif, voici les variables classées par ordre d'importance :

Variables	AIC Retiré
CLASSE_AGE	2 090,93
REGION	104,11
NB_SALARIE_SOINS	13,19
TX_CHOMAGE	4,67
INDICE_VIEILLISSEMENT	3,71
NIV_VIE_MEDIAN	1,15

FIGURE 10.8 – Methode forward - Selection de variables (Actes d'imagerie et de radiologie)

L'ajout des variables externes ne semble pas viable pour cette sous-famille de soin, mais les mêmes conclusions ont été obtenues en effectuant les tests sur les autres segments de tarification. De plus, d'un point de vue opérationnel, un assureur lambda ne demanderait pas au souscripteur des informations telles que son taux de chômage, son indice de vieillissement, etc. En effet, ce dernier n'en aura pas connaissance, et même si l'information était incorporée ultérieurement par correspondance, l'ajout de ces informations ne procurerait pas un réel bénéfice au modèle.

De plus, il est difficile de justifier la différence d'un tarif entre une personne au chômage et une personne en emploi. Il serait plus approprié d'inclure la proportion de ces individus, semblable à l'approche utilisée pour la variable du sexe, et d'établir un tarif uniforme pour les deux catégories. Cette démarche pourrait être réalisée en employant une analyse univariée ou bivariée des variables externes. Toutefois, comme mis en évidence dans le chapitre précédent, cette analyse n'a pas abouti à des conclusions significatives quant à leur impact.

Il a donc été décidé d'écarter ces variables du modèle et de ne conserver que la classe d'âge et la région.

La sélection des variables a été présentée pour le segment de tarification "Actes d'imagerie et de radiologie", mais les mêmes variables seront conservées pour l'ensemble des autres segments de tarification.

	Assureur DPT	Assureur REGION
Fréquence	CLASSE_AGE DPT	CLASSE_AGE REGION
Coût	CLASSE_AGE DPT	CLASSE_AGE REGION

FIGURE 10.9 – Variables explicatives retenues

10.1.4 Validation des modèles

Afin de valider le modèle, la base initiale est séparée en deux bases :

- une base d'apprentissage qui entraîne le modèle sélectionné ;
- une base de test qui permet comme son nom l'indique de tester le modèle et de le valider.

Il existe plusieurs méthodes de séparation des données. La première méthode consiste à créer la base d'apprentissage en sélectionnant aléatoirement un échantillon de données dont la taille représente généralement 80% de la base de données initiale. Les données restantes, soit 20% dans ce cas, sont alors utilisées pour former la base de test.

Cependant, en appliquant cette méthode, il a été observé que certains départements étaient manquants dans la base de test. Par exemple, toutes les lignes du département "92" étaient dans la base d'apprentissage et aucune n'était présente dans la base de test. Cela pose un problème, car il n'est alors pas possible d'effectuer une prédiction pour le département absent, empêchant ainsi une tarification pour ce dernier.

Ainsi, la méthode de séparation stratifiée a été choisie dans ce mémoire. En effet, elle permet de conserver la répartition des modalités de la variable "DPT" afin de garantir la présence de tous les départements à la fois dans l'ensemble d'apprentissage (train) et dans l'ensemble de test.

La qualité d'ajustement du modèle est vérifiée si la déviance résiduelle est inférieure au quantile d'une loi du chi-deux d'ordre 0,95 avec les degrés de liberté associés. Le test est détaillé en annexe 12.5.6.

FREQUENCE			
Variables explicatives	Degré de liberté	Quantile à 0,95 (χ^2)	Deviance
CLASSE_AGE DPT	1102	1180,34	1123,7
CLASSE_AGE REGION	1184	1265,16	1217,2
COÛT			
Variables explicatives	Degré de liberté	Quantile à 0,95 (χ^2)	Deviance
CLASSE_AGE DPT	1102	1180,34	231,8
CLASSE_AGE REGION	1184	1265,16	314,9

FIGURE 10.10 – Validation des modèles par test de déviance

Les déviations résiduelles des modèles pour la fréquence et le coût sont bien inférieures au quantile d'une loi du chi-deux d'ordre 0,95 avec les degrés de liberté associés. Ils sont donc validés.

Les résultats de prédiction pour cette sous-famille sont présentés dans la figure ci-dessous :

Variables explicatives	Entité	Base test		
		Réel	Prédit	Écart
CLASSE_AGE + DPT	Fréquence moyenne (Unitaire)	1,95	1,97	0,84%
	Coût moyen unitaire (Garantie V1)	15,11	15,68	3,79%
CLASSE_AGE + REGION	Fréquence moyenne (Unitaire)	1,95	2,03	3,96%
	Coût moyen unitaire (Garantie V1)	15,11	15,46	2,31%

FIGURE 10.11 – Résultats modélisation par 2 GLM (Actes d'imagerie et de radiologie)

D'après les résultats, le modèle semble prédire à la hausse la fréquence et le coût, mais que les écarts relatifs ne dépassent pas 4 % en valeur absolue. Il est conclu que les modèles sont satisfaisants.

10.2 La modélisation par un modèle : Tweedie

Lorsque les résultats des tests de corrélation ne sont pas en faveur d'une modélisation par deux modèles GLM, alors la méthode "Consommation" est mise en avant. Cela revient à modéliser la prime pure directement par une "loi de Tweedie". Il s'agit d'une famille de distributions issue de la famille exponentielle. La définition est détaillée en annexe 12.5.3.

La démarche est similaire à celle utilisée pour la modélisation à deux modèles, ce qui permettra de présenter les résultats de manière plus concise.

Ainsi, la sous-famille "Actes techniques médicaux" pour la garantie V2 présente les résultats suivants pour les tests de corrélation :

		Coefficient V2
Actes techniques médicaux	Kendall	0,44
	Pearson	0,39
	Spearman	0,64

FIGURE 10.12 – Résultats des tests de corrélations (Actes médicaux techniques)

Les coefficients dépassant le seuil de 0,30 en valeur absolue, une modélisation par la loi de Tweedie est envisagée pour cette sous-famille. Concrètement, cette loi intègre la loi Gamma et la loi Poisson-Gamma afin de modéliser la prime pure directement.

Ainsi, la variable à expliquer est le montant de remboursement de la mutuelle (MT_RC_V1 ou MT_RC_V2), et l'exposition totale est traitée comme un offset dans le modèle. Le paramètre de la loi est estimé par la fonction *tweedie.profile* et s'élève à **2,28** et la densité théorique est tracée sur le graphique ci-dessous :

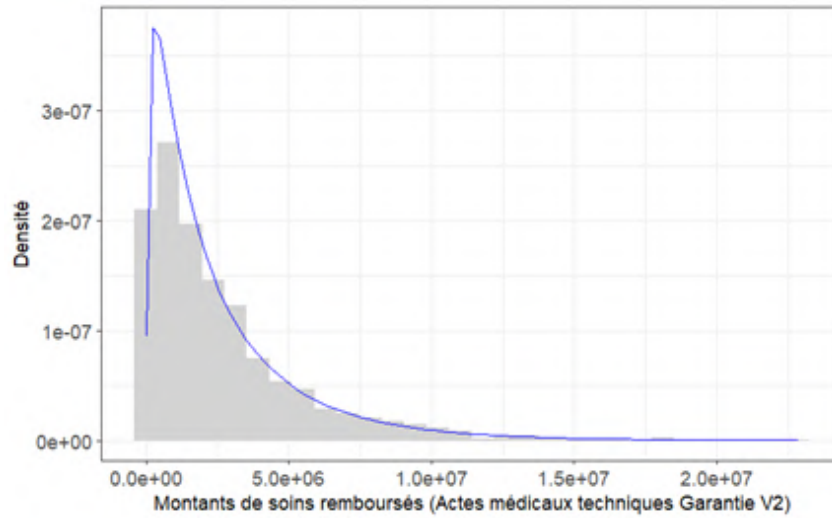


FIGURE 10.13 – Histogramme empirique VS densité théorique (Actes médicaux techniques - Garantie V2)

D'après le graphique, il semble avoir une bonne adéquation de la loi Tweedie. De plus, elle pourrait se comporter comme une loi Gamma ou une distribution positive stable car le paramètre se rapproche de 2.

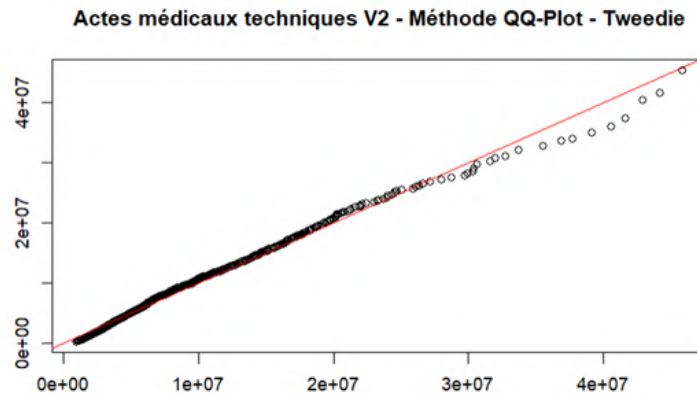


FIGURE 10.14 – QQ-plot Tweedie (Actes médicaux techniques - Garantie V2)

Le QQ-plot présente une adéquation de loi satisfaisante. Si tel n'était pas le cas, cela indiquerait que la sous-famille "actes médicaux techniques" pour la garantie V2 ne serait adaptée ni à une modélisation par deux GLM ni à une modélisation par un GLM. Dans ce cas, l'approche statistique aurait été mise en œuvre. Les résidus sont très inférieurs au quantile d'une loi du chi-deux d'ordre 0,95 et la consommation moyenne empirique de la base test s'élève à 23,33 contre 22,60. L'écart relatif en valeur absolue est de 3,13 %, il est décidé que le modèle prédictif est satisfaisant.

Synthèse de la troisième partie

Au départ, afin d'obtenir une meilleure compréhension des évolutions du marché, une analyse univariée a été entreprise dans le but de repérer les éventuelles influences des variables externes sur les fréquences et les montants de remboursements au sein de la mutuelle communale. Toutefois, en raison de la variabilité des comportements des habitants, aussi bien entre différents départements que à l'intérieur d'un même département, il s'est avéré difficile de mettre en évidence un impact significatif sur les tendances de consommation.

Pour établir les tarifs des garanties V1 et V2 pour la mutuelle communale, trois approches sont mises en œuvre. La décision concernant la méthode à adopter est guidée par des tests de corrélation préliminaires qui visent à évaluer la pertinence du choix de méthode pour chaque modèle. Si une indépendance est observée entre la fréquence et les coûts, la préférence va vers l'application d'un modèle linéaire généralisé (GLM) à la variable de fréquence ainsi qu'à la variable de coût. En cas de corrélation, un autre modèle linéaire généralisé appliqué directement à la prime pure est utilisé. Enfin, lorsque les deux premières approches ne donnent pas de résultats concluants ou ne sont pas adaptées aux données, une méthode de calcul direct est mise en place en se basant sur les statistiques extraites des données.

Parmi les 18 segments tarifaires définis dans la deuxième partie du mémoire, une sélection de 7 segments a été faite pour l'application des GLM. Ainsi, il convient de procéder à des tests pour évaluer la pertinence des lois mathématiques en accord avec les GLM. En général, la loi Binomiale Négative a été choisie pour modéliser la fréquence, tandis que la loi Gamma a été retenue pour modéliser les coûts dans le cadre de deux GLM distincts. En ce qui concerne le GLM sur la prime pure, la loi Tweedie a constamment été adoptée.

Un des objectifs du mémoire est de développer deux tarifs distincts, l'un pour chaque département et l'autre pour chaque région. Les variables tarifaires retenues étaient soit le département, soit la région en combinaison avec la classe d'âge des adhérents.

Ainsi, pour élaborer le tarif par département, il était essentiel de maintenir la variable départementale en tant que variable explicative. Cela a eu pour conséquence que les variables externes introduites en fonction du département étaient fortement corrélées avec celle-ci, ce qui les a empêchées d'être intégrées dans la modélisation. Quant au tarif par région, il a été observé que les variables externes avaient une influence relativement limitée en comparaison avec la variable région et la classe d'âge.

En termes opérationnels, il était également compliqué pour un assureur de solliciter des informations telles que l'indice de vieillissement, le salaire d'une personne et la disponibilité d'offres de soins à proximité. En pratique, exiger un paiement plus élevé pour l'assurance d'une personne au chômage soulève des questions de justification. L'objectif visait davantage à intégrer ces divers indicateurs en vue de créer un tarif global, à l'instar de la tarification entre les hommes et femmes, où le tarif est uniforme tout en prenant en considération la proportion d'hommes et de femmes au sein d'un portefeuille d'assurance. Cela n'a pas conduit à une conclusion satisfaisante.

Quatrième partie

Les tarifs et le marché concurrentiel des mutuelles communales

La quatrième section est consacrée à la présentation des tarifs obtenus. Elle met en premier lieu en avant les résultats des tarifs pour l'assureur régional, puis elle se penche sur la présentation des tarifs par département. Ces derniers seront ensuite comparés aux tarifs régionaux afin de déterminer la pertinence potentielle de l'établissement d'une mutuelle régionale pour certains territoires.

Dans le cas où cette pertinence serait établie, il est envisageable que des mutuelles départementales soient déjà présentes sur le territoire. Dans ce contexte, la concurrence serait simulée à travers un marché concurrentiel fictif opposant l'assureur régional à l'assureur départemental, ce qui permettrait d'estimer les parts de marché réellement accessibles avec la mise en place d'un partenariat à l'échelle régionale.

CHAPITRE

11

LES RÉSULTATS

Ce chapitre présente les tarifs globaux obtenus pour l'ensemble des postes de soins. Tout d'abord, le tarif pour l'assureur régional est exposé, suivi du tarif de l'assureur départemental. Par la suite, une analyse comparative des tarifs sera effectuée non seulement entre les départements de la région Île-de-France, mais également entre les départements de la région Nouvelle-Aquitaine.

11.1 Les tarifs de l'assureur régional

Les tarifs sont présentés sous forme de primes pures mensuelles pour la couverture globale, ce qui signifie qu'il s'agit de la somme totale des sous-primes calculées pour chaque sous-famille.

11.1.1 La garantie entrée de gamme V1

Les résultats de la garantie V1 sont présentés

GARANTIE V1	CLASSE AGE							
REGION	0	20	30	40	50	60	70	80
Île-de-France	17,5	21,2	26,8	31,1	41,1	54,8	68,5	104,9
Centre-Val de Loire	14,1	16,2	19,9	23,8	32,8	42,8	58,4	88,0
Bourgogne-Franche-Comté	13,2	16,3	20,0	23,7	31,2	42,1	58,2	87,1
Normandie	14,4	16,5	19,7	23,7	32,7	41,6	57,5	81,7
Hauts-de-France	14,7	16,2	20,8	25,7	34,6	44,7	64,1	101,2
Grand Est	15,0	17,6	23,1	26,7	36,0	48,1	57,4	91,2
Pays de la Loire	14,9	17,1	21,2	25,4	32,6	41,2	63,1	84,0
Bretagne	14,8	16,5	21,6	24,3	32,9	42,9	58,0	88,8
Nouvelle-Aquitaine	13,7	16,6	22,8	25,9	34,0	43,6	58,6	91,3
Occitanie	16,4	20,7	25,0	28,8	38,3	49,0	65,7	112,0
Auvergne-Rhône-Alpes	14,3	17,5	23,1	26,7	34,9	44,7	60,7	89,8
Provence-Alpes-Côte d'Azur	18,5	23,6	27,2	31,7	44,2	57,4	76,6	134,3

FIGURE 11.1 – Primes pures mensuelles Garantie V1

Deux effets principaux se manifestent. Tout d'abord, le tarif connaît une augmentation significative en fonction de l'âge. De plus, un effet lié à la région est observé, bien que son impact ne soit pas considérable. En d'autres termes, les tarifs varient d'une région à l'autre, mais cette variation n'est pas excessive. Cette situation peut être attribuée au niveau de la garantie (entrée de gamme), où les plafonds de remboursement sont atteints, restreignant ainsi les possibilités de remboursement par la complémentaire santé et, par conséquent, limitant la variation des tarifs.

