

**Mémoire présenté le :**

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA  
et l'admission à l'Institut des Actuaires**

Par : **Edouard PLUNET**

Titre **Modification d'un zonier et conséquences opérationnelles – Application  
sur une gamme de produits individuels en complémentaire santé**

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

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

Membres présents du jury de l'Institut des Actuaires

Membres présents du jury de l'ISFA

Nom : AESIO Mutuelle

Signature :

Directeur de mémoire en entreprise :

Nom : Thomas SENNE

Signature :

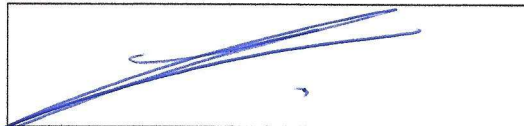
Invité :

Nom :

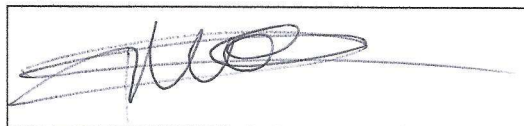
Signature :

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

Signature du responsable entreprise



Signature du candidat



## Résumé

La différenciation des primes pures par zone géographique est un sujet régulièrement évoqué dans la littérature. En effet, autant en IARD qu'en complémentaire santé nombre de mémoires d'actuariat, entre autres, traitent des méthodes de mise en œuvre ou de comparaison de zoniers. Il est désormais communément accepté que la localisation géographique est un facteur déterminant pour une tarification sur de nombreux risques. Cela étant, la mise à jour et le suivi d'un zonier existant est un sujet qui semble plus rarement évoqué. Dans le cadre de ce mémoire, après la mise en œuvre de méthodes de construction de zoniers usuelles, des comparaisons de zoniers seront réalisées par le biais d'analyses moins proposées dans la littérature relative à la complémentaire santé. Ainsi, après l'utilisation des modèles linéaires généralisés et la présentation des méthodes de classification comme la CAH ou les K-moyennes, nous sélectionnerons le zonage le plus adapté par la réintégration dans les modèles des variables de zones créées d'une part et par l'utilisation de critères statistiques d'autre part. Enfin, ce mémoire s'intéressera également aux conséquences opérationnelles associées à la mise à jour d'un zonier existant en particulier sur la commercialisation et l'étude du risque. La modification en profondeur d'un zonier peut être assez difficile à comprendre notamment commercialement et les conséquences sur les modifications de la structure du portefeuille, potentiellement difficiles à anticiper tant sur les résiliations que sur les adhésions. Aujourd'hui, la mise à jour du zonier en complémentaire santé se fait plutôt de manière empirique par le biais, par exemple, d'analyse des P/C par département. En outre, un P/C dégradé par département n'est pas nécessairement en lien avec le zonage réalisé mais peut s'expliquer par d'autres facteurs. Ainsi, l'utilisation de méthodes plus théoriques pourrait constituer une alternative intéressante au suivi de zonier et à la compréhension des différents facteurs sur le risque.

## Abstract

The differentiation of pure premiums by geographical area is a subject regularly raised in the literature. Indeed, in both P&C and supplementary health insurance sectors, a number of actuarial dissertations, among others, deal with methods of implementation or comparison of segmentations. It is now commonly accepted that geographic location is a determining factor in the pricing of many risks. However, the updating and monitoring of an existing segmentation is a subject that seems to be less frequently discussed. In the framework of this thesis, after the implementation of standard methods for segmentations, comparisons of segmentations will be made through analyses less proposed in the literature on supplementary health insurance. Thus, after the use of generalized linear models and the presentation of classification methods such as hierarchical clustering or K-means, we will select the most suitable segmentation by reintegrating obtained geographical variables into the models on the one hand and by using statistical criteria on the other hand. Finally, this dissertation will also look at the operational consequences associated with the updating of an existing segmentation system, in particular with respect to marketing and risk analysis. The in-depth modification of a segmentation can be quite difficult to understand, notably for commercial purposes, and consequences on changes in the portfolio structure potentially difficult to anticipate both on terminations and memberships. Today, the updating of the complementary health insurance segmentation is done rather empirically through, for example, P/C analysis by department. In addition, a degraded P/C by department is not necessarily related to the segmentation carried out but can be explained by other factors. Thus, the use of more theoretical methods could be an interesting alternative to segmentation monitoring and for understanding the various factors associated with risk.

## Remerciements

Je souhaite avant tout remercier mon directeur de mémoire, Thomas SENNE pour son aide dans le choix du sujet, l'orientation de mes réflexions au cours de la rédaction, pour ses diverses relectures ainsi que pour son soutien et pour la motivation qu'il a su me donner. Un grand merci également à Laurianne PLAUD qui m'a beaucoup aidé pour l'ensemble de la conception de ce mémoire et pour le temps qu'elle a pu y consacrer.

Plus largement, je remercie également mon entreprise ADREA Mutuelle devenue AESIO mutuelle et en particulier mon responsable direct Florian LEULLIER ainsi que Corine BENOIT, directrice technique d'AESIO Mutuelle, pour le temps accordé permettant la réalisation de ce mémoire.

Je présente également mes remerciements à l'ISFA tant aux personnels administratifs qu'aux enseignants.

Enfin, j'aimerais exprimer ma gratitude à ma famille, Anaïs et Armand, pour leur soutien quotidien et leurs encouragements ainsi qu'à mes amis, en particulier Gautier pour ses éclaircissements sur certaines particularités techniques et linguistiques. Je tiens à remercier également mes parents pour leur disponibilité.

# Sommaire

Introduction .....	6
PARTIE I : ENVIRONNEMENT, CONTEXTE ET ENJEU .....	8
1. La protection sociale en France – Zoom sur l’assurance Santé .....	8
1.1. Le régime obligatoire .....	9
1.2. La branche maladie .....	10
1.3. La complémentaire santé et ses acteurs .....	16
1.4. Le cadre juridique de la complémentaire santé .....	18
1.5. Contexte spécifique des mutuelles .....	23
2. Présentation d’ADREA Mutuelle .....	24
2.1. Chiffres et dates clés .....	24
2.2. Présentation du portefeuille global d’ADREA Mutuelle .....	27
PARTIE II : LA GAMME FLEX’ADREA .....	31
1. Présentation de la base de données .....	31
1.1. Qualité des données .....	31
1.2. Extrapolation des données et choix de la période d’analyse .....	32
2. Généralités sur la gamme .....	34
2.1. Présentation des différents produits de la gamme Flex’ADREA .....	34
2.2. Résultats globaux 2019 .....	37
3. Analyse descriptive et présentation du zonier en place .....	42
3.1. Statistiques descriptives .....	42
3.2. Zonier en place .....	58
PARTIE III : UTILISATION DES MODELES LINEAIRES GENERALISES (GLM) EN SANTE ET METHODE DE CONSTRUCTION DE ZONIER .....	63
1. Présentation théorique des GLM .....	63
2. Loi de Tweedie .....	65
3. Méthode de construction d’un zonier en utilisant les GLM .....	67
PARTIE IV : APPLICATION - MODELISATION DU RISQUE SANTE .....	70
1. Modèle complet .....	70
1.1. Etapes préliminaires .....	70
1.2. Construction du modèle .....	74
1.3. Validation et analyse du modèle complet .....	78
2. Création d’un zonier .....	83
2.1. Présentation des méthodes de mise en place d’un zonier .....	83
2.2. Construction d’une base de résidus par département .....	83
3. Méthodes de classification .....	84
3.1. Classification ascendante hiérarchique (CAH) .....	84
3.2. Classification avec la méthode des K-moyennes .....	88

4. Comparaison des zoniers obtenus .....	90
PARTIE V : CONSEQUENCES OPERATIONNELLES DE LA MODIFICATION D'UN ZONIER .....	98
Conclusion .....	107
Bibliographie .....	108
Annexes .....	109
Annexe 1 : Droit et mutualité .....	109
Annexe 2 : Modélisation Gamma .....	111
Annexe 3 : Lissage Whittaker Henderson .....	111
Annexe 4 : Zonier avec 3 classes .....	113
Annexe 5 : Hypothèse des scénarios de sensibilités .....	114
Table des figures .....	116

## Introduction

Le marché de l'assurance maladie est devenu extrêmement concurrentiel au cours de ces dernières années. D'une part, les fortes contraintes réglementaires ont conduit nombre d'acteurs à fusionner pour former des groupes de taille très importante, d'autre part, la montée en puissance de l'assurance collective au détriment de l'assurance individuelle, portée par des textes de loi récents comme l'Accord National Interprofessionnel, a permis une importante prise de part de marché par les sociétés d'assurance au détriment des mutuelles.

En outre, l'expérience désormais très significative de nombreux acteurs sur le marché de la complémentaire santé mais aussi la sensibilité au prix de l'ensemble des consommateurs, impose aux assureurs de réaliser des tarifications de plus en plus proches du profil réel des consommateurs, tout en essayant de maintenir une mutualisation relative afin de permettre aux personnes les plus consommatrices de ne pas se couvrir à des prix trop importants.

C'est dans ce cadre que depuis plusieurs années maintenant, les zoniers de tarification ont été mis en place. Il est désormais avéré que la zone d'habitation est un facteur déterminant dans le niveau de consommation des personnes protégées pour la complémentaire santé. Aussi, la zone d'habitation étant un critère accepté dans la segmentation tarifaire, les assureurs ont mis en place des zoniers tarifaires qui vont venir modifier la cotisation estimée sur une consommation globale pour l'adapter à la zone de soins de l'assuré. Ainsi, si certaines zones très urbaines vont être potentiellement plus coûteuses en prestations pour l'assureur, du fait entre autres de la présence d'un plus grand nombre de médecins et hôpitaux pratiquant des dépassements d'honoraires ou de divers facteurs sociaux, d'autres plus rurales seront estimées moins coûteuses et pourront permettre d'appliquer des correctifs baissiers sur les prix.

Il est important de comprendre que le sujet est d'autant plus prégnant pour les mutuelles qu'il est relativement plus récent. Historiquement, les mutuelles étaient plutôt localisées et avaient, par conséquent, une connaissance précise du profil des consommateurs dans sa région d'activité. Cela rendait potentiellement inutile l'application des zoniers car les forces de ventes de ces structures étant concentrées dans la même zone, segmenter les tarifs sur ce critère semblait moins justifié. Cependant, avec les nombreuses fusions évoquées précédemment mais aussi avec la volonté d'expansion exprimée par les très grands groupes mutualistes, l'approche tarifaire par zone est devenue une nécessité commerciale et technique.

Il existe aujourd'hui un zonier, mis en place lors de la fusion de huit mutuelles ayant mené à la création d'ADREA Mutuelle. Dans un premier temps, ce mémoire s'intéressera à l'exploitation de l'expérience propre de la mutuelle afin de valider si elle est adaptée à la création d'un nouveau zonier ou au minimum peut-elle permettre de remettre en cause le zonier en place. Il semble assez intuitif de penser que la mise à jour du zonier, au moins sur les zones d'exploitations des mutuelles historiques, sera possible. C'est pourquoi, dans un second temps, l'objectif sera de mesurer les avantages et inconvénients potentiels d'une mise à jour de zonier.

C'est en anticipation de la création du groupe AESIO Mutuelle que s'inscrit ce mémoire. Ainsi, la fusion d'ADREA Mutuelle, d'APREVA Mutuelle et d'EOVI-MCD Mutuelle va permettre à la base d'expérience utilisée dans ce mémoire de s'étoffer de l'expérience historique d'autres mutuelles présentes sur des régions dans lesquelles l'expérience de notre mutuelle n'est pas ou peu significative.

Dans cette optique, ce mémoire se décomposera en cinq parties. Dans la première partie, nous rappellerons le cadre de l'assurance maladie et les mécanismes existants entre le régime obligatoire et les complémentaires santé puis nous présenterons plus en détail la mutuelle. Dans une deuxième partie, nous ferons un focus sur le produit principalement

commercialisé par ADREA Mutuelle, Flex'ADREA et présenterons le zonier en place. Ensuite, nous rappellerons les méthodes théoriques utilisées pour la création des zoniers, avant de présenter l'application réalisée et la mise en place d'un nouveau zonier basé sur l'expérience portefeuille. Enfin, nous analyserons les conséquences de la modification d'un zonier sous différents aspects opérationnels.

## PARTIE I : ENVIRONNEMENT, CONTEXTE ET ENJEU

Le risque santé, s'il est désormais relativement connu et maîtrisé par un grand nombre d'acteur, présente certaines spécificités et un cadre particulier explicité dans cette partie. Le portefeuille global d'ADREA Mutuelle sera également présenté de manière synthétique avant un zoom sur l'offre Flex'ADREA.

### 1. La protection sociale en France – Zoom sur l'assurance Santé

Selon le rapport de la DREES, en 2019, la consommation de soins et de biens médicaux (CSBM) a été évaluée à 208 milliards d'euros. Si la croissance avait été modérée en 2017 et 2018 (respectivement de +1,7% et +1,6%), notamment du fait du fort ralentissement des soins hospitaliers, la croissance en 2019 s'élève à +2,1%. Elle s'explique notamment par l'augmentation de ces mêmes soins hospitaliers. En 2019, la France consacre au total 11,3 % de son PIB à la santé, soit 1,4 point de plus que la moyenne de l'Union Européenne à 28<sup>1</sup>. La CSBM a augmenté de près de 30 milliards d'euros en 8 ans (178 milliards d'euros en 2011) soit une augmentation de près de 17% sur la période, équivalente à une évolution linéaire théorique de 2% par an. En comparaison, l'inflation a cru sur la période de moins de 8 %, soit une évolution linéaire théorique de moins de 1% par an.

Pour faire face à ces dépenses, les ménages sont généralement couverts par deux biais :

- le régime général de la Sécurité Sociale,
- les complémentaires santé.

En 2019, la Sécurité Sociale finance 78,2 % de la CSBM (contre 78,0% en 2018 et 76,2% en 2011), et les organismes complémentaires 13,4%. La part restant à la charge des ménages continue de reculer pour s'établir à 6,9% en 2019. En outre, pour les populations les plus démunies, l'Etat peut intervenir, notamment via la Couverture Maladie Universelle (CMU) ou l'Aide à la Complémentaire Santé (ACS) ayant fusionné fin 2019 pour devenir la Complémentaire Santé Solidaire (1,5% de la CSBM en 2019).

Cependant, si la part des dépenses totales prise en charge par les complémentaires santé est assez stable en pourcentage depuis maintenant plusieurs années comme nous pouvons le voir dans la figure 1, l'augmentation de la dépense globale au cours du temps entraîne une augmentation proportionnelle des prises en charge pour les complémentaires santé.

<b>Ensemble de la CSBM (en %)</b>	<b>2008</b>	<b>2011</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>
<i>Sécurité sociale</i>	76,2	76,2	77,8	78,1	78,2
<i>Etat, CMU-C org. de base</i>	1,3	1,3	1,5	1,5	1,5
<i>Organismes complémentaires</i>	13,2	13,4	13,2	13,4	13,4
<i>Ménages</i>	9,4	9,2	7,5	7,0	6,9

Figure 1 – Structure de financement de la CSBM – source DREES

Ainsi, le montant remboursé par les complémentaires santé est passé d'environ 22 milliards d'euros en 2008 à près de 28 milliards d'euros en 2019.

<sup>1</sup> [https://drees.solidarites-sante.gouv.fr/IMG/pdf/cns\\_2020\\_29092020.pdf](https://drees.solidarites-sante.gouv.fr/IMG/pdf/cns_2020_29092020.pdf)



Nous présenterons dans un premier temps le régime obligatoire et en particulier le fonctionnement de la branche maladie qui nous intéresse dans ce mémoire puis, dans un deuxième temps, nous présenterons le marché des complémentaires santé, en commençant par le détail des acteurs de la branche santé et en précisant de manière synthétique le cadre juridique et réglementaire.

### 1.1. Le régime obligatoire

La création de la Sécurité Sociale fait suite à l'ordonnance du 4 octobre 1945 dont quelques extraits sont repris ci-dessous :

« La Sécurité Sociale est la garantie donnée à chacun qu'en toutes circonstances il disposera des moyens nécessaires pour assurer sa subsistance et celle de sa famille dans des conditions décentes. Trouvant sa justification dans un souci élémentaire de justice sociale, elle répond à la préoccupation de débarrasser les travailleurs de l'incertitude du lendemain, de cette incertitude constante qui crée chez eux un sentiment d'infériorité et qui est la base réelle et profonde de la distinction des classes entre les possédants sûrs d'eux-mêmes et de leur avenir et les travailleurs sur qui pèse, à tout moment, la menace de la misère. [...] Envisagée sous cet angle, la Sécurité Sociale appelle l'aménagement d'une vaste organisation nationale d'entraide obligatoire qui ne peut atteindre sa pleine efficacité que si elle présente un caractère de très grande généralité à la fois quant aux personnes qu'elle englobe et quant aux risques qu'elle couvre. Le but à atteindre est la réalisation d'un plan qui couvre l'ensemble de la population du pays contre l'ensemble des facteurs d'insécurité ; un tel résultat ne s'obtiendra qu'au prix de longues années d'efforts persévérants, mais ce qu'il est possible de faire aujourd'hui, c'est d'organiser le cadre dans lequel se réalisera progressivement ce plan. »

La création de la Sécurité Sociale prévoyait un réseau coordonné de caisses se substituant à de multiples organismes en place à l'époque. Cependant, l'unité administrative ne sera pas réalisée et ne l'est toujours pas. Les professions agricoles vont conserver leurs institutions spécifiques dans le cadre de la Mutualité Sociale Agricole. Ainsi, les salariés des régimes spéciaux vont refuser de s'intégrer dans le régime général et conserver dans un cadre " transitoire " qui dure encore, leurs régimes spéciaux (les fonctionnaires qui peuvent relever de différents statuts en fonction de leur type de rattachement (état, territoire, activité, etc.), les marins, les cheminots, les mineurs, etc.). Il existe aujourd'hui une douzaine de régimes spéciaux.

Le financement de la Sécurité Sociale s'est aussi modifié depuis 1945. Bien que les cotisations assises sur la masse salariale et composées d'une part salariale et d'une part patronale représentent encore la principale ressource des régimes, la part des autres recettes telles que les taxes fiscales, la contribution sociale généralisée (CSG) assise sur l'ensemble des revenus d'activité, de remplacement, du patrimoine et du capital (instituée par la loi du 22 juillet 1993), la contribution sociale de solidarité à la charge des entreprises et enfin la contribution au remboursement de la dette sociale (CRDS), reposant sur la même assiette que la CSG, instituée par l'ordonnance du 24 janvier 1996 croît rapidement.

Les risques couverts par le régime général de la Sécurité Sociale sont découpés en cinq branches :

- l'assurance maladie,
- les accidents de travail et maladies professionnelles,
- l'assurance retraite,
- l'assurance famille,
- la branche recouvrement.

Il est à noter que dans le cadre du PLFSS 2020, une sixième branche relative à la prise en charge dépendance vient d'être créé.

L'objet de ce mémoire ne portant que sur la complémentaire santé, nous ne détaillerons que l'assurance maladie par la suite.

## 1.2. La branche maladie

Après avoir rappelé le fonctionnement général de l'assurance maladie, nous présenterons en détail les remboursements du Régime Obligatoire (RO) sur lesquels s'adosent les produits de complémentaire santé dans leur fonctionnement et leurs valeurs. Le fonctionnement et les particularités des régimes spéciaux ne seront pas abordés dans ce mémoire.

### 1.2.1. Généralités

La branche maladie de la Sécurité Sociale assure la prise en charge des dépenses de santé des assurés malades et garantit l'accès aux soins. Elle mène également différents programmes de prévention et favorise l'accès à la santé des plus démunis grâce à son action sociale. Elle contribue également au fonctionnement d'établissements médico-sociaux.

Dans la plupart des régimes, la branche maladie gère aussi, lorsqu'ils existent, les risques :

- maternité : prise en charge des examens pré et postnataux et accorde des indemnités journalières pendant le congé de maternité
- invalidité : octroi d'une pension lorsque l'assuré présente une invalidité réduisant sa capacité de travail ou de gains dans des proportions déterminées.
- décès : paiement d'un capital égal à un multiple de gain journalier de base en cas de décès de l'assuré.

La branche maladie est gérée principalement par les trois principaux régimes d'assurance maladie :

- le régime général, géré par la CNAMTS (Caisse Nationale d'Assurance Maladie des Travailleurs Salariés) qui pilote le réseau des CPAM (Caisses Primaires d'Assurance Maladie),
- le régime agricole (MSA - Mutualité Sociale Agricole),
- le Régime Social des Indépendants (RSI),

tous trois regroupés au sein de l'Union nationale des caisses d'assurance maladie (Uncam).

Il est à noter que les travailleurs indépendants ne cotisent plus au RSI depuis le 1<sup>er</sup> janvier 2018 et sont soumis désormais au régime général.

Il existe à l'intérieur du régime général une spécificité pour la région Alsace-Moselle. Le régime local d'assurance maladie d'Alsace-Moselle est issu de l'histoire des départements du Bas-Rhin, du Haut-Rhin et de la Moselle. Pendant le rattachement de ces territoires à l'Empire Allemand, de 1871 à 1918, la population a bénéficié du système d'assurance maladie des lois de Bismarck. Ces lois allemandes assuraient une très forte socialisation de la prise en charge des dépenses de soins, laissant une fraction modeste à la charge des assurés. Ces lois ont été maintenues en vigueur jusqu'en 1946. Après la Seconde Guerre mondiale, lors de la création du régime général de Sécurité Sociale, la population locale s'est mobilisée pour conserver son régime particulier. Un décret du 12 juin 1946 l'a maintenu à titre provisoire en attendant que le régime général s'aligne sur son haut niveau de solidarité. Cette perspective ne s'est pas réalisée. Le Régime Local a été pérennisé par une loi du 31 décembre 1991.

Le Régime Local d'assurance maladie d'Alsace-Moselle verse à ses bénéficiaires un complément de remboursement, en plus de ce que le régime de base de la Sécurité Sociale prend en charge (hospitalisation, pharmacie et soins courants). Le Régime Local finance également des actions de prévention en santé publique pour lutter contre deux pathologies graves qui touchent particulièrement l'Alsace et la Moselle : les cancers et les maladies cardiovasculaires.

L'assuré ouvre droits aux prestations en nature de l'assurance maladie et à la maternité à :

- son conjoint, concubin ou pacsé lorsqu'il ne bénéficie pas d'un régime de protection sociale.
- les enfants à sa charge ou à la charge de son conjoint, concubin, pacsé jusqu'à 20 ans maximum.
- toute personne à la charge effective et permanente de l'assuré ne bénéficiant pas d'un régime de protection sociale.

L'assurance maladie reste marquée par son histoire et le régime dont dépend un assuré dépend de son activité professionnelle présente ou passée. Toutefois, pour les personnes ne remplissant pas les conditions d'affiliation à un régime sur la base professionnelle mais résidant en France depuis au moins trois mois en situation régulière, il existe une couverture maladie universelle.

Les prestations prise en charge par l'assurance maladie sont de deux types :

- soit des prestations en espèces,
- soit des prestations en nature.

Les prestations en espèces compensent la perte d'un revenu pour les personnes devant cesser leur activité professionnelle pour raison de santé. Ces revenus de remplacement correspondent :

- aux indemnités journalières en cas d'arrêt maladie ou durant les congés maternité et paternité, à un capital en cas de décès,
- à une pension d'invalidité, lorsque l'assuré présente une invalidité réduisant sa capacité de travail ou de gains dans des proportions déterminées.

Il est à noter que ces prestations ne font pas l'objet de ce mémoire.

Les prestations en nature correspondent au remboursement des frais de santé suivants :

- les frais de médecine générale et spécialisée,
- les frais de soins et de prothèses dentaires,
- les frais pharmaceutiques et d'appareillage,
- les frais d'analyses et d'examens de laboratoire,
- les frais afférents aux vaccinations dont la liste est fixée par arrêté,
- les frais relatifs aux examens de dépistage effectué dans le cadre de programmes de santé publique,
- les frais d'hébergement et de traitement des enfants ou adolescents handicapés dans les établissements d'éducation spéciale et professionnelle,
- les frais de transport des malades dans des conditions et limites tenant compte de l'état du malade et du coût du transport.

### 1.2.2. Principes de remboursements

L'assurance maladie intervient sur la base de tarifs fixés par convention ou d'autorité. Cette base est appelée base de remboursement (BR) et est définie pour chaque acte. A ces bases de remboursements s'associent des taux de remboursements. Le remboursement de la Sécurité Sociale correspondra au produit entre la base de remboursement

et le tarif de convention. Des exemples de bases de remboursements et de taux de remboursements sont présentés ci-dessous.

*Tarifs de consultations du médecin traitant (tarifs applicables au 1<sup>er</sup> juillet 2017)*

<b>Médecin consulté</b>	<b>Tarif*</b>	<b>Base du remboursement</b>	<b>Taux de remboursement</b>	<b>Montant remboursé RSS**</b>
<b>Généraliste</b>				
<i>Secteur 1</i>	25,00 €	25,00 €	70%	16,50 €
<i>Adhérent à l'option de pratique tarifaire maîtrisée</i>	Honoraires avec dépassement maîtrisé	25,00 €	70%	16,50 €
<i>Secteur 2</i>	Honoraires libres	23,00 €	70%	15,10 €
<b>Spécialiste</b>				
<i>Secteur 1</i>	25,00 €	25,00 €	70%	16,50 €
<i>Adhérent à l'option de pratique tarifaire maîtrisée</i>	Honoraires avec dépassement maîtrisé	25,00 €	70%	16,50 €
<i>Secteur 2</i>	Honoraires libres	23,00 €	70%	15,10 €
<b>Psychiatre Neuropsychiatre Neurologue</b>				
<i>Secteur 1</i>	41,70 €	41,70 €	70%	28,19 €
<i>Adhérent à l'option de pratique tarifaire maîtrisée</i>	Honoraires avec dépassement maîtrisé	41,70 €	70%	28,19 €
<i>Secteur 2</i>	Honoraires libres	39,00 €	70%	26,30 €

\*Le tarif prend en compte la consultation et les éventuelles majorations appliquées par le médecin.

\*\*Les montants remboursés indiqués tiennent compte de la participation forfaitaire de 1 € retenue sur chaque consultation ou acte réalisé par un médecin, sauf pour les personnes de moins de 18 ans, les femmes enceintes à partir du 6<sup>ème</sup> mois de grossesse jusqu'à 12 jours après l'accouchement, les bénéficiaires de la CMU complémentaire, d'une aide au paiement d'une complémentaire santé (ACS) ou de l'aide médicale de l'État (AME).

Figure 2 - Exemple de bases et taux de remboursements pour médecins généralistes et spécialistes<sup>2</sup>

La partie (%) de la BR qui n'est pas remboursée par la Sécurité Sociale et qui est appelée Ticket Modérateur (TM) et est laissée à la charge de l'assuré (pouvant être couvert par la complémentaire santé). Il est à noter que les taux de remboursements appliqués sur le régime local Alsace-Moselle sont plus importants sur certains postes.

La participation forfaitaire correspond à un reste à charge obligatoire pour le patient. Ainsi, la participation forfaitaire de 1 euro s'applique :

- pour toute consultation ou acte réalisé par un médecin généraliste ou spécialiste : que vous respectiez le parcours de soins ou non, que la consultation ait lieu à son cabinet, à votre domicile, dans un dispensaire, dans un centre de soins, en consultation externe ou aux urgences à l'hôpital,
- lors d'examens de radiologie,
- lors d'analyses de biologie médicale.

La franchise médicale existe aussi sur d'autres postes comme par exemple :

- 0,50 euros par boîte de médicaments,
- 0,50 euros par acte paramédical (infirmiers, masseurs kinésithérapeutes, orthoptistes, orthophonistes et les pédicures/podologues) limité à 2 euros maximum par jour par acte paramédical,
- 2 euros par transport réalisé en transport sanitaire limité à 4 euros par jour.

<sup>2</sup> <https://www.ameli.fr/paris/assure/remboursements/rembourse/consultations/metropole>

Ce reste à charge est limité à 50 euros annuel par l'assurance maladie<sup>3</sup>.