Il convient de souligner qu'il existe tout de même un saut tarifaire entre les tranches d'âge de 70 et 80 ans pour les régions de l'Occitanie et de la Provence-Alpes-Côte-d'Azur. Il est envisageable que dans ces régions, certains départements soient insuffisamment représentés en termes de bénéficiaires, ce qui pourrait induire des biais dans les modèles. Par ailleurs, une autre explication à cet écart pourrait découler des coefficients de ventilation utilisés lors de l'établissement de la base de données. À titre de rappel, la "méthode 2" appliquait des coefficients par département, sans prendre en considération les tendances de consommation liées à l'âge, ce qui pourrait influencer les montants calculés.

Par conséquent, des tarifs ont été calculés en utilisant exclusivement la méthode "Statistique Directe", c'est-à-dire sans recourir à la modélisation, dans le but d'évaluer si l'écart était en partie attribuable à la modélisation ou si le biais éventuel provenait directement de la base de données. Ainsi, après avoir effectué les tests, il a été observé que les tarifs obtenus étaient légèrement inférieurs à ceux affichés dans la figure 11.1, mais que les sauts tarifaires restaient présents. De plus, il est notable que dans les régions de l'Occitanie et de la Provence-Alpes-Côte d'Azur, des tarifs excessivement élevés sont constatés pour certains départements qui présentent une faible exposition. Cette situation sera examinée dans la section réservée aux tarifs par département.

11.1.2 La garantie moyenne de gamme V2

Les résultats de la garantie V2 sont présentés dans la figure ci-dessous :

GARANTIE V2	CLASSE AGE							
REGION	0	20	30	40	50	60	70	80
Île-de-France	24,3	28,3	34,1	44,1	55,5	68,2	88,7	124,8
Centre-Val de Loire	20,1	24,4	28,5	34,7	48,3	58,2	78,8	105,1
Bourgogne-Franche-Comté	19,1	23,7	27,9	33,5	42,8	56,8	69,1	100,3
Normandie	20,4	23,2	25,4	31,8	44,3	55,9	74,3	102,7
Hauts-de-France	19,4	21,1	26,7	35,2	47,8	60,3	84,6	121,4
Grand Est	21,5	24,9	30,7	37,9	50,6	67,7	74,6	112,9
Pays de la Loire	20,1	24,8	28,6	34,6	44,4	56,8	83,1	103,5
Bretagne	20,6	23,4	29,2	32,2	44,1	55,8	74,8	99,0
Nouvelle-Aquitaine	19,2	23,9	31,5	36,7	46,0	58,2	75,7	110,8
Occitanie	23,3	29,2	34,6	41,7	56,3	68,8	93,9	137,8
Auvergne-Rhône-Alpes	20,9	24,2	29,9	36,8	47,3	58,7	76,2	110,1
Provence-Alpes-Côte d'Azur	26,5	35,2	43,3	52,5	68,1	80,2	91,6	155,2

FIGURE 11.2 – Primes pures mensuelles Garantie V2

Les mêmes effets que ceux observés pour la garantie V1 se manifestent également ici, à savoir un effet lié à l'âge et un effet lié à la région. Toutefois, des disparités plus marquées entre les régions sont constatées, atteignant jusqu'à 71 % pour la tranche d'âge 30-39 ans entre la Normandie et la Provence-Alpes-Côte d'Azur. Cette accentuation des disparités peut être attribuée au fait que cette garantie offre des niveaux de remboursement plus élevés, ce qui pourrait entraîner une augmentation des tarifs pour les régions consommant davantage.

Enfin, les sauts tarifaires identifiés pour la garantie V1 sont également présents pour la garantie V2, mais de manière plus accentuée. Ainsi, dans le monde de l'assurance santé, une pratique courante est effectuée lorsque des écarts de tarifs entre les tranches d'âges extrêmes sont trop importants. Il s'agit du transfert intergénérationnel, où les assureurs diminuent délibérément les primes pour les individus âgés et augmentent les primes pour les autres classes d'âge plus jeunes. Cette approche représente un système de solidarité où les jeunes actifs apportent leur soutien aux seniors retraités. Cependant, cette approche n'a pas été mise en œuvre dans le cadre du mémoire, car sa réalisation dépend de la stratégie adoptée par l'assureur.

11.1.3 La comparaison des garanties

La garantie V2 présente des tarifs supérieurs à ceux de la garantie V1 pour toutes les tranches d'âge, ce qui est cohérent étant donné qu'elle offre des remboursements plus avantageux. Les constatations sont similaires pour l'ensemble des régions, c'est pourquoi seules les régions de l'Ile-de-France et de la Nouvelle-Aquitaine sont présentées afin d'illustrer les variations tarifaires.

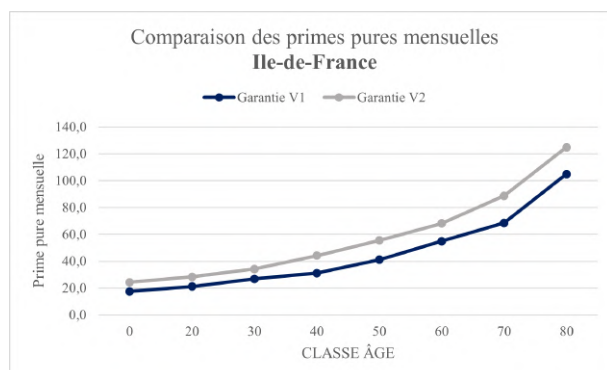


FIGURE 11.3 – Comparaison des primes pures mensuelles (Ile-de-France)

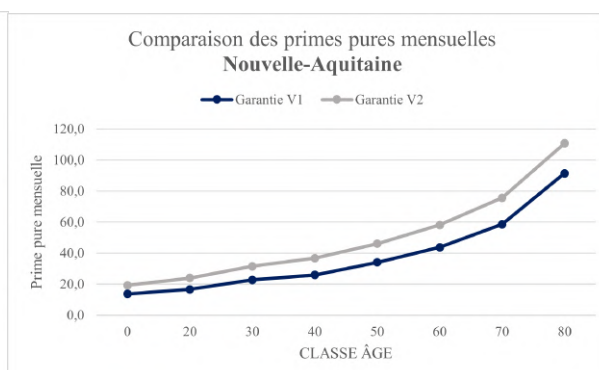


FIGURE 11.4 – Comparaison des primes pures mensuelles (Nouvelle-Aquitaine)

À partir de 60 ans, les écarts entre les tarifs se creusent davantage, ce qui est conforme à la dynamique observée sur le marché.

Il est complexe de tirer des conclusions définitives quant à la qualité des résultats, étant donné l'absence de tarif de référence. Bien qu'une comparaison ait été effectuée avec les résultats présentés dans le mémoire de Sandrine [11], il ne sera pas possible de divulguer les résultats pour des raisons d'anonymisation. De plus, il convient de noter que les grilles de garanties utilisées différaient.

Ainsi, pour mieux appréhender les tendances de consommation et les tarifs en vigueur dans la réalité, l'avis de personnes ayant une expérience approfondie a été sollicité. Dans l'ensemble, les avis s'accordaient sur le fait que les résultats semblaient cohérents pour les tarifs régionaux.

11.2 Les tarifs de l'assureur départemental

Les résultats des tarifs par département sont disponibles en annexe 12.5.8, évitant ainsi de présenter plusieurs pages consécutives de tarifs. Cette section se concentrera davantage sur les limites des résultats. Ainsi, la cartographie des tarifs des départements est présentée ci-dessous pour la classe d'âge 60-69 ans. Le nom des départements est disponible dans l'annexe 12.5.6.

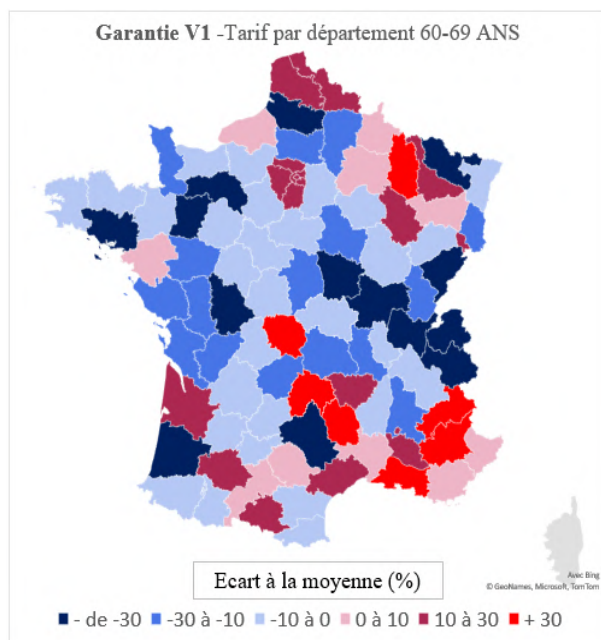


FIGURE 11.5 – Tarif par département Garantie V1

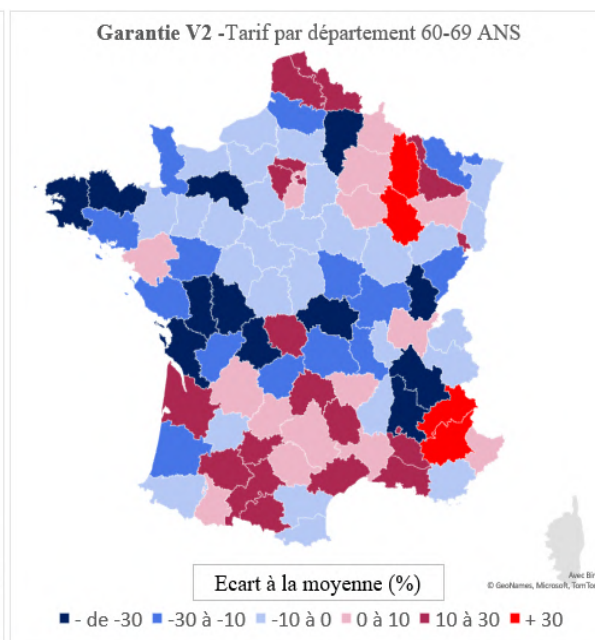


FIGURE 11.6 – Tarif par département Garantie V2

Les tarifs des départements d'Île-de-France ainsi que de la région Provence-Alpes-Côte-d'Azur sont constatés comme étant supérieurs à la moyenne, en accord avec leurs tendances de consommation qui sont également supérieures à la moyenne. Néanmoins, des résultats surprenants sont également observables. En effet, pour les deux niveaux de garanties, des pics de tarifs sont observés dans les départements classés en rouge, ce qui signifie un écart à la moyenne de plus de 30 %, pour les départements suivants : Lozère (48), Creuse (23), Meuse (55), Haute-Marne (52), Hautes-Alpes (5) et Alpes-de-Haute-Provence (4). Tous ces départements sont caractérisés par une exposition très limitée en termes de bénéficiaires, c'est-à-dire une faible population dans la tranche d'âge 60-69 ans.

De plus, il semble que les départements éligibles au régime de l'Alsace-Moselle semblent être sous-tarifés. Ceci peut s'expliquer par le fait que les données dans la base pour ces départements sont celles du régime de la Sécurité Sociale. Cependant, sur ce territoire, il existe un régime local auquel une partie des habitants adhèrent. Néanmoins, l'exposition suppose que l'intégralité de la population adhère au régime de la Sécurité Sociale, créant ainsi une disparité. Il aurait fallu prendre en compte cette différence d'exposition lors de l'étape de tarification.

11.3 La comparaison des tarifs départementaux et régionaux

Pour rappel, l'objectif final du mémoire consiste à permettre à un assureur de prendre une décision quant à l'approche à privilégier, soit en démarchant les élus locaux des départements, soit en s'adressant aux élus de la région.

Ainsi, cette section entreprend une comparaison entre les tarifs départementaux et les tarifs régionaux dans le but de parvenir à la décision la plus adaptée.

Dans le cadre de ce mémoire, le critère permettant de décider s'il faut solliciter les élus de la région consistera à examiner la répartition des départements présentant des tarifs supérieurs et inférieurs à ceux de la région. Par conséquent, si plus de la moitié des départements affichent des tarifs plus élevés que celui de la région, il sera considéré comme opportun d'initier une démarche auprès des élus de la région.

11.3.1 Garantie V1

GARANTIE V1		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Paris	75	19,4	23,4	28,2	33,9	44,9	60,1	75,1	117,5
Seine-et-Marne	77	15,8	19,1	25,5	28,4	37,5	44,8	63,2	98,1
Yvelines	78	17,7	22,0	27,0	32,1	41,9	58,0	70,8	108,1
Essonne	91	16,1	20,1	26,5	30,2	40,9	53,1	64,4	101,0
Hauts-de-Seine	92	16,7	19,9	26,3	30,6	40,0	52,2	66,4	98,4
Seine-Saint-Denis	93	16,5	20,8	25,2	30,7	39,8	51,8	65,9	100,2
Val-de-Marne	94	18,3	22,1	26,9	31,7	41,3	58,8	69,3	109,9
Val-d'Oise	95	18,0	21,7	27,8	31,5	41,3	58,7	71,9	106,3

GARANTIE V1		CLASSE AGE							
REGION		0	20	30	40	50	60	70	80
Île-de-France		17,5	21,2	26,8	31,1	41,1	54,8	68,5	104,9

FIGURE 11.7 – Comparaison des tarifs de l'assureur départemental et régional pour la région Île-de-France

Les tarifs pratiqués dans les départements de la région Île-de-France semblent présenter un certain équilibre, avec la moitié d'entre eux ayant des tarifs supérieurs au tarif régional et l'autre moitié ayant des tarifs inférieurs. Les variations de tarifs dépassant 10 % en valeur absolue sont rares, et les exemples sont affichés ci-dessous :

GARANTIE V1		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Paris	75	11,0%	10,5%	5,4%	9,2%	9,1%	9,7%	9,6%	12,0%
Seine-et-Marne	77	-9,6%	-9,8%	-4,7%	-8,5%	-8,8%	-18,3%	-7,7%	-6,4%
Yvelines	78	1,0%	3,7%	0,6%	3,2%	1,9%	5,8%	3,3%	3,0%
Essonne	91	-8,0%	-5,0%	-1,1%	-2,9%	-0,6%	-3,1%	-6,0%	-3,7%
Hauts-de-Seine	92	-4,2%	-5,9%	-2,0%	-1,6%	-2,8%	-4,7%	-3,1%	-6,2%
Seine-Saint-Denis	93	-5,4%	-1,9%	-5,8%	-1,2%	-3,2%	-5,5%	-3,8%	-4,5%
Val-de-Marne	94	4,9%	4,1%	0,6%	2,0%	0,5%	7,3%	1,2%	4,7%
Val-d'Oise	95	3,1%	2,4%	3,8%	1,3%	0,4%	7,2%	4,9%	1,3%

FIGURE 11.8 – Ecart relatif entre l'assureur départemental et régional pour la région Ile-de-France

Pour la classe d'âge 20, le tarif de l'assureur départemental à Paris est 10,5 % plus élevé que celui de la région. De plus, compte tenu de la faible disparité entre les départements, l'implémentation d'une mutuelle régionale dans cette région semble réalisable.