Les conditions de prise en charge sont définies par le respect du parcours du soins. Le taux de remboursement du régime obligatoire est réduit lorsque le parcours de soins n'est pas respecté :

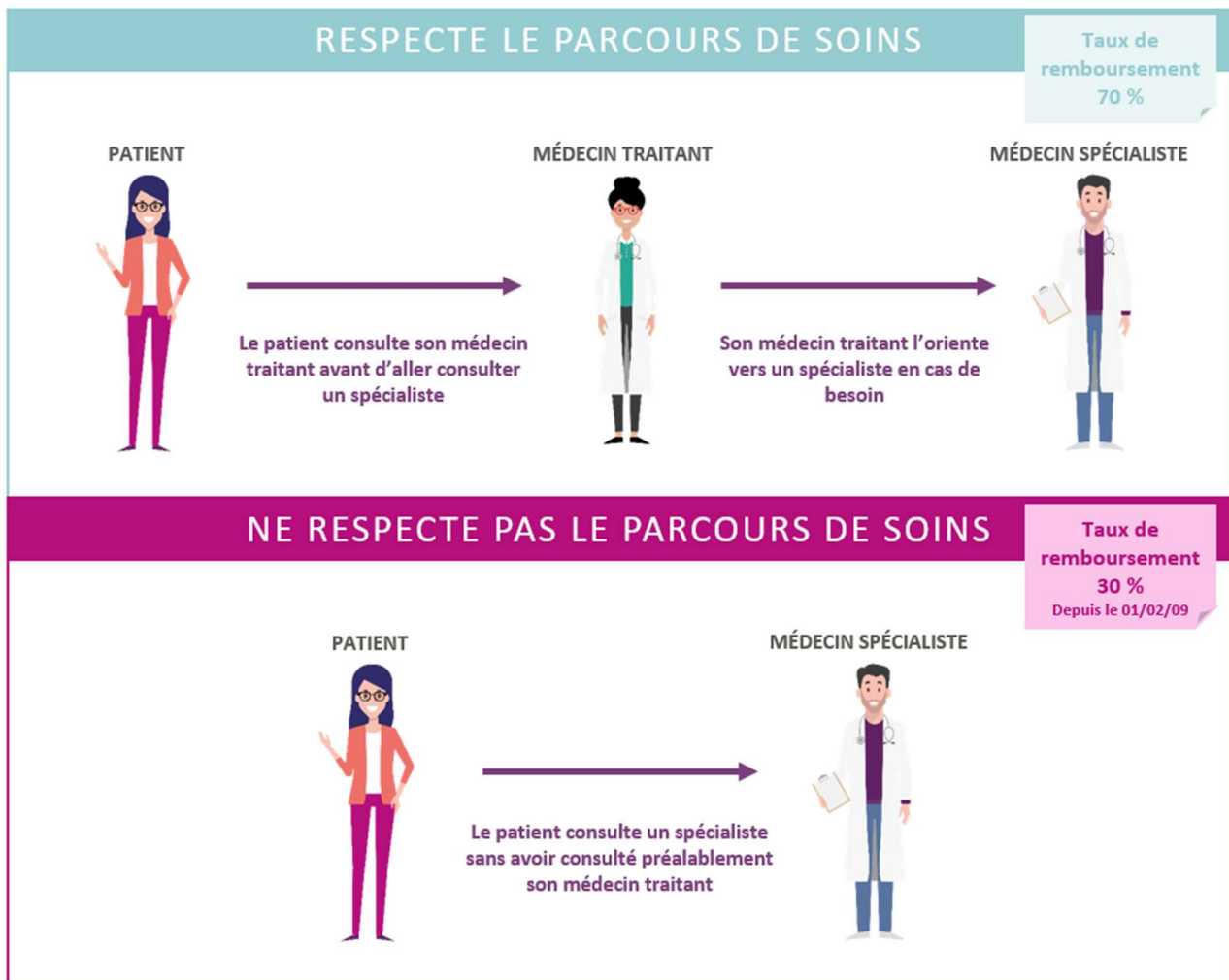


Figure 3 - Le parcours de soins

Il existe un certain nombre d'exonérations possible du ticket modérateur. Les principales sont reprises ci-dessous :

- les frais de santé médicaux remboursables (pharmaceutiques, d'analyses, d'examens de laboratoire, d'hospitalisation), en rapport ou non avec la grossesse, à partir du 6ème mois de grossesse jusqu'au 12e jour après l'accouchement,
- les frais d'hospitalisation ainsi que tous les soins dispensés en établissement de santé pour les nouveau-nés, dans les 30 jours suivant leur naissance,
- les soins liés à un accident du travail ou une maladie professionnelle,
- les soins et traitements visés par le protocole de soins établi pour une affection de longue durée,
- la plupart des soins médicaux, examens et médicaments pour la maladie et la maternité des personnes titulaires d'une pension d'invalidité,
- les soins en lien direct avec un acte de terrorisme pour les victimes figurant sur la liste officielle des victimes établie par le Parquet de Paris,
- les frais de santé des personnes bénéficiaires de la législation des pensions militaires d'invalidité.

<sup>3</sup> Les contrats de complémentaire santé dit responsables ne peuvent prendre en charge cette participation – voir I-1.4)

Ainsi, le principe de remboursement peut se schématiser comme suit :

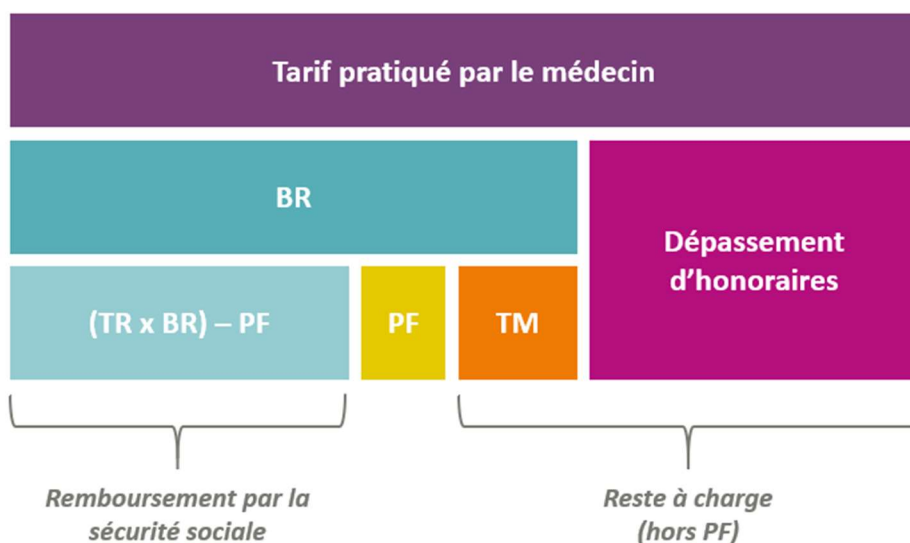


Figure 4 - Principe de remboursement de la Sécurité Sociale

Un exemple de remboursement est présenté à la figure 5 ci-dessous :

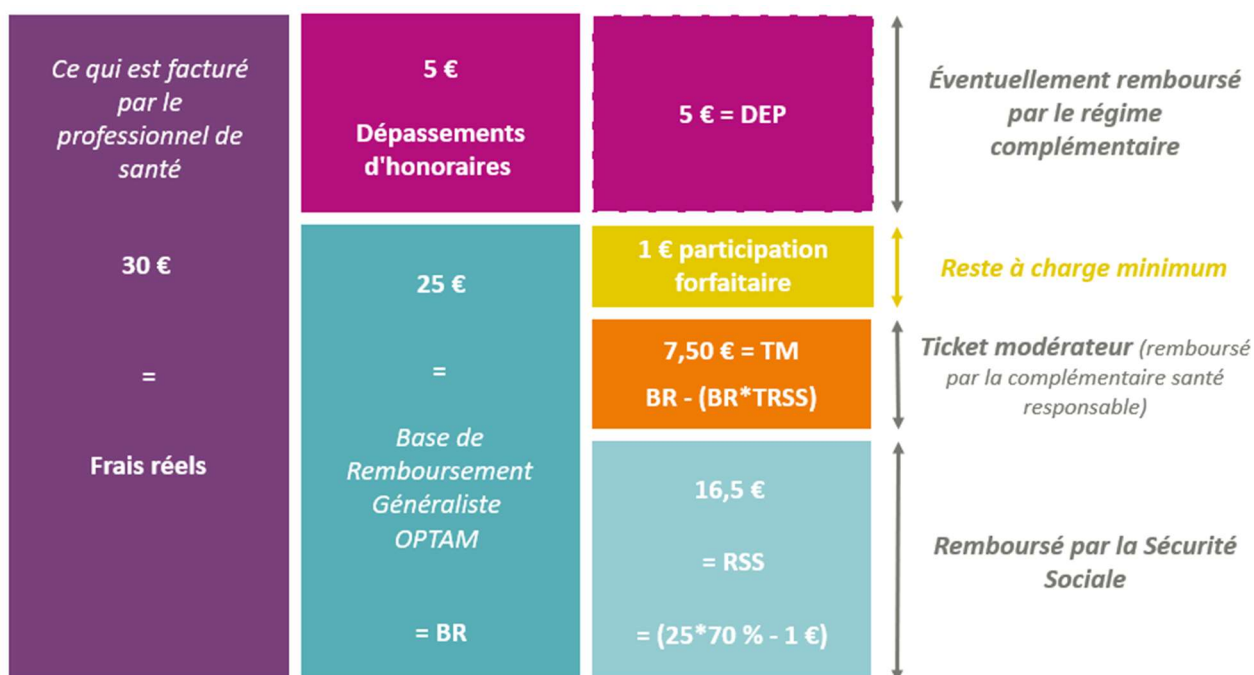


Figure 5 - Exemple de remboursement

### 1.2.3. Quelques chiffres

Le tableau ci-dessous (figure 6), représente la répartition en milliards d'euros de la dépense par grand poste pour chacun de ses financeurs en 2019.

En millions d'euros

Financiers	Publics		Privés				TOTAL
	Sécurité sociale*	Etat, collectivités locales et CMU-C organismes de base	Mutuelles**	Sociétés d'assurances**	Institutions de prévoyance**	Ménages	
<b>Postes de dépense</b>							
<b>1 - Soins hospitaliers</b>	<b>89 010</b>	<b>1 152</b>	<b>2 685</b>	<b>1 583</b>	<b>735</b>	<b>1 961</b>	<b>97 127</b>
Hôpitaux du secteur public	69 360	1 031	1 947	1 092	500	962	74 892
Hôpitaux du secteur privé	19 650	121	738	491	235	999	22 234
<b>2 - Soins ambulatoires</b>	<b>73 699</b>	<b>1 952</b>	<b>10 871</b>	<b>7 337</b>	<b>4 709</b>	<b>12 340</b>	<b>110 908</b>
Soins de ville	37 208	1 256	5 755	3 852	2 527	5 792	56 389
Soins de médecins et de sages-femmes***	16 473	431	1 950	1 463	842	1 884	23 045
Soins d'auxiliaires médicaux**	13 510	173	1 126	562	319	1 342	17 034
Soins de dentistes	4 003	550	2 137	1 460	1 156	2 481	11 786
Laboratoires d'analyses	3 222	102	541	367	210	85	4 525
Contrats et autres soins	0	0	0	0	0	0	0
Transports sanitaires	4 695	49	122	55	18	111	5 051
Médicaments en ambulatoire	24 220	477	2 093	1 245	687	3 860	32 592
Autres biens médicaux****	7 498	170	2 901	2 185	1 467	2 576	16 798
<b>Consommation de soins et de biens médicaux (1+2)</b>	<b>162 708</b>	<b>3 104</b>	<b>13 557</b>	<b>8 920</b>	<b>5 445</b>	<b>14 301</b>	<b>2 018 035</b>

Figure 6 - Structure de la CSBM et de la dépense de chacun de ses financeurs en 2019

Il est intéressant de constater que pour certains postes, lorsque la part de remboursement de la dépense est faible pour la Sécurité Sociale, la part de remboursement des organismes complémentaires ainsi que la part dans le reste à charge des ménages est important (exemple pour les « soins de dentistes »).

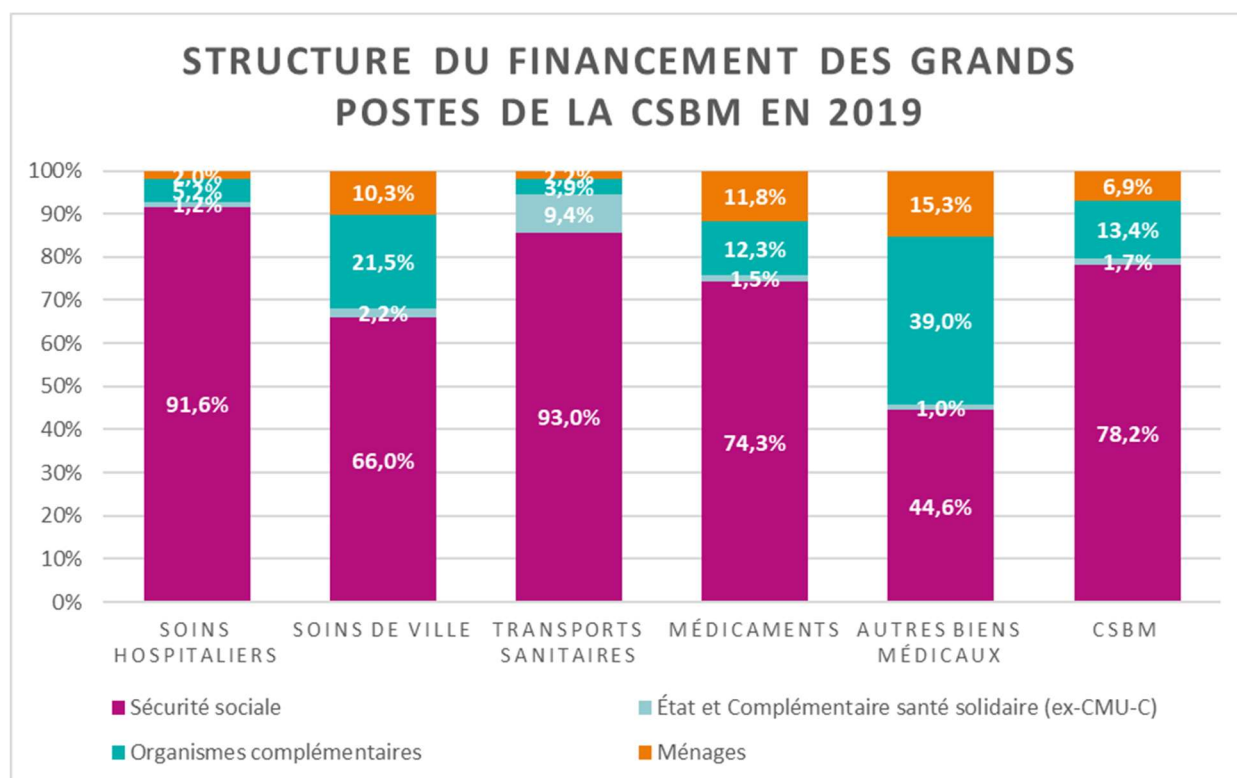


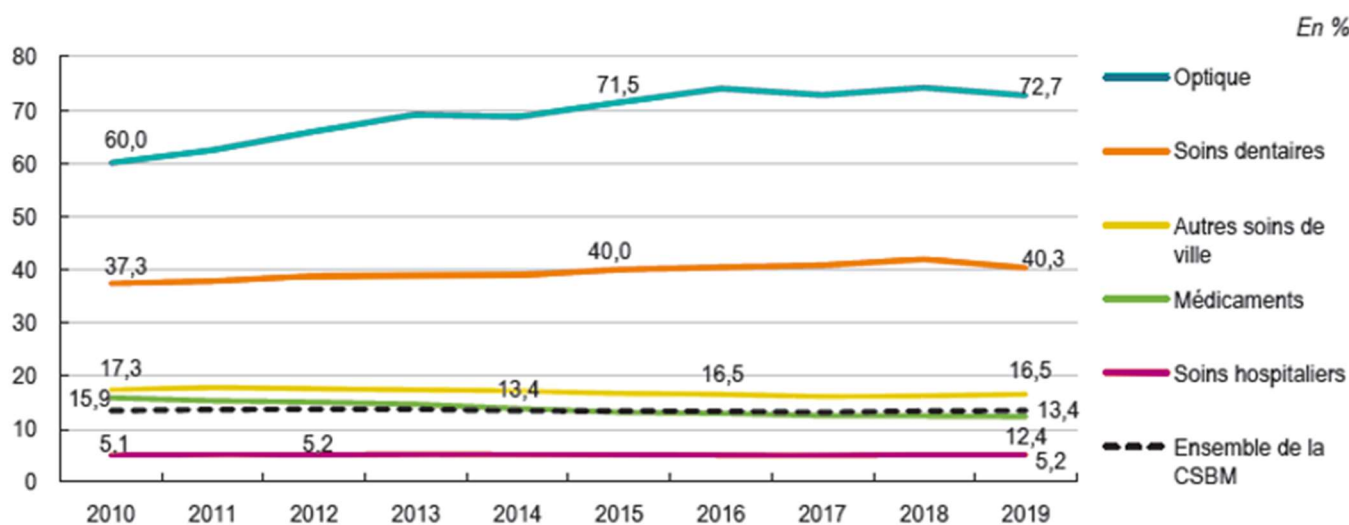
Figure 7 - Structure du financement des grands postes de la CSBM en 2019

Comme évoqué dans les éléments du tableau précédent (figure 6), la figure 7 met en évidence les postes pour lesquels la Sécurité Sociale intervient de manière importante et les postes (« Autres biens médicaux ») où sa participation est plus limitée (part des dépassements importante). Ainsi, les soins hospitaliers sont très largement financés par le régime obligatoire (à hauteur de 91%). En revanche, sur les soins de ville (incluant le dentaire) et les autres biens médicaux

(optique), le financement est proche de 60% laissant ainsi, la place aux organismes complémentaires d'assurance maladie (OCAM). La réforme du 100% santé, intégrant la prise en charge de biens médicaux sans reste à charge sur les postes dentaire, audioprothèse et optique (voir Partie I.1.4)) a pour but de modifier les constats précédents (moins de reste à charge sur les postes où il est potentiellement important aujourd'hui).

### 1.3. La complémentaire santé et ses acteurs

Comme présenté aux figures 6 et 7, l'analyse de la dépense par poste montre que le régime obligatoire est peu engagé sur des postes avec des dépenses significatives. Ainsi, sur les postes dentaire et optique, les graphiques suivants présentent la répartition de la dépense entre les ménages, le régime obligatoire et le régime complémentaire.



**Lecture** > Les prestations versées par les organismes complémentaires représentent 72,7 % de la consommation de soins et de biens médicaux en optique en 2019.

**Source** > DREES, comptes de la santé

Figure 8 - Part des organismes complémentaires dans les principaux postes de 2010 à 2019

La figure 8 met en évidence la prépondérance des complémentaires santé dans les remboursements en optique et dans une moindre mesure sur le poste dentaire. En plus de compléter les prestations de la Sécurité Sociale sur les grands postes présentés ci-dessus, les complémentaires santé peuvent prendre en charge un certain nombre de prestations non remboursées par le régime général. Ainsi, en hospitalisation, en dehors des prestations en complément de la Sécurité Sociale tels les frais de séjour, les honoraires de chirurgie (par exemple), des prestations dites de confort comme le lit d'accompagnement et la chambre particulière peuvent être pris en charge. En optique, les opérations de la kératotomie et les lentilles sont souvent remboursées par les complémentaires santé. Enfin, la prise en charge d'autres prestations comme la médecine douce (ostéopathie, chiropractie, etc.) peut être proposée.

Trois différents types d'organismes proposent des complémentaires santé, et sont soumises à des règles juridiques spécifiques :

- les compagnies d'assurance et les sociétés d'assurances mutuelles :  
Elles sont régies par le code des Assurances. Les compagnies d'assurance sont à but lucratif et les sociétés d'assurances mutuelles sont à but non lucratif. En outre, les assurés des sociétés d'assurance mutualistes sont appelés sociétaires et peuvent voter pour leurs représentants en assemblée générale.
- les institutions de prévoyance :



Elles n'assurent que les salariés, les anciens salariés et leurs ayants droits éventuels. Elles sont gérées de manière paritaire par les représentants des entreprises souscriptrices (partenaires sociaux). Elles sont régies par le code de la Sécurité Sociale et sont à but non lucratif.

- les mutuelles :

Elles sont régies par le code de la mutualité et sont à but non lucratif. L'assemblée générale est constituée des assurés de la mutuelle.

En 2019, 49% des prestations des organismes complémentaires ont été versées par des mutuelles, 32% par des sociétés d'assurances et 20% par des institutions de prévoyance (figure 9). Les mutuelles perdent des parts de marché depuis plusieurs années au profit principalement des sociétés d'assurances. En 2016, la généralisation de la complémentaire santé d'entreprise a plutôt profité à ces dernières, qui ont nettement accru leur part de marché de 2016 à 2019.

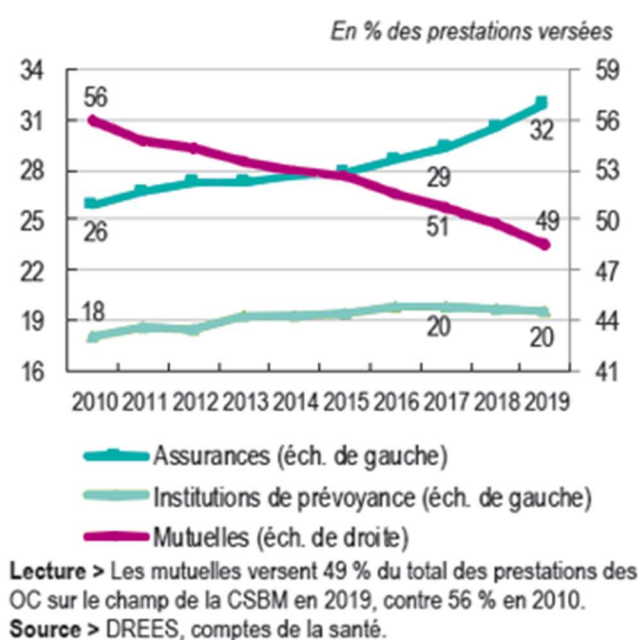


Figure 9 - Part de marché par type d'organismes entre 2010 et 2019

La part de marché des mutuelles est passée en une dizaine d'année de près de 56% à moins de 50%.

Les cotisations collectées sont la principale ressource des organismes complémentaires. La plus grande partie de ces cotisations est reversée aux assurés sous forme de prestations. Le reste regroupe principalement les frais de fonctionnement des organismes. Au cours du temps, la répartition du marché de la complémentaire santé a évolué comme suit :

	2006	2016	2017	Variations 2017 vs 2006
<b>Nombre d'organismes d'assurance</b>				
Sociétés d'assurance et succursales de pays tiers	378	277	271	-107
Sociétés de réassurance	29	14	14	-15
Institutions de prévoyance	66	37	36	-30
Mutuelles livre II - pratiquant des opérations d'assurance	1 158	446	421	-737
<b>Total</b>	<b>1 631</b>	<b>774</b>	<b>742</b>	<b>-889</b>

Figure 10 - Evolution du nombre d'organismes sur le marché de l'assurance entre 2006 et 2017

Parmi les organismes contrôlés par l’Autorité de Contrôle Prudentiel et de Résolution (ACPR), il y avait en 2017, 285 sociétés régies par le code des assurances, 36 institutions de prévoyance régies par le code de la Sécurité Sociale et 421 mutuelles régies par le code de la mutualité dont 111 mutuelles substituées (mutuelles ayant transféré tout ou partie de son risque assurantiel à un autre organisme, tout en conservant une autonomie juridique et la relation avec ses adhérents). Cela étant, depuis la réforme de la substitution lors de la dernière mise à jour du code de la Mutualité, l’autonomie des mutuelles substituées a été plus limitée. Le nombre d’organismes continue de diminuer, avec un mouvement de concentration des organismes qui concerne particulièrement les mutuelles. En 2017, les organismes soumis au code de la mutualité sont au nombre de 421 soit 737 de moins qu’en 2006 (-64 %), ce qui représente la grande majorité, relativement à la baisse totale de 889 organismes. Cette diminution est en grande partie due aux nombreuses fusions et absorptions.

#### 1.4. Le cadre juridique de la complémentaire santé

Le contexte législatif et réglementaire encadrant l’activité d’assurance des produits de complémentaire santé est riche. Cette partie présente les principaux textes ou principes fondamentaux.

##### 1.4.1. Les différents types de régime de complémentaire santé

La couverture d’assurance proposée est soit individuelle soit collective. Un contrat de complémentaire santé individuel est souscrit directement par un particulier, tandis qu’un contrat collectif est souscrit par une personne morale (entreprise ou association) au profit d’un groupe de salariés ou d’adhérents. En 2019, un peu moins de la moitié des prestations servies par les organismes complémentaires sur le champ de la CSBM l’ont été au titre de contrats individuels (47%), le reste des prestations servies ayant été réalisées dans le cadre de contrats collectifs.

Dans un régime individuel, l’assuré souscrit un contrat qui le lie directement à l’organisme d’assurance santé. Dans un régime collectif, le souscripteur est l’entreprise ou l’association. Les assurés sont les salariés ou adhérents (dans le cadre d’une association), et éventuellement leurs ayants droits. Si la couverture est majoritairement obligatoire dans le cadre d’un contrat d’entreprise (les salariés ont l’obligation d’adhérer au régime proposé par l’entreprise sauf s’il justifie une exclusion autorisée telle que la couverture via un conjoint par exemple), elle est facultative dans le cadre des associations.

Les garanties des contrats collectifs sont en général plus avantageuses pour les assurés, notamment en optique et en dentaire. Ainsi, pour 100 euros de cotisation, les contrats collectifs reversent en moyenne 17 euros en optique et 16 euros en dentaire, contre respectivement 8 et 9 euros en individuel (figure 11). Les contrats individuels reversent en revanche plus de prestations en soins hospitaliers et médicaments que les contrats collectifs, car les retraités, plus consommateurs de ces deux types de soins, sont couverts par des contrats individuels, tandis que les contrats collectifs couvrent principalement des salariés d’entreprises et leurs ayants droit. En collectif, la meilleure mutualisation permet la souscription de garanties plus haut de gamme. En individuel, l’âge des adhérents étant plus élevé, le recours à des postes comme l’hospitalisation est plus marqué.

En outre, les contrats collectifs peuvent contenir différents niveaux de garanties, souvent un socle de garantie obligatoire et des options facultatives dont l’adhésion est laissée au choix des salariés.

Si les contrats collectifs obligatoire réduisent fortement l’antisélection<sup>4</sup>, les régimes facultatifs (options ou contrats d’associations) au contraire sont souscrites par des salariés ayant souvent la nécessité de consommer.

Il est à noter, compte tenu des économies d’échelle, que les contrats collectifs (obligatoires en particulier) permettent de reverser une plus grande part des cotisations sous forme de prestations que les contrats individuels (meilleure mutualisation des risques

---

<sup>4</sup> L’antisélection en assurance santé correspond au fait que l’adhérent ayant une connaissance de ses besoins plus précise que l’assureur, il choisira une garantie plus en adéquation avec ses besoins. En collectif avec adhésion obligatoire, l’antisélection est plus limitée.

et des coûts de gestion). Ce point est explicité sur la figure 12 qui présente la répartition par grand poste des prestations versées en fonction du type de contrat.

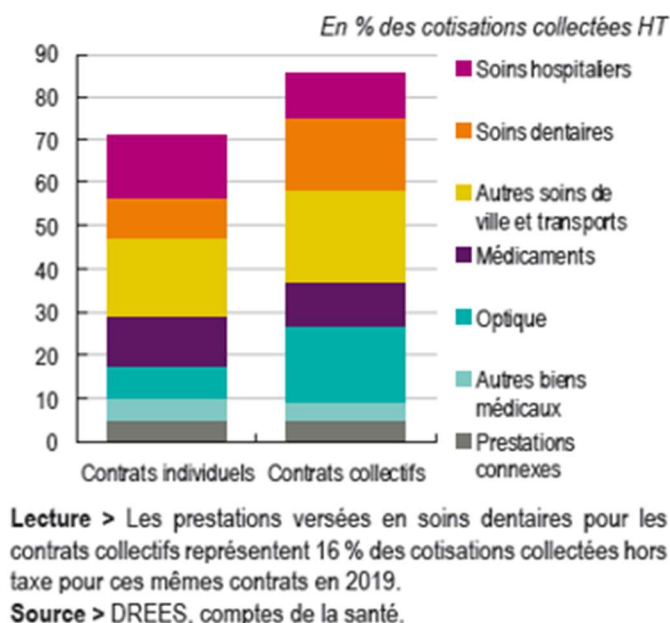


Figure 11 - Prestations versées selon le type de soins et de contrat en 2019

#### 1.4.2. Contexte législatif de la complémentaire santé : les principaux textes

Nous allons présenter dans l'ordre chronologique les principaux textes encadrant l'activité de complémentaire santé. Cette présentation n'est évidemment pas exhaustive.

##### LA LOI EVIN

La loi N°89-1009 du 31 décembre 1989, dite Loi Evin, est une loi dont l'objectif était de renforcer les garanties offertes aux personnes assurées contre certains risques. Elle est le texte fondamental qui régit la complémentaire santé en France. Les deux principaux articles ayant des incidences sur l'activité d'assurance sont les articles 4 et 6.

L'article 4 de la loi du 31 décembre 1989 impose aux organismes assureurs d'organiser les modalités de maintien de la complémentaire santé, afin de permettre aux anciens salariés bénéficiaires d'une garantie collective de conserver leur couverture complémentaire à un tarif encadré. L'article 1er du décret du 30 août 1990 prévoyait que les tarifs ne pouvaient être supérieurs de plus de 50 % aux tarifs globaux applicables aux salariés actifs. Le décret numéro 2017-372 du 21 mars 2017 modifie cette tarification en organisant un plafonnement progressif des tarifs, échelonné sur trois ans. La première année qui fait suite à la sortie du contrat d'assurance collectif, il est prévu que les tarifs restent identiques aux tarifs globaux applicables aux salariés actifs. La deuxième année, il est prévu que ces tarifs ne peuvent être supérieurs de plus de 25 % aux tarifs globaux applicables aux salariés actifs. La troisième année qui suit la sortie du contrat d'assurance collectif, ces tarifs ne peuvent être supérieurs de plus de 50 % aux tarifs globaux applicables aux salariés actifs.

L'article impose aux assureurs ayant mis en place un contrat collectif facultatif ou individuel de maintenir la garantie aux assurés tant que les cotisations sont versées. Il indique également que les évolutions de tarifs concernant ces types

de régime ne peuvent être basées sur l'évolution de l'état de santé de l'individu mais doivent être les mêmes pour tous les assurés du contrat collectif ou de groupe de contrats individuels<sup>5</sup>.

L'article 6 impose notamment de ne pas exclure, résilier ou tarifer différemment les adhérents en fonction de leur état de santé sur une garantie ou un contrat donné.

### LES CONTRATS RESPONSABLES ET SOLIDAIRES

Depuis la loi n°2004-810 du 13 août 2004 relative à l'assurance maladie, les contrats de complémentaire santé peuvent bénéficier d'avantages fiscaux et sociaux, sous conditions du respect des règles des contrats responsables définies à l'article L.871-1 du code de la Sécurité Sociale et par le décret (n°2005-1226 du 29 septembre 2005) en vigueur depuis le 1er janvier 2006.

Ainsi, pour que le contrat soit responsable, il fallait répondre à certaines conditions définies par la Sécurité Sociale :

- respecter l'ensemble des mesures liées au parcours de soins coordonné autour du médecin traitant : ne pas prendre en charge les pénalités financières en cas de non-choix du médecin traitant et de consultation hors parcours de soins,
- garantir de bons niveaux de prise en charge sur les prestations effectuées dans le cadre du parcours de soins : le remboursement des consultations et des prescriptions de médecins généralistes et spécialistes, des médicaments à vignettes blanches ou encore des analyses médicales,
- prendre en charge au minimum deux actes de prévention parmi une liste,
- ne pas prendre en charge la participation forfaitaire de l'assuré,
- ne pas prendre en charge la franchise applicable pour tous les médicaments (actuellement 0,50 euros par boîte de médicament).

Dans un objectif d'amélioration de l'accès aux soins et de renforcement de la régulation, les pouvoirs publics ont fait le choix de faire évoluer les règles. Ainsi, la loi de Financement de la Sécurité Sociale (LFSS) pour 2014, amendée par la loi de financement rectificative de la Sécurité Sociale pour 2014 (LFSSR), instaure un renforcement des règles. Le décret n° 2014-1374 du 18 novembre 2014, relatif au contenu des contrats d'assurance maladie complémentaire bénéficiant d'aides fiscales et sociales, précise les modalités de prises en charge que les contrats devront respecter pour être qualifiés de responsables.

Ainsi, les nouvelles règles fixées sont les suivantes :

- la prise en charge obligatoire à 100% du ticket modérateur (TM) pour les actes pris en charge par l'assurance maladie (à l'exception des médicaments ayant un taux de remboursement à 15% et 30% par l'assurance maladie et l'homéopathie),
- les frais d'hospitalisation, avec ou sans hébergement y compris les consultations et les actes externes sont obligatoirement couverts ainsi que l'intégralité du forfait journalier des établissements de santé,
- limitation des dépassements d'honoraires avec l'introduction du contrat d'accès aux soins (CAS) pour les médecins (devenu OPTAM par la suite) limitant les remboursements des complémentaires santé pour les médecins non-signataires à 100% du tarif opposable (125% en 2015 et 2016) et imposant un écart minimum de -20% BR avec les médecins adhérents au CAS.
- limitation des garanties en optique proposées par les complémentaires :

---

<sup>5</sup> <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000034258687&categorieLien=id>

si l'obligation de prendre en charge le TM (très faible en optique) a été introduite, les remboursements sont limités à un équipement tous les 2 ans à l'exception des mineurs ou en cas d'évolution de la vue et sont compris entre un minimum obligatoire (incluant le TM) et un maximum autorisé en fonction des déficits de la vision.

Ainsi, des postes de dépenses, jusqu'à aujourd'hui non concernés, font l'objet d'une obligation de prise en charge minimale et d'autres champs de prestations, dont la prise en charge reste optionnelle, devront respecter des plafonds et/ou des planchers de prise en charge. L'ensemble des contrats complémentaires santé, individuels et collectifs, sont concernés par ces nouvelles règles. S'agissant des contrats collectifs obligatoires, ces règles sont à articuler avec celles du panier minimum des garanties qui s'appliquent aux salariés bénéficiant d'une couverture santé complémentaire à adhésion obligatoire, à compter du 1er janvier 2016, et dont les règles sont définies par le décret n° 2014-1025 du 8 septembre 2014.

Depuis 2019, les contrats responsables ont l'obligation de respecter la réforme du 100% Santé.

#### **Remarque sur les taxes appliquées aux contrats de complémentaires santé :**

Les contrats de complémentaire santé ne sont plus exonérés de taxes sur les contrats d'assurance depuis 2011. Le taux de la taxe est toutefois réduit si le contrat est solidaire est responsable (taxe de 7% au lieu de 14%). En outre, la contribution CMU a été transformée en taxe en 2011 (taxe de 6,27%). En 2017, les contrats santé responsables sont soumis à une taxe de 13,27 %. Les contrats santé non responsables sont soumis à une taxe de 20,27%.

#### **L'ACCORD NATIONAL INTERPROFESSIONNEL DU 11 JANVIER 2013**

L'Accord National Interprofessionnel du 11 janvier 2013 (ANI) a, en matière de protection sociale, créé de nouveaux droits individuels :

- généralisation de la complémentaire santé à tous les salariés du privé au 01/01/2016,
- définition d'un panier de soins minimum (qui s'articule avec les contrats responsables définis précédemment),
- fixation d'un taux de prise en charge minimal pour l'employeur,

Il a également aménagé le dispositif de portabilité existant (article 14 de l'ANI du 11 janvier 2008). L'ANI a été transposé dans le cadre de la loi de sécurisation de l'emploi du 14 juin 2013.

#### **LA MISE EN PLACE DU 100% SANTE**

Le gouvernement a lancé une nouvelle réforme du système de la santé et s'est fixé comme objectif de parvenir d'ici 2022 à un reste à charge zéro dans les domaines de l'optique, du dentaire et de l'audioprothèse. Quelques extraits du site du gouvernement sont repris ci-après<sup>6</sup> :

« Constat : parmi les 20% de Français aux revenus les plus bas, près d'1 sur 5 d'entre eux renonce à s'équiper en optique, et près d'1 sur 3 renonce à des soins dentaires, pour des raisons financières. [...] Les restes à charge des secteurs du dentaire, l'audiologie et l'optique sont élevés. C'est pourquoi les pouvoirs publics proposent, tout en préservant la liberté de choix de chacun, des paniers de soins sans reste à charge qui répondent aux besoins de santé nécessaires, avec une garantie de qualité, "par exemple en termes d'aminicissement ou d'antireflet pour les verres, ou, pour les audioprothèses, de discrétion des appareils", a précisé Agnès Buzyn, la ministre des Solidarités et de la Santé. [...] Cette offre sera accessible à tous les Français disposant d'une complémentaire santé (et sera prise en charge

<sup>6</sup> <https://www.gouvernement.fr/argumentaire/100-sante-remboursement-integral-dans-les-domaines-de-l-optique-de-l-audiologie-et-du>

à 100 % par la Sécurité Sociale et les complémentaires santé). Les contrats d'assurance complémentaire proposeront tous le panier "100% santé", sans augmentation liée à cette réforme du tarif des complémentaires. »

La réforme du 100% santé s'étalera sur trois ans avec des impacts à prévoir pour probablement une période plus importante. Le calendrier de mise en œuvre est précisé ci-dessous.

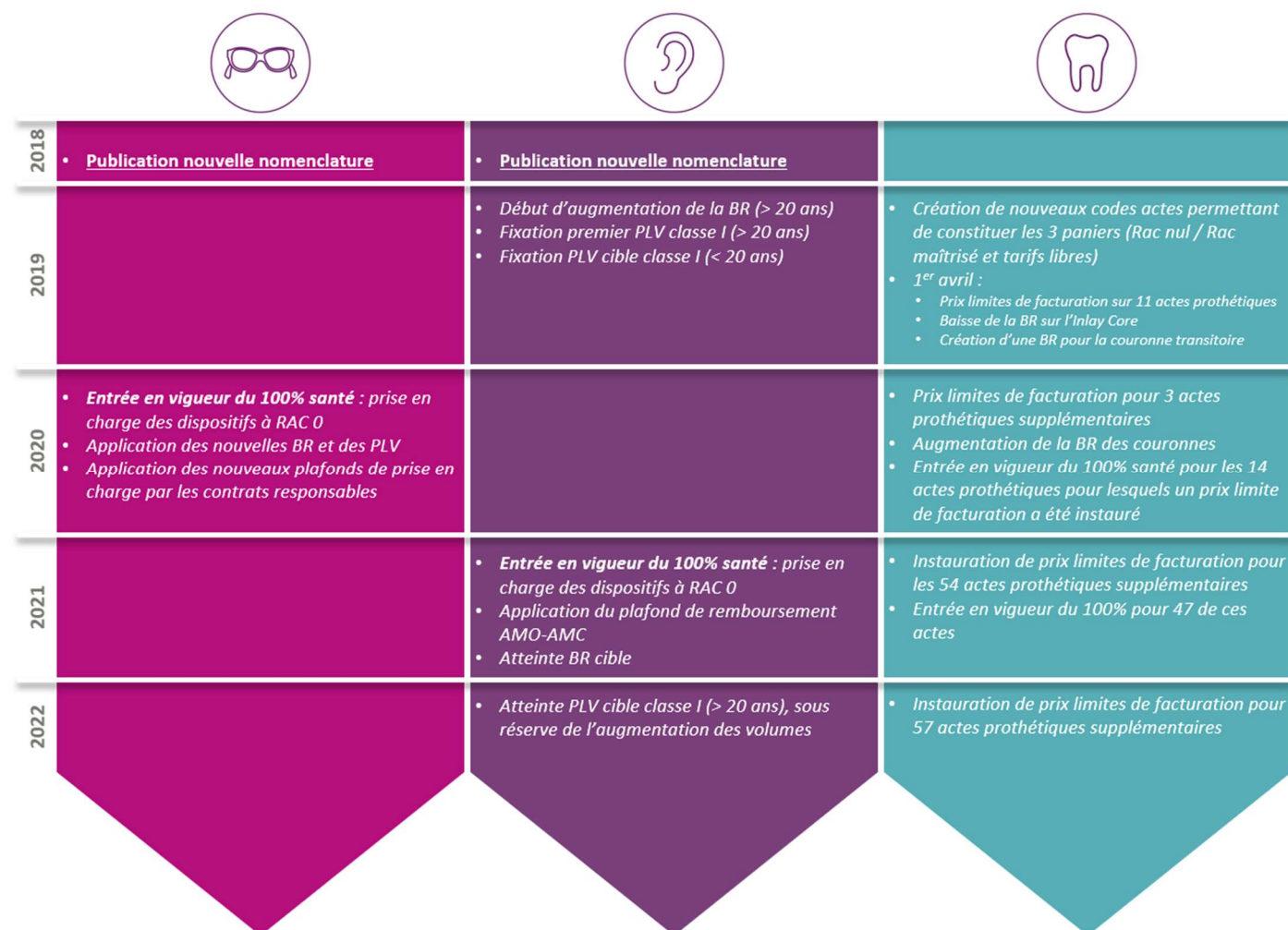


Figure 12 - Calendrier de mise en place de la réforme du 100% Santé<sup>7</sup>

#### FUSION ACS – CMU :

Jusqu'en Novembre 2019, un certain nombre d'assureurs (mutuelles, sociétés d'assurance mutualiste ou sociétés d'assurance) avaient des offres labellisées pour les personnes pouvant bénéficier de l'aide à la complémentaire santé (ACS). Les bénéficiaires de l'ACS, disposaient d'un chèque ACS permettant de financer une partie de la cotisation d'un régime inclus dans une offre labellisée. En parallèle, certains organismes assureurs avaient accepté de gérer la complémentaire santé CMU. Il n'y avait, pour les bénéficiaires de la CMU aucune cotisation à verser et les prestations étaient remboursées à l'organisme complémentaire sur la base de la dépense réelle prévu dans le respect des garanties accordées dans le cadre de la CMU-C dans la limite d'un plafond par personne. Il n'y avait pas de rémunération supplémentaire prévue pour les organismes réalisant la gestion.

<sup>7</sup> Source : Support présentation FNMF – 100% Santé – 23 Juillet 2018

A partir du 1er novembre 2019, les dispositifs de la CMU-C et de l'ACS ont été modifiés et fusionnés pour devenir la Complémentaire Santé Solidaire (CSS), avec ou sans participation financière selon le niveau de ressources du bénéficiaire. La couverture santé de la Complémentaire Santé Solidaire devient identique pour l'ensemble des bénéficiaires. Pour les personnes ayant des ressources situées entre les plafonds de la CMU-C actuelle et ces plafonds majorés de 35%, la Complémentaire Santé Solidaire donne lieu à une participation financière variable en fonction de l'âge des bénéficiaires (entre 8 euros et 30 euros par mois). Le Fonds CMU-C remboursera les sommes correspondantes à la prise en charge des dépenses des organismes complémentaire santé solidaire. A compter du 1er novembre 2019, le remboursement se fait au réel (et non plus au réel dans la limite d'un plafond forfaitaire). Pour les bénéficiaires de la Complémentaire Santé Solidaire avec participation financière, le Fonds CMU remboursera les dépenses de prestations des personnes adhérant à la Complémentaire Santé Solidaire majorées d'un montant forfaitaire au titre des frais de gestion et minorées du montant des participations dues à l'organisme complémentaire. Ce montant forfaitaire au titre de la gestion sera trimestriel et fixé pour chaque bénéficiaire à 8 euros pour les dépenses engagées en 2019 et en 2020, 7,5 euros pour celles engagées en 2021 et 7 euros pour les dépenses engagées en 2022.

### RESILIATION INFRA-ANNUELLE :

Instaurée par la loi n° 2019-733 du 14 juillet 2019, loi du 14 juillet 2019 relative au droit de résiliation sans frais de contrats de complémentaire santé permet aux assurés de résilier, après un an de souscription, leur contrat de complémentaire santé, à tout moment, sans frais ni pénalité. Auparavant, cette faculté ne pouvait être exercée qu'une fois par an, avant la date d'anniversaire du contrat. Cette disposition entre en vigueur le 1er décembre prochain. À compter du 1er décembre 2020, les assurés auront la possibilité de résilier un contrat de complémentaire santé dès lors que leur contrat a été souscrit depuis au moins une année. Le décret précise les contrats concernés sont :

- des contrats visant à couvrir les risques liés à la santé (frais occasionnés par une maladie, une maternité ou un accident) ;
- des contrats visant à couvrir les risques liés à la santé et comprenant d'autres garanties non strictement liées au remboursement des frais de santé, limitativement énumérés (risques décès, incapacité de travail ou invalidité, garanties d'assistance, de protection juridique, de responsabilité civile, de nuptialité-natalité ou d'indemnités en cas d'hospitalisation).

S'agissant des modalités de mise en œuvre, les assurés qui souhaitent résilier leur contrat peuvent entreprendre eux-mêmes les démarches ou bien se tourner vers leur nouvel assureur qui devra prendre en charge les formalités de résiliation. Cette nouvelle loi est rétroactive. Elle s'applique à la fois aux contrats souscrits après son application et à ceux signés avant.

#### 1.5. Contexte spécifique des mutuelles

Comme vu précédemment, il existe trois types d'organisme complémentaire. La section ci-dessous présente les particularités appliquées aux mutuelles.

Le principe fondateur des mutuelles est l'instauration d'une mutualisation des risques entre ses membres participants et leurs bénéficiaires. Les caractéristiques des contrats traduisent ce principe. Le principal élément constitutif du contrat en mutuelle est le règlement mutualiste. Il définit les droits et obligations respectifs des membres participants et de la mutuelle en matière de cotisation et de prestations. Il est utilisé dans les opérations individuelles comme les opérations collectives, même si dans le dernier cas il n'est pas obligatoire. Chez ADREA Mutuelle, trois types de documents le composent, le règlement lui-même contenant les règles applicables à tous, les fiches gammes (règles spécifiques à une offre) et les fiches garanties (tableaux de prestations d'un produit appartenant à une offre). Le

règlement est adopté dans le cadre d'un processus démocratique par l'Assemblée Générale et s'impose aux adhérents. L'Assemblée Générale a donc le pouvoir de modifier unilatéralement les engagements des parties. Ce pouvoir unilatéral et puissant trouve sa contrepartie dans l'obligation d'informer tous les membres participants des évolutions retenues. Le membre participant a la possibilité de refuser ces modifications et cela se traduit par la résiliation de son contrat. L'acte d'adhésion se traduit par la signature d'un bulletin d'adhésion qui formalise un contrat d'adhésion. Le contrat d'adhésion est conclu entre deux parties dont l'une ne peut, en fait, discuter les différentes clauses et n'a pour liberté que d'accepter ou de refuser le contenu global de la proposition.

En complémentaire santé, ce contrat d'adhésion est :

- solidaire (pas de sélection médicale),
- illimité (impossibilité pour la mutuelle de résilier unilatéralement l'adhésion à l'échéance tant que l'adhérent paye la cotisation),
- responsable ou non.

L'acte d'adhésion entraîne pour l'adhérent le paiement d'une cotisation. Les critères pouvant être pris en compte pour la fixation des cotisations en complémentaire santé, sont déterminés par le Code de la Mutualité :

- le revenu,
- la durée d'appartenance à la mutuelle,
- le régime de Sécurité Sociale d'affiliation,
- le lieu de résidence (si l'impact sur la consommation médicale est avéré),
- le nombre d'ayants droits,
- l'âge.

En mutualité, les droits et obligations de chaque contractant s'organisent autour des deux types de régime (Art L 221-1 et 221-2 du code de la mutualité) :

- les opérations individuelles,
- les opérations collectives.

Des éléments complémentaires relatifs au droit des mutuelles sont présentés dans (Senné, 2015) et présentées en Annexe 1.

## 2. Présentation d'ADREA Mutuelle

### 2.1. Chiffres et dates clés

Compte tenu de la fusion des trois mutuelles ADREA Mutuelle, APREVA Mutuelle et EOVI MCD Mutuelle afin de former le groupe AESIO au 31/12/2020, et compte tenu de la disponibilité des données au niveau d'ADREA Mutuelle, nous présentons les éléments de la mutuelle au 31/12/2018.

ADREA Mutuelle, mutuelle de Livre II, a été créée en 1999 en région Rhône-Alpes par des mutuelles interprofessionnelles appartenant à une Union de mutuelles. En décembre 2011, les huit mutuelles appartenant à cette Union fusionnent en une seule.



ADREA Mutuelle possède les agréments pour exercer des activités d'assurance dans les branches suivantes :

- Branche 21 Nuptialité-natalité,
- Branche 1 : Accidents,
- Branche 2 : Maladie,
- Branche 20 : Vie-Décès.

ADREA Mutuelle porte principalement des garanties en santé et, de manière moins significative, des garanties allocations obsèques. Elle compte plus d'un million de personnes protégées pour un chiffre d'affaires net de réassurance de 683 millions d'euros au 31/12/2018. Ses activités sont réparties entre l'individuel et le collectif (voir figure 13).

	<b>Nombre de Personnes Protégées (PP) par Adrea Mutuelle*</b>
<i>Marché des particuliers</i>	601 255
<i>Marché des entreprises et des professionnels</i>	588 232
<i>Substitution</i>	10 154
<b>Total (sans double compte)</b>	<b>1 103 415</b>

*\*pondéré au temps de présence sur 2018*

*Figure 13 - Répartition des personnes protégées (PP) en 2018*

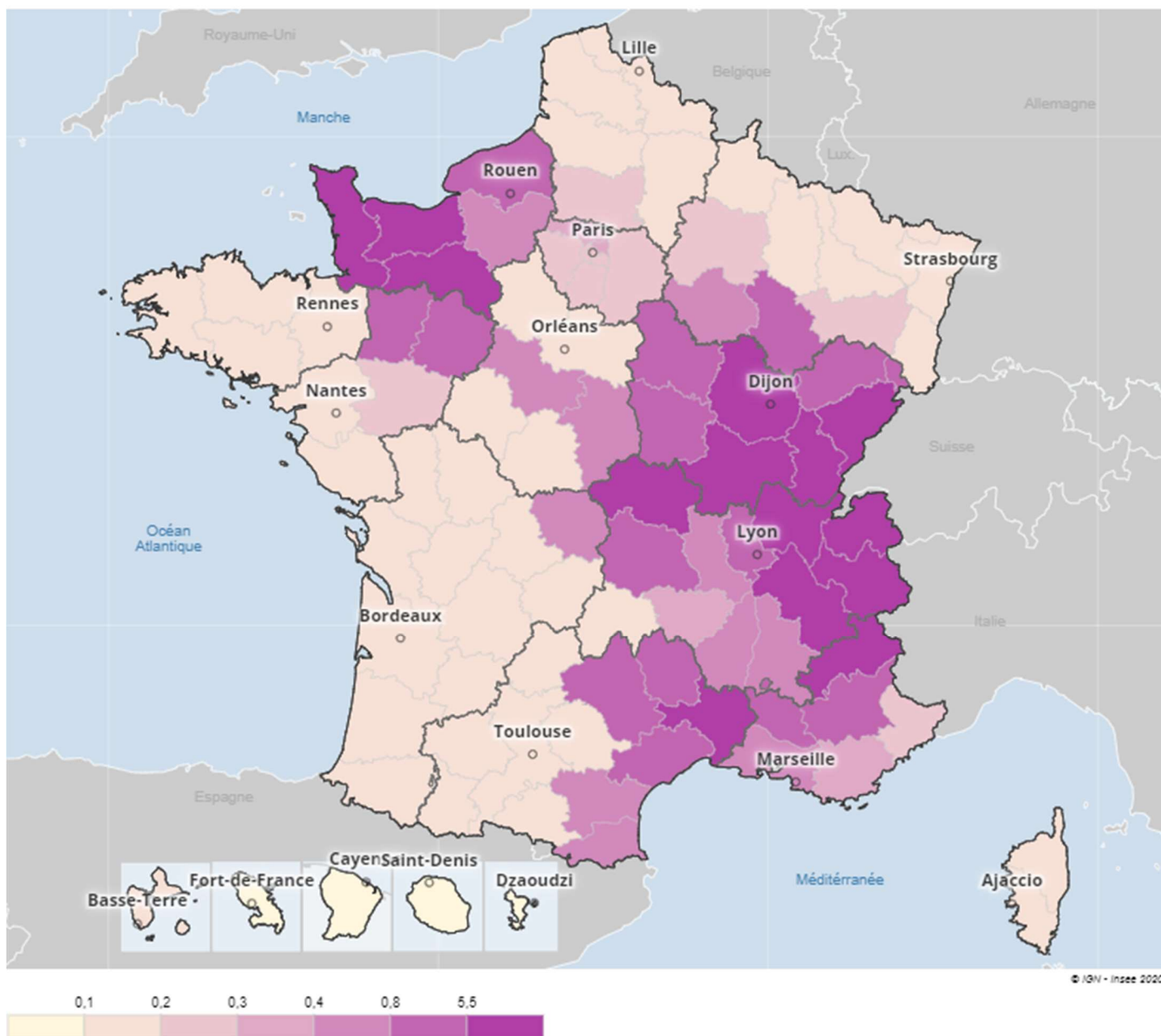


Figure 14 - Pourcentage de pénétration d'ADREA Mutuelle par département - Nombre de personnes protégées rapportées à la population du département

Comme nous pouvons le voir sur la figure 14, le portefeuille d'ADREA Mutuelle est particulièrement localisé en Basse-Normandie, dans le Centre-Sud et dans le Centre-Est et le Sud-Est, le plus souvent en lien avec les régions de création des anciennes mutuelles.

Quelques indicateurs utilisés dans le cadre de Solvabilité 2<sup>8</sup>, sont présentés ci-dessous. Ainsi, l'évolution du ratio des fonds propres, du capital de solvabilité requis et du ratio de couverture (Fonds propre / Capital requis) au cours du temps est présenté en figure 15.

<sup>8</sup> Les notions de solvabilité n'étant pas l'objet de ce mémoire, les définitions de ces indicateurs ne sont pas développées ici. Pour plus d'information : <https://acpr.banque-france.fr/europe-et-international/assurances/reglementation-europeenne/solvabilite-ii>

	2018	2017	2016	2015
<i>Fonds propres éligibles</i>	552 462 K€	562 253 K€	550 231 K€	591 124 K€
<i>Capital requis (SCR)</i>	195 822 K€	172 401 K€	170 496 K€	161 200 K€
<b>Ratio de couverture</b>	<b>282,12%</b>	<b>326,13%</b>	<b>322,72%</b>	<b>366,70%</b>

Figure 15 - Evolution des indicateurs de solvabilité 2 – ADREA Mutuelle

Le ratio de couverture reste très important en 2018, bien que la tendance au cours des quatre dernières années soit décroissante.

## 2.2. Présentation du portefeuille global d'ADREA Mutuelle

L'objectif de ce paragraphe est de présenter succinctement la structure du portefeuille d'ADREA Mutuelle. Les deux graphiques suivants présentent la structure par grande zone de portefeuille. Les informations ci-dessous sont extraites du rapport sur la solvabilité et la situation financière d'ADREA Mutuelle 2018 (ADREA, 2018).

<b>2018</b>	<b>Cotisations acquises brutes de réassurance</b>	<b>Prestations brutes de réassurance hors variation de PSAP</b>
<i>Données du compte de résultat technique non-vie</i>	789 176 K€	668 074 K€
<i>Données du compte de résultat technique vie</i>	8 936 K€	7 257 K€

Figure 16 – Ventilation des cotisations et des prestations 2018 entre Vie et Non-Vie

Les cotisations présentées sur la figure 16 correspondent à l'ensemble des cotisations hors taxes reçues par la mutuelle relative à ses activités d'assurance (y compris acceptation en réassurance et substitution éventuelle de mutuelles mais hors gestion pour compte de tiers). Ainsi, nous constatons que le risque santé représente 98,9 % du profil de risque de la mutuelle et la part des « cotisations brutes vie » représente 1,1 % de l'ensemble des cotisations.