En ce qui concerne les tarifs établis pour les départements de la région Nouvelle-Aquitaine, ils sont affichés dans la figure ci-dessous :

GARANTIE V1		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Charente	16	11,7	14,1	19,4	22,4	29,1	37,4	51,1	84,2
Charente-Maritime	17	13,0	15,1	20,9	23,6	31,1	37,9	50,9	74,5
Corrèze	19	12,9	15,5	21,3	23,5	30,9	38,5	51,8	87,7
Creuse	23	20,4	26,5	37,6	40,4	49,7	63,0	82,9	126,4
Dordogne	24	14,2	18,1	24,0	26,1	35,4	44,1	60,3	91,6
Gironde	33	18,3	18,8	26,1	32,2	41,8	56,3	79,8	118,7
Landes	40	12,4	14,2	19,5	23,1	30,3	39,8	50,5	76,2
Lot-et-Garonne	47	14,2	16,8	23,5	26,8	34,7	45,2	60,7	94,2
Pyrénées-Atlantiques	64	12,7	15,1	20,4	25,1	33,1	41,9	54,5	85,3
Deux-Sèvres	79	11,2	13,7	18,9	21,6	28,2	37,6	50,0	78,6
Vienne	86	11,5	13,4	18,4	23,5	30,9	41,6	56,4	88,9
Haute-Vienne	87	12,7	15,1	20,7	24,4	33,2	42,6	56,0	90,6

GARANTIE V1		CLASSE AGE							
REGION		0	20	30	40	50	60	70	80
Nouvelle-Aquitaine		13,7	16,6	22,8	25,9	34,0	43,6	58,6	91,3

FIGURE 11.9 – Comparaison des tarifs de l'assureur départemental et régional pour la région Nouvelle-Aquitaine

Tous les départements de la Nouvelle-Aquitaine affichent des tarifs inférieurs à ceux de la région Île-de-France, à l'exception de la Creuse et de la Gironde, où les tarifs semblent être plus élevés. En ce qui concerne la Gironde, des similitudes sont observées avec les départements de la région Île-de-France, notamment en ce qui concerne une offre de soins très étendue et une population relativement jeune. Il n'est donc pas surprenant que des similarités soient constatées au niveau des tarifs avec ceux de la région Île-de-France. En revanche, dans le cas du département de la Creuse, une différence est clairement marquée. Ce département se caractérise par une offre de

11.3. LA COMPARAISON DES TARIFS DÉPARTEMENTAUX ET RÉGIONAUX

soins parmi les plus limitées et une population qui présente un indice de vieillissement dépassant 140. Par conséquent, il est probable que les tarifs établis dans ce département soient excessifs.

Dans la suite du mémoire, malgré la probabilité que le tarif soit excessif par rapport à la réalité, l'accent est mis sur la démarche réflexive. Par exemple, quelle serait la décision prise par l'assureur si le tarif affiché reflétait effectivement la réalité.

Pour compléter l'analyse, le schéma ci-dessous présente les écarts relatifs entre les tarifs de la région Nouvelle-Aquitaine et ceux des départements qui la constituent.

GARANTIE VI		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Charente	16	-15%	-15%	-15%	-13%	-14%	-14%	-13%	-8%
Charente-Maritime	17	-5%	-9%	-8%	-9%	-8%	-13%	-13%	-18%
Corrèze	19	-5%	-6%	-6%	-9%	-9%	-12%	-12%	-4%
Creuse	23	49%	60%	65%	56%	46%	44%	42%	38%
Dordogne	24	4%	9%	5%	1%	4%	1%	3%	0%
Gironde	33	34%	14%	15%	24%	23%	29%	36%	30%
Landes	40	-9%	-14%	-14%	-11%	-11%	-9%	-14%	-17%
Lot-et-Garonne	47	4%	2%	3%	4%	2%	4%	4%	3%
Pyrénées-Atlantiques	64	-7%	-9%	-10%	-3%	-2%	-4%	-7%	-7%
Deux-Sèvres	79	-18%	-17%	-17%	-17%	-17%	-14%	-15%	-14%
Vienne	86	-16%	-19%	-19%	-9%	-9%	-5%	-4%	-3%
Haute-Vienne	87	-7%	-9%	-9%	-6%	-2%	-2%	-4%	-1%

FIGURE 11.10 – Ecart relatif entre l'assureur départemental et régional pour la région Nouvelle-Aquitaine

En Nouvelle-Aquitaine, les variations relatives dépassant 10 % en valeur absolue sont plus fréquentes, ce qui met en évidence une certaine volatilité au niveau des tarifs départementaux. De plus, la majorité des départements affichent des tarifs inférieurs à ceux de la région. Par conséquent, si une mutuelle communale/départementale proposait des tarifs plus avantageux, il semblerait peu probable que les habitants de la plupart des départements soient enclins à opter pour un produit régional, même si ce dernier était mis en place.

11.3.2 Garantie V2

Les tarifs pour la garantie V2 sont également présentés et étant donné leur similitude avec les résultats précédents, ils seront abordés de manière plus succincte dans l'analyse.

Les résultats concernant la région Île-de-France sont présentés ci-dessous, de même que les variations relatives entre les tarifs départementaux et régionaux :

GARANTIE V2		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Paris	75	29,6	32,2	39,3	51,0	64,4	77,7	100,8	138,5
Seine-et-Marne	77	20,3	24,5	31,5	36,4	46,5	62,4	75,6	111,7
Yvelines	78	25,5	30,4	36,3	47,4	59,0	69,3	96,7	141,3
Essonne	91	22,9	27,5	31,3	41,7	54,3	67,1	86,2	118,8
Hauts-de-Seine	92	23,2	26,5	29,9	41,5	52,3	64,2	81,1	114,1
Seine-Saint-Denis	93	21,7	25,7	33,3	43,4	54,6	63,4	80,0	115,2
Val-de-Marne	94	25,0	29,0	34,8	45,3	56,3	70,4	92,3	130,4
Val-d'Oise	95	24,6	29,4	35,5	45,4	57,4	70,9	95,5	128,7

GARANTIE V2		CLASSE AGE							
REGION		0	20	30	40	50	60	70	80
Île-de-France		24,3	28,3	34,1	44,1	55,5	68,2	88,7	124,8

GARANTIE V2		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Paris	75	21,8%	14,0%	15,4%	15,6%	16,0%	14,0%	13,7%	10,9%
Seine-et-Marne	77	-16,2%	-13,2%	-7,7%	-17,4%	-16,2%	-8,4%	-14,7%	-10,5%
Yvelines	78	5,0%	7,7%	6,6%	7,5%	6,4%	1,7%	9,1%	13,2%
Essonne	91	-5,7%	-2,6%	-8,3%	-5,5%	-2,1%	-1,6%	-2,8%	-4,8%
Hauts-de-Seine	92	-4,5%	-6,3%	-12,3%	-6,0%	-5,8%	-5,9%	-8,6%	-8,6%
Seine-Saint-Denis	93	-10,6%	-9,2%	-2,3%	-1,6%	-1,5%	-7,0%	-9,8%	-7,7%
Val-de-Marne	94	2,8%	2,5%	2,0%	2,7%	1,5%	3,2%	4,1%	4,5%
Val-d'Oise	95	1,3%	4,2%	4,2%	2,9%	3,6%	4,0%	7,8%	3,1%

FIGURE 11.11 – Comparaison entre les tarifs départementaux et régionaux - Ile-de-France - Garantie V2

Les écarts relatifs dépassant 10 % en valeur absolue sont plus fréquents par rapport à la garantie V1, ce phénomène pouvant s'expliquer à nouveau par le niveau de garantie plus élevé. Les tendances de consommation pourraient se refléter de manière plus marquée, contribuant ainsi à accroître les disparités entre les tarifs. Néanmoins, ces fluctuations ne dépassent pas des niveaux excessivement élevés, et la répartition entre les départements affichant des tarifs supérieurs et ceux présentant des tarifs inférieurs à la région reste identique à celle observée pour la garantie V1. Sur la base de ce critère de répartition, la mise en place d'une mutuelle régionale reste envisageable.

Il convient de souligner que cette distribution équilibrée selon les départements et les tranches d'âge constitue un cas particulier, et ne s'applique pas à tous les départements.

11.3. LA COMPARAISON DES TARIFS DÉPARTEMENTAUX ET RÉGIONAUX

Du côté de la Nouvelle-Aquitaine, les écarts de tarifs s'accroissent également pour des raisons similaires. Il en est de même pour la répartition des départements qui ne s'équilibrent pas. En effet, la majorité des départements restent en dessous du tarif de la région.

GARANTIE V2		CLASSE AGE							
L DPT	DPT	0	20	30	40	50	60	70	80
Charente	16	18,1	21,6	27,9	32,4	41,2	52,0	66,6	104,6
Charente-Maritime	17	16,8	19,9	25,7	29,2	37,5	42,7	55,6	84,2
Corrèze	19	18,4	23,4	30,4	32,5	41,5	50,0	71,9	103,4
Creuse	23	27,4	39,0	52,9	59,2	64,7	81,0	94,1	141,7
Dordogne	24	21,6	28,1	37,2	42,5	49,4	63,3	83,9	117,0
Gironde	33	24,6	28,3	36,0	41,3	51,7	74,3	104,4	141,8
Landes	40	17,9	21,6	28,9	33,3	43,7	49,6	69,6	101,4
Lot-et-Garonne	47	19,7	25,4	33,8	39,7	50,9	62,1	83,1	113,6
Pyrénées-Atlantiques	64	17,5	20,0	26,4	34,3	44,2	57,8	73,2	101,4
Deux-Sèvres	79	16,0	20,3	26,2	30,0	39,1	53,8	66,8	101,5
Vienne	86	16,0	18,1	23,2	33,4	42,7	55,9	70,2	108,9
Haute-Vienne	87	18,3	21,3	27,4	34,3	45,0	54,0	70,9	109,9

GARANTIE V2		CLASSE AGE							
REGION		0	20	30	40	50	60	70	80
Nouvelle-Aquitaine		19,2	23,9	31,5	36,7	46,0	58,2	75,7	110,8

GARANTIE V2		CLASSE AGE							
L DPT	DPT	0	20	30	40	50	60	70	80
Charente	16	-6%	-10%	-11%	-12%	-10%	-11%	-12%	-6%
Charente-Maritime	17	-13%	-17%	-18%	-20%	-18%	-27%	-27%	-24%
Corrèze	19	-4%	-2%	-3%	-12%	-10%	-14%	-5%	-7%
Creuse	23	43%	63%	68%	61%	41%	39%	24%	28%
Dordogne	24	13%	17%	18%	16%	7%	9%	11%	6%
Gironde	33	28%	19%	14%	12%	12%	28%	38%	28%
Landes	40	-7%	-10%	-8%	-9%	-5%	-15%	-8%	-8%
Lot-et-Garonne	47	3%	6%	7%	8%	11%	7%	10%	3%
Pyrénées-Atlantiques	64	-9%	-16%	-16%	-7%	-4%	-1%	-3%	-8%
Deux-Sèvres	79	-16%	-15%	-17%	-18%	-15%	-8%	-12%	-8%
Vienne	86	-17%	-24%	-26%	-9%	-7%	-4%	-7%	-2%
Haute-Vienne	87	-4%	-11%	-13%	-7%	-2%	-7%	-6%	-1%

FIGURE 11.12 – Comparaison entre les tarifs départementaux et régionaux - Nouvelle-Aquitaine - Garantie V2

Ainsi, même avec un niveau de garantie plus élevé, il ne semble pas opportun d'opter pour le démarchage des élus de la région Nouvelle-Aquitaine.

Intuitivement, l'extension à une échelle plus vaste pourrait faciliter la mutualisation des risques, mais à la suite de l'analyse des régions, il semblerait que l'implémentation d'une mutuelle régionale ne sera pas nécessairement bénéfique pour l'assureur. Cette décision requiert une évaluation au cas par cas, qui sera entrepris dans le chapitre suivant.

CHAPITRE

12

LE MARCHÉ CONCURRENTIEL

Pour rappel, le terme "commune" a été employé de façon interchangeable avec "département" tout au long du mémoire. Ceci s'explique par le fait que, d'une part, les données au niveau communal n'étaient pas disponibles, et d'autre part, le principe sous-jacent entre les mutuelles départementales et communales reste identique : il s'agit d'un partenariat entre une collectivité locale (commune ou département) et un organisme assureur.

Une hypothèse qui pourrait être avancée consiste à affirmer que les conclusions tirées de la concurrence entre une mutuelle départementale et une mutuelle régionale seraient analogues à celles résultant de la concurrence entre une mutuelle communale et une mutuelle à plus grande échelle.

Ainsi, la décision de démarcher les élus de certaines régions a été adoptée. L'introduction d'une mutuelle régionale ne signifierait pas la suppression des mutuelles communales ou départementales déjà existantes. En conséquence, il faudrait tenir compte de l'existence préalable d'une mutuelle communale/départementale dans la région ciblée.

Par conséquent, l'objectif de ce chapitre réside dans la validation des assertions exposées précédemment. Cela sera accompli en instaurant un marché concurrentiel simulé entre un assureur visant à mutualiser son risque au niveau communal/départemental, et un autre visant à le faire à l'échelle régionale. Les parts de marché potentielles susceptibles d'être acquises sont exposées, et les résultats sont calculés pour les deux assureurs. Deux cas d'application sont présentés : celui de la région Île-de-France et celui de la Nouvelle-Aquitaine.

Il est essentiel de souligner que les conclusions issues de cette section sont influencées par les retraitements réalisés, nécessitant ainsi de garder un regard critique quant à la fiabilité des résultats. Par conséquent, une analyse sera exposée concernant l'écart par rapport à la réalité.

12.1 Les adhérents potentiels de l'offre

Les individus susceptibles d'adhérer à la mutuelle régionale ou départementale sont normalement représentés par l'ensemble des résidents de la région, car la seule condition requise pour bénéficier de l'offre est la résidence sur le territoire. Il est supposé que les individus aient facilement accès aux tarifs des deux assureurs grâce à un comparateur en ligne, et qu'il n'y a pas d'autres assureurs.

Cependant, comme exposé dans la première partie du mémoire, la majorité des individus intéressés par les mutuelles sont principalement des personnes âgées de plus de 60 ans. Par conséquent, pour s'aligner avec le concept des mutuelles communales, la décision a été prise de créer une base de données fictive regroupant les résidents potentiellement enclins à s'engager dans ce genre de partenariat. Ainsi, la constitution des bases fictives se basera principalement sur le critère suivant : les individus âgés de 60 ans et plus représenteront au moins 60 % de la population intéressée dans chaque région.

Pour rappel, la base de données NEW_DAMIR offre une vue agrégée des données selon les sous-familles d'actes. Ainsi, il n'est pas encore possible d'attribuer à chaque ligne le montant de **prime pure annuelle globale** calculée dans le chapitre précédent. C'est pour cette raison qu'il est nécessaire d'unifier les montants de remboursements de chaque sous-famille des garanties V1 ou V2, tout en conservant la granularité de l'information au niveau départemental. Ensuite, il convient de sélectionner un échantillon de cette nouvelle base agrégée, car il n'est pas possible de faire tendre l'ensemble de la population française vers une tranche d'âge de 60 ans.

Dans cette perspective, la proportion d'individus âgés de plus de 60 ans a été calculée pour les régions Île-de-France et Nouvelle-Aquitaine. Cette proportion déterminera la taille de l'échantillon qui sera prélevé de la base NEW_DAMIR. La composition de cet échantillon sera ensuite ajustée afin de satisfaire le critère énoncé précédemment, à savoir une structure démographique constituée à 60 % de personnes âgées de plus de 60 ans. De plus une comparaison est faite entre la composition initiale de NEW_DAMIR et les nouvelles bases fictives. Le cas de la région Nouvelle-Aquitaine est présenté ci-dessous et celui de l'Île-de-France est disponible en annexe :

Nouvelle-Aquitaine		Base complète NEW_DAMIR	Base fictive modifiée
Population totale de la région		6 032 438	939 783
Composition démographique	0	22%	5%
	20	10%	9%
	30	11%	6%
	40	13%	8%
	50	14%	12%
	60	14%	27%
	70	9%	18%
	80	8%	16%
V1	Prime mensuelle moyenne	33,74 €	46,20 €
V2	Prime mensuelle moyenne	44,73 €	59,83 €

FIGURE 12.1 – Composition démographique de la base fictive

La proportion des individus âgés de 60 ans et plus au sein de la Nouvelle-Aquitaine s'élevait à 31 %¹ et elle constitue maintenant 60 % de l'exposition totale de la base fictive.