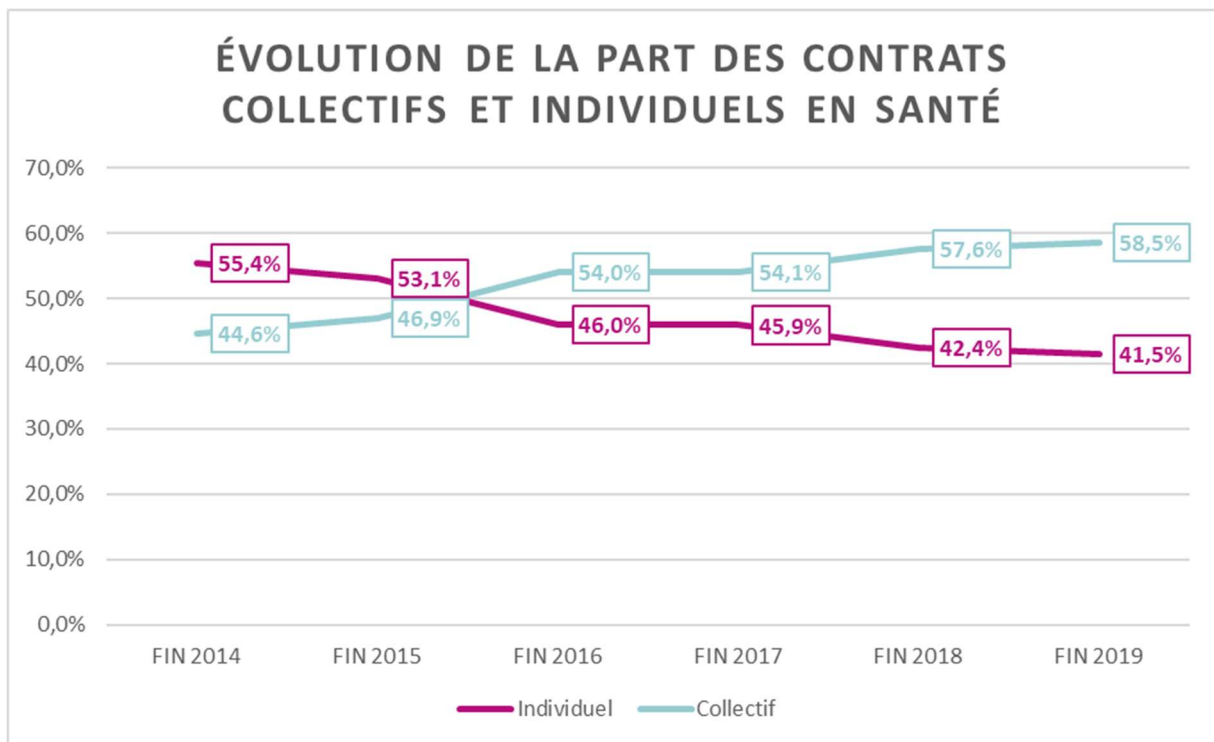


Figure 17 – Evolution de la part des contrats collectifs et individuels en santé (en nombre de personnes protégées)

Nous constatons en figure 17 que le portefeuille d'ADREA Mutuelle, sur le modèle de la plupart des mutuelles sur le marché, était historiquement majoritairement individuel et a amorcé depuis 2014 la transition de l'individuel vers le collectif. La proportion du collectif augmente et représente plus de la moitié du portefeuille (en nombre de personnes protégées).

L'un des éléments d'analyse retenu par ADREA Mutuelle pour assurer le suivi et le pilotage du portefeuille est la marge économique brute. Cette marge correspond à la différence entre les cotisations hors taxes et les prestations après effet des schémas d'assurance et de réassurance (y compris dotations éventuelles aux réserves de stabilité). Cela correspond à la marge technique brute générée lors de l'exercice considéré. Elle permettra de financer les charges techniques diverses (frais de gestion et de fonctionnement de la mutuelle entre autres), les cotisations diverses reversées pour des prestations présentes dans les garanties santé mais non assurées par ADREA Mutuelle (par exemple l'assistance) et les charges diverses liées à la délégation de la gestion ou au courtage. La marge technique économique brute est ajustée de la charge du forfait médecin traitant, des prestations et indus en attente de paiement identifiés et intégrés en comptabilité sur l'exercice mais qui ne sont pas encore remontés dans les systèmes.

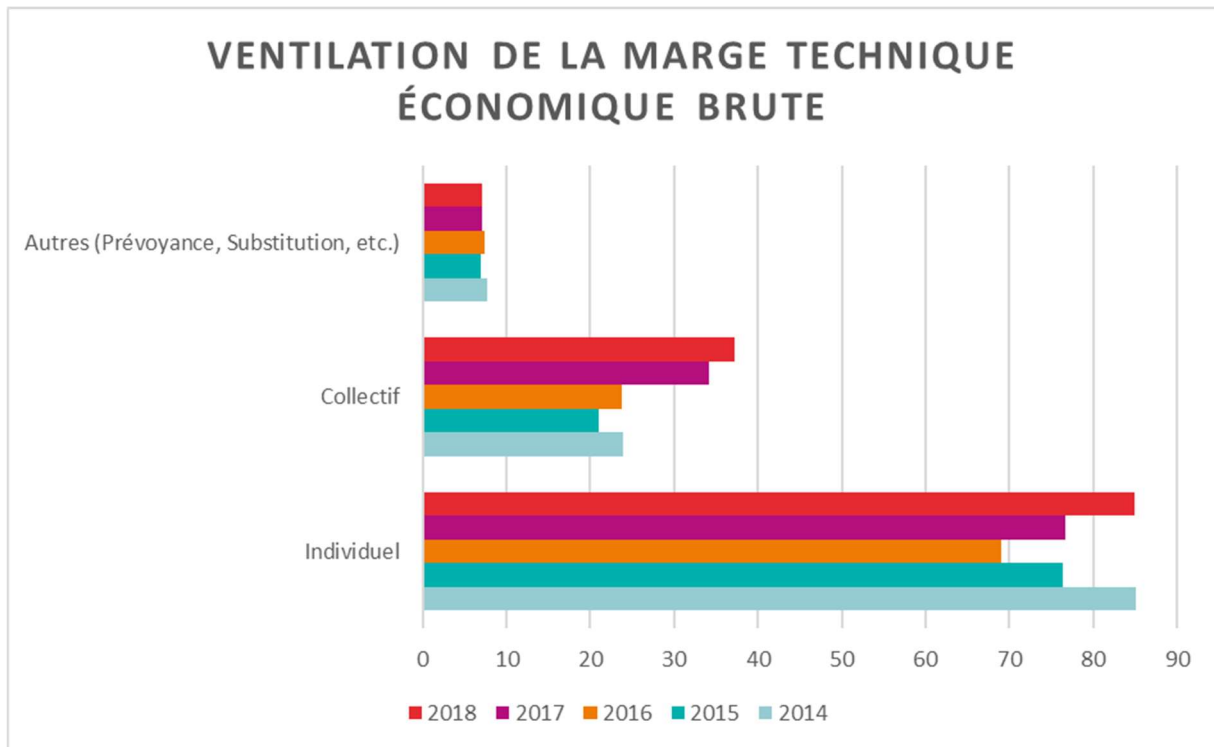


Figure 18 - Ventilation de la marge technique économique brute (en millions d'euros (M€))

La figure 18 présente la ventilation de la marge technique économique brute entre individuel et collectif. La lecture de ce graphique montre, comme pour beaucoup de mutuelle, que le principal contributeur au chiffre d'affaires d'ADREA Mutuelle, et encore plus à la marge technique, est la zone constituée des contrats individuels.

En particulier, comme le montre la figure 19 ci-dessous, la marge du portefeuille individuel est principalement portée par des produits historiques. Ainsi, l'équilibre économique d'ADREA Mutuelle dépend aujourd'hui principalement de produits d'assurance fermés à la commercialisation et dont les effectifs sont vieillissants et en diminution (les décès et radiations n'étant pas compensés par des souscriptions nouvelles).

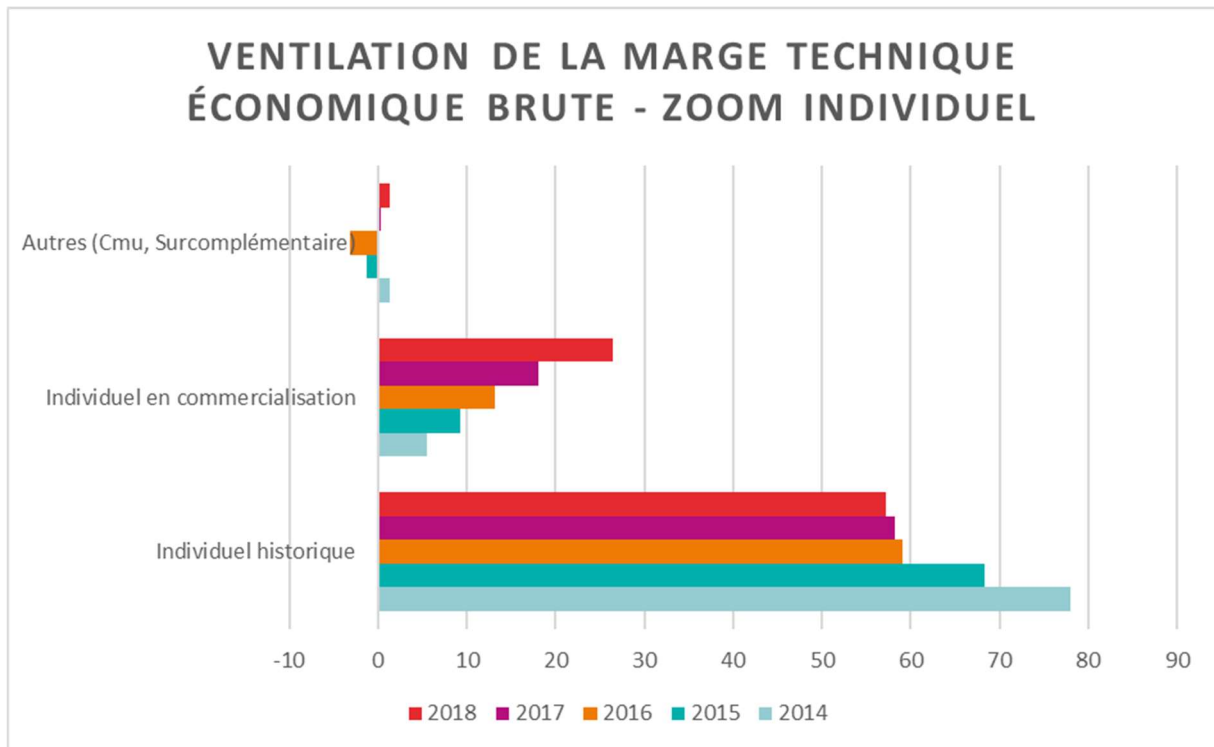


Figure 19 – Zoom sur la marge économique brute en individuel (en M€)

Cependant, l'amélioration de la marge sur les produits individuels en commercialisation au cours du temps est justifiée par l'attention particulière portée au suivi de ces produits à travers notamment la révision des paramètres permettant une segmentation plus adaptée des différentes tarifications comme par exemple l'ajustement des zoniers lié à la région de souscription des adhérents.

Cette structure de portefeuille, combinée à un contexte réglementaire particulièrement mouvant, n'est donc pas sans exiger d'ADREA Mutuelle un pilotage technique particulièrement précis afin d'assurer sa solvabilité permanente et donc de garantir les engagements pris auprès des adhérents.

## PARTIE II : LA GAMME FLEX'ADREA

La gamme Flex'ADREA sera au cœur de l'analyse réalisée dans ce mémoire. Elle correspond à la gamme individuelle principale d'ADREA Mutuelle actuellement en commercialisation. Après avoir présenté succinctement les données que nous utiliserons, nous analyserons les résultats techniques de la gamme Flex'ADREA et nous nous pencherons plus particulièrement sur les facteurs qui semblent les plus influents sur le niveau de risque.

### 1. Présentation de la base de données

#### 1.1. Qualité des données

Les données relatives au portefeuille de bénéficiaires couverts par un contrat de complémentaire frais de santé, ainsi que les données de consommation et de cotisations, ont été extraites sous la forme de fichiers Excel depuis le Datawarehouse (DWH) ACTUARIAT.

Le DWH ACTUARIAT est un entrepôt de données alimenté par les différents systèmes informatiques de gestion d'ADREA Mutuelle. Il est destiné aux analyses techniques et actuarielles sur le portefeuille. De nombreuses règles de gestion ont été mises en œuvre afin d'uniformiser les données et leur interprétation. Les principales règles de gestion ayant un impact sur la qualité des données utilisées dans l'étude sont :

- 1) L'homogénéisation des dates d'adhésion et des dates de résiliation afin que l'interprétation soit identique quel que soit le centre de gestion, et donc le système d'information.
- 2) Une transcodification des nomenclatures d'actes de remboursements complémentaires : pour chaque acte de remboursement utilisé dans les systèmes d'information sources, une correspondance avec la table unique d'actes utilisée dans le DWH ACTUARIAT a été retenue. Cela permet donc une unicité de l'analyse sur les données de consommation.

Par ailleurs, le DWH ACTUARIAT fait l'objet d'analyses récurrentes quant aux erreurs de paramétrage possibles ou de qualification qui pourraient conduire à une dégradation de la qualité des données contenues dans l'entrepôt.

Compte tenu du fonctionnement du décisionnel DWH ACTUARIAT, nous avons dû procéder par étape pour constituer les bases d'analyse nécessaires à notre étude. Ainsi, les informations définissant la personne protégée étant liées à la base des effectifs nous avons, une fois cette information récupérée dû la lier aux informations de cotisations d'une part (pour l'analyse globale sur le produit) et de prestations d'autre part (pour l'analyse de la prime pure).

Ainsi, la base de données relative aux prestations extraites contiendra les informations suivantes sur la gamme Flex'ADREA par numéro de personne protégée (PP) sur la période 2017 – 2019 :

- les prestations payées,
- l'année de soins,
- l'année de paiement,
- le produit Flex'ADREA associé,
- la date de naissance de l'assuré,
- le critère géographique lié au contrat (le département en l'occurrence),
- le sexe : bien que le critère « sexe » ne soit plus autorisé comme critère de différenciation dans la tarification depuis le 31/12/2012, il n'en demeure pas moins un facteur potentiellement explicatif du niveau de prestations

constaté. L'objectif de ce mémoire n'étant pas la mise en place d'un outil de tarification mais d'isoler le caractère géographique, il nous a semblé pertinent d'envisager son utilisation dans le modèle.

- la date d'adhésion au produit,
- la date de résiliation au produit,
- la Catégorie Socio Professionnelle (CSP),
- le type de personne protégée (adhérent principal, conjoint, etc.).

Une première analyse des données nous a imposé de retirer les informations relatives à la CSP. En effet, ces informations n'étant pas nécessairement intégrées dans la base décisionnelle, les données sont manquantes pour plus de 80% des personnes protégées pour la CSP.

La base d'étude comporte initialement 571 608 lignes.

Suite à l'extraction de ces données, nous avons pu introduire l'exposition annuelle qui représente le temps de présence de chacune des personnes protégées par an dans notre base.

Le premier retraitement consiste au retrait des éléments ayant une exposition nulle. Il s'agit notamment de retraitements de prestations réalisées a posteriori sur des adhérents ayant résiliés.

Il est à noter que des informations sur le département sont manquantes. Nous avons pris le choix de supprimer ces lignes, les réintégrer s'avérant être une tâche délicate. En effet, nous aurions pu envisager afin de ne pas supprimer ces lignes de les associer par exemple au département le plus représenté. Cela étant, cette solution ne nous ayant pas semblée satisfaisante et compte tenu du nombre de lignes supprimées relativement faible (cela revient à supprimer 473 lignes), nous avons estimé la perte d'information potentielle non significative.

Enfin, les prestations négatives ont été retirées de l'analyse, s'agissant de régularisations de prestations qui pourraient nuire à la modélisation souhaitée.

Ces retraitements sont synthétisés à la figure 20 :

<b>Retraitements réalisés</b>		<b>Nombre de lignes concernées</b>	<b>(en %)</b>
<i>Nombre de ligne sur la base initiale</i>		<b>571 608</b>	
exposition nulle		27	0,0%
information département non alimenté		473	-0,1%
prestations < 0		143	0,0%
<b><i>Nombre de ligne sur la base d'étude après retraitement</i></b>		<b>570 965</b>	<b>-0,1%</b>

Figure 20 - Premiers retraitements sur la base d'étude

## 1.2. Extrapolation des données et choix de la période d'analyse

Les données ont été extraites au 31 juillet 2020. Aussi, à cette date, l'observation des prestations payées au titre des soins 2018 et 2019 ne reflète a priori pas la charge finale de prestations pour respectivement 2018 et 2019. Il reste encore des paiements qui seront effectués, soit pour des soins connus à cette date mais non payés, soit pour des soins pas encore connus. En outre, sur le risque santé, le délai de forclusion est de deux ans à compter du remboursement de l'assurance maladie. La provision pour sinistres à payer (PSAP) représente l'estimation de cette charge de



prestations restant à payer. Les méthodes de provisionnement ne faisant pas l'objet de ce mémoire, la méthode d'extrapolation est présentée brièvement ci-dessous.

L'objectif de l'extrapolation dans le cadre du suivi des résultats est d'estimer le plus justement possible la charge ultime de chacune des années de soins étudiées. Pour ce faire, nous utilisons l'écoulement constaté sur les prestations de l'année de soins 2017 entre le 1<sup>er</sup> janvier 2017 et le 31 juillet 2020. L'hypothèse d'une atteinte de la charge ultime de 2017 au 31 juillet 2020 semble raisonnable compte tenu du délai de forclusion de deux ans d'une part et de la faible représentativité des prestations non soumises au remboursement de l'assurance maladie d'autre part.

Le taux de PSAP au titre de l'année de soins 2018 à appliquer sur les prestations constatées jusqu'au 31 juillet 2020, sera obtenu en divisant la charge ultime 2017 par la charge de prestations au titre des soins 2017 constatée jusqu'au 31/07/2019.

De même, le taux de PSAP au titre de l'année de soins 2019 à appliquer sur les prestations constatées jusqu'au 31/07/2020, sera obtenu en divisant la charge ultime 2017 par la charge de prestations 2017 constatée au 31/07/2018. Pour obtenir le taux de PSAP 2019, nous pouvons aussi, une fois la charge ultime calculée sur 2018, utiliser l'expérience 2018 en divisant la charge ultime 2018 par la charge de prestations 2018 constatée au 30/07/2019. En l'occurrence, nous avons fait le choix de prendre la moyenne des coefficients calculés pour 2019 sur 2017 et 2018.

Par soucis de cohérence, une analyse sur la stabilité des délais de liquidation est réalisée dans le cadre général du calcul des provisions. Ainsi, après échange avec le service de gestion en charge du paiement des prestations, s'il nous a été confirmé que la liquidation n'avait pas présenté de spécificité de liquidation (par exemple, régularisation de prestations importante sur un poste), la crise de la COVID-19 semble avoir permis la régularisation massive de prestations de survenance 2019 et antérieures tant, de par la disponibilité des gestionnaires pour le traitement de ces prestations que par la volonté des adhérents de régulariser leur factures pendant la période de confinement.

La figure 21, présente les coefficients d'extrapolation obtenus sur les prestations d'une part au titre des soins 2018, d'autre part au titre des soins 2019.

<b>Année de soins</b>	<b>Coefficient d'extrapolation à appliquer sur les prestations constatées au 31.07.2020</b>
2018	100,0%
2019	100,5%

Figure 21 - Coefficient d'extrapolation à appliquer sur les prestations constatées au 31.07.2020 pour 2018 et 2019

Il est à noter que la charge ultime 2018 a été estimée atteinte au 31/07/2020, sur la base de l'expérience 2017. En outre, compte tenu de la remarque liée à la crise de la COVID-19, et de la faiblesse du coefficient constatée, nous avons décidé de ne pas extrapoler les prestations au-delà de Juillet 2019. Ce choix se justifie notamment par le fait qu'une erreur d'extrapolation de l'ordre de 0,3% n'aura qu'une conséquence mineure sur notre analyse compte tenu du choix de conserver les trois années de données à notre disposition (voir ci-dessous).

Après analyse de la sinistralité sur les années 2017, 2018 et 2019, il nous a semblé judicieux de conserver les trois années d'informations à notre disposition. En effet, la segmentation que nous avons à réaliser étant proche de celle des départements, il nous a semblé important de garder un maximum de données d'expérience.

En outre, si la santé est souvent un risque analysé sur un an, notamment lors des phases de renouvellement, l'expérience sur trois ans, permettra de lisser un certain nombre de prestations exceptionnelles qui pourraient être constatées si nous ne retenions qu'une année d'expérience. Nous avons au préalable vérifié qu'aucune modification de garanties notables ne soient intervenues au cours de la période. Si la réforme du 100% Santé s'est effectivement mise en place au 1<sup>er</sup> janvier 2019, les modifications pour l'année 2019 sont mineures.

Cependant, il est important de noter qu'il existe une inflation sur les prestations de soins annuelle qui peut être expliquée à la fois par les réformes successives, mais aussi par une évolution des coûts naturelle des soins (matériels médicaux, médicaments, recours aux soins, etc.). Pour ne pas nuire à la qualité du raisonnement, nous avons pris la décision de faire évoluer les prestations de 1,5% par an. Ce coefficient d'1,5% correspond à l'inflation moyenne retenue par an au cours de la période 2017 – 2019. Ainsi, nous appliquons la formule suivante sur nos prestations :

$$Prestation = Prestation\_extrapolée_{année\ i} * (1 + inflation)^{(2019 - année\ i)}$$

Ainsi, nous pouvons raisonner par ligne. Il est à noter qu'avec ce raisonnement, nous acceptons l'hypothèse de non-saisonnalité des données de santé. Le choix d'avoir retenu les trois années de prestations devrait permettre d'écarter la saisonnalité des données et semble donc pouvoir rendre acceptable cette hypothèse. Cependant, nous n'étudierons pas d'avantage ce point dans ce mémoire.

## 2. Généralités sur la gamme

La gamme des produits Flex'ADREA, objet de la base d'étude sera présentée dans cette partie.

### 2.1. Présentation des différents produits de la gamme Flex'ADREA

La gamme Flex'ADREA est une gamme modulaire. Ainsi, l'adhérent aura le choix entre six niveaux de garanties en hospitalisation / soins à associer à six choix de garanties Optique, Dentaire et Prothèses et appareillages.

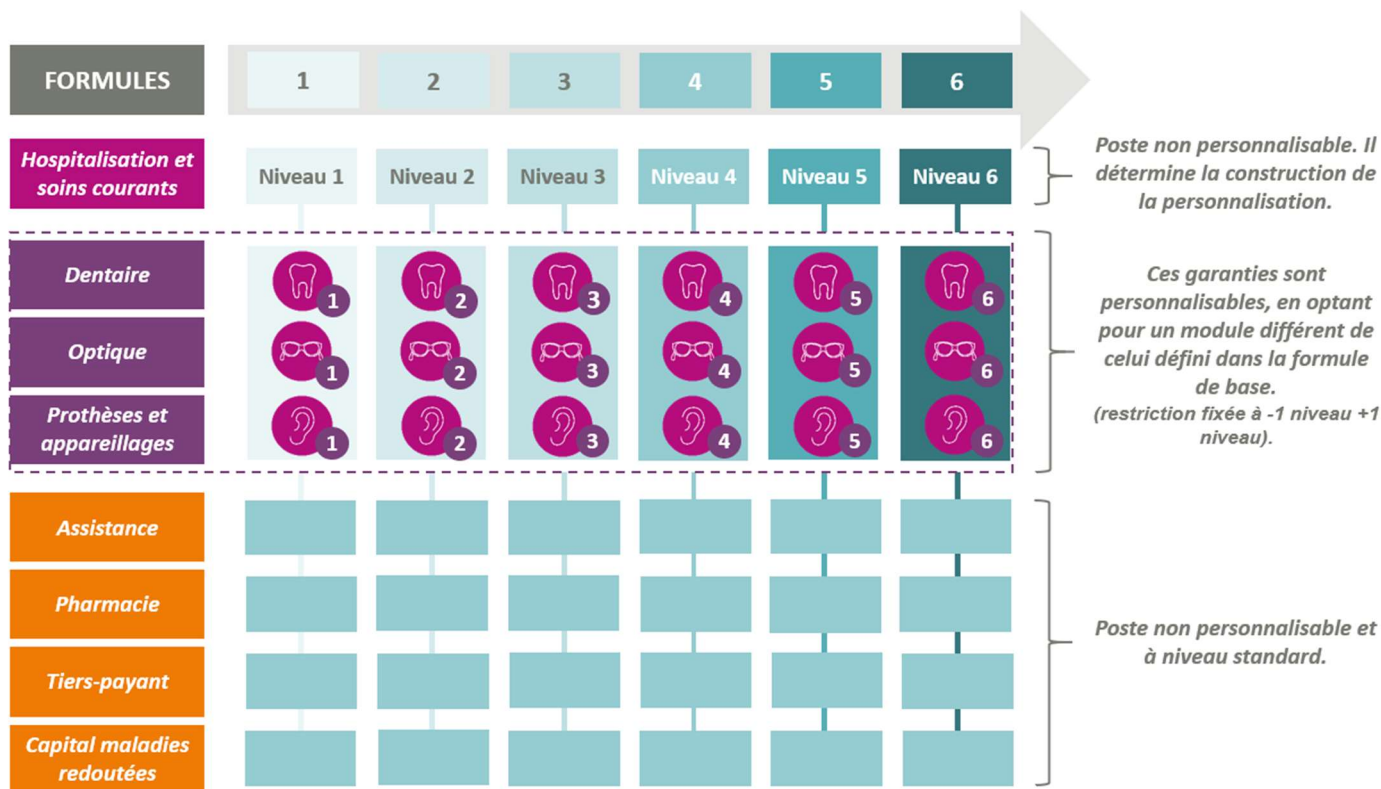


Figure 22 - Modularité de la gamme Flex'ADREA

La modularité étant de plus ou moins un, un choix de niveaux 3 en Hospitalisation / Soins, laissera le choix de trois niveaux en Optique / Dentaire (niveau 2,3 ou 4 – voir figure 22).

Ainsi, seize combinaisons sont possibles pour l'adhérent (figure 23). Par la suite, les combinaisons seront exprimées selon la règle suivante, Flex'ADREA Niveau 6.6 sera exprimé comme F66.

		Combinaison avec Blocs Dent/Opt/App		
		-1	=	+1
NIVEAU 1 Hospi/Soins 1			Dent/Opt/App 1 1.1	Dent/Opt/App 2 1.2
NIVEAU 2 Hospi/Soins 2	Dent/Opt/App 1 2.1	Dent/Opt/App 2 2.2	Dent/Opt/App 3 2.3	
NIVEAU 3 Hospi/Soins 3	Dent/Opt/App 2 3.2	Dent/Opt/App 3 3.3	Dent/Opt/App 4 3.4	
NIVEAU 4 Hospi/Soins 4	Dent/Opt/App 3 4.3	Dent/Opt/App 4 4.4	Dent/Opt/App 5 4.5	
NIVEAU 5 Hospi/Soins 5	Dent/Opt/App 4 5.4	Dent/Opt/App 5 5.5	Dent/Opt/App 6 5.6	
NIVEAU 6 Hospi/Soins 6	Dent/Opt/App 5 6.5	Dent/Opt/App 6 6.6		

*Restriction à un niveau inférieur ou un niveau supérieur dans le choix du module.*

**16 combinaisons possibles**

Figure 23 - Les seize combinaisons de la Gamme Flex'ADREA

L'offre Flex'ADREA contient une offre adaptée au régime Alsace Moselle à travers cinq produits en lien avec les produits précédents (F21AM, F23AM, F34AM, F45AM, F66AM).

Les cotisations sont fixées chaque année par l'assemblée générale, conformément aux critères du code de la mutualité. Elles présentent toutes des garanties en conformité avec le contrat responsable et intègrent donc une taxe de 13,27%.

Les cotisations varient en fonction :

- de l'âge du bénéficiaire, apprécié au 1<sup>er</sup> janvier de chaque année variant âge par âge de 0 à 80 ans et plus, avec une linéarité entre 0 et 18 ans et à partir de 80 ans.
- de l'appartenance au régime général ou au régime Alsace Moselle,
- de la garantie choisie dans une des seize combinaisons possibles en régime général et de cinq combinaisons en régime Alsace Moselle,
- en cas d'appartenance au régime général, de la zone d'habitation (il existe sept zones basées sur le département d'habitation).

Il est à noter que le zonier est appliqué sur les tarifs TTC de la gamme Flex'ADREA. En outre, le changement d'adresse d'un adhérent peut entraîner une modification de la tarification en cas de changement de zone. Dans ce cas, la mutuelle recalculera la nouvelle cotisation du membre participant et celle de ses ayants droit. La modification prendra effet au premier jour du mois suivant la date de réception de l'information par la mutuelle.

Enfin, il existe quelques conditions de gratuités telle que celle pour la souscription d'un troisième enfant que nous n'énumérons pas ici.

Les tarifs sont calculés par âge avec une mutualisation sur les âges inférieurs à 20 ans et ceux supérieurs à 85 ans.

## 2.2. Résultats globaux 2019

Dans le cadre du suivi de portefeuille, l'analyse des résultats Flex'ADREA est réalisée au global et par produit. Tout d'abord, il peut être intéressant d'analyser la pyramide des âges et son évolution (figure 24).