1. A titre indicatif, la part des 60 ans et plus pour l'Île-de-France était de 19 %

Les autres catégories d'âge ont été établies en fonction de divers critères choisis de manière arbitraire :

- Les 0-19 ans ont une proportion plus faible, car il est supposé que les individus de plus de 60 ans ont probabilité réduite d'avoir une descendance importante au-delà de cet âge ;
- Les 20-29 ans sont considérées comme étant des étudiants susceptibles de faire face à des contraintes financières. Par conséquent, ils pourraient davantage s'orienter vers les offres de mutuelles communales ;
- Les 30-49 ans sont considérés majoritairement comme des salariés et ont pour la plupart déjà leur complémentaire santé d'entreprise ;
- Les 50-59 ans sont également identifiées comme des travailleurs salariés, et leur inclination à démissionner est moindre, en grande partie parce qu'ils se rapprochent de l'âge de la retraite. Néanmoins, une proportion plus élevée est assignée dans la base par rapport aux individus âgés de 30 à 49 ans, en raison de l'augmentation des besoins en matière de soins de santé ;
- Un saut est observé entre la tranche d'âge de 50 et 60 ans en raison de l'âge de départ à la retraite à 64 ans.

De plus, il est constaté que la prime mensuelle moyenne connaît une augmentation dans la base fictive modifiée. En effet, elle passe de 33,74 € à 46,20 € pour la garantie V1. Cette hausse s'explique par le fait que la population est plus âgée.

Ainsi, l'hypothèse adoptée dans ce mémoire concernant le choix des individus sur leur offre de couverture santé indique que, pour des garanties équivalentes, la préférence est accordée à l'option la moins coûteuse.

12.2 Les assureurs

Dans la réalité, l'adhérent aurait la possibilité de choisir entre la garantie V1 et la garantie V2. Néanmoins, afin de faciliter la mise en place du marché concurrentiel, il est supposé que les assureurs ne proposent que soit la garantie V1, soit la garantie V2.

Dans le cadre de la commercialisation de ce produit, un taux de frais doit normalement être déterminé par les assureurs pour couvrir à la fois les charges de gestion et d'acquisition. En menant une recherche sur les frais de gestion couramment adoptés par les mutuelles communales, il a été constaté que la plupart des assureurs optaient pour un taux de frais se situant entre 15 % et 25 %. Cependant, dans le cadre de ce mémoire, l'aspect des frais en tant que tel est absent de la base de données. Il sera donc supposé que la marge de frais est neutre par rapport aux frais que l'assureur pourrait rencontrer. Ainsi, seule une marge de bénéfice de 3 % a été choisie et sera appliquée aux primes pures afin de constituer une prime commerciale simplifiée. Les hypothèses retenues ne sont pas absolues et peuvent être discutées.

12.3 Application à la région Nouvelle-Aquitaine

Ainsi, il a été conclu dans le chapitre précédent que la Nouvelle-Aquitaine ne semble pas être l'environnement optimal pour la mise en place d'une mutuelle régionale. Les résultats sont présentés dans la figure ci-dessous afin de vérifier l'affirmation :

Garantie V1 Nouvelle-Aquitaine	3 % de marge	
	Résultat	Part de marché
Assureur départemental	12 M€	60%
Assureur régional	-35 M€	40%
Garantie V2 Nouvelle-Aquitaine	3 % de marge	
	Résultat	Part de marché
Assureur départemental	14 M€	60%
Assureur régional	-45 M€	40%

FIGURE 12.2 – Résultats du marché concurrentiel pour la Nouvelle-Aquitaine

Malgré le caractère agrégé des données de la base, il a été possible de calculer la part de marché, qui reflète le nombre de personnes ayant souscrit à l'offre départementale ou régionale. Ainsi, il est observé que 60 % de la population de la région Nouvelle-Aquitaine est couverte par l'assureur départemental, tandis que l'offre de l'assureur régional a été choisie par le reste de la population pour les deux niveaux de garanties.

Ainsi, il est constaté que des résultats négatifs sont présentés par l'assureur régional pour les deux niveaux de garanties, en raison de la manifestation d'un problème d'anti-sélection. En effet, l'adhésion à l'offre régionale n'a pas été souscrite par les habitants des départements où les tarifs étaient inférieurs à ceux de la région. Simultanément, une sous-tarification de l'offre a été constatée pour les autres départements, ce qui a entraîné l'attraction des résidents de ces zones. Le taux de marge de 3 % est largement insuffisant pour contrebalancer ce problème d'anti-sélection. Cela implique alors un résultat négatif de -35 M€ pour la garantie V1 et -45 M€ pour la garantie V2.

En conclusion et dans le cadre de ce mémoire, il apparaît que pour la région Nouvelle-Aquitaine, une approche plus pertinente pour l'assureur consisterait à rester au niveau de la commune/département. En effet, les habitudes de consommation au sein de la région semblent révéler des disparités significatives.

12.4 Application à la région Ile-de-France

Pour la région Île-de-France, une opportunité semblait se dessiner en faveur de la mise en place d'une mutuelle régionale :

Garantie V1 Ile-de-France	3 % de marge	
	Résultat	Part de marché
Assureur départemental	8 M€	47%
Assureur régional	-8 M€	53%
Garantie V2 Ile-de-France	3 % de marge	
	Résultat	Part de marché
Assureur départemental	11 M€	47%
Assureur régional	-15 M€	53%

FIGURE 12.3 – Résultats du marché concurrentiel pour Ile-de-France

Dans le chapitre précédent, il a été évoqué que la répartition des départements affichant des tarifs égaux ou supérieurs à ceux du département était équilibrée dans la région Île-de-France. Néanmoins, les parts de marché penchent en faveur de l'assureur régional. En effet, cette situation trouve son explication dans le fait que, dans la région Île-de-France, l'ensemble du département de Paris, représentant la population la plus importante, a opté pour l'offre régionale. Cependant, cette part de marché majoritaire ne s'accompagne pas de performances supérieures en ce qui concerne le résultat de l'assureur régional. Celui-ci enregistre un déficit de -8 M€ et -15 M€ respectivement pour les garanties V1 et V2. Tout comme dans le cas de la région Nouvelle-Aquitaine, la marge bénéficiaire de 3 % n'est pas suffisante pour compenser l'effet d'antisélection.

Ainsi, il semble que la viabilité de la mutuelle régionale soit compromise étant donné qu'un problème d'antisélection se manifeste de manière systématique. Cependant, après avoir consulté des professionnels travaillant au sein des organismes ayant établi un partenariat avec la région Île-de-France, ces derniers ont affirmé que l'activité était rentable dans la zone de mutualisation de cette région. La différence peut s'expliquer par plusieurs raisons, dans un premier temps, le critère établi dans le cadre de ce marché concurrentiel repose sur l'hypothèse d'un coefficient d'antisélection de 100 %, ce qui signifie que tous les adhérents connaissent parfaitement le produit le moins cher et adoptent un comportement rationnel en fonction du prix. Cependant, en pratique, il est envisageable que certaines personnes ne consacrent pas nécessairement du temps à la comparaison des tarifs, tandis que d'autres accordent davantage de confiance à une entité régionale qu'à une entité départementale, ou bien que certaines personnes ne soient informées que de l'existence de la mutuelle régionale et non de la mutuelle communale/départementale.

Afin de s'aligner davantage avec la réalité, il serait nécessaire de réduire le coefficient d'antisélection, de sorte qu'il ne soit pas fixé à 100 %. Cette modification aurait le potentiel d'améliorer les performances de l'assureur régional, du moins dans le cadre théorique de cette étude. En pratique, cela impliquerait de prendre en compte, pour chaque ligne de la base fictive, un groupe d'individus susceptibles de ne pas opter pour le prix le plus bas.

12.5 Réflexion sur les résultats obtenus

L'objectif de cette section est de mettre en évidence des pratiques réelles qui n'ont pas été incluses dans la rédaction de ce mémoire.

Les analyses effectuées dans les sections précédentes suggèrent de maintenir une position sur le marché des mutuelles communales/départementales en raison du phénomène d'antisélection observé pour la mutuelle régionale. De plus, d'autres facteurs n'ont pas pu être pris en compte dans ce modèle concurrentiel simplifié, notamment le mécanisme de transfert intergénérationnel mentionné dans le chapitre précédent. En effet, en fonction de la stratégie adoptée par l'assureur, il pourrait choisir de réduire les tarifs pour les seniors tout en augmentant ceux des autres tranches d'âge plus jeunes. Cependant, cette tarification pourrait potentiellement aggraver le problème d'antisélection.

Par ailleurs, dans la pratique, l'établissement de tarifs distincts pour chaque classe d'âge et chaque département pourrait se révéler complexe et coûteux en termes de suivi. Une alternative envisageable serait d'explorer la possibilité de regrouper certains départements géographiquement proches et de proposer des tarifs communs pour ces groupes à l'image des zoniers. Toutefois, cette approche pourrait soulever des questions et des préoccupations de la part des départements non inclus dans ces regroupements. Il apparaît ainsi qu'une approche qui pourrait être pertinente serait d'adopter une stratégie de démarchage axée sur la mutualisation par zones géographiques en incluant des départements ou communes de différentes régions. Le concept de zonier en assurance santé n'est pas nécessairement lié à la géographie. Par exemple, les régions Île-de-France et Provence-Alpes-Côte d'Azur présentent fréquemment des tarifs similaires malgré leur éloignement géographique.

Bien que la possibilité de créer un partenariat avec une région puisse être envisagée, il est essentiel de se pencher sur la capacité à assumer des risques à l'échelle régionale. Cela découle de l'exigence d'avoir des fonds propres nettement supérieurs à ceux nécessaires pour une commune ou un département. Il est donc envisageable d'envisager un partenariat entre plusieurs assureurs afin de couvrir une région. Cette situation est observée dans le cas de la mutuelle régionale de l'Île-de-France, qui a établi un partenariat avec trois assureurs, aboutissant à l'élimination de toute concurrence des autres assureurs désireux de pénétrer le marché des mutuelles régionales².

De plus, il n'y a pas de concurrence entre les assureurs ayant conclu un contrat avec la région, étant donné que les adhérents sont répartis en fonction des zones. Chaque assureur se voit allouer une zone de couverture et si, par exemple, un assureur se trouve dans l'incapacité d'accepter de nouveaux adhérents en raison de contraintes telles qu'un manque de fonds propres, les adhérents seraient redirigés vers l'un des autres assureurs partenaires.

2. Toutefois, il est à noter que d'autres acteurs privés demeurent présents.

Synthèse de la quatrième partie

Cette partie a dans un premier temps présenté les tarifs de l'assureur régional et ceux de l'assureur départemental pour les garanties V1 et V2. Deux effets majeurs sont observés. Tout d'abord, une augmentation significative du tarif en fonction de l'âge est constatée. De plus, un effet lié à la région est relevé, bien que son impact ne soit pas considérable. En d'autres termes, les tarifs varient en fonction des régions, mais cette variation n'est pas excessive. En examinant ensuite le niveau de garantie supérieur, il semble que les écarts de tarifs entre les régions tendent à s'accroître, bien qu'ils ne soient pas non plus excessifs. Dans l'ensemble, les tarifs des différentes régions ne présentent pas d'incohérences notables et ont été validés en sollicitant l'avis d'experts possédant de l'expérience dans le domaine de la tarification en santé.

En ce qui concerne l'analyse des résultats pour les départements, il a été observé que des tarifs excessivement élevés étaient présents dans certains départements, et il a été remarqué que cela concernait systématiquement les départements les moins exposés en termes de population. D'autre part, une sous-tarification semble également être enregistrée dans les départements éligibles au régime de l'Alsace-Moselle. Cette situation peut s'expliquer par le fait qu'au sein de ces départements, la majorité de la population adhère au régime local plutôt qu'au régime de la Sécurité Sociale. Cependant, dans le cadre de ce mémoire, il a été supposé que l'ensemble des habitants de ces territoires adhéraient au régime de la Sécurité Sociale, ce qui constitue un biais dans l'établissement des tarifs.

Ainsi, pour les départements, l'accent n'a pas été mis sur la comparaison entre tous les départements les uns par rapport aux autres, mais plutôt sur la comparaison d'une région avec les départements qui la constituent. En effet, les tarifs des départements au sein de la région Île-de-France, constituée principalement de zones urbaines, ont affiché une similarité pour les deux niveaux de garantie. En revanche, pour la région Nouvelle-Aquitaine, englobant à la fois des zones urbaines et rurales, les tarifs présentaient des différences plus marquées, s'accroissant avec l'augmentation du niveau de garantie.

Cette analyse des tarifs a permis de déterminer, pour les régions examinées, s'il était préférable de solliciter davantage les élus des départements ou plutôt ceux des régions en vue d'établir un partenariat dans le contexte des mutuelles communales/départementales ou régionales.

Ainsi, la simulation d'un marché concurrentiel fictif pour les deux régions étudiées a conduit à des conclusions concernant la pertinence d'une mutuelle régionale, qui s'avère ne pas être la meilleure option dans le cadre de ce mémoire.

Pour conclure, une réflexion a été proposée concernant les pratiques qui n'ont pas été abordées dans ce mémoire.

CONCLUSION

Les *Open Data* représentent une ressource précieuse dans le monde de l'assurance santé. Elles peuvent fournir une opportunité aux organisations envisageant de lancer de nouveaux projets tout en n'ayant pas accès à une base de données privée exhaustive. Dans ce cadre, ce mémoire avait pour ambition d'examiner le marché de niche que constituent les mutuelles communales en se basant uniquement sur les données ouvertes. Les objectifs de cette étude étaient donc de déterminer les tarifs pour une offre de mutuelle communale et d'évaluer si le passage à une mutuelle régionale serait avantageux pour un assureur opérant dans ce domaine spécifique.

Pour ce faire, les travaux ont été initiés en utilisant une base de données interne au cabinet Actuelia, extraite de l'*Open DAMIR*, constituée dans le cadre d'un mémoire antérieur. Cependant, cette base présentait des lignes agrégées et présentait l'information uniquement à la maille régionale. Il fallait donc répartir ces informations en une granularité plus fine, correspondant à celle des départements. Il a été décidé de se concentrer sur cette échelle du fait de l'indisponibilité des données à l'échelle communale. Par conséquent, dans le cadre de ce mémoire, la commune a été confondue avec le département. De plus, cette base de données était une version de sauvegarde intermédiaire, ce qui a requis des vérifications de la qualité des données, conduisant à de nouveaux traitements pour les montants de remboursements de la Sécurité Sociale qui n'étaient pas complets.

Ainsi, dans un premier temps, en utilisant d'autres ensembles de données ouvertes, ce mémoire a abouti à la création d'une nouvelle base de données extraite de l'*Open DAMIR*, fournissant les informations au niveau départemental. Elle est appelée *NEW_DAMIR*. Par la suite, la base de données a été complétée par l'ajout de variables externes telles que le taux de chômage, le nombre de salariés dans les établissements de santé par département. L'objectif initial de l'introduction de ces variables était de repérer les tendances de consommation qui pourraient être liées à leur inclusion. Par exemple, il a été étudié si une disponibilité élevée en termes d'établissements de santé était corrélée avec une augmentation de la fréquence des consultations médicales. Néanmoins, les études n'ont pas permis de dégager une relation précise quant à l'effet de ces variables

externes. Elles n'ont donc pas été incluses dans l'élaboration des tarifs.

La mise en place de deux tarifs distincts pour deux niveaux de garanties a ouvert la voie à l'analyse des disparités de consommation observées entre les différentes régions. En examinant les tarifs de la garantie entrée de gamme entre les régions, il est ressorti que les tarifs augmentent de manière significative avec l'âge, avec des tarifs plus élevés constatés dans les régions Île-de-France, Occitanie et Provence-Alpes-Côte d'Azur. Toutefois, malgré une certaine variabilité régionale, celle-ci demeure modérée. Elle peut être en partie limitée par les plafonds de remboursement établis au niveau des garanties étudiées. Par ailleurs, les tarifs de la garantie moyenne gamme ont présenté des écarts plus significatifs entre les régions, suggérant que des plafonds de garantie plus élevés ont permis de mieux mettre en évidence les tendances de consommation. Pour l'analyse des tarifs départementaux, l'approche s'est concentrée sur la comparaison d'une région avec les départements qui la composent. Par exemple, les tarifs des départements au sein de la région Île-de-France, principalement constituée de zones urbaines, ont affiché une similitude pour les deux niveaux de garantie. En revanche, dans la région Nouvelle-Aquitaine, qui englobe à la fois des zones urbaines et rurales, les tarifs présentaient des écarts plus marqués, lesquels s'accroissaient avec l'augmentation du niveau de garantie.

Dans la réalité, l'introduction d'une mutuelle régionale ne signifierait pas la suppression des mutuelles communales ou départementales déjà existantes. C'est pourquoi ces analyses ont permis de définir un critère permettant de déterminer si la mise en place d'une mutuelle régionale serait systématiquement avantageuse pour toutes les régions. Ce critère ainsi établi a été mis à l'épreuve grâce à la simulation d'un marché concurrentiel fictif qui oppose un assureur cherchant à mutualiser ses risques au niveau départemental à un autre optant pour la mutualisation régionale. Cette démarche a conduit à des conclusions quant à la pertinence d'une mutuelle régionale. Cependant, l'apparition d'un phénomène d'antisélection souligne que cette approche n'est pas toujours la meilleure alternative.

Ces travaux ont démontré qu'il était possible de ventiler les informations extraites de la base *Open DAMIR* à une granularité plus fine. Malgré quelques limitations, les résultats ont affiché une certaine cohérence avec les tendances observées dans le domaine de l'assurance santé.

Pour approfondir davantage les recherches de ce mémoire, il pourrait s'avérer utile de partir de la version la plus récente de la base de données *Open DAMIR*. Cette mise à jour plus récente pourrait permettre une meilleure saisie des effets du 100% Santé, qui ont été pris en compte dans ce mémoire uniquement pour les coûts, mais pas pour la fréquence.

Par ailleurs, il était à noter que la base de données ne comportait pas d'informations sur le nombre réel de bénéficiaires. Par conséquent, il a fallu faire l'hypothèse que tous les résidents de la France étaient affiliés au régime de la Sécurité Sociale, ce qui a engendré une surexposition de certains départements, principalement ceux relevant du régime local de l'Alsace-Moselle. Dans ces départements, la plupart des résidents ne sont pas affiliés au régime de Sécurité Sociale traditionnel, mais plutôt au régime local. Il serait donc judicieux de prendre en compte cette information sur l'exposition pour de futures études approfondies.

En ce qui concerne les tarifs, il convient de noter qu'il a été observé des écarts parfois significatifs entre les tranches d'âge senior. Il pourrait être pertinent de considérer la mise en œuvre de techniques de lissage, que ce soit au niveau départemental, en fonction de l'âge, voire les deux simultanément.

Finalement, les variables externes n'ont pas été intégrées dans les modèles de tarification en raison de leur faible significativité en comparaison des autres variables explicatives, à savoir la tranche d'âge et la variable de localisation (département ou région). Une alternative pourrait consister à exclure la variable de localisation des modèles et à envisager la tarification en se basant exclusivement sur les tranches d'âge et les variables externes. De plus, la création d'un zonier pourrait également être envisagée, basée sur les primes pures obtenues au sein de ce mémoire.

Les mutuelles communales, bien qu'ayant gagné en popularité ces dernières années, continuent de représenter un marché de niche. Cette forme de partenariat entre les collectivités locales et les assureurs demeure largement méconnue du grand public. De plus, bien que les conclusions de ce mémoire revêtent un caractère pessimiste à l'égard de l'avenir des mutuelles régionales, il est important de reconnaître que les comportements des résidents dans la réalité peuvent différer considérablement du contexte examiné dans ce mémoire. Par conséquent, il demeure envisageable de considérer les mutuelles régionales comme une opportunité commerciale, tout en améliorant l'accès aux soins pour les populations vulnérables. La région Île-de-France illustre cette dynamique avec la création de la première mutuelle régionale en France, qui semble susciter un intérêt croissant parmi les jeunes actifs d'après les témoignages des professionnels œuvrant dans ces organismes. Cette évolution laisse entrevoir la perspective d'une diffusion plus étendue de ces offres, potentiellement encourageante pour d'autres régions à suivre cette voie.

ANNEXES

Annexe A : Grilles de garanties

Grands postes	Sous-familles d'actes	%SS	%V1	%V2
Soins courants	Consultations/visites	70%	100%	130%
	Actes médicaux (techniques)	70%	100%	130%
	Auxiliaires médicaux (kiné, infirmiers, ...)	60%	100%	130%
	Analyse médicale et examens laboratoire	60%	100%	130%
	Actes d'imagerie, de radiologie et ostéodensitométrie	70%	100%	130%
	Transport	65%	100%	100%
	Petit appareillage	60%	100%	100% + 70€/an
	Pharmacie	15/30/65%	100%	100%
Hospitalisation	Frais de séjour	80%	100%	100%
	Honoraires et actes chirurgicaux	80%	100%	100%
	Forfait journalier	-	Frais réels	Frais réels
	Chambre particulière	-	-	40€/an
	Lit accompagnant	-	-	150€/an
Dentaire	Prothèse	70%	100%	170%
	Soins dentaires	70%	100%	100%
	Parodontologie	70%	100%	100%
	Implantologie	70%	100%	170%
	Orthodontie	70%	100%	150%
Optique	Monture	60%	50 €	70 €
	Verres	Variables selon les types de verres		
	Chirurgie œil	-	-	100€/an
	Lentilles	60%	100%	100%+100€/an
Aides auditives	Audioprothèse (par oreille)	60%	100%	100% + 200€/prothèse
	Pile, accessoire	60%	100%	100%
Autres	Cure thermique	65%	100%	100%
	Ostéopathie, diététique, chiropractie, ...	-	-	75€/an
	Prévention	65%	100%	100% + 30€
	Maternité	-	-	40€/jour

FIGURE 12.4 – Grilles de garanties pour tarification

Annexe B : Les variables de l'Open DAMIR

Nom variable	Libellé
BENEFICIAIRE	
AGE_BEN_SNDS	Tranche d'Age Bénéficiaire au moment des soins
BEN_CMU_TOP	Top Bénéficiaire CMU-C
BEN_QLT_COD	Qualité du Bénéficiaire
BEN_RES_REG	Région de Résidence du Bénéficiaire à partir de 2015 (tables A)
BEN_RES_ZEAT	ZEAT de Résidence du Bénéficiaire jusqu'à 2014 (tables P)
BEN_SEX_COD	Sexe du Bénéficiaire
MTM_NAT	Modulation du Ticket Modérateur
EXECUTANT	
DDP_SPE_COD	Discipline de Prestation Etb Exécutant
ETE_CAT_SNDS	Catégorie Etb Exécutant
ETE_REG_COD	Région d'Implantation Etb Exécutant à partir de 2015 (tables A)
ETE_TYP_SNDS	Type Etb Exécutant
ETE_ZEAT_COD	ZEAT d'Implantation Etb Exécutant jusqu'à 2014 (tables P)
EXE_INS_REG	Région du PS Exécutant à partir de 2015 (tables A)
EXE_INS_ZEAT	ZEAT du PS Exécutant jusqu'à 2014 (tables P)
MDT_TYP_COD	Mode de Traitement Etb Exécutant
MFT_COD	Mode de Fixation des Tarifs Etb Exécutant
PSE_ACT_CAT	Catégorie de l' Exécutant
PSE_ACT_SNDS	Nature d'Activité PS Exécutant
PSE_SPE_SNDS	Spécialité Médicale PS Exécutant
PSE_STJ_SNDS	Statut Juridique PS Exécutant

FIGURE 12.5 – Les variables de l'Open DAMIR partie 1

INDICATEURS BRUTS	
PRS_ACT_COG	Coefficient Global
PRS_ACT_NBR	Dénombrement
PRS_ACT_QTE	Quantité
PRS_DEP_MNT	Montant du Dépassement
PRS_PAI_MNT	Montant de la Dépense
PRS_REM_BSE	Base de Remboursement
PRS_REM_MNT	Montant Versé/Remboursé
INDICATEURS PREFILTRES	
FLT_ACT_COG	Coefficient Global de la Prestation Préfiltré
FLT_ACT_NBR	Dénombrement de la Prestation Préfiltré
FLT_ACT_QTE	Quantité de la Prestation Préfiltrée
FLT_DEP_MNT	Montant du Dépassement de la Prestation Préfiltré
FLT_PAI_MNT	Montant de la Dépense de la Prestation Préfiltrée
FLT_REM_MNT	Montant Versé/Remboursé Préfiltré
ORGANISME	
ORG_CLE_REG	Région de l'Organisme de Liquidation à partir de 2015 (tables A)
ORG_CLE_ZEAT	ZEAT de l'Organisme de Liquidation jusqu'à 2014 (tables P)
PERIODE	
SOI_ANN	Année de Soins
SOI_MOI	Mois de Soins
PERIODE DE TRAITEMENT	
FLX_ANN_MOI	Année et Mois de Traitement

FIGURE 12.6 – Les variables de l'Open DAMIR partie 2

PRESCRIPTEUR	
ETP_CAT_SNDS	Catégorie Etb Prescripteur
ETP_REG_COD	Région d'Implantation Etb Prescripteur à partir de 2015 (tables A)
ETP_ZEAT_COD	ZEAT d'Implantation Etb Prescripteur jusqu'à 2014 (tables P)
PRE_INS_REG	Région du PS Prescripteur à partir de 2015 (tables A)
PRE_INS_ZEAT	ZEAT du PS Prescripteur jusqu'à 2014 (tables P)
PSP_ACT_CAT	Catégorie du Prescripteur
PSP_ACT_SNDS	Nature d'Activité PS Prescripteur
PSP_SPE_SNDS	Spécialité Médicale PS Exécutant
PSP_STJ_SNDS	Statut Juridique PS Prescripteur
PRESTATION	
ASU_NAT	Nature d'Assurance
ATT_NAT	Nature de l'Accident du Travail
CPL_COD	Complément d'Acte
CPT_ENV_TYP	Type d'Enveloppe
DRG_AFF_NAT	Nature du Destinataire de Règlement affiné
ETE_IND_TAA	Indicateur TAA Privé/Public
EXO_MTF	Motif d'Exonération du Ticket Modérateur
PRS_FJH_TYP	Type de Prise en Charge Forfait Journalier
PRS_NAT	Nature de Prestation
PRS_PDS_QCP	Code Qualificatif Parcours de Soins (sortie)
PRS_PPU_SEC	Code Secteur Privé/Public
PRS_REM_TAU	Taux de Remboursement
PRS_REM_TYP	Type de Remboursement

FIGURE 12.7 – Les variables de l'Open DAMIR partie 3

Annexe C : Backtesting de la classification des codes actes

PRS_NAT	Libellé Nature de Prestation	Sous-famille d'acte	CHECK
1096	TELECONSULTATION MEDECIN TRAITANT AVEC EHPAD	Consultations/visites	OK
1164	TÉLÉ CONSULTATION - ALD ET / OU EHPAD	Consultations/visites	OK
1312	ACTES DE SPECIALITE EN K	Actes médicaux (techniques)	OK
1975	FRANCHISE HORS TIERS PAYANT SUR MEDICAMENT	Autres (aides, inclassable, ...)	OK
2112	FRAIS DE SEJOUR SUPPLEMENTAIRE AU GHS	Frais de séjour	OK
1171	RÉMUNÉRATION MÉDECIN TRAITANT CENTRES DE SANTÉ	Autres (aides, inclassable, ...)	OK
1408	COURONNE ZIRCONNE RAC MODERE	Prothèse	OK
1432	ACTES EN DC	Soins dentaires	OK
1612	CONTRAT DE BONNES PRATIQUES ZONE URBAINE	Autres (aides, inclassable, ...)	OK
1721	FORFAIT PROFESSIONNEL (F01) FILIERES ET RESEAUX	Autres (aides, inclassable, ...)	OK
3112	ACTES INFIRMIERS DE SOINS (AMI3-AMI13,AMI16)	Auxiliaires médicaux (kiné, infirmiers, ...)	OK
1914	FORFAIT PEDIATRIQUE	Consultations/visites	OK
4145	FORFAIT THERMAL 72 SEANCES AVEC KINE	Cure thermique	OK
4214	TAXIS	Transport	OK
3546	ACCESSOIRES DE PROTHESES ET D ORTHOPEDIE (CENTRES	Petit appareillage	OK
5117	FORFAIT OPTIQUE -ENFANT-N° 9 UNIFOCAUX (CMU)	Verres	OK
6131	IJ NORMALES POUR CURE THERMALE	Cure thermique	OK
6135	IJ MALADIE PAMC PLUS DE 3 MOIS	Autres (aides, inclassable, ...)	OK
8120	MAJORATION FAUTE INEXCUSABLE RENTE ORPHELIN	Autres (aides, inclassable, ...)	OK
9141	VISITE D URGENCE	Consultations/visites	OK
9211	VACCIN (MILITAIRES) / VACCIN GRIPPE (CRPCEN)	Autres (aides, inclassable, ...)	NOK
9311	DEPISTAGE DU CANCER COLO-RECTAL	Prévention	OK
4208	FORFAIT D'URGENCE SUR APPEL DU SAMU EXPERIMENTATI	Autres (aides, inclassable, ...)	OK
3542	PROTHESES EXTERNES NON ORTHOPEDIQUES (CHAP. 4)	Petit appareillage	OK
3522	DIVERS ORTHESES	Petit appareillage	OK

FIGURE 12.8 – Backtesting de la classification des codes actes partie 1

PRS_NAT	Libellé Nature de Prestation	Sous-famille d'acte	CHECK
2173	DIFFERENTIEL PSY REGLEMENTAIRE	Consultations/visites	OK
2127	DIALYSE PERITONEALE AUTOMATISEE (DPA)	Autres - hospitalisation	OK
1960	SUPPLEMENT DEROGATOIRE SG SUR ACTE PROFESSIONNEL	Autres (aides, inclassable, ...)	OK
1904	MAJORATION ENFANT PEDIATRE	Consultations/visites	OK
1767	FORFAIT FORMATION (FF6) FILIERES ET RESEAUX	Autres (aides, inclassable, ...)	OK
1704	CDS DENTAIRE OPTANT FORF INITIAL	Autres (aides, inclassable, ...)	OK
1612	CONTRAT DE BONNES PRATIQUES ZONE URBAINE	Autres (aides, inclassable, ...)	OK
1424	TRAITEMENTS D ORTHODONTIE PRATIQUES PAR LE CHIRU	Orthodontie	OK
1407	COURONNE TRANSITOIRE RAC MODERE	Prothèse	OK
1194	TELE EXPERTISE DE NIVEAU 2	Consultations/visites	OK
1910	Plan personnalisé de santé	Autres (aides, inclassable, ...)	OK
2181	PRELEVEMENT D ORGANE 1	Hospitalisation	OK
4366	AIDE DEMARRAGE TELETRANSMISSION	Autres (aides, inclassable, ...)	OK
2252	FORFAIT JOURNALIER DE SORTIE	Forfait journalier	OK
2351	FORFAIT TECHNIQUE NORMAL IRMN -SCANNERS	Actes d'imagerie, de radiologie et ostéodensitométrie	OK
1466	TRAITEMENT ORTHOPEDIE DENTO FACIALE CCAM	Autres - dentaire	OK
4311	FRAIS DE DEPLACEMENT - ENQU'TE LEGALE AT	Autres (aides, inclassable, ...)	OK
4353	INDEMNITE POUR PERTE DE SALAIRE (MALADIE, AT)	Autres (aides, inclassable, ...)	OK
1176	CONTRAT INDIVIDUEL EMBAUCHE - AVANCE	Autres (aides, inclassable, ...)	OK
6014	COMPLEMENT IJ >PLAFOND (CRPCEN)	Autres (aides, inclassable, ...)	OK
5105	VERRES UNIFOCAUX OP5(CMU)	Verres	OK
5110	FORFAIT OPTIQUE -ENFANT-N° 2 UNIFOCAUX (CMU)	Verres	OK
6232	IJ CONGE ADOPTION PAMC	Autres (aides, inclassable, ...)	OK
1431	ACTES EN D (ET OCC POUR LA CRPCEN)	Soins dentaires	OK
8116	RENTES D ASCENDANT	Autres (aides, inclassable, ...)	OK

FIGURE 12.9 – Backtesting de la classification des codes actes partie 2

Annexe D : Retraitement des classes d'âges

La méthode de retraitement des classes d'âges est explicité pour la région Ile-de-France. Il s'agit de la même méthode pour toutes les autres régions.