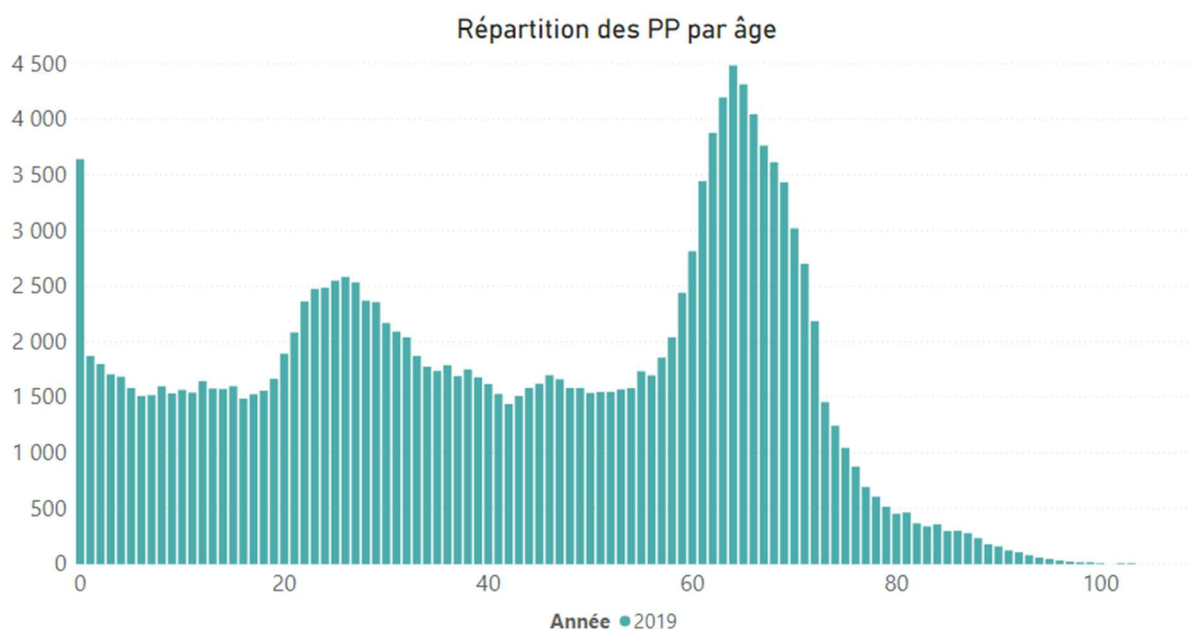


Figure 24 - Pyramide des âges - Offre Flex'ADREA

L'âge moyen sur la gamme est de 43,6 ans en 2019, contre 41,5 ans en 2017. En 2019, la part des plus de 60 ans représente 36,6% des personnes protégées. Sur une offre individuelle, il est courant de voir ces deux pics de souscription, le premier entre 20 et 28 ans et le second à partir de 60 ans. Le premier pic correspond à la fin de validité de couverture des enfants sur la complémentaire santé de leurs parents. A partir de 25 ans en revanche les souscriptions individuelles sont moins nombreuses car elles coïncident avec l'entrée dans le monde professionnel et la bascule vers des couvertures collectives. Le second pic correspond à la période de bascule en retraite où, avec la fin de la prise en charge par l'employeur, il peut être intéressant pour les personnes de souscrire à une couverture individuelle.

Les figures 25 à 28 présentent les éléments de suivi généraux de l'offre Flex'ADREA.

Les éléments présentés sur les figures qui suivent sont brièvement définis ci-dessous :

- PP moyen : nombre de personnes protégées pondéré par le temps de présence sur l'année 2019 (ou exposition 2019),
- cotisations brutes HT : cotisations appelées nettes de taxes (13,27%),
- P/C Net : Afin d'analyser de manière plus fine la bonne adéquation entre les cotisations et les prestations, les frais de gestion propre estimés par ADREA Mutuelle sont déduits des Cotisations HT. Ce taux de frais est de

20,1% en 2019 et de 19,3% en 2017 et 2018 pour le portefeuille individuel. Ainsi, le P/C Net correspond à la formule suivante : Prestations par survenance / (Cotisations brutes HT\*(1-Frais de gestion d'ADREA Mutuelle)).

- marge nette : Cotisations nettes de frais de gestion – Prestations par survenance.

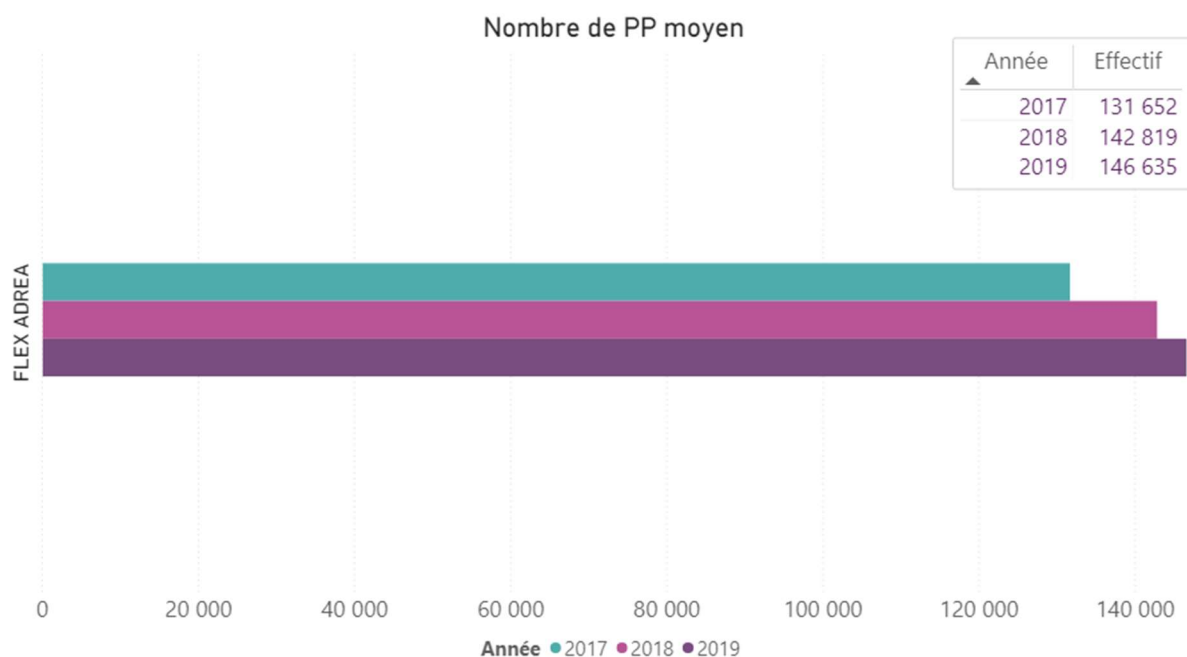


Figure 25 - Suivi des personnes protégées – Offre Flex'ADREA

Le nombre de PP moyen a augmenté de 8,5% entre 2017 et 2018 et de 2,7% entre 2018 et 2019. Le nombre de PP est en hausse de +11,4% en 2019 par rapport à 2017. Si l'offre est encore en croissance, la part du stock est devenue largement prédominante.

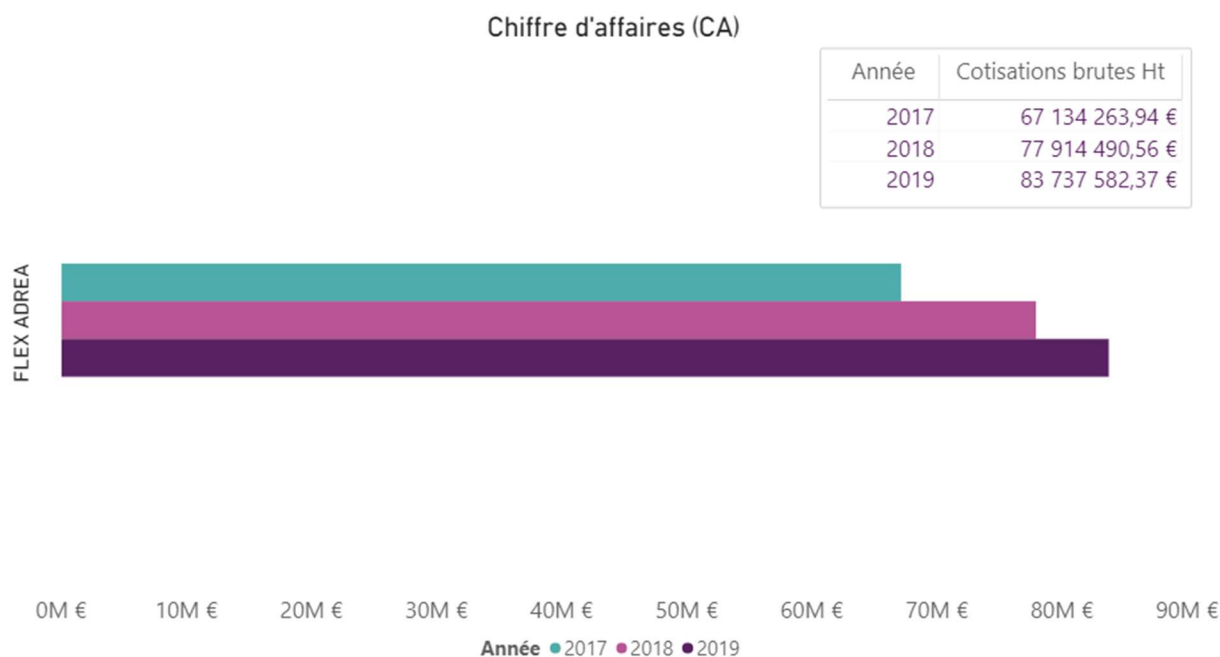


Figure 26 - Suivi du chiffre d'affaires – Offre Flex'ADREA

Le chiffre d'affaires a augmenté de 16,1% entre 2017 et 2018 et de 7,5% entre 2018 et 2019. L'évolution 2017-2019 du chiffre d'affaires est de +24,7%. La cotisation moyenne est passée de 509,9€ en 2017 à 545,6€ en 2018 (+7,0%) et à 571,1€ en 2019 (+4,7%), soit une évolution entre 2017 et 2019 de +12,0%. L'évolution de la masse cotisations se justifie à la fois par l'évolution des effectifs couverts, le vieillissement de la population (indexation naturelle avec l'âge en individuel) et les taux directeurs appliqués sur l'offre en 2017 et 2018.

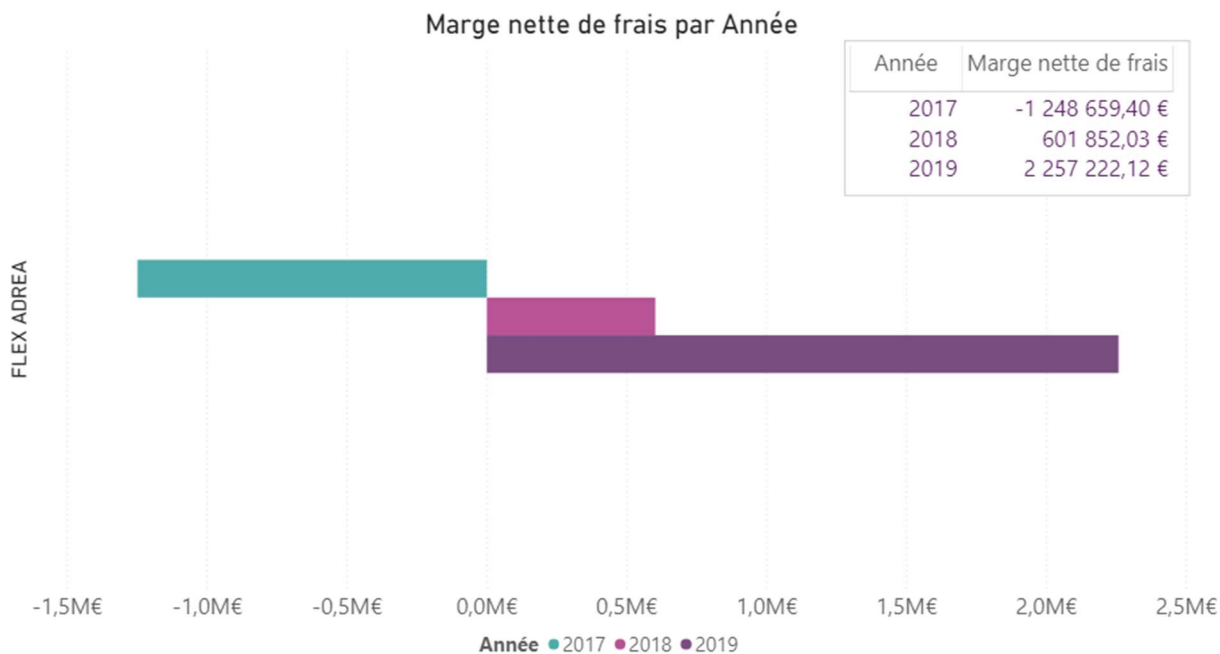


Figure 27 - Suivi de la marge - Offre Flex'ADREA

La marge nette est en forte augmentation entre 2017 et 2019 (+3,5 millions d'euros). L'évolution de la marge nette entre 2017 et 2018 est très proche de celle entre 2018 et 2019 (respectivement +1,9 millions d'euros et +1,7 millions d'euros).

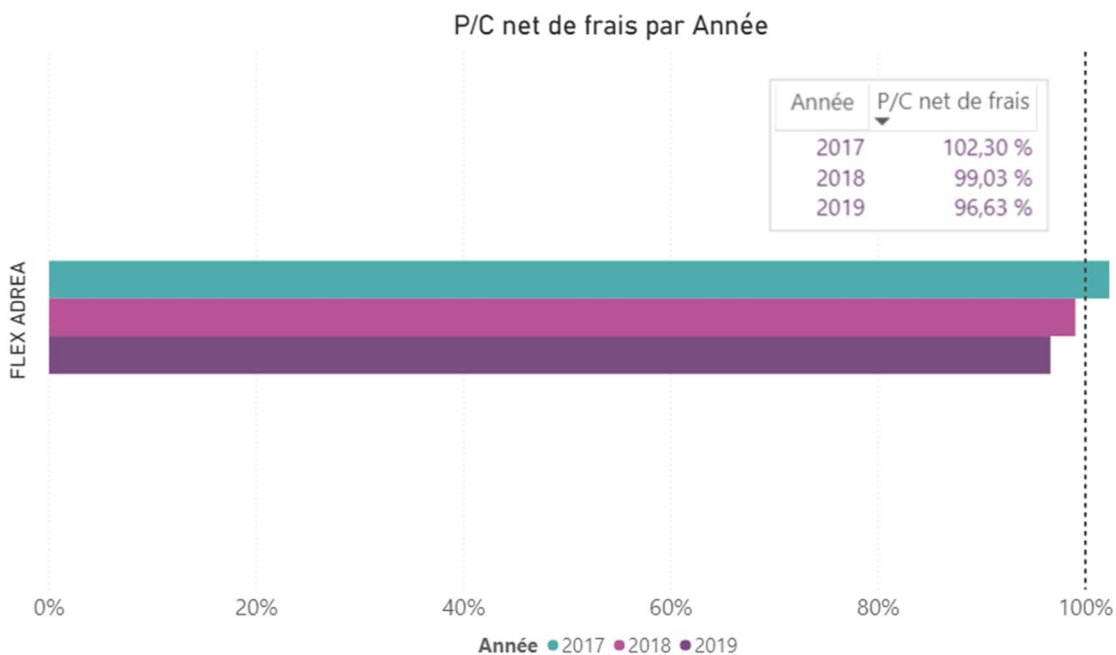


Figure 28 - Suivi Prestations / Cotisations nettes – Offre Flex'ADREA

En cohérence avec l'évolution de la marge nette, le P/C net est en forte amélioration entre 2017 et 2019 (de près de 6 points). L'amélioration du P/C entre 2017 et 2018 était légèrement plus marquée que celle entre 2018 et 2019 (respectivement -3,3 points et -2,4 points).

Après avoir présenté les indicateurs sur l'offre, nous pouvons faire un zoom sur les produits principaux.

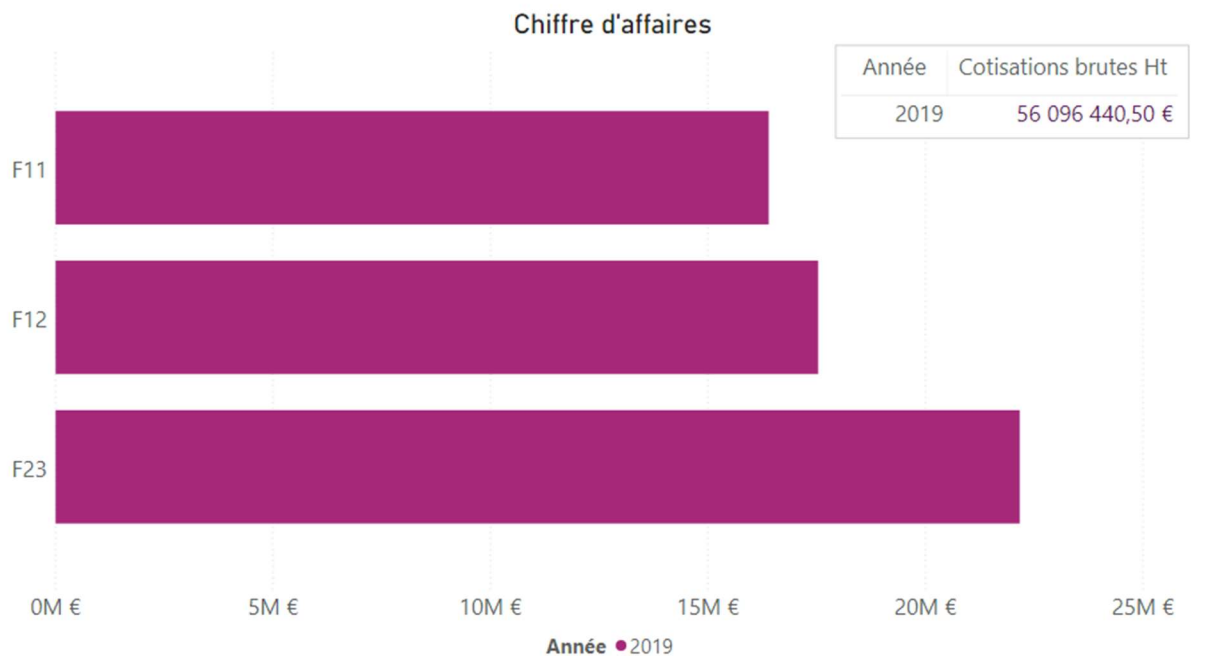


Figure 29 - CA des produits les plus souscrits – Offre FlexADREA

Les produits les plus souscrits en 2019 (F11 / F12 / F23) représentent 74,7% du nombre de PP total et 67,0% du CA total. Le produit Flex 11 est le produit le plus souscrit (nombre de PP) et le produit Flex 23 est celui présentant le CA le plus important en 2019.

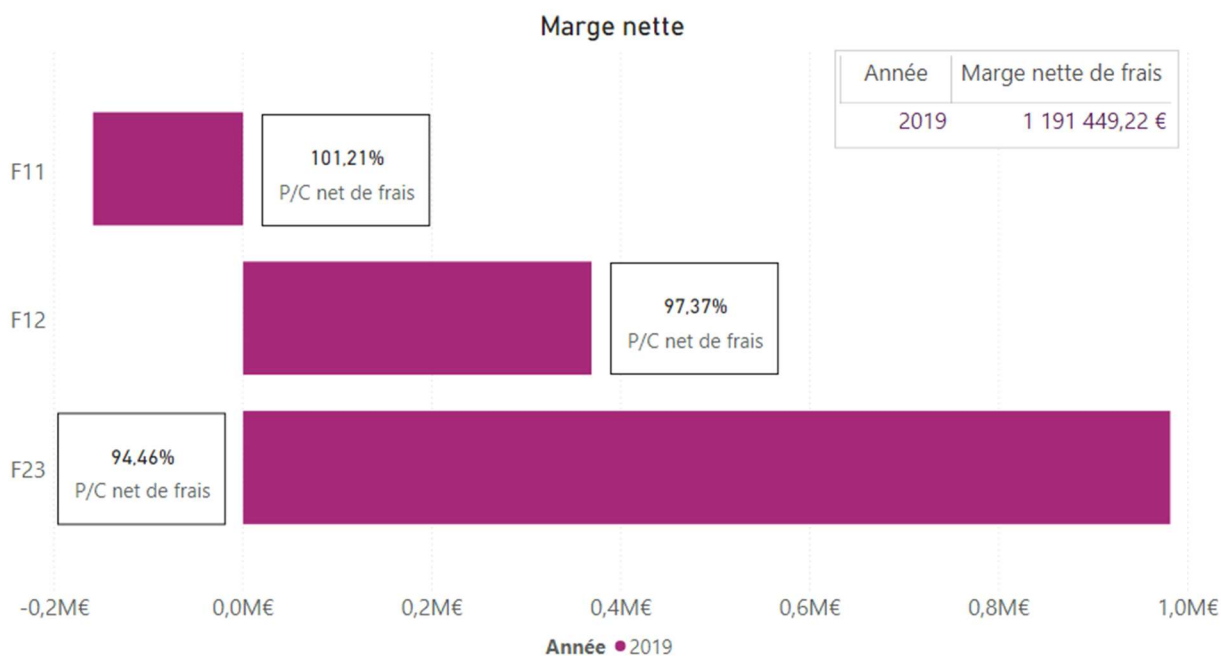


Figure 30 - Marge des produits les plus souscrits – Offre Flex'ADREA



Il est à noter que la marge nette du produit F11 s'améliore en 2019 mais reste déficitaire (passage de -672 026 euros en 2017 à -158 897 euros). Le P/C net 2019 est en amélioration sur ces trois produits de près de 3 points en cohérence avec la gamme au global (P/C net de 96,6% en 2019 contre 99,0% en 2018).

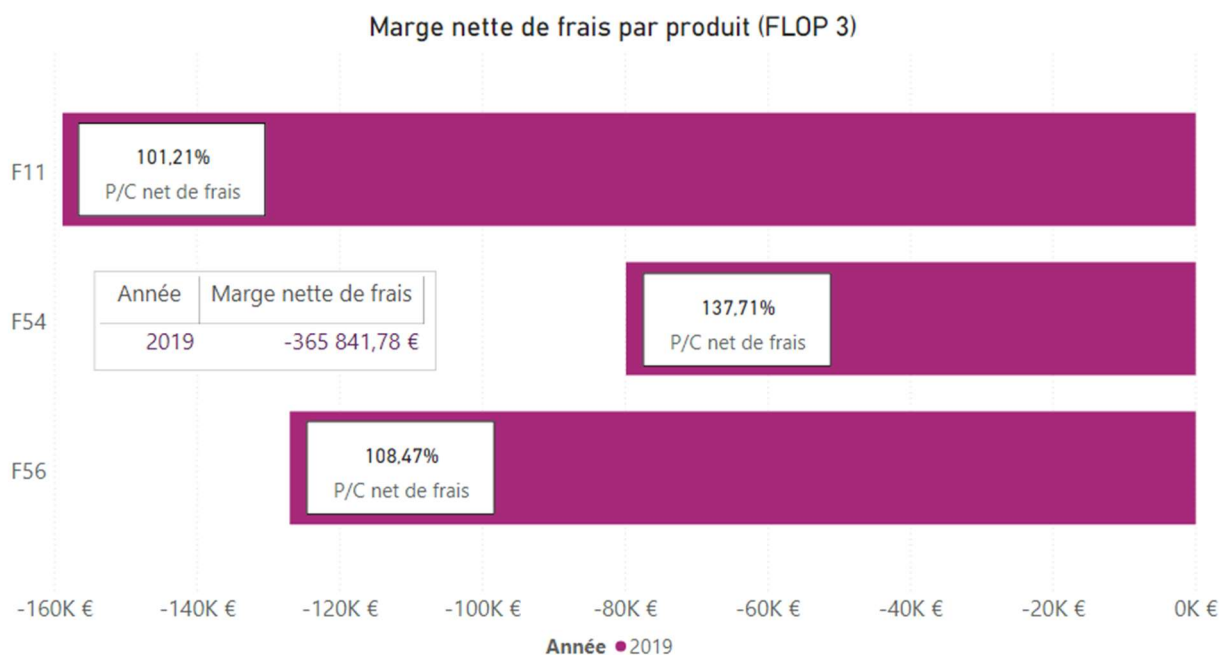


Figure 31 – Zoom sur les produits les plus déficitaires – Offre Flex'ADREA

Les trois principaux produits déficitaires sont les produits F54, F56 et F65 qui représentent seulement 1,6% des PP et 2,2% du CA total en 2019. Le P/C net de ces produits est de 112,0% en 2019 mais est en amélioration de plus de 5 points (P/C net 2018 : 116,6%).

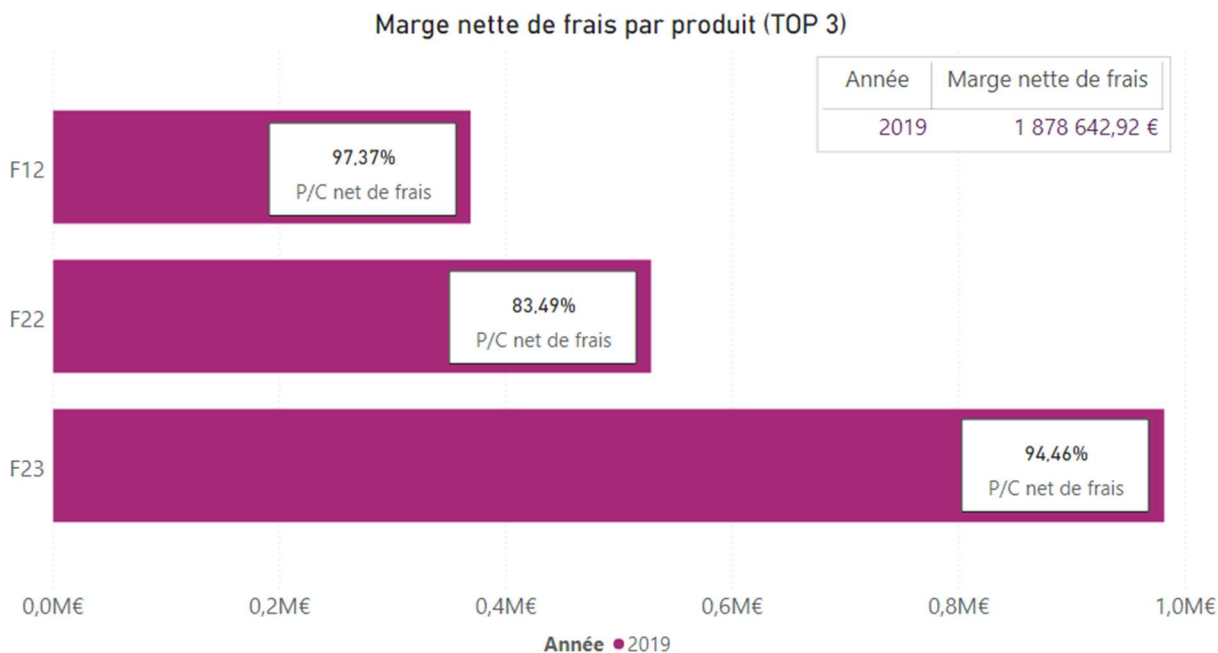


Figure 32 - Zoom sur les produits les plus excédentaires – Offre Flex'ADREA

Les trois principaux produits excédentaires sont les produits F12, F22 et F23. Ils représentent 50,8% des PP et 41,7% du CA total en 2019. Le P/C net de ces produits est de 94,6% en 2019 et est en amélioration d'environ 3 points (P/C net 2018 : 97,3%).

### 3. Analyse descriptive et présentation du zonier en place

L'objectif de cette section est de présenter, dans un premier temps, quelques statistiques descriptives sur la base de travail et en particulier les relations entre les différentes variables de la base avec la charge de prestations par personne protégée. Puis, dans un deuxième temps, de présenter une analyse plus détaillée du zonier actuellement en place et mis en œuvre au sein d'ADREA Mutuelle.

#### 3.1. Statistiques descriptives

Nous devons dans un premier temps réaliser un certain nombre de retraitements sur notre base pour permettre une analyse fonctionnelle tout en essayant de la maintenir réaliste.

La variable principale qui fera l'objet d'étude de ce mémoire est la prestation par personne protégée.

Dans un premier temps, compte tenu de ce qui a été évoqué précédemment, nous avons pris la position de ne pas estimer de zonier sur les départements Alsace-Moselle ainsi que sur les DOM. Sur l'Alsace-Moselle, la différence d'intervention de la Sécurité Sociale, peut engendrer, en particulier sur des garanties assez faibles, de grosses différences dans l'analyse de la prestation constatée. Compte tenu de la faible exposition, moins de 300 PP en moyenne sur les départements 57, 67, 68 (respectivement Moselle, Haut-Rhin, Bas-Rhin), nous avons préféré exclure ces éléments de l'étude. Sur les DOM, l'expérience constatée dans nos données est trop faible pour bien représenter les particularités de consommation dans les départements d'Outre-Mer. Le retraitement de ces départements entraîne la suppression de 824 lignes.

Il est important de noter que la prestation étudiée afin de raisonner en année pleine, l'analyse portera sur une variable prenant en compte l'exposition. Nous pouvons alors définir la prime pure, en notant N l'année de soins :

$$\mathbf{prime\ pure}_i = \frac{\mathbf{prestation}_i}{\mathbf{exposition}_i}$$

Ainsi, pour les personnes protégées pour lesquelles l'exposition n'est pas complète, nous nous ramenons en vision annualisée.

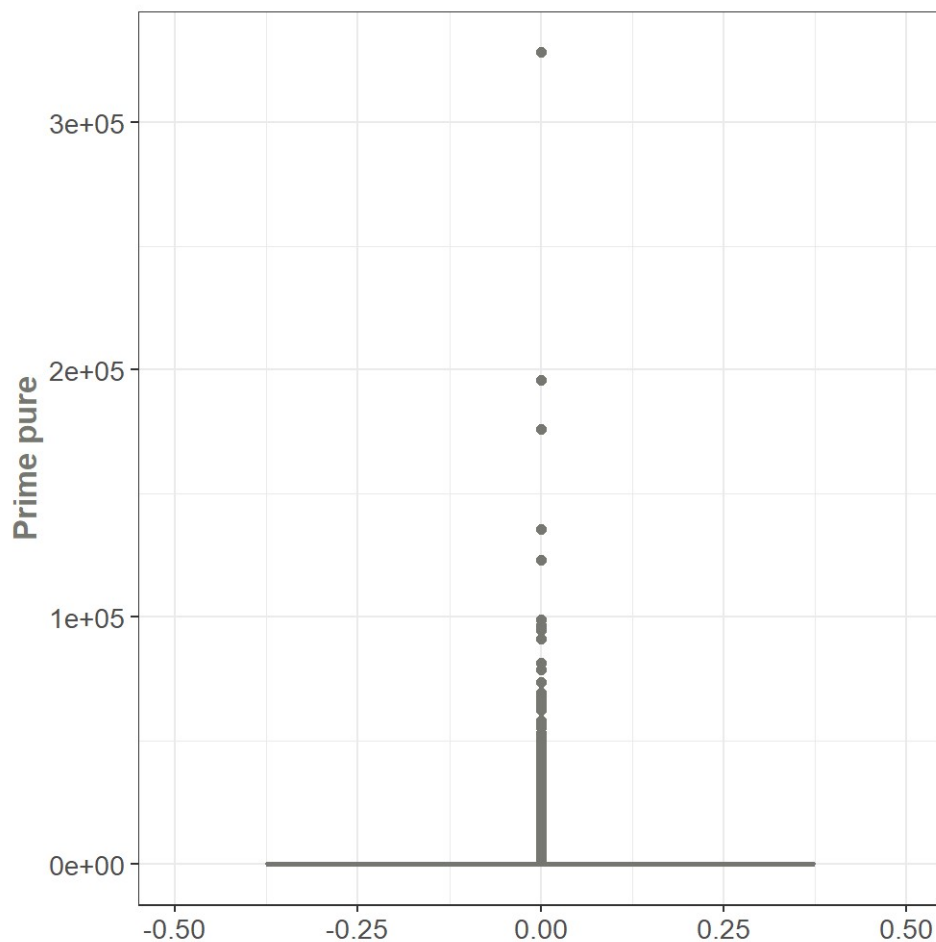


Figure 33 – Boîte à moustache de la prime pure – avant retraitement

L'analyse de la boîte à moustache de la prime pure (figure 33) permet de constater que les 10% de données les plus élevées sont assez difficiles à traiter intégralement tant leur montant est différent. En effet, le montant des 10% des prestations les plus élevées varient d'environ 1 300 € à plus de 200 000 €. Nous prenons la position d'écarter de notre modèle les lignes de prestations les plus élevée.

Tout d'abord, le calcul de la prime pure et son lien avec l'exposition peut engendrer quelques problèmes d'analyse. En effet, pour une exposition très faible, les données corrigées peuvent être à tort exceptionnellement élevées. Supprimer les lignes extrêmes permettra de corriger cette limite liée à la construction de la variable d'analyse. En outre, les modèles couramment utilisés en tarification et notamment ceux utilisés dans ce mémoire, n'étant pas très adaptés à expliquer des données de sinistres dit « graves », nous prenons le choix de fixer le seuil de suppression au quantile à 99,75% et excluons les lignes présentant des sinistres supérieurs à 9000 €. Cela revient malgré tout à supprimer 8,2% de la masse de sinistres de notre base et un peu plus de 1 400 lignes. Il est à noter que dans une étude de tarification, l'impact sur le calcul de la prime aurait été trop significatif pour ne pas être traité par exemple par le biais d'un traitement des sinistres graves ou par une correction des expositions faibles. Cependant, notre étude étant assez sensible aux sinistres extrêmes et n'ayant pas vocation à calculer une prime pure directement, ce retraitement nous a semblé acceptable. Cependant, une étude approfondie de ces sinistres extrêmes sur les départements concernés n'a pas été réalisée et pourra constituer une des limites à notre analyse.

La boîte à moustache intégrant ces retraitements est présentée sur la figure ci-après.

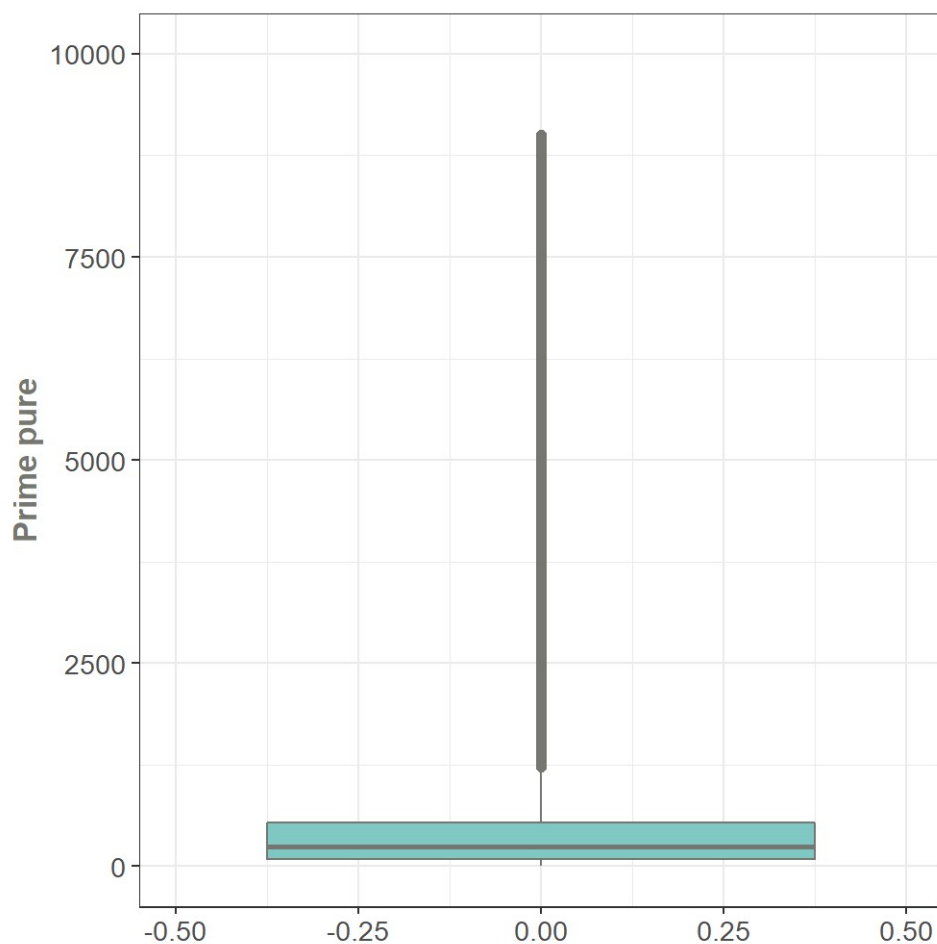


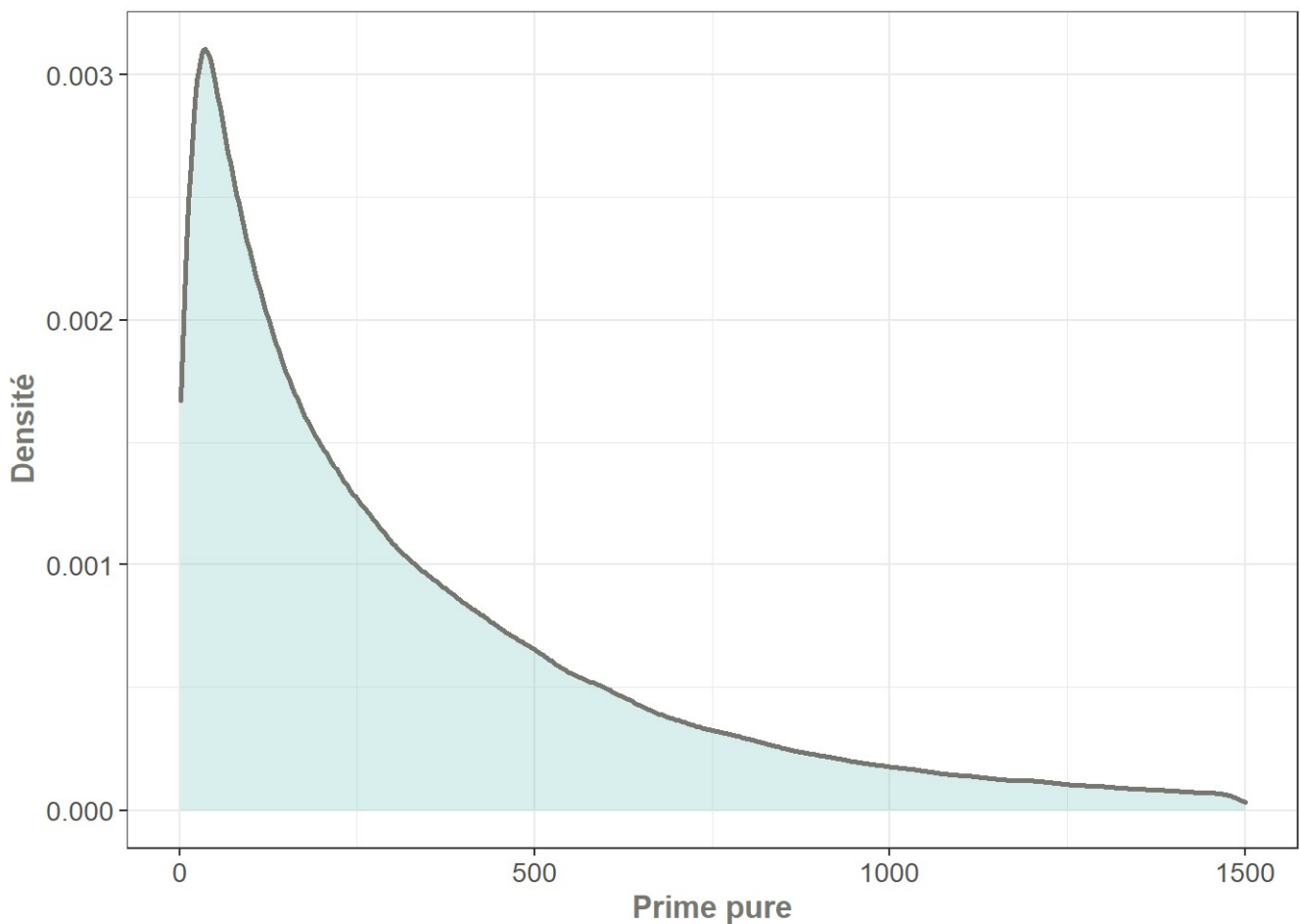
Figure 34 – Boite à moustache – Prime pure

En figure 35, une synthèse des retraitements réalisés est présentée.

<b>Retraitements réalisés</b>		<b>Nombre de lignes concernées</b>	<b>(en %)</b>
<b>Nombre de ligne sur la base</b>		<b>570 965</b>	
départements Alsace-Moselle (67, 68, 57) et DOM-TOM		1 824	-0,3%
prestations > 9000		1 424	-0,3%
<b>Nombre de ligne sur la base d'étude après retraitement</b>		<b>567 717</b>	<b>-0,6%</b>

Figure 35 - Deuxième retraitement sur la base d'étude

Nous avons représenté en figure 36 la distribution de la prime pure.



*Figure 36 – Distribution de la prime pure*

Nous constatons une masse de bénéficiaires sans prestations, ce qui devra être pris en compte dans l'analyse et la modélisation. Nous reviendrons sur ce point dans les parties suivantes.

Afin de commencer à réfléchir sur l'apport des différentes variables sur la charge de prestation par personne protégée, les impacts de chacune des variables sur la prime pure par personne protégée est représenté ci-dessous.

**Produit :**

Dans les données extraites, il existe 21 produits différents en comptant les produits liés au régime local. Cela étant nous avons évoqué, en partie II.1.1, l'exclusion des produits du régime local de notre analyse.

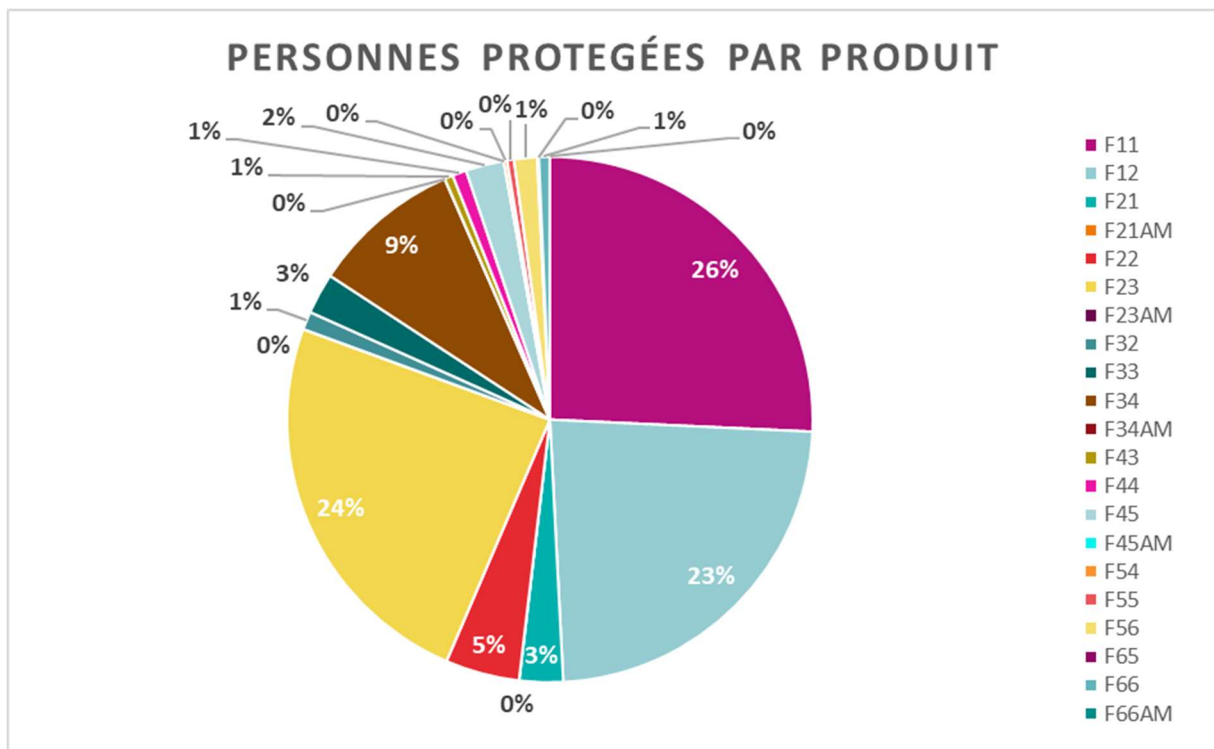


Figure 37 - Répartition des effectifs par produit avant regroupement des modalités

Comme nous pouvons le constater, un certain nombre de ces produits semble sous représenté. Malgré l'exclusion des départements associés au régime local, la répartition des effectifs par produit (figure 37), montre que sur les produits associés au régime local il reste quelques personnes protégées. Il est probable qu'il s'agisse de personnes protégées, assuré principal ou ayants-droit, affiliés au régime Alsace-Moselle mais n'y résidant pas. Cependant, nous ne pouvons exclure qu'il puisse s'agir d'anomalies de paramétrage ou de personnes protégées en cours de changement de situation à date.

L'analyse de l'impact du produit sur la prime pure (figure 38), semble conduire à certaines observations difficiles à exploiter. Ainsi, si la tendance d'évolution de la prime pure en fonction du produit semble croissante, la prime pure associée à certains produits, du fait notamment de la structure modulaire de la gamme Flex'ADREA, présente quelques particularités et notamment des résultats contraires aux attentes. C'est le cas par exemple sur les produits F43 et F54 qui présentent une prime pure plus élevée que des produits avec des garanties plus importantes. Cependant, ces deux produits ne représentant respectivement que 0,5% et 0,2% de personnes protégées, ils pourront largement nuire à l'analyse des modèles et méthodes statistiques réalisées dans ce mémoire.

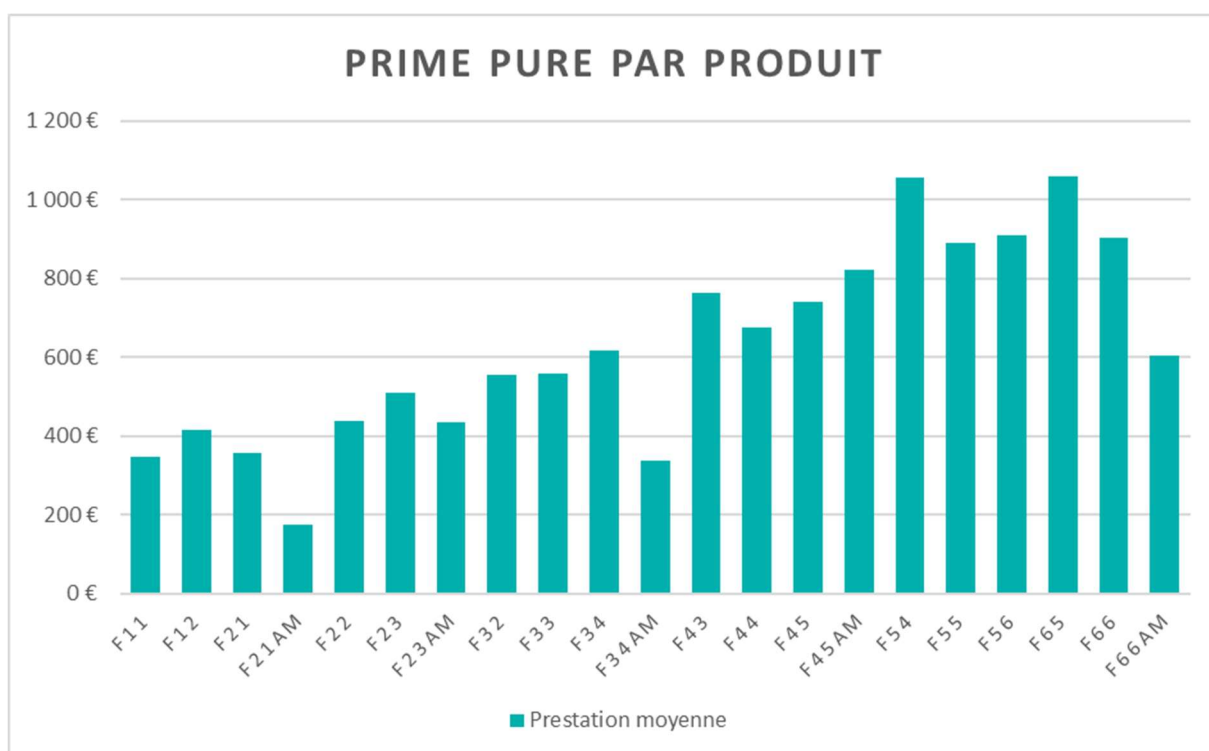


Figure 38 - Evolution de la prime pure en fonction du produit

Compte tenu de ces observations, nous avons appliqué deux modifications successives.

Afin de ne pas nuire à l'analyse de notre zonier qui ne prendra pas en compte le régime local Alsace Moselle, nous avons fait le choix de supprimer les produits Alsace-Moselle de notre base de données (230 lignes).

Remarque : Il est à noter que ce retraitement est le dernier réalisé sur la base d'étude. Comme nous pouvons le voir à la figure 39, la base d'étude finale contient 567 487 lignes.

Retraitement réalisé		Nombre de lignes concernées	(en %)
produits AM		230	0,0%
<b>Nombre de ligne sur la base d'étude après retraitement</b>		<b>567 487</b>	<b>0,0%</b>

Figure 39 - Dernier retraitement sur la base d'étude

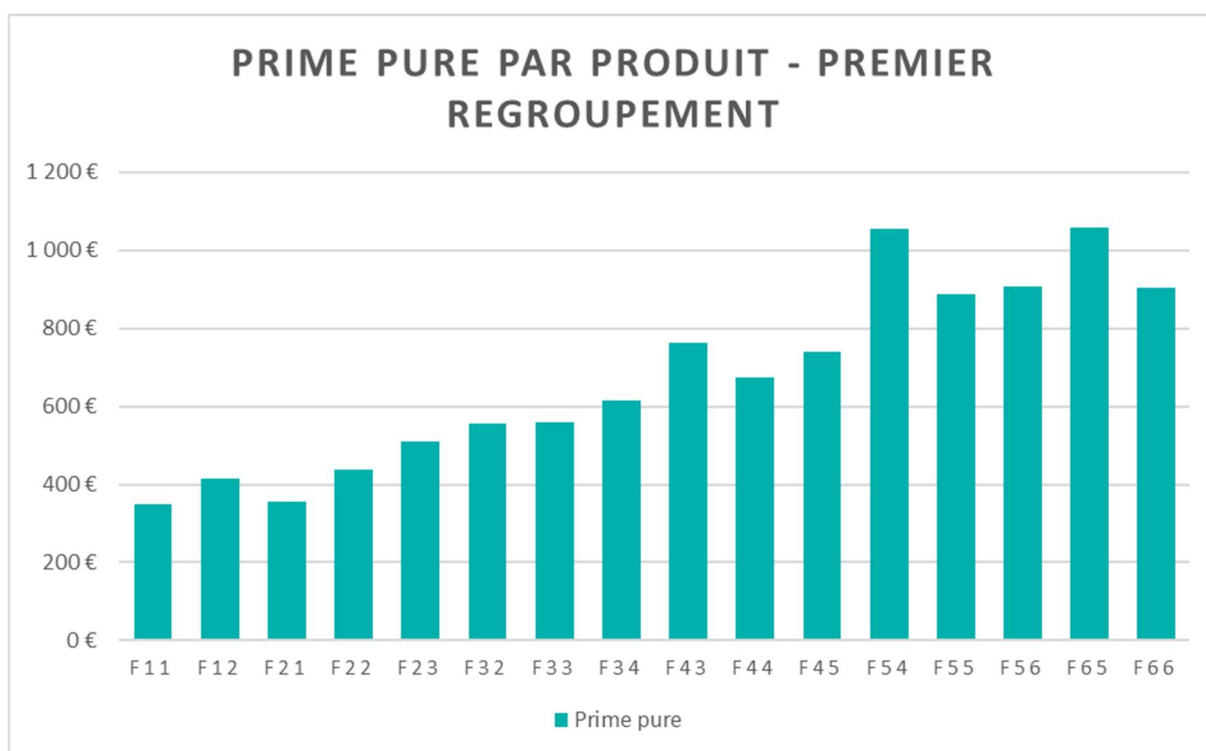


Figure 40 - Evolution de la prime pure en fonction du produit après exclusion des données Alsace-Moselle

Compte tenu du constat de la relative tendance croissante de la prime pure moyenne en fonction du produit et de la particularité de la gamme Flex'ADREA lié à la modularité, l'analyse d'un produit par rapport à un autre peut être rendue délicate. Ainsi, le lien entre la gamme F32 qui permet un choix 3 en Hospitalisation / Médecine courante et un choix 2 en Optique / Dentaire, et entre la gamme F23 qui permet un choix 2 en Hospitalisation / Médecine courante et un choix 3 en Optique / Dentaire est difficile à analyser. Ce type de retraitement de nos variables permet d'alléger la phase de conception de modèle qui imposera des regroupements de certaines catégories dont les écarts d'impacts explicatif seront non significatifs. Cependant, il est important de préciser que nous pourrions également perdre en précision d'information.

Nous avons décidé de rapprocher les contrats en 3 grands types de produit, selon les niveaux de prime pure mais aussi de la représentativité des produits :

- Bas de gamme : regroupe les quatre produits F11 à F22, dont la prime pure moyenne est de 383 €
- Milieu de gamme : regroupe les quatre produits F23 à F34, dont la prime pure moyenne est de 541 €,
- Haut de gamme : regroupe les huit produits de F43 à F66, dont la prime pure moyenne est de 811 €.

L'impact du produit sur la prime pure avec la nouvelle répartition retenue est présenté en figure 41.



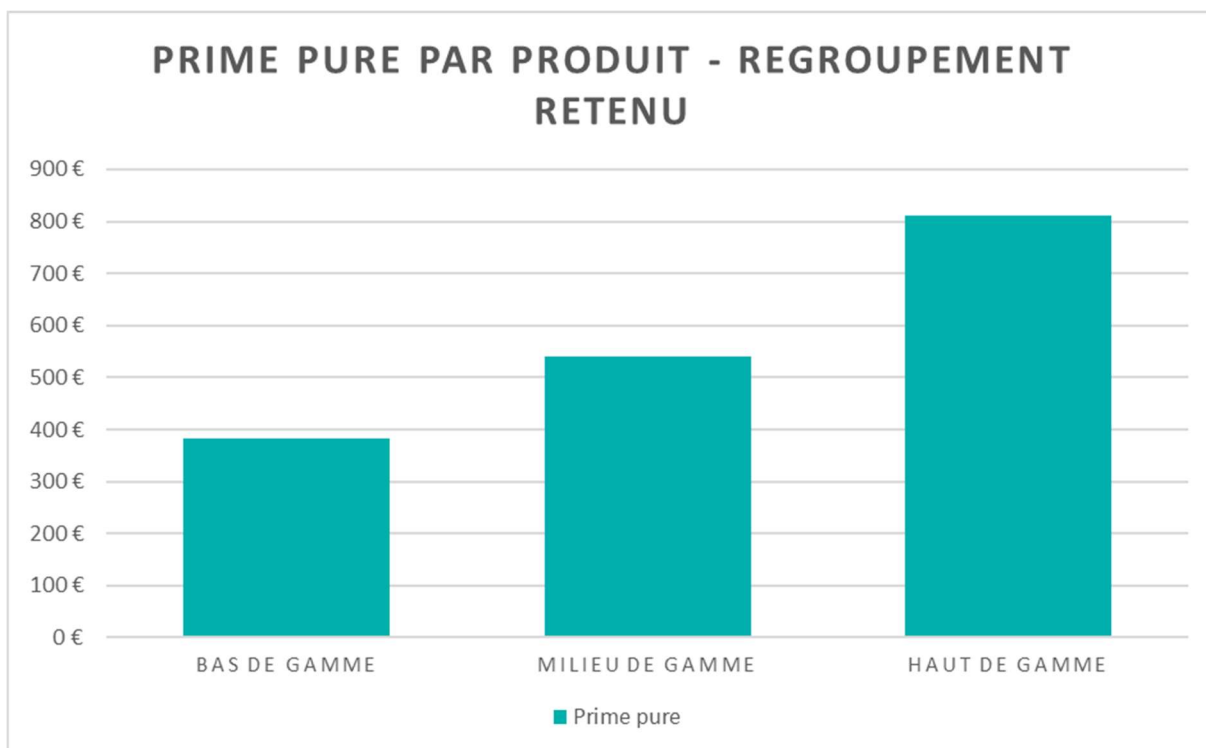


Figure 41 - Evolution de la prime pure moyenne selon le choix final de regroupement des produits

De même que précédemment, et comme attendu, l'augmentation du niveau des garanties entraîne une augmentation des prestations payées.

L'effectif par type de produit est présenté en figure 42. Les produits hauts de gamme sont moins représentés dans le portefeuille.