2018	Ile-de-France	Ensemble					
		1	2	3	4	5	6
Dpt	Nom	0 à 19 ans	20 à 39 ans	40 à 59 ans	60 à 74 ans	75 ans et plus	Total
75	Paris	422 752	736 749	550 917	309 022	168 086	2 187 526
77	Seine-et-Marne	392 026	358 391	381 646	186 117	85 817	1 403 997
78	Yvelines	388 195	348 287	393 546	200 398	107 840	1 438 266
91	Essonne	356 565	336 198	347 644	167 806	87 917	1 296 130
92	Hauts-de-Seine	402 708	461 396	426 784	200 869	117 549	1 609 306
93	Seine-Saint-Denis	473 039	471 049	413 251	182 706	83 066	1 623 111
94	Val-de-Marne	359 413	388 307	366 355	176 682	97 169	1 387 926
95	Val-d'Oise	351 331	328 428	322 682	154 421	71 756	1 228 618
	Total	3 146 029	3 428 805	3 202 825	1 578 021	819 200	12 174 880

FIGURE 12.10 – Répartition de la population par tranche d'âge INSEE

L'outil en ligne [13] permet de sélectionner des tranches d'âges par région afin d'obtenir le pourcentage de la classe d'âge sélectionné.

Tranches d'âge	Millions	Pourcentage
30+	7,3	60%
20-29	1,7	14%
0-19	3,1	26%
Total	12,2	100%

FIGURE 12.11 – Résultat outil en ligne pour une tranche d'âge

L'étape est répétée pour toutes les tranches d'âges et toutes les régions afin d'obtenir la ligne colorée en vert. Une règle de trois permet de déduire le nombre d'habitants de la classe d'âge correspondante.

Ile-de-France		0 à 19 ans	20 à 29 ans	30 à 39 ans	40 à 49 ans	50 à 59 ans	60 à 69 ans	70 à 74 ans	75 à 79 ans	80 ans et plus	Total
Comparaison outil		26%	14%	14%	14%	13%	10%	3%	2%	4%	100%
75	Paris	422 752	368 375	368 375	296 648	275 459	237 709	71 313	48 025	96 049	2 184 703
77	Seine-et-Marne	392 026	179 196	179 196	205 502	190 823	143 167	42 950	24 519	49 038	1 406 416
78	Yvelines	388 195	174 144	174 144	211 909	196 773	154 152	46 246	30 811	61 623	1 437 997
91	Essonne	356 565	168 099	168 099	187 193	173 822	129 082	38 724	25 119	50 238	1 296 941
92	Hauts-de-Seine	402 708	230 698	230 698	229 807	213 392	154 515	46 354	33 585	67 171	1 608 928
93	Seine-Saint-Denis	473 039	235 525	235 525	222 520	206 626	140 543	42 163	23 733	47 466	1 627 139
94	Val-de-Marne	359 413	194 154	194 154	197 268	183 178	135 909	40 773	27 763	55 525	1 388 135
95	Val-d'Oise	351 331	164 214	164 214	173 752	161 341	118 785	35 636	20 502	41 003	1 230 778

FIGURE 12.12 – Résultat après retraitement des classes d'âges

L'outil affiche un pourcentage arrondi, ce qui a pour conséquence une légère augmentation 0,051%

Annexe E : Modèle linéaire généralisé (GLM)

Les GLM (Generalized Linear Model) introduits en 1972 par Nelder et Wedderburn, permettent de résoudre les limitations liées à l'application des modèles linéaires. En effet, ces derniers fonctionnent uniquement dans certaines limitations. Il est nécessaire que la distribution de la variable à expliquer soit gaussienne ou que l'homoscédasticité (les variances des erreurs stochastiques d'une régression linéaire doivent être égales) soit vérifiée.

12.5.1 Généralités sur les GLM

Le cadre d'utilisation des GLM

Globalement, les modèles linéaires généralisés sont utilisés lorsque les résidus sont suspectés de ne pas suivre une loi normale. Ce qui arrive assez fréquemment.

Par exemple, un *datascientist* cherche à prédire si un décès est causé par une maladie cardiaque. La variable réponse est binaire (oui/non). Les résidus ne seront alors pas distribués selon une loi normale. Les GLM seront alors plus adaptés que les modèles linéaires classiques pour aider le *datascientist*.

La forme du modèle

En supposant que la variable Y est expliqué par un ensemble de variables explicatives $X = (X_1, \dots, X_p)$. Dans le cadre du mémoire, les quantités d'actes peuvent être expliquées par la classe d'âge et le département.

La forme du modèle s'écrit alors :

$$g(E(Y)) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (12.1)$$

Les paramètres sont estimés par la méthode de « maximum de vraisemblance » contrairement aux modèles linéaires classiques où ces derniers peuvent également être estimés par la méthode des « moindres carrés ordinaires »

Les composants d'un GLM

La classe générale des modèles linéaires est composée de 3 composants :

- **composante aléatoire** : il s'agit de la variable cible Y et sa loi ne suit plus obligatoirement une distribution gaussienne, mais elle doit appartenir à la famille exponentielle (gamma, poisson,...)
Exemple : Y peut représenter le montant remboursé par une mutuelle sur une année. ;
- **composante systématique** : identifie l'ensemble des variables explicatives (X_1, X_2, \dots, X_k) ;
- **fonction de lien** : il s'agit d'une fonction qui « délinéarise » ou non, la relation mathématique entre la composante systématique et aléatoire. Elle est strictement monotone et dérivable sur \mathbb{R} .
Exemple : g peut être la fonction logarithmique ou identité dans le cas d'un modèle linéaire classique.

12.5.2 La famille exponentielle

$$f_{\theta, \phi}(y) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right), \quad y \in \mathcal{S} \quad (12.2)$$

Avec :

- \mathcal{S} est un sous-ensemble de \mathbb{N} ou de \mathbb{R} ;
- $\theta \in \mathbb{R}$ est appelé le « paramètre canonique » ou le « paramètre de la moyenne » et est inconnu ;
- $\phi \in \mathbb{R}$ est le « paramètre de dispersion » supposé connu ;
- $a(\cdot)$ est une fonction définie sur \mathbb{R} et non nulle ;
- $b(\cdot)$ est une fonction définie sur \mathbb{R} , non nulle, deux fois dérivable et à dérivée première injective.
- $c(\cdot)$ est une fonction définie sur \mathbb{R}^2 .

Bien que ϕ soit supposé connu, ce n'est pas toujours le cas et ce paramètre est alors désigné comme le « paramètre de nuisance ». Ce paramètre sera ensuite estimé et considéré comme étant connu. En général, il est habituel de prendre en compte que :

$$a(\phi) = \frac{\phi}{\omega_i}$$

où ω_i représente un poids connu a priori attribué aux observations Y_i de Y . Par la suite, il est supposé que $\omega_i = 1$.

Quelques exemples de lois dans la famille exponentielle et leurs cas d'usage pour une composante systématique qui s'écrit $Y = b_0 + b_1 X$:

Lois	Cas d'usage	Fonction de lien
Normal	Modèle linéaire classique	$b_0 + b_1 X$ Identité
Régression logistique	Prédiction de variable binaire	$\frac{e^{b_0 + b_1 X}}{1 + e^{b_0 + b_1 X}}$
Poisson	Variable de comptage discrète	$e^{b_0 + b_1 X}$ Logarithme
Gamma	Variable continue strictement positive	$\frac{1}{b_0 + b_1 X}$ Inverse

TABLE 12.1 – Cas d'usage et fonctions de lien pour différentes lois

12.5.3 Estimation des paramètres par la méthode du maximum de vraisemblance

Le vecteur $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ est estimé en utilisant le principe du maximum de vraisemblance. En supposant que f soit la fonction de distribution choisie pour un échantillon de n variables indépendantes Y_i pour $i \in [1, \dots, n]$, ayant respectivement une espérance m_i avec $g(m_i) = \nu_i$. La forme de la densité pour chaque observation i est la suivante :

$$f_{\theta, \phi}(y_i) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right)$$

L'expression de la fonction de vraisemblance $\mathcal{L}(\theta(\beta), y, \phi)$ s'écrit alors de la manière suivante :

$$\begin{aligned} \mathcal{L}(\theta(\beta), y, \phi) &= \ln \left(\prod_{i=1}^n f_{\theta, \phi}(y_i) \right) \\ &= \sum_{i=1}^n \ln (f_{\theta, \phi}(y_i)) \\ &= \sum_{i=1}^n \omega_i \frac{y_i \theta_i - b(\theta_i)}{\phi} + \sum_{i=1}^n c(y_i, \phi) \end{aligned} \tag{12.3}$$

Il faut donc rechercher les valeurs du vecteur β qui maximisent la vraisemblance de l'équation précédente. Cela correspond à la recherche des paramètres β_i tels que :

$$\frac{\partial \mathcal{L}(\theta(\beta), y, \phi)}{\partial \beta_{ij}} = 0, \quad j = 0, \dots, p$$

En appliquant les hypothèses d'indépendance pour chaque i , il est possible d'écrire les équations de vraisemblance de la manière suivante :

$$\frac{\partial \mathcal{L}(\theta(\beta), y, \phi)}{\partial \beta_{ij}} = \sum_{i=1}^n \frac{\omega_i (y_i - b'(\theta_i)) x_{ij}}{\phi b''(\theta_i) g'(\mu_i)} = 0, \quad j = 0, \dots, p$$

Par la suite, il n'est pas possible de fournir une expression précise pour les estimateurs. Ensuite, différentes méthodes peuvent être envisagées pour déterminer la solution de ces équations. Deux de ces méthodes sont l'algorithme de Newton-Raphson et l'algorithme du score de Fisher. Cependant, ces approches dépassent le cadre de ce mémoire et ne seront pas détaillées ici.

Annexe F : La distribution Tweedie

La distribution de Tweedie (1984) est une sous-famille de la famille exponentielle avec une relation particulière entre l'espérance et la variance

$$f(y, \mu, \phi) = A(y, \phi) e^{\frac{1}{\phi}(y\theta(\mu) - \kappa(\theta(\mu)))}$$

Avec :

$$\theta(\mu) = \begin{cases} \frac{\mu}{1-p}, & p \neq 1 \\ \ln(\mu), & p = 1 \end{cases} \quad \kappa(\theta(\mu)) = \begin{cases} \frac{\mu}{2-p}, & p \neq 2 \\ \ln(\mu), & p = 2 \end{cases}$$

La distribution de la variable aléatoire Y est exprimée comme une loi de Poisson composée, avec des sauts Gamma :

$$Y \sim \mathcal{CPoi} \left(\mu^{2-p} \phi (2-p), \Gamma \left(-\frac{2-p}{\phi(1-p)}, \phi 2 - p\mu^{p-1} \right) \right)$$

Où $p \in [1, 2]$; $\mu = E(Y)$ et $V(\mu) = \phi\mu^p$.

Les distributions de Tweedie sont entièrement caractérisées par deux paramètres, à savoir l'espérance et la variance. En outre, en fonction du paramètre p , des distributions classiques peuvent être retrouvées :

- Lorsque $p \rightarrow 0$, la distribution de Tweedie se comporte comme une loi normale ;
- Lorsque $p \rightarrow 1$ (ou $\alpha \rightarrow \infty$), la distribution de Tweedie se comporte comme une loi de Poisson ;
- Lorsque $p \rightarrow 2$ (ou $\alpha \rightarrow 0$), la distribution de Tweedie se comporte comme une loi Gamma.

De plus :

- Lorsque $1 < p < 2$, la distribution de Tweedie se comporte comme une loi composée Poisson-Gamma ;
- Lorsque $2 < p < 3$, la distribution de Tweedie se comporte comme une distribution positive stable ;
- Lorsque $p \rightarrow 3$, la distribution de Tweedie se comporte comme une loi Gaussienne inverse.

Annexe G : Mesure de corrélation

Les trois coefficients qui vont être présentés sont compris dans l'intervalle $[-1;1]$ et en fonction de l'intensité des coefficients, ils sont classés dans des catégories :

Valeur absolue du coefficient	Intensité de la corrélation
$0,0 < r < 0,1$	Pas de corrélation
$0,1 < r < 0,3$	Faible corrélation
$0,3 < r < 0,5$	Moyenne corrélation
$0,5 < r < 0,7$	Forte corrélation
$0,7 < r < 1$	Très forte corrélation

TABLE 12.2 – Valeur absolue du coefficient de corrélation et intensité de la corrélation

12.5.4 Coefficient de Pearson

$$c(\widehat{X}, \widehat{Y}) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (12.4)$$

Avec :

- r : le coefficient de corrélation de Pearson
- x_i : les valeurs individuelles de la première variable (ex : fréquence)
- \bar{x} : moyenne de la variable x (moyenne des fréquences)
- y_i : les valeurs individuelles de la seconde variable (ex : coût)
- \bar{y} : moyenne de la variable y (moyenne des coûts)

Le coefficient de corrélation de Pearson présente plusieurs limitations. Il est sensible uniquement aux corrélations linéaires, ne tenant pas compte d'autres types de relations. De plus, il requiert que les variables X et Y aient des variances finies. Il peut également donner un faible coefficient même en présence d'une corrélation substantielle, particulièrement lorsque la variance de X est élevée mais celle de Y est faible. Enfin, ce coefficient est influencé par les lois marginales de X et Y , ce qui peut limiter son applicabilité.

12.5.5 Tau de Kendall

Contrairement au coefficient de Pearson, le tau de Kendall est un test non paramétrique. Les données ne doivent pas être distribuées selon une loi normale. X et Y sont deux variables aléatoires (par exemple, X est la fréquence et Y le coût moyen) ; (\tilde{X}, \tilde{Y}) est une copie indépendante de (X, Y) .

$$\tau(X, Y) = \underbrace{\mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) > 0]}_{\text{probabilité de concordance}} - \underbrace{\mathbb{P}[(X - \tilde{X})(Y - \tilde{Y}) < 0]}_{\text{probabilité de discordance}} \quad (12.5)$$

La version empirique s'écrit :

$$\tau = \frac{C - D}{C + D} \quad (12.6)$$

où :

-
- C représente le nombre de paires concordantes,
 - D représente le nombre de paires discordantes.

Deux paires d'observations (X, Y) et (\tilde{X}, \tilde{Y}) sont dites concordantes lorsque, par exemple, $X > \tilde{X}$ et $Y > \tilde{Y}$. Elles sont discordantes lorsque, par exemple, $X > \tilde{X}$ et $Y < \tilde{Y}$.

Il est possible qu'un problème survienne si, par exemple, X est égal à \tilde{X} , ce qui pourrait sous-estimer la corrélation.

12.5.6 Rho de Spearman

Il s'agit de la version non paramétrique du test de Pearson. Pour chaque valeur X et Y , un rang est associé aux valeurs de X et Y . Ainsi, le test du rho de Spearman est le même que le test de Pearson, mais il s'applique aux rangs et non aux valeurs directement.

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (12.7)$$

où :

- r_s est le coefficient de corrélation
- n est le nombre d'observations,
- d_i représente la différence en rang entre les deux variables.

Le coefficient de Spearman est utilisé pour identifier des tendances monotones. Lorsque la tendance est linéaire, il se comporte de manière similaire au coefficient de Pearson.

Annexe H : La déviance

Dans le cadre de ce mémoire, c'est le test de la déviance qui est utilisé pour évaluer la pertinence de l'ajustement d'un modèle linéaire généralisé. L'approche implique la comparaison entre le modèle estimé et un modèle de référence parfaitement ajusté, connu sous le nom de modèle saturé.

La comparaison entre le modèle estimé à l'aide du GLM et le modèle saturé repose sur l'évaluation de leurs vraisemblances respectives, notées \mathcal{L} et $\mathcal{L}_{\text{saturé}}$. Plus ces valeurs sont proches, meilleure est l'ajustement du modèle. Le rapport de vraisemblance qui en découle est le suivant :

$$\lambda = \frac{\mathcal{L}_{\text{saturé}}}{\mathcal{L}}$$

Avec la log-vraisemblance :

$$\ln(\lambda) = \ln(\mathcal{L}_{\text{saturé}}) - \ln(\mathcal{L})$$

En particulier, pour évaluer l'ajustement, deux statistiques sont généralement prises en compte :

- La déviance réduite : $D = 2 \ln(\lambda)$ (la statistique d'intérêt)
- La déviance non réduite : $D^* = \phi \cdot D$, où ϕ représente le paramètre de dispersion de la famille exponentielle associée à la variable expliquée Y dans le modèle estimé considéré.

Une déviance proche de zéro (c'est-à-dire, la déviance du modèle saturé proche de la déviance du modèle estimé) indique que le modèle saturé n'apporte que peu plus d'informations que le modèle estimé. En revanche, une déviance élevée suggère que le modèle estimé diffère considérablement du modèle saturé et qu'il n'est donc pas approprié.

Par ailleurs, sous l'hypothèse nulle H_0 , la statistique D suit une loi du Khi-deux avec $n - p - 1$ degrés de liberté ($D \sim \chi_{n-p-1}^2$), où n est le nombre d'observations et p est le nombre de variables explicatives ou de modalités de variables explicatives. En conséquence, il est admis que la qualité d'ajustement du modèle estimé est insatisfaisante si la condition suivante est vérifiée :

$$D_{\text{obs}} > \chi_{n-p-1, 1-\alpha}^2$$

où :

- D_{obs} représente la valeur empirique observée de D ,
- $\chi_{n-p-1, 1-\alpha}^2$ est le quantile d'ordre $1 - \alpha$ de la distribution du Khi-deux à $n - p - 1$ degrés de liberté.

Annexe I : La carte des départements et région de France

12.5.7 Les régions

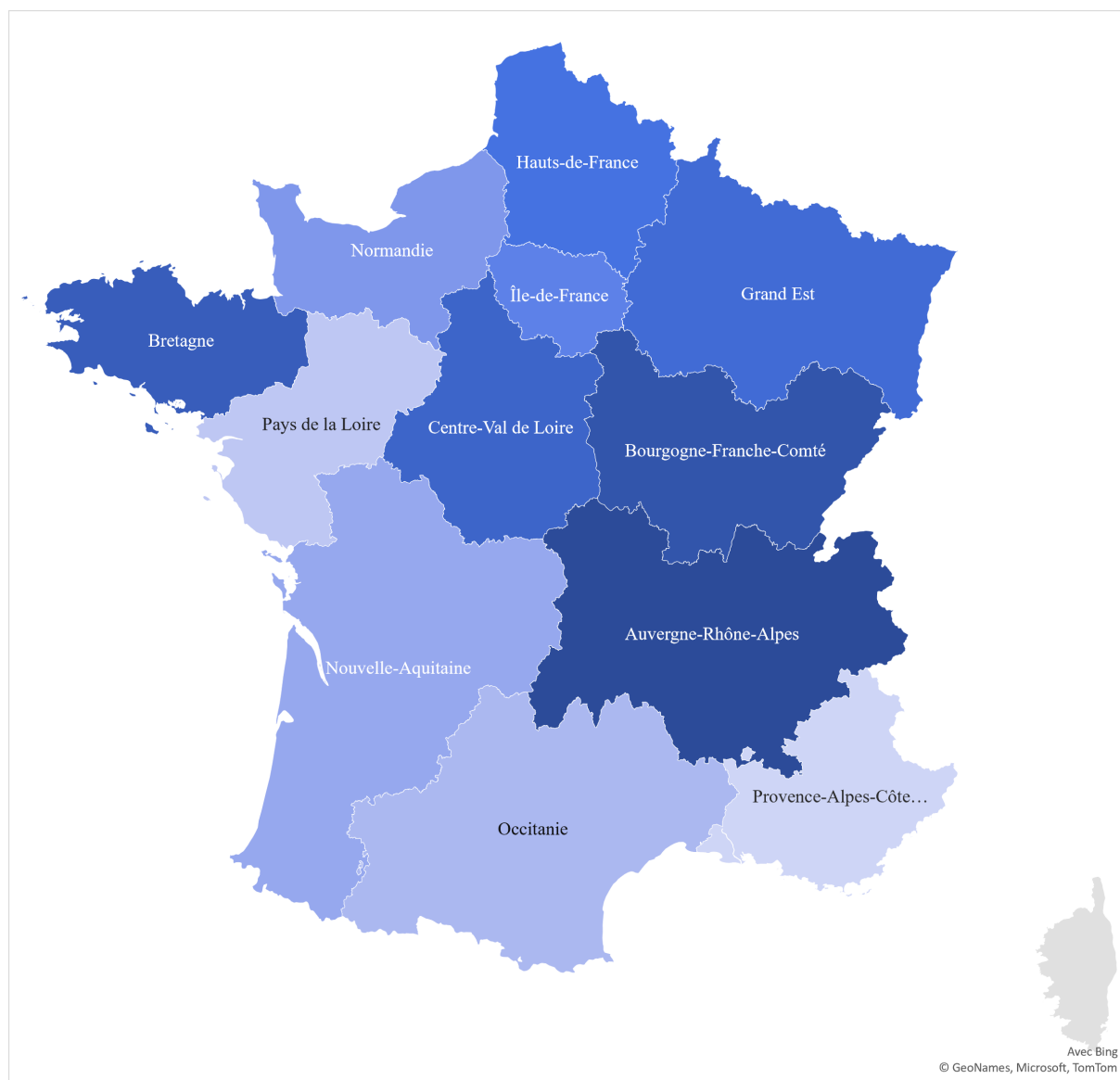


FIGURE 12.13 – Carte des régions de France

12.5.8 Les départements



FIGURE 12.14 – Carte des départements de France

Annexe J : Les tarifs par département

12.5.9 Les tarifs par département - Garantie V1

GARANTIE V1		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Ain	1	11,2	15,2	19,8	23,2	29,8	41,5	58,6	93,2
Aisne	2	12,3	14,6	18,9	22,4	30,6	37,9	55,2	90,0
Allier	3	15,3	18,6	24,1	27,4	37,0	43,9	58,0	82,1
Alpes-de-Haute-Provence	4	23,7	29,6	32,6	37,0	51,5	62,2	97,4	157,5
Hautes-Alpes	5	24,9	31,1	34,2	38,9	54,1	65,3	102,2	160,5
Alpes-Maritimes	6	16,2	20,0	24,3	27,9	38,5	50,6	58,6	106,9
Ardèche	7	14,6	19,5	25,8	26,5	35,5	43,2	59,0	91,2
Ardennes	8	14,4	17,8	23,6	26,4	36,0	48,3	59,4	107,2
Ariège	9	17,5	22,3	28,0	31,6	42,3	53,3	63,2	107,8
Aube	10	13,5	15,6	21,1	24,4	32,3	41,9	51,6	87,4
Aude	11	13,0	16,1	19,7	23,5	31,9	42,0	58,1	108,0
Aveyron	12	15,4	18,4	22,8	24,1	33,2	40,9	55,6	102,4
Bouches-du-Rhône	13	17,3	20,5	24,6	31,0	42,7	63,3	74,5	125,8
Calvados	14	14,7	16,3	19,8	25,2	34,3	43,1	62,2	83,4
Cantal	15	23,8	29,3	40,3	43,8	54,9	65,7	82,1	126,8
Charente	16	11,7	14,1	19,4	22,4	29,1	37,4	51,1	84,2
Charente-Maritime	17	13,0	15,1	20,9	23,6	31,1	37,9	50,9	74,5
Cher	18	13,6	16,0	19,3	22,7	31,3	38,3	53,1	81,8
Corrèze	19	12,9	15,5	21,3	23,5	30,9	38,5	51,8	87,7
Côte-d'Or	21	13,9	15,8	19,2	24,1	31,5	43,1	58,1	86,2
Côtes-d'Armor	22	16,3	18,6	24,1	26,2	34,3	41,8	55,9	87,4
Creuse	23	20,4	26,5	37,6	40,4	49,7	63,0	82,9	126,4
Dordogne	24	14,2	18,1	24,0	26,1	35,4	44,1	60,3	91,6
Doubs	25	11,9	13,3	16,2	21,6	28,4	40,4	57,1	85,7
Drôme	26	12,7	16,2	21,5	24,1	30,6	38,3	51,4	73,2
Eure	27	13,5	15,9	19,4	22,6	31,7	42,1	61,2	86,8
Eure-et-Loir	28	12,8	15,0	18,2	22,1	30,4	42,0	58,9	86,4
Finistère	29	15,4	16,7	21,7	24,1	33,4	43,0	58,4	85,8
Gard	30	16,4	20,7	24,9	28,1	38,4	48,3	67,9	104,6
Haute-Garonne	31	15,6	16,9	20,6	26,7	35,9	50,6	70,2	132,5
Gers	32	19,1	23,5	28,6	33,2	43,0	53,1	64,5	104,7
Gironde	33	18,3	18,8	26,1	32,2	41,8	56,3	79,8	118,7
Hérault	34	16,4	18,9	23,0	30,4	41,4	59,8	81,2	134,8
Ille-et-Vilaine	35	13,5	14,6	18,8	23,2	30,5	45,5	61,0	95,7
Indre	36	17,2	19,5	24,5	27,1	37,3	43,7	58,5	89,2
Indre-et-Loire	37	13,6	14,4	17,7	22,9	31,1	45,7	59,8	95,2
Isère	38	13,4	15,3	20,6	26,4	33,6	45,5	62,1	92,7
Jura	39	11,6	16,4	18,3	21,5	28,2	38,8	54,9	84,6
Landes	40	12,4	14,2	19,5	23,1	30,3	39,8	50,5	76,2
Loir-et-Cher	41	14,9	17,5	21,6	24,7	34,1	43,0	58,2	84,8
Loire	42	12,3	14,8	19,8	24,3	31,0	38,3	50,5	71,5
Haute-Loire	43	17,1	22,5	28,0	31,4	42,3	52,8	72,4	109,4
Loire-Atlantique	44	16,2	16,8	20,8	26,8	35,1	46,4	73,6	89,4
Loiret	45	12,0	14,4	17,8	23,3	32,1	43,7	61,7	90,8
Lot	46	19,0	27,2	31,3	34,7	40,0	46,0	59,1	94,3
Lot-et-Garonne	47	14,2	16,8	23,5	26,8	34,7	45,2	60,7	94,2
Lozère	48	24,1	29,5	34,1	39,3	50,6	62,7	84,9	138,3

FIGURE 12.15 – Primes pures mensuelles par département (Garantie V1) page 1/2

GARANTIE V1		CLASSE AGE							
L DPT	DPT	0	20	30	40	50	60	70	80
Maine-et-Loire	49	13,3	14,8	18,1	23,3	30,4	39,3	63,0	86,5
Manche	50	13,3	14,6	17,8	20,0	28,5	34,7	47,9	65,2
Marne	51	13,5	16,1	19,6	25,2	34,8	47,1	62,1	99,0
Haute-Marne	52	19,7	23,6	31,3	33,9	45,8	58,0	65,8	98,8
Mayenne	53	15,7	18,9	23,7	25,9	32,4	39,7	56,6	86,2
Meurthe-et-Moselle	54	15,5	17,6	22,7	27,0	39,0	53,9	61,9	91,1
Meuse	55	18,9	22,3	31,6	36,6	46,6	61,0	70,7	115,9
Morbihan	56	14,3	16,6	21,8	24,2	33,3	41,7	56,9	85,8
Moselle	57	12,5	13,6	17,6	21,3	28,1	39,7	47,2	72,2
Nièvre	58	15,0	19,4	23,5	25,3	33,3	39,9	53,1	78,6
Nord	59	16,3	17,7	23,4	29,2	39,6	52,4	72,1	110,8
Oise	60	12,4	14,4	17,9	21,1	27,9	36,8	58,3	92,2
Orne	61	15,1	17,4	21,4	23,9	33,2	39,8	55,0	77,7
Pas-de-Calais	62	18,3	19,6	24,8	31,7	42,9	55,0	74,4	120,1
Puy-de-Dôme	63	12,8	15,1	19,7	23,9	31,0	39,4	53,2	72,3
Pyrénées-Atlantiques	64	12,7	15,1	20,4	25,1	33,1	41,9	54,5	85,3
Hautes-Pyrénées	65	14,0	18,3	21,7	25,1	34,7	43,1	67,5	115,6
Pyrénées-Orientales	66	14,7	18,3	22,1	26,3	36,6	45,1	57,3	106,2
Bas-Rhin	67	14,4	15,4	19,9	23,5	31,9	44,5	53,8	83,9
Haut-Rhin	68	12,3	14,1	18,2	20,5	27,2	38,3	46,1	70,8
Rhône	69	13,4	14,5	19,8	25,6	32,4	43,5	57,7	81,3
Haute-Saône	70	13,4	16,9	20,8	23,6	31,3	43,3	63,0	91,8
Saône-et-Loire	71	13,5	17,5	20,7	22,9	30,4	39,5	51,9	71,6
Sarthe	72	14,3	17,6	21,7	26,7	34,5	44,0	61,1	81,9
Savoie	73	13,1	14,9	20,0	22,4	30,6	41,7	61,9	96,9
Haute-Savoie	74	11,3	13,1	17,4	21,8	28,7	41,2	60,3	86,9
Paris	75	19,4	23,4	28,2	33,9	44,9	60,1	75,1	117,5
Seine-Maritime	76	15,3	17,2	21,1	26,8	36,6	47,6	61,0	95,8
Seine-et-Marne	77	15,8	19,1	25,5	28,4	37,5	44,8	63,2	98,1
Yvelines	78	17,7	22,0	27,0	32,1	41,9	58,0	70,8	108,1
Deux-Sèvres	79	11,2	13,7	18,9	21,6	28,2	37,6	50,0	78,6
Somme	80	14,4	15,5	19,6	23,9	32,7	41,3	59,5	92,9
Tarn	81	15,5	20,4	24,5	27,1	37,5	47,6	64,4	99,8
Tarn-et-Garonne	82	14,0	19,1	22,8	24,6	32,9	43,4	60,6	106,2
Var	83	16,3	20,9	24,8	29,0	40,3	50,5	60,5	116,8
Vaucluse	84	13,4	18,6	22,1	26,1	37,7	51,4	66,4	137,6
Vendée	85	15,6	18,2	22,2	24,0	31,3	37,2	61,0	76,9
Vienne	86	11,5	13,4	18,4	23,5	30,9	41,6	56,4	88,9
Haute-Vienne	87	12,7	15,1	20,7	24,4	33,2	42,6	56,0	90,6
Vosges	88	15,4	19,5	25,5	28,5	37,7	48,5	55,9	85,3
Yonne	89	11,9	15,6	18,4	22,3	29,6	39,2	52,4	76,1
Territoire de Belfort	90	15,3	16,7	22,0	28,1	37,1	52,1	74,0	121,4
Essonne	91	16,1	20,1	26,5	30,2	40,9	53,1	64,4	101,0
Hauts-de-Seine	92	16,7	19,9	26,3	30,6	40,0	52,2	66,4	98,4
Seine-Saint-Denis	93	16,5	20,8	25,2	30,7	39,8	51,8	65,9	100,2
Val-de-Marne	94	18,3	22,1	26,9	31,7	41,3	58,8	69,3	109,9
Val-d'Oise	95	18,0	21,7	27,8	31,5	41,3	58,7	71,9	106,3