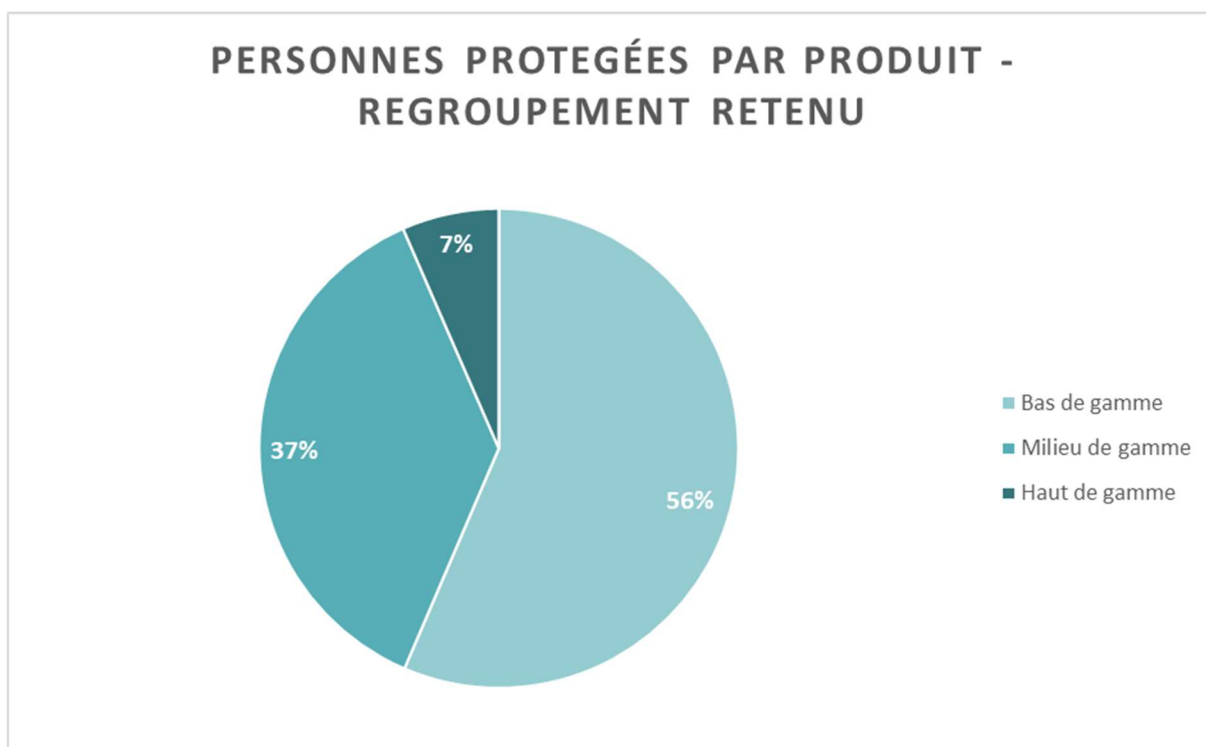
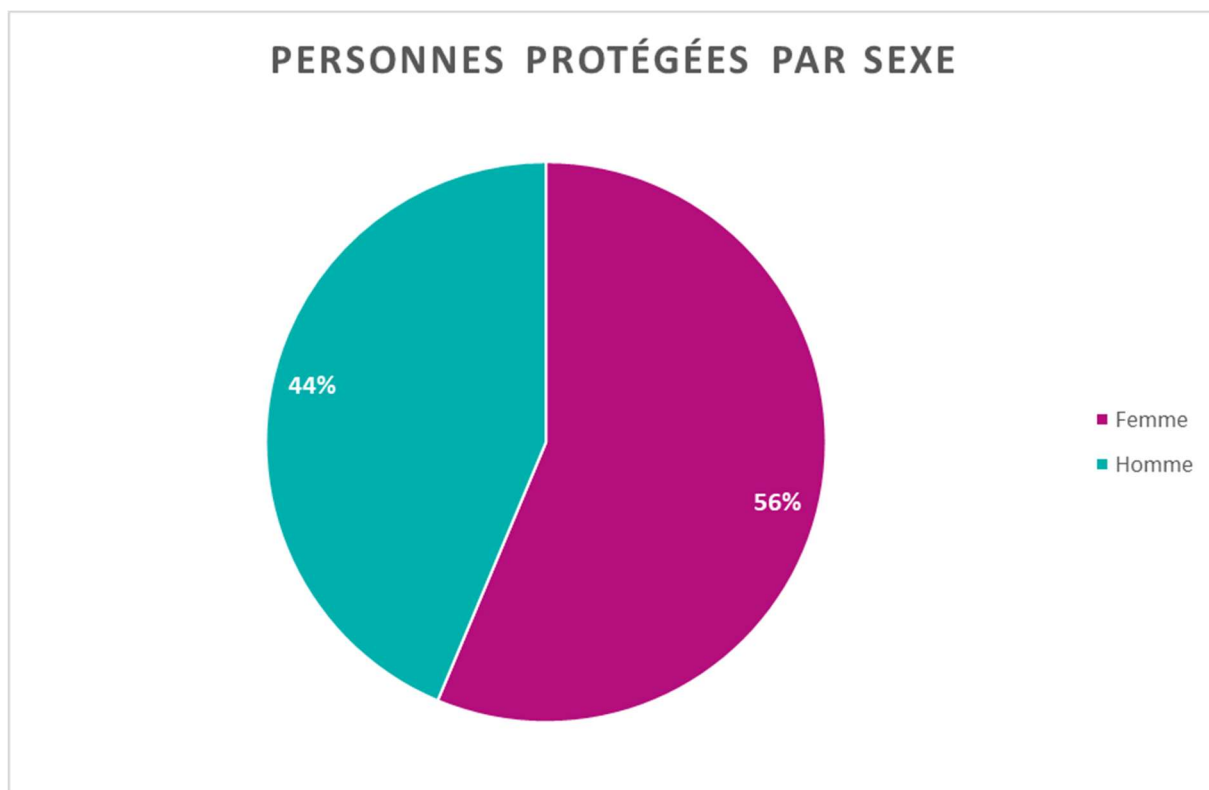


Figure 42 - Répartition des effectifs par produit – Regroupement retenu

**Sexe :**

La base d'étude comporte 57 % de femmes et 43 % d'hommes (figure 43).



*Figure 43 - Répartition Femme / Homme*

En figure 44, nous nous intéressons à la prime pure moyenne par sexe. Cela correspond, d'une part pour les femmes à la somme des primes pures constatées divisée par le nombre de femmes dans le portefeuille et d'autre part pour les hommes à la somme des primes pures constatées, divisées par le nombre d'hommes dans le portefeuille. Cette analyse montre une prime pure moyenne plus élevée pour les femmes que pour les hommes. Cela semble confirmer que le sexe a un impact sur le niveau de la prime pure.

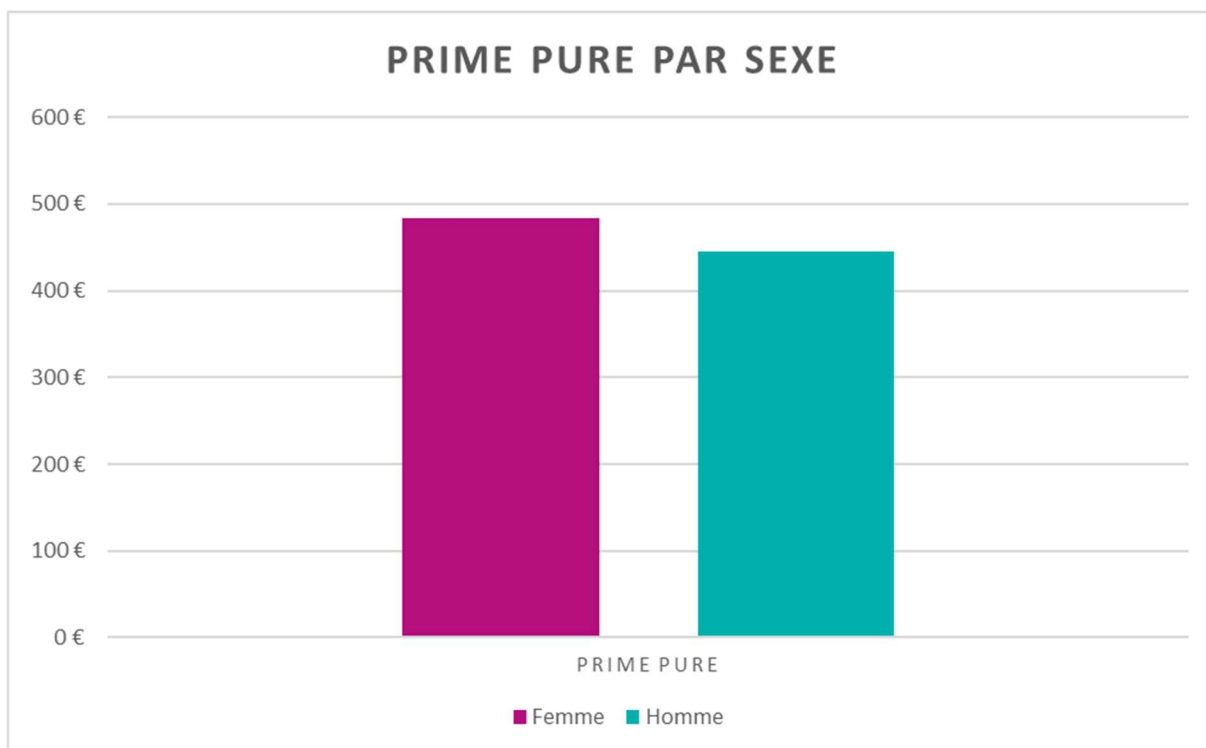


Figure 44 – Prime pure par sexe

**Tranche d'âge :**

Après le calcul de l'âge par génération, nous avons procédé à une segmentation par tranche d'âge. La répartition des individus par tranche d'âge est présentée en figure 45.

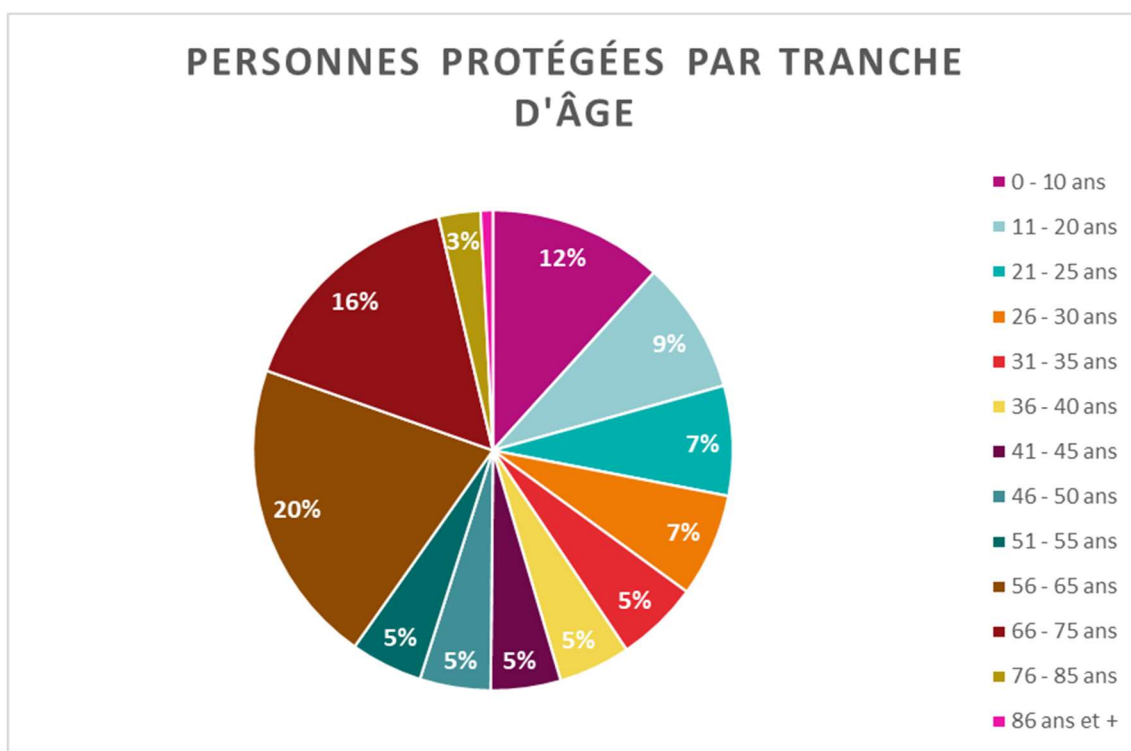


Figure 45 - Répartition des PP par tranche d'âge

Nous constatons que la tranche d'âge la plus représentée est celle des 55 – 65 ans. La moyenne d'âge des personnes protégées est de 44 ans.

En outre, l'analyse de l'évolution de la prime pure moyenne en fonction de la tranche d'âge est présentée en figure 46.

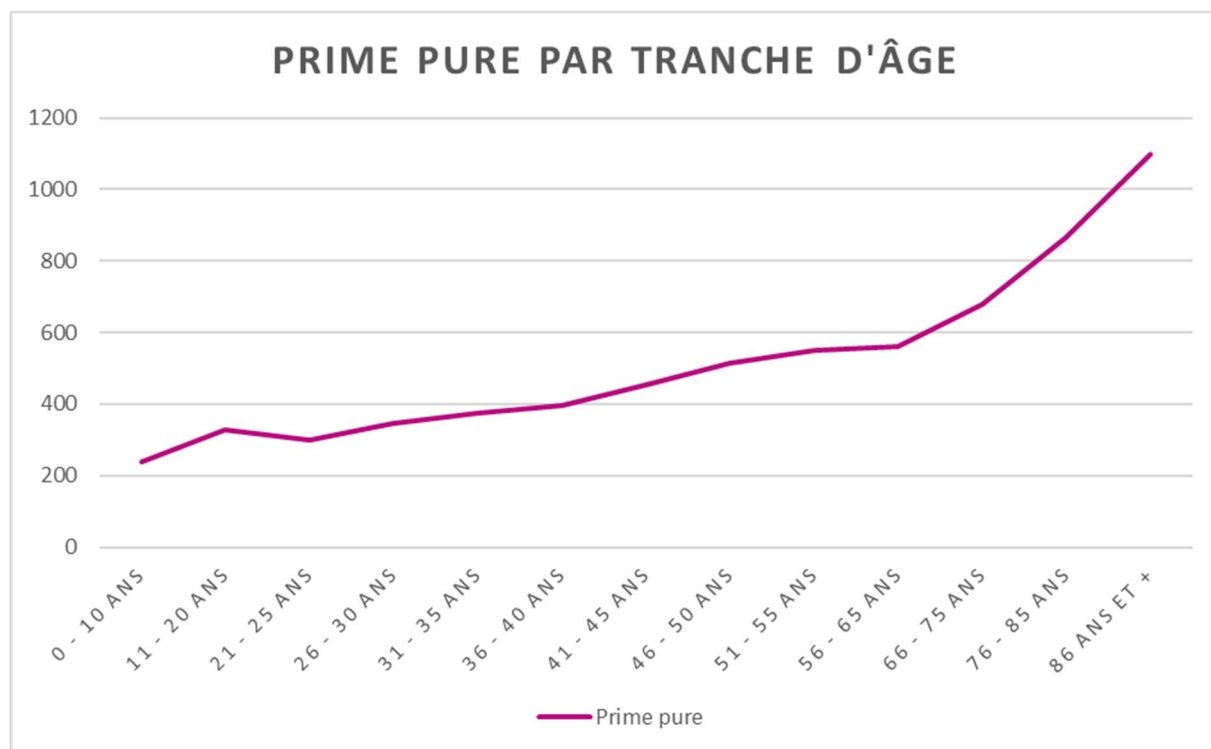


Figure 46 - Evolution de la prime pure en fonction de l'âge

Nous pouvons constater que la croissance de l'âge semble engendrer une croissance des prestations payées à partir de 20 ans. Un saut sur la prime pure est constaté sur la tranche d'âge entre 10 et 20 ans. Cela correspond notamment à l'augmentation des frais d'orthodontie et de l'optique à ces âges.

### **Adhésion :**

Nous avons réalisé une étude de la sinistralité par cohorte afin de déterminer si la durée d'adhésion a un impact sur la prime pure moyenne. En assurance individuelle, il existe une antisélection plus marquée qu'en assurance collective, les adhérents souscrivant aux niveaux les plus adaptés à leurs besoins et à leurs moyens. Aussi, intuitivement, le choix d'une couverture semble pouvoir être associé à un besoin de consommation à court terme. C'est pourquoi, nous avons réalisé une analyse du coût technique 2019 en fonction de l'année d'adhésion.

Nous avons, pour l'année de survenance 2019, calculé la prime pure (également appelé coût technique) en fonction de l'année d'adhésion des personnes protégées. Si nous avons fait l'analyse au global de l'offre nous l'avons également réalisé par produit. Nous avons présenté à la figure 47 les effectifs concernés, non seulement pour se rendre compte de la significativité des produits présentés (97% de l'exposition de l'offre), mais aussi pour analyser le stock de personnes protégées par cohorte. Nous pouvons ainsi déduire que la cohorte 2019 représente 16% de l'exposition globale 2019 et que la cohorte 2016 représente moins de 7% de l'exposition globale.

Cohorte	Flex 11	Flex 12	Flex 21	Flex 22	Flex 23	Flex 32	Flex 33	Flex 34	Flex 43	Flex 44	Flex 45	Flex 54	Flex 55	Flex 56	Flex 65	Flex 66	Total
2013	1 960	2 330	214	521	2 577	97	349	936	59	101	268	20	43	137	4	57	9 673
2014	3 289	3 948	374	808	4 328	160	478	1 614	80	184	383	39	65	233	14	83	16 083
2015	4 121	4 637	490	937	5 125	245	621	1 993	109	272	520	46	115	275	34	235	19 775
2016	5 360	4 930	537	952	5 352	214	507	1 844	74	138	404	29	82	218	13	111	20 763
2017	7 623	5 943	744	1 083	5 813	244	466	2 130	112	166	474	33	83	333	7	144	25 399
2018	10 169	6 701	906	1 139	6 010	320	492	2 560	191	152	620	51	87	438	19	117	29 971
2019	8 214	5 079	672	855	4 377	239	321	1 830	112	100	445	35	49	315	13	55	22 712
<b>Total</b>	<b>40 737</b>	<b>33 568</b>	<b>3 937</b>	<b>6 296</b>	<b>33 583</b>	<b>1 518</b>	<b>3 235</b>	<b>12 906</b>	<b>736</b>	<b>1 111</b>	<b>3 114</b>	<b>254</b>	<b>523</b>	<b>1 950</b>	<b>104</b>	<b>802</b>	<b>144 376</b>

Figure 47 - Répartition du personnes protégées par cohorte - année de soins 2019

En figure 48, nous avons présenté le coût technique par cohorte.

## Coût technique

Cohorte	Flex 11	Flex 12	Flex 21	Flex 22	Flex 23	Flex 32	Flex 33	Flex 34	Flex 43	Flex 44	Flex 45	Flex 54	Flex 55	Flex 56	Flex 65	Flex 66	Total
2012	305 €	370 €	290 €	360 €	475 €	382 €	543 €	569 €	485 €	601 €	594 €	1 042 €	1 030 €	768 €	667 €	823 €	430 €
2013	322 €	420 €	223 €	374 €	495 €	432 €	565 €	574 €	729 €	653 €	624 €	870 €	856 €	921 €	462 €	836 €	454 €
2014	299 €	395 €	289 €	427 €	496 €	316 €	468 €	590 €	552 €	566 €	679 €	850 €	732 €	668 €	675 €	863 €	444 €
2015	308 €	419 €	338 €	417 €	518 €	413 €	572 €	587 €	855 €	613 €	723 €	795 €	972 €	660 €	1 391 €	1 019 €	450 €
2016	313 €	393 €	346 €	425 €	488 €	480 €	542 €	567 €	649 €	693 €	686 €	1 067 €	777 €	680 €	369 €	757 €	426 €
2017	323 €	381 €	351 €	420 €	461 €	582 €	562 €	558 €	585 €	673 €	716 €	1 089 €	640 €	861 €	495 €	736 €	418 €
2018	354 €	426 €	429 €	450 €	513 €	949 €	679 €	665 €	1 128 €	820 €	871 €	2 164 €	1 398 €	1 063 €	3 094 €	1 167 €	477 €
<b>Total</b>	<b>322 €</b>	<b>401 €</b>	<b>338 €</b>	<b>415 €</b>	<b>492 €</b>	<b>528 €</b>	<b>553 €</b>	<b>586 €</b>	<b>707 €</b>	<b>646 €</b>	<b>707 €</b>	<b>1 121 €</b>	<b>864 €</b>	<b>814 €</b>	<b>995 €</b>	<b>862 €</b>	<b>442 €</b>

Figure 48 - Coût technique par cohorte - année de soins 2019

Ainsi, au global de l'offre le coût technique par cohorte plus important en 2019 que pour les autres années. En revanche, la croissance n'est pas marquée sur l'ensemble des années. Ainsi, les cohortes 2014 et 2016 seraient a priori plus consommatrices que les autres cohortes (hormis 2019) sur la période d'étude. Si une tendance se dégage, tant sur l'offre au global que sur chacun des produits, c'est que les personnes protégées ayant adhérées en 2019 ont une consommation 2019 significativement plus élevée que les personnes protégées ayant souscrit les années précédentes (de l'ordre de +9% par rapport à la moyenne arithmétique des autres années). De ce fait, le coût technique semble bien être plus important lors de la première année d'adhésion.

Si effectivement cette étude ne permet pas de conclure directement à un impact certain de la durée d'adhésion sur le coût technique, de même que pour les variables présentées précédemment, compte tenu des effets croisés potentiels (répartition homme / femme, âge moyen par cohorte, etc.), il semble intéressant de prendre en compte dans notre modélisation future le fait d'être en première année de souscription ou non (afin notamment de neutraliser cet impact sur les départements qui auraient connu une forte souscription).

Aussi, nous avons créé une variable indiquant si la consommation associée à un PP représente sa première année de consommation ou non. Ainsi la variable créée est la suivante :

$$Adhésion = \begin{cases} \text{adhésion} \leq 1 \text{ an si Date adhésion} \geq 31/12/(\text{année} - 1) \\ \text{adhésion} > 1 \text{ an si Date adhésion} < 31/12/(\text{année} - 1) \end{cases}$$

la variable « année », représentant l'année de soins à laquelle la prestation est constatée. La part d'adhésion inférieure ou égal à 1 an, représente 19% des personnes protégées.

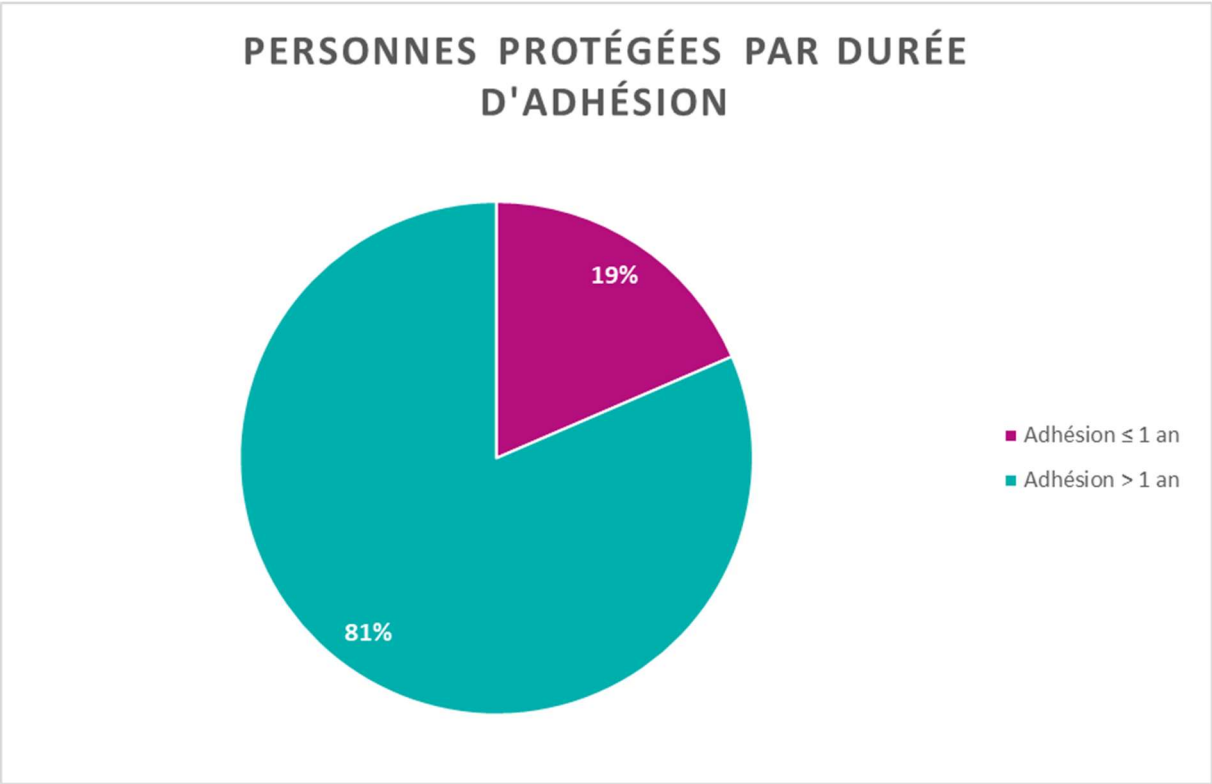


Figure 49 – Répartition Adhésion > à 1 an / ≤ à 1 an

En figure 50 nous nous intéressons à la prime pure moyenne en fonction de l'adhésion. Ce graphique confirme la tendance constatée dans l'étude par cohorte. Ainsi, l'adhésion récente présente une prime pure moyenne supérieure de près de 20% par rapport à la prime pure des personnes protégées ayant adhérées depuis plus longtemps.

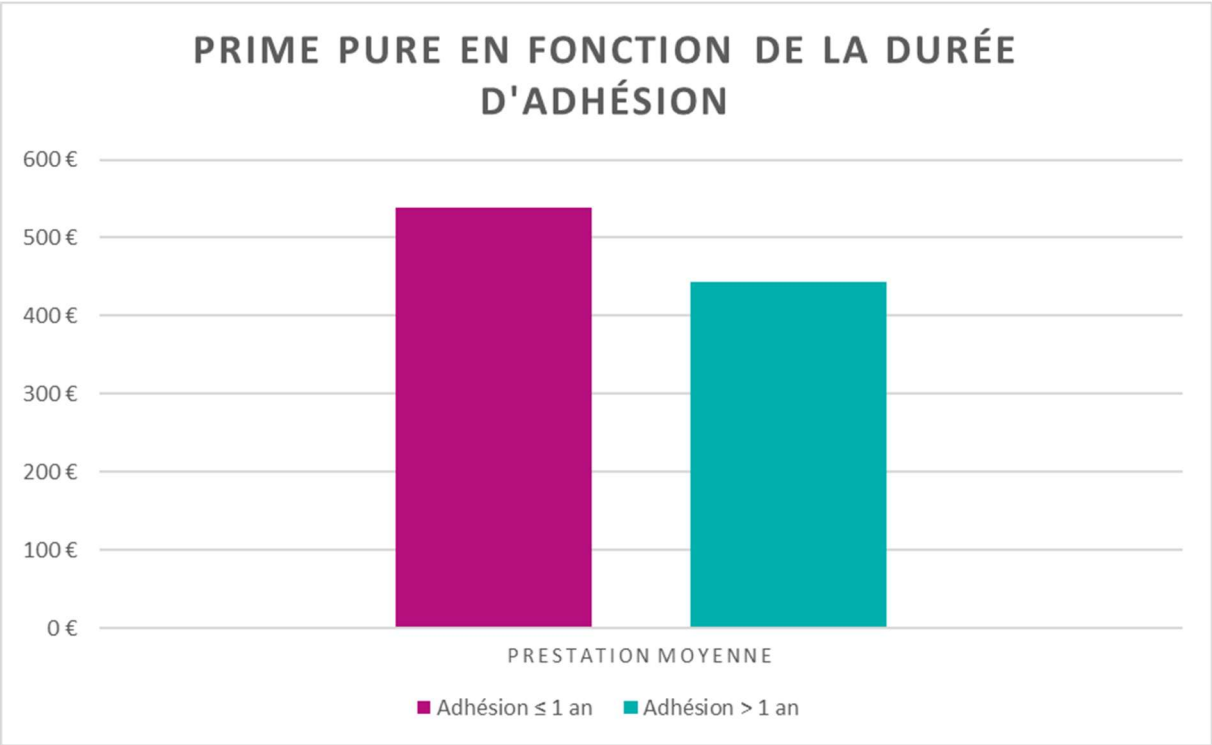


Figure 50 – Prime pure en fonction de la durée d'adhésion

### Type PP :

Nous avons aussi regardé comment le type de personnes protégées pouvait influencer le coût technique.