FIGURE 12.16 – Primes pures mensuelles par département (Garantie V1) page 2/2

12.5.10 Les tarifs par département - Garantie V2

GARANTIE V2		CLASSE AGE							
L DPT	DPT	0	20	30	40	50	60	70	80
Ain	1	18,7	21,5	28,7	34,4	44,9	64,6	79,1	114,3
Aisne	2	19,4	22,0	27,8	33,8	46,3	54,1	77,6	115,7
Allier	3	25,4	27,3	34,5	39,5	52,5	56,0	71,5	106,9
Alpes-de-Haute-Provence	4	33,9	47,6	63,1	74,3	85,6	96,3	102,3	175,7
Hautes-Alpes	5	35,0	50,7	67,8	75,1	89,3	98,1	108,6	180,9
Alpes-Maritimes	6	23,6	30,8	33,5	41,3	58,8	67,9	74,2	133,5
Ardèche	7	24,0	31,8	35,9	44,7	51,2	62,4	70,9	110,2
Ardennes	8	22,6	29,7	35,3	42,2	56,4	67,5	83,4	132,9
Ariège	9	25,5	30,1	35,8	42,2	58,2	71,4	94,3	140,2
Aube	10	21,6	26,6	33,4	38,8	50,6	65,6	71,1	124,1
Aude	11	22,7	29,7	34,4	37,8	52,7	61,1	84,0	131,7
Aveyron	12	26,1	28,1	33,2	38,4	53,9	63,7	82,1	128,4
Bouches-du-Rhône	13	21,9	26,3	28,7	40,8	57,2	80,0	102,8	148,1
Calvados	14	19,8	20,7	22,9	31,7	44,5	57,1	75,7	101,9
Cantal	15	28,2	31,0	43,5	46,6	68,9	73,9	91,0	132,1
Charente	16	18,1	21,6	27,9	32,4	41,2	52,0	66,6	104,6
Charente-Maritime	17	16,8	19,9	25,7	29,2	37,5	42,7	55,6	84,2
Cher	18	19,3	25,8	29,4	32,2	45,2	56,9	76,2	100,3
Corrèze	19	18,4	23,4	30,4	32,5	41,5	50,0	71,9	103,4
Côte-d'Or	21	17,0	19,2	21,7	29,8	39,0	56,4	64,4	96,2
Côtes-d'Armor	22	22,6	26,8	34,0	35,2	47,0	54,3	72,3	97,6
Creuse	23	27,4	39,0	52,9	59,2	64,7	81,0	94,1	141,7
Dordogne	24	21,6	28,1	37,2	42,5	49,4	63,3	83,9	117,0
Doubs	25	14,1	16,1	18,2	26,6	35,1	53,1	63,9	95,7
Drôme	26	20,7	29,2	34,9	45,4	46,2	54,7	73,6	110,6
Eure	27	19,3	23,4	26,0	31,5	45,3	61,2	77,8	118,0
Eure-et-Loir	28	19,2	26,1	29,7	35,0	48,0	57,6	91,1	109,6
Finistère	29	20,6	21,6	27,4	30,9	42,9	55,3	75,3	98,1
Gard	30	22,7	30,0	34,4	39,3	54,5	67,4	94,4	145,1
Haute-Garonne	31	18,3	22,0	26,8	32,9	45,3	69,4	97,7	153,7
Gers	32	26,8	30,8	35,3	46,3	63,4	69,7	83,1	131,0
Gironde	33	24,6	28,3	36,0	41,3	51,7	74,3	104,4	141,8
Hérault	34	21,0	23,2	26,7	40,1	55,6	79,2	118,6	167,8
Ille-et-Vilaine	35	18,7	21,9	26,6	31,8	42,8	59,9	77,7	101,4
Indre	36	24,4	27,0	34,3	37,2	51,7	57,0	73,1	94,7
Indre-et-Loire	37	18,2	20,6	23,7	33,1	45,7	60,7	79,2	105,9
Isère	38	17,0	18,2	22,6	31,1	40,0	54,5	72,9	106,0
Jura	39	17,6	23,7	27,6	30,2	39,7	56,0	66,2	101,2
Landes	40	17,9	21,6	28,9	33,3	43,7	49,6	69,6	101,4
Loir-et-Cher	41	22,2	25,5	29,3	36,9	52,1	58,1	77,7	109,2
Loire	42	17,2	20,0	24,7	31,8	41,0	48,0	61,4	84,6
Haute-Loire	43	25,2	27,8	35,9	41,8	55,5	64,8	91,1	129,9
Loire-Atlantique	44	16,6	19,5	22,6	32,4	42,1	63,3	101,6	112,8
Loiret	45	17,5	22,3	25,6	34,9	47,5	58,4	76,3	109,7
Lot	46	27,1	37,2	34,1	48,9	61,9	68,9	76,0	140,9
Lot-et-Garonne	47	19,7	25,4	33,8	39,7	50,9	62,1	83,1	113,6
Lozère	48	27,7	36,0	55,3	61,0	73,4	80,9	121,2	169,0

FIGURE 12.17 – Primes pures mensuelles par département (Garantie V2) page 1/2

GARANTIE V2		CLASSE AGE							
L_DPT	DPT	0	20	30	40	50	60	70	80
Maine-et-Loire	49	18,4	20,5	23,5	31,3	40,5	53,1	72,7	110,1
Manche	50	20,0	21,2	23,5	26,6	38,9	45,6	64,6	80,2
Marne	51	18,6	21,4	25,0	35,8	47,8	67,4	76,6	122,5
Haute-Marne	52	33,1	33,6	42,6	52,7	70,5	86,8	87,0	130,6
Mayenne	53	23,1	30,0	35,1	38,5	48,1	60,5	80,4	97,1
Meurthe-et-Moselle	54	20,2	21,4	26,7	37,7	50,3	71,5	74,0	112,9
Meuse	55	25,1	33,9	42,9	47,4	67,1	86,8	93,2	130,1
Morbihan	56	20,1	22,6	28,9	31,1	43,1	53,0	73,4	99,7
Moselle	57	15,9	16,2	20,4	25,8	34,1	50,7	61,9	85,4
Nièvre	58	24,5	30,3	34,4	42,9	48,4	53,0	57,0	83,5
Nord	59	20,2	21,1	27,1	38,7	52,6	69,7	99,3	130,8
Oise	60	17,3	20,1	25,2	31,4	42,2	56,5	76,7	113,1
Orne	61	22,2	28,8	32,1	36,1	45,6	55,1	77,3	97,5
Pas-de-Calais	62	22,4	24,1	30,0	40,5	55,2	69,5	93,2	133,3
Puy-de-Dôme	63	19,5	22,5	24,7	31,8	41,7	50,4	68,3	95,9
Pyrénées-Atlantiques	64	17,5	20,0	26,4	34,3	44,2	57,8	73,2	101,4
Hautes-Pyrénées	65	23,5	27,2	31,8	39,2	54,0	62,8	99,6	128,0
Pyrénées-Orientales	66	21,5	28,5	32,2	37,2	52,4	61,7	72,7	113,2
Bas-Rhin	67	16,5	16,9	21,6	29,2	38,6	58,3	66,7	94,4
Haut-Rhin	68	18,0	21,2	26,6	30,1	39,9	59,1	67,2	94,2
Rhône	69	16,8	17,4	23,4	31,7	40,8	56,4	71,6	112,8
Haute-Saône	70	18,6	24,3	29,2	31,5	41,8	59,0	73,9	114,8
Saône-et-Loire	71	17,2	23,2	27,1	28,9	38,2	48,9	53,7	74,8
Sarthe	72	22,4	27,9	32,0	38,2	49,1	59,0	75,9	88,0
Savoie	73	21,7	26,4	29,4	33,8	45,4	59,5	79,8	107,0
Haute-Savoie	74	16,4	19,0	21,2	28,8	38,5	58,2	82,0	111,1
Paris	75	29,6	32,2	39,3	51,0	64,4	77,7	100,8	138,5
Seine-Maritime	76	20,1	21,2	23,4	33,5	47,2	61,2	76,0	116,6
Seine-et-Marne	77	20,3	24,5	31,5	36,4	46,5	62,4	75,6	111,7
Yvelines	78	25,5	30,4	36,3	47,4	59,0	69,3	96,7	141,3
Deux-Sèvres	79	16,0	20,3	26,2	30,0	39,1	53,8	66,8	101,5
Somme	80	18,5	18,8	23,5	31,1	43,1	52,2	75,2	114,6
Tarn	81	22,8	28,1	33,5	39,8	55,5	68,6	89,5	115,6
Tarn-et-Garonne	82	20,2	27,6	33,3	37,7	52,6	70,3	106,3	127,0
Var	83	22,2	29,8	33,0	39,3	56,0	61,7	67,3	125,3
Vaucluse	84	21,4	27,1	33,2	44,7	63,0	76,6	95,0	168,3
Vendée	85	19,6	26,2	30,3	32,8	42,6	48,0	84,0	108,4
Vienne	86	16,0	18,1	23,2	33,4	42,7	55,9	70,2	108,9
Haute-Vienne	87	18,3	21,3	27,4	34,3	45,0	54,0	70,9	109,9
Vosges	88	22,2	28,9	32,3	38,3	50,5	63,5	66,0	102,2
Yonne	89	19,7	25,4	30,5	34,1	44,8	59,3	66,5	99,4
Territoire de Belfort	90	24,2	26,6	33,0	43,9	56,5	69,2	106,6	135,9
Essonne	91	22,9	27,5	31,3	41,7	54,3	67,1	86,2	118,8
Hauts-de-Seine	92	23,2	26,5	29,9	41,5	52,3	64,2	81,1	114,1
Seine-Saint-Denis	93	21,7	25,7	33,3	43,4	54,6	63,4	80,0	115,2
Val-de-Marne	94	25,0	29,0	34,8	45,3	56,3	70,4	92,3	130,4
Val-d'Oise	95	24,6	29,4	35,5	45,4	57,4	70,9	95,5	128,7

FIGURE 12.18 – Primes pures mensuelles par département (Garantie V2) page 2/2

Annexe K : Base fictive Ile-de-France

Ile-de-France		Base complète NEW_DAMIR	Base fictive modifiée
Population totale de la région		12 215 245	677 420
Composition des classes d'âges	0	26%	5%
	20	14%	10%
	30	14%	5%
	40	14%	7%
	50	13%	10%
	60	10%	34%
	70	5%	17%
	80	4%	13%
V1	Prime mensuelle moyenne	33,83 €	51,49 €
V2	Prime mensuelle moyenne	44,45 €	65,11 €

FIGURE 12.19 – Base fictif Ile-de-France

LISTE DES ACRONYMES

- AIC : Akaike Information Criterion
- ACS : Aide à la Complémentaire Santé
- AME : Aide Médicale d'État
- AMI : Acte Médico-Infirmier
- AMO : Assurance Maladie Obligatoire
- BRSS : Base de remboursement de la Sécurité Sociale
- CCAM : Classification Commune des Actes Médicaux
- CCAS : Centre Communal d'Action Sociale
- CMU-C : Couverture Maladie Universelle Complémentaire
- CSS : Complémentaire santé solidaire
- CNAM : Caisse Nationale d'Assurance Maladie
- DAMIR : Dépenses d'Assurance Maladie Inter Régimes
- DREES : Direction de la Recherche, des Études, de l'Évaluation et des Statistiques
- GLM : Generalized linear model
- MT_RO : Montant remboursé par la Sécurité Sociale
- NGAP : Nomenclature Générale des Actes Professionnels
- ONDAM : Objectif National de Dépenses d'Assurance Maladie
- PLFSS : Projet de loi de financement de la Sécurité sociale
- PUMA : Protection Universelle Maladie
- RAC : Reste à charge
- RGPD : Règlement Général sur la Protection des Données
- SNIIRAM : Système National Inter Régimes d'Assurance Maladie
- TR : Taux de remboursement de la Sécurité Sociale

BIBLIOGRAPHIE

Articles

- [3] Brigitte DORMONT. *Liberté ou solidarité : le dilemme des complémentaires*. 2010.
- [8] Thomas Frinault et RÉMY LE SAOUT. *Communes, intercommunalité et action sociale. Les contraintes d'un changement d'échelle territoriale*. 2011.
- [9] Christophe DEFEUILLEY. *Légitimité de l'intervention privée dans un service public*. 2014.

Mémoires

- [11] Sandrine HUYNH. *Open data et Assurance santé : l'union fait de la force ?*, https://www.actuelia.fr/_files/ugd/e2a094_c7e2a57acbe4428892b11625855fb249.pdf. 2021.
- [15] Fatemeh ABDOLLAHI. *Tarifification d'une complémentaire santé à destination des séniors, modulaire par poste de garanties et l'impact sur la solvabilité*, <https://www.institutdesactuaires.com/docs/mem/46dd866818e6465b650f8cec38a75327.pdf>. 2017.

Sites internet

- [1] AMELI. *Participation forfaitaire*, <https://www.ameli.fr/val-de-marne/assure/remboursements/reste-charge/participation-forfaitaire-1-euro>. 2023.
- [2] AMELI. *Franchise médicale*, <https://www.ameli.fr/val-de-marne/assure/remboursements/reste-charge/franchise-medicale>.
- [4] AMF. *Les aspects juridiques liés à la mise en place de mutuelles communales*, <https://www.amf62.fr/les-aspects-juridiques-lies-a-la-mise-en-place-de-mutuelles-communales/>.

-
- [5] Catherine MAISONNEUVE. *Les aspects juridiques liés à la mise en place de mutuelles communales*, <https://www.lagazettedescommunes.com/195646/caumont-sur-durance-dans-le-vacluse-lance-la-1ere-mutuelle-de-village/>.
- [6] Le PARISIEN. *Montreuil mutuelle communale*, <https://www.leparisien.fr/seine-saint-denis-93/montreuil-a-desormais-sa-mutuelle-communale-25-09-2018-7902375.php/>.
- [7] Région IDF. *IDF mutuelle régionale*, <https://www.iledefrance.fr/ile-de-france-mutuelle-sante-la-premiere-mutuelle-regionale-pour-tous-les-franciliens>.
- [10] DATAGOUV. *Open DAMIR : base complète sur les dépenses d'assurance maladie inter régimes*, <https://www.data.gouv.fr/fr/datasets/open-damir-base-complete-sur-les-depenses-dassurance-maladie-inter-regimes/>.
- [12] DATAGOUV. *Dépenses d'assurance maladie hors prestations hospitalières par caisse primaire/département*, <https://www.data.gouv.fr/fr/datasets/depenses-d-assurance-maladie-hors-prestations-hospitalieres-par-caisse-primaire-departement/>. 2023.
- [13] INSEE. *Pyramide Age*, <https://www.insee.fr/fr/outil-interactif/5014911/pyramide.htm?#!y=2019&a=20,30&v=2&t=2&c=11>. 2023.
- [14] DATAGOUV. *Base SIRENE SIRET*, <https://www.data.gouv.fr/fr/datasets/base-sirene-des-entreprises-et-de-leurs-etablissements-siren-siret/>.
- [17] INSEE. *Taux de chômage*, https://www.insee.fr/fr/statistiques/2012804#tableau-TCRD_025_tab1_regions2016. 2023.
- [18] ANCT. *Indice de vieillissement*, <https://www.observatoire-des-territoires.gouv.fr/indice-de-vieillissement>. 2023.
- [19] DATAGOUV. *Etablissement de sante*, <https://www.data.gouv.fr/fr/datasets/finess-extraction-du-fichier-des-etablissements/>. 2023.
- [20] INSEE. *Niveau de vie median*, https://www.insee.fr/fr/statistiques/6436484?sommaire=6036904#figure2_radio2. 2023.

Formations

- [22] ACTUELIA. *Formation sur la tarification santé*. 2019.