Les personnes protégées sont de quatre types différents :

- AP : Assuré Principal,
- CJ : Conjoint,
- EF : Enfant,
- AU : Autres (ascendant notamment).

Comme nous pouvons le voir sur la figure 51, le type de pp « Autres » est trop peu représenté pour être analysé.

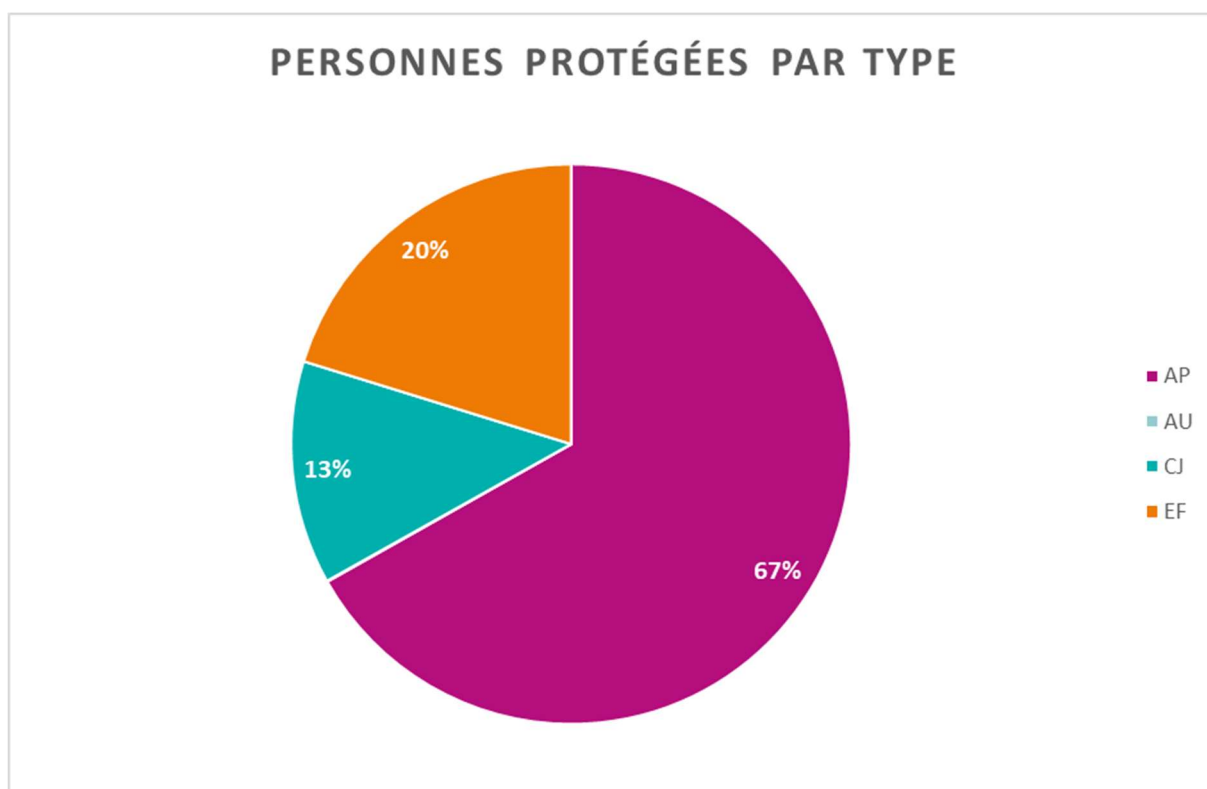


Figure 51 - Répartition par type de personnes protégées

La figure 52 représente l'impact du type de personnes protégées sur la prime pure. Si l'impact est marqué entre les enfants et les autres types, cet impact est probablement déjà porté par l'âge. Effectivement, la différence entre assuré principal et conjoint n'est que peu marquée.

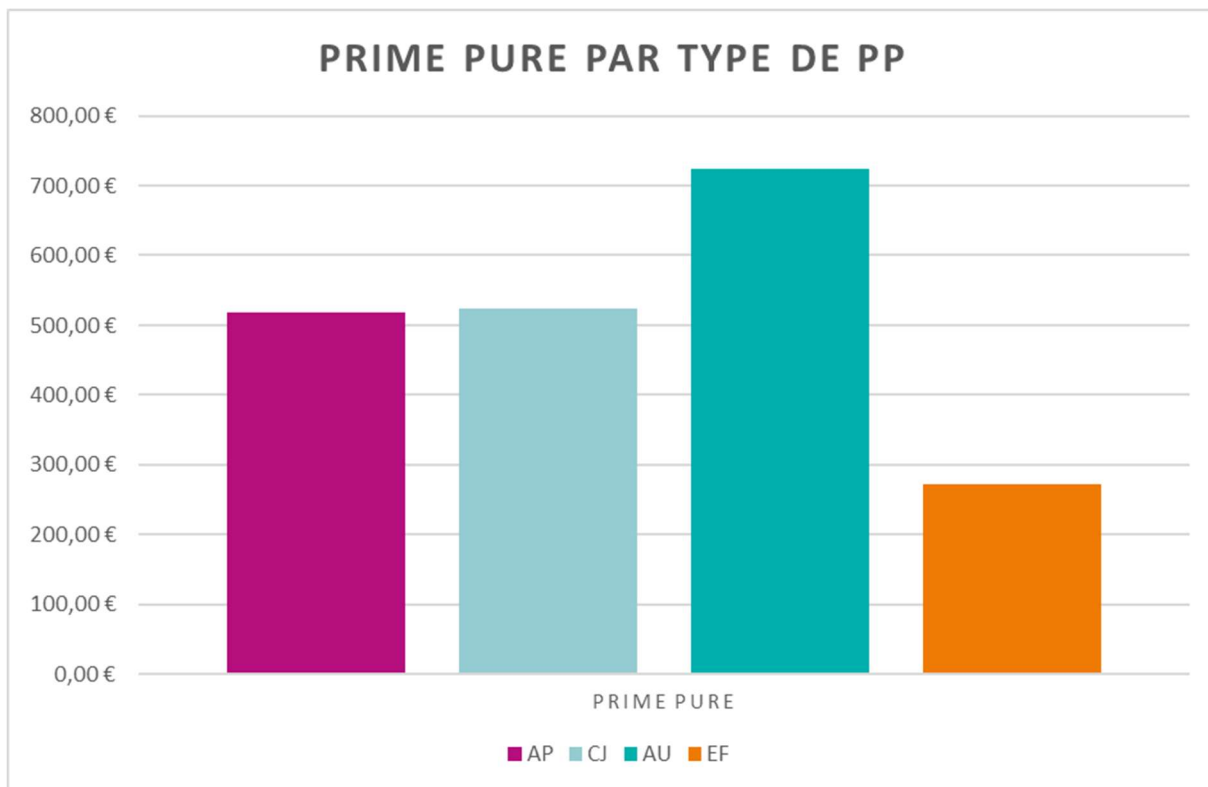


Figure 52 – Evolution de la prime pure en fonction du type de personne protégées

L'analyse des corrélations entre les variables semble être indispensable pour choisir de retenir ou non le type de PP dans la modélisation future.

**Département :**

En figure 53, nous avons analysé l'impact des départements sur la prime pure moyenne

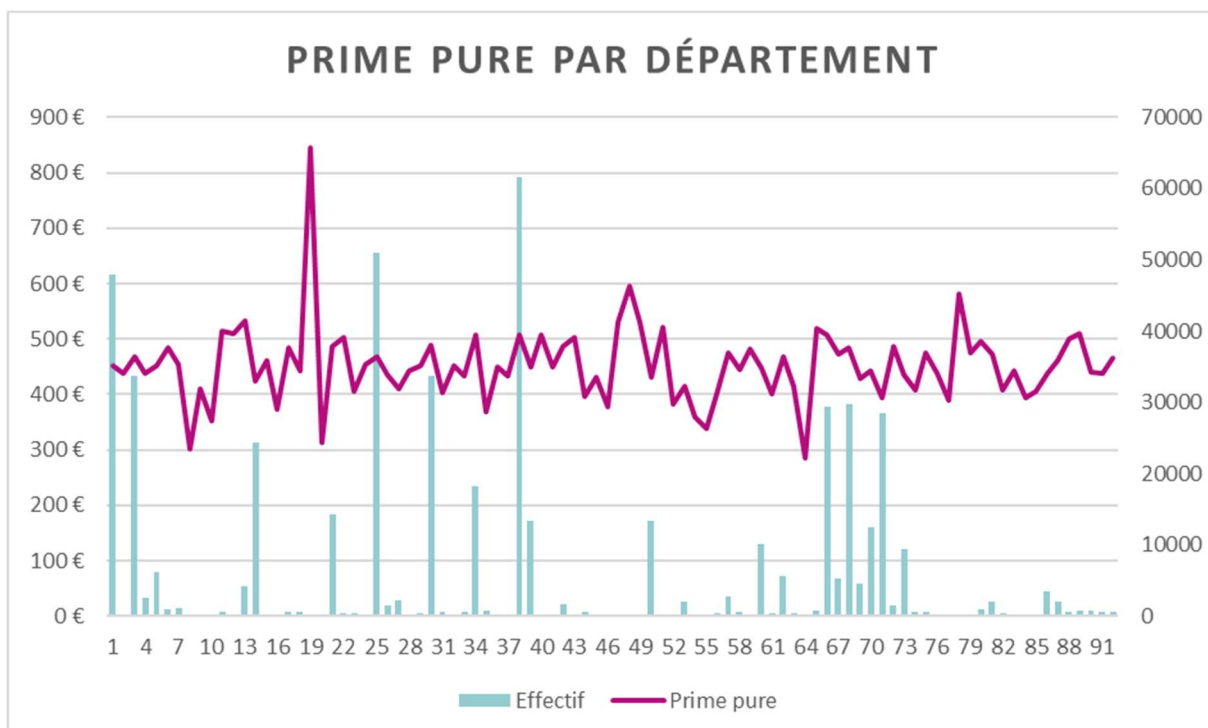




Figure 53 – Evolution de la prime pure en fonction du département ou groupe de département

L'analyse de la prime pure par département montre une variation allant de 200€ à plus de 750 €. En outre, l'histogramme des effectifs associés présente une grande disparité entre les départements du fait notamment de la faible exposition sur certains départements. Aussi, compte tenu des données, la sous-représentation des informations pour certains départements ne permettra pas, sans une analyse plus détaillée, de tirer des conclusions suffisamment robustes.

C'est pourquoi, nous avons pris la décision de réaliser un premier rapprochement pour les départements ayant moins de 300 PP. Ces départements ont été rattaché par ordre de priorité :

- si dans une région (au sens des régions avant 2016), tous les départements présentent moins de 300 PP, ils sont rattachés entre eux,
- si dans une région, plusieurs départements présentent moins de 300 pp ils sont rattachés au département présentant le plus de pp dans la région,
- la Corse étant faiblement représentée elle a été rattachée au département de la région voisine présentant le plus de pp (département 13).

Pour ne pas perdre la notion de ces rattachements, les départements recodées seront nommés par le numéro des départements reliés par un espace. Une fois ces rapprochements réalisés, il semble se dégager une différence assez nette de consommation entre les différents départements ou groupe de départements créés. Un impact du département sur la prime pure se dessinant, cela justifie l'approfondissement de notre analyse sur ce point.

Dans ce contexte, le regroupement des départements par zone semble pouvoir permettre une correction, au moins partielle de cette problématique de manque de données comme nous pouvons le voir sur la figure 54.

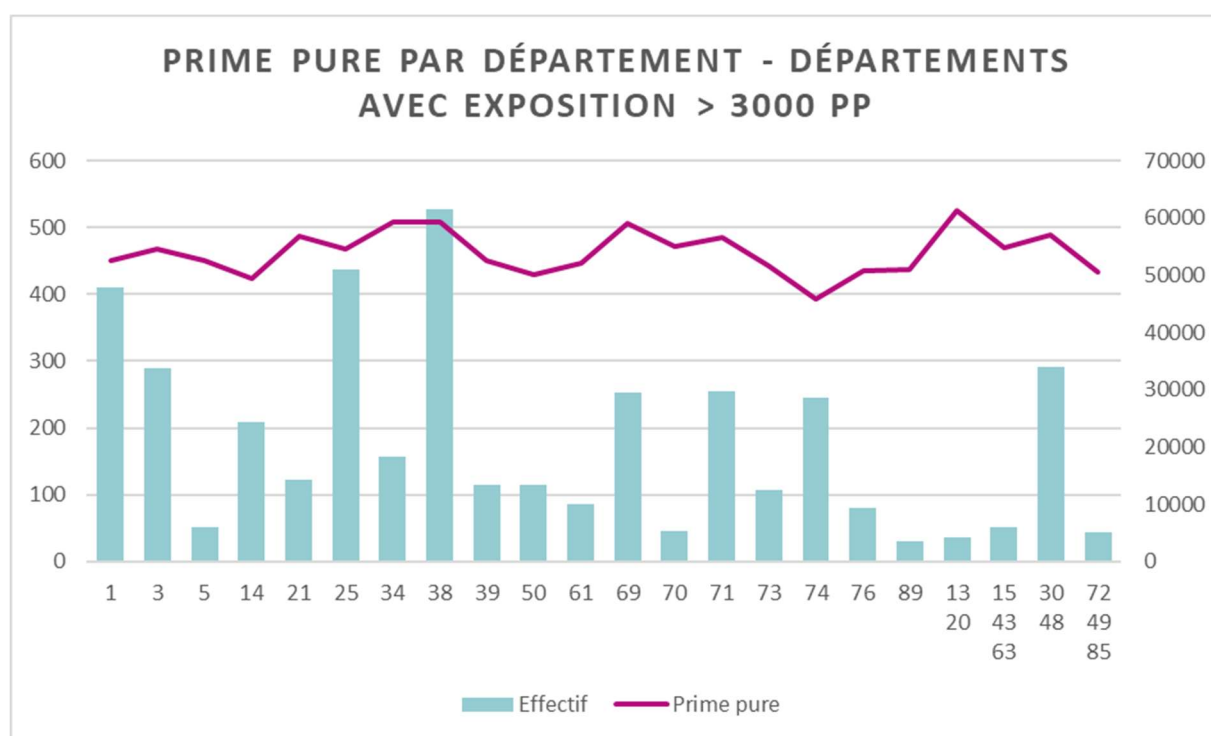


Figure 54 – Evolution de la prime pure en fonction du département ou groupe de département (département avec exposition > 3000 PP)

Nous pouvons remarquer qu'en général, pour la constitution des zoniers il existe plusieurs niveaux d'agrégation pour les variables géographiques (région / départements / communes). En effet, cela permet de créer des zones mixtes toutes suffisamment représentées en choisissant le bon niveau de segmentation. Cependant, si des écarts de tarifs peuvent aisément se justifier entre ville et campagne, il est moins clair de pouvoir justifier auprès d'un adhérent que son déménagement dans le centre de l'agglomération impacte fortement à la hausse son tarif de complémentaire santé<sup>9</sup>. Aussi, nous avons fait le choix de rester sur une décomposition sur des groupements de départements comme nous le verrons dans la partie suivante.

### 3.2. Zonier en place

L'objectif d'un zonier est de mettre en place un tarif segmenté tout en conservant la mutualisation.

Sur la gamme Flex'ADREA, un zonier est mis en œuvre afin de prendre en compte les différences de consommation entre différentes zones géographiques d'une part et, d'autre part, d'intégrer les orientations commerciales définies sur certains départements et permettre d'atteindre certains objectifs de vente.

Le zonier consiste en un coefficient qui sera appliqué sur la cotisation par âge et par produit présenté dans la partie II.2.1). Nous présenterons d'abord le zonier technique dans lequel quatre zones ont été identifiées.

---

<sup>9</sup> Il est à noter que si la différenciation du tarif commercial à une maille aussi fine que la commune, par exemple, est difficile à justifier en complémentaire santé, c'est moins le cas en assurance IARD.

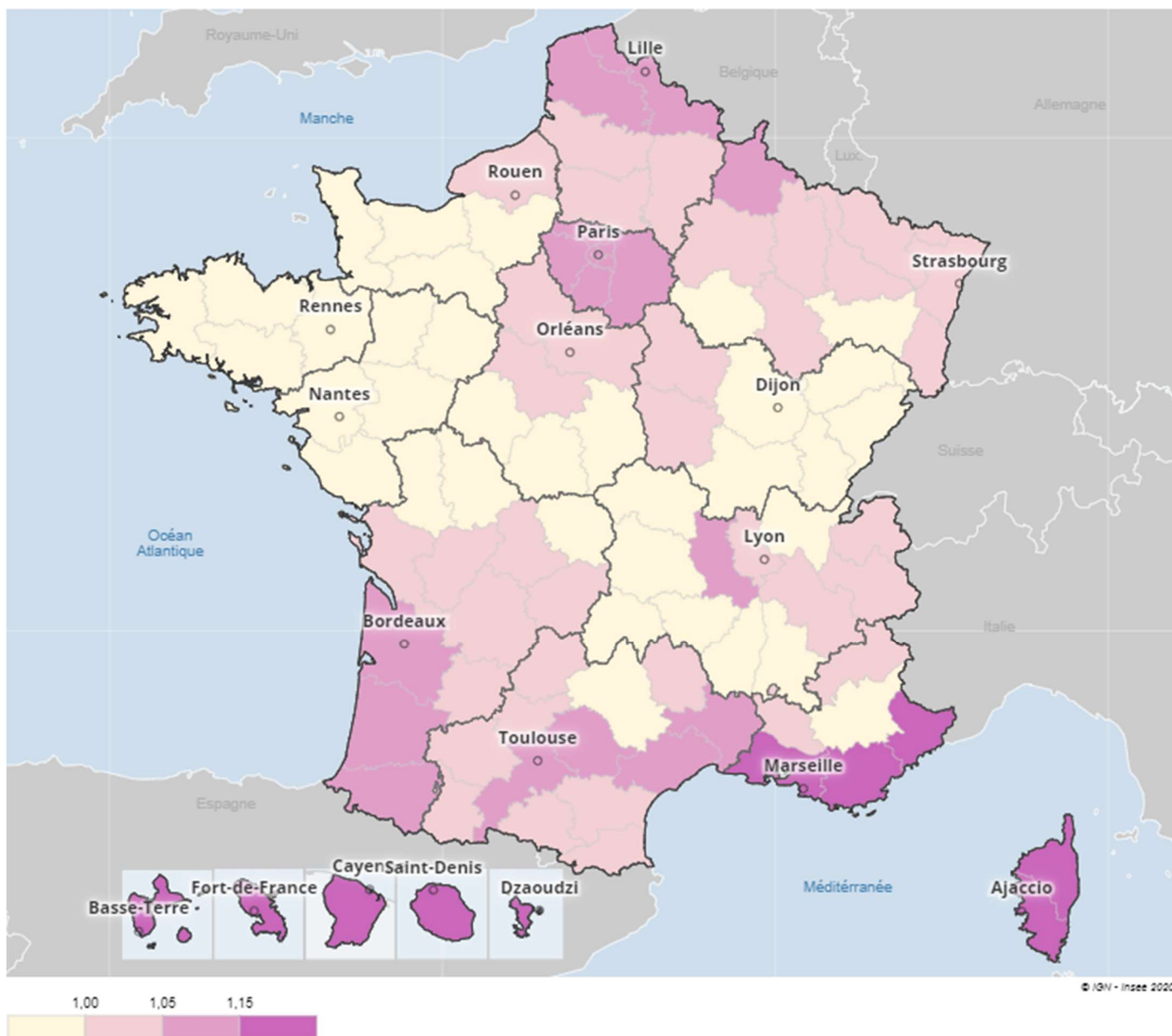


Figure 55 - Présentation du zonier actuel – Zonier technique

Le coefficient varie de 0,95 à 1,15. Par exemple, en considérant la consommation associée à un coefficient de 100%, il a été estimé que dans les Bouches-du Rhône ou les départements de la Corse par exemple, le niveau de consommation était plus élevé de 15%. Ceci peut se justifier par un recours plus régulier aux prestations de soins ou par une présence plus importante de soins plus coûteux (médecins conventionnés de secteur 2 avec autorisation de dépassement d'honoraires par exemple). Le zonier technique, objet d'étude de ce mémoire est présenté en figure 55<sup>10</sup>.

Cependant, comme évoqué précédemment, certaines zones ont été considérées comme cible prioritaire de développement commercial ou au contraire de zones non prioritaires et peuvent se voir appliquer un coefficient complémentaire conduisant à l'élaboration des 7 zones détaillées ci-dessus. Une fois la modification en lien avec les objectifs de vente appliquée, les coefficients associés au zonier varient de 0,90 à 1,15.

<sup>10</sup> Le zonier en place constitue une approche par barème à appliquer et ne distingue pas comme pour un modèle coût / fréquence les deux effets.

Le découpage actuel et le coefficient appliqué sur la prime associée est présentée en figure 56.

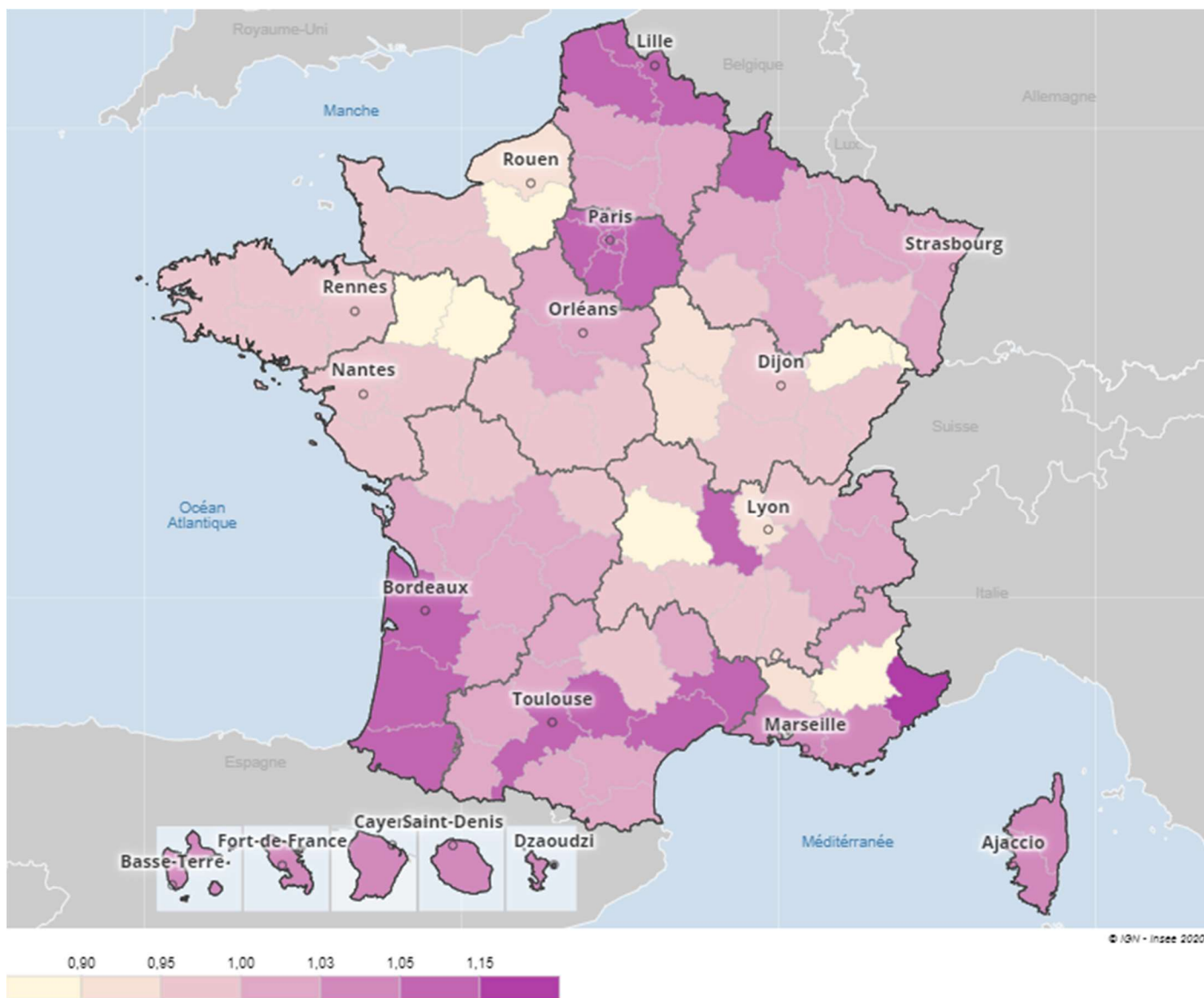


Figure 56 - Présentation du zonier actuel – Zonier commercial

Il existe aujourd’hui sept niveaux de tarifs commerciaux :

- Zone 1 : 04 – 27 – 53 – 63 – 70 – 72 – 90,
- Zone 2 : 58 – 69 – 76 – 84 – 89,
- Zone 3 : 01 – 03 – 07 – 10 – 12 – 14 – 15 – 18 – 21 – 22 – 23 – 25 – 26 – 29 – 35 – 36 – 37 – 39 – 43 – 44 – 49 – 50 – 56 – 61 – 71 – 79 – 85 – 86 – 88,
- Zone 4 : 02 – 05 – 09 – 11 – 16 – 17 – 19 – 24 – 28 – 32 – 38 – 41 – 45 – 46 – 47 – 48 – 51 – 52 – 54 – 55 – 57 – 60 – 65 – 66 – 67 – 68 – 73 – 74 – 80 – 82 – 87,
- Zone 5 : 13,
- Zone 6 : 08 – 30 – 31 – 33 – 34 – 40 – 42 – 59 – 62 – 64 – 75 – 77 – 78 – 81 – 91 – 92 – 93 – 94 – 95
- Zone 7 : 06 – 83 – DOM / TOM et étranger.

Aujourd'hui, la mise à jour du zonier est réalisée en fonction de deux types de demande :

- d'une part, la direction technique, en cas de constat d'une zone ou d'un département dans une zone plus ou moins sinistrée que la prévision en lien avec le coefficient appliqué, peut proposer la mise à jour du zonier.
- d'autre part, la direction du développement, en cas de constat d'un « dépositionnement » tarifaire justifié soit par l'absence de souscription sur une zone soit par un benchmark prouvant le tarif trop cher pratiqué, peut demander l'étude d'une révision du zonier.

La figure 57, présente un extrait de la dernière présentation en date des conclusions d'une analyse réalisée sur le zonier ayant conduit à sa modification.

- ▶ 1. Zonier par produit (sur les 5 principaux produit, soit 87% des PP):
  - Le P/C moyen est de 81%. Pour les zones 0.95, 1 et 1.05, le P/C est très proche de ces 0.81. Donc le zonier est au global bien adapté pour ces trois zones.
  - Le P/C moyen de la zone 1.15 (qui correspond à Île de France) est de 0.76. Donc une sur-tarification qui pourrait être portée à 1.1 au lieu de 1.15.
- ▶ 2. Adéquation du zonier par département (sur les 14 principaux départements, soit 83 % des PP) :
  - Les écarts du P/C des départements avec le P/C global conduisent à préconiser techniquement les changements de zone suivants:
    - Départements 25, 71 (0,95 actuellement), 38 (1) et 34 (1,05) doivent monter d'un niveau ;
    - Département 74 peut baisser d'un niveau.
- ▶ 3. Simulation des changements de zones identifiés :
  - Après changements pour les 5 départements et après bascule de la zone 1.15 en 1.1, le P/C général s'améliore à 0,80 au lieu de 0,81.
  - Tous les produits voient leur P/C amélioré (tenant compte des changements de zonier à la hausse ou à la baisse).

Figure 57 - Extrait de décision : Mise à jour du zonier

Comme nous pouvons le voir, les critères de décisions appliqués ce jour se basent sur l'analyse du ratio entre les prestations et les cotisations hors taxes par zone et par département. Cela permet à la fois de valider ou d'ajuster le positionnement de certaines zones mais aussi de réaliser des changements de zone pour certains départements.

Cependant, ces mises à jour se basant sur l'expérience portefeuille, le point évoqué précédemment sur l'absence d'information sur certains départements n'est pas corrigé par ces analyses. Ainsi, des études d'évolutions sont mises en œuvre sur des départements dans lesquels le taux de transformation, correspondant au nombre de souscription relativement au nombre total de devis, est relativement bas par exemple. Un taux de transformation plus bas dans un département en particulier peut se justifier, entre autres, par un tarif potentiellement hors marché. Dans ce cas, le zonier peut être modifié pour permettre une amélioration des ventes. En outre, le suivi technique en considérant le P/C par département, peut être expliqué par d'autres éléments et doit être utilisé avec précautions. Ainsi, un P/C dégradé dans un département pourrait se justifier par une souscription importante de produit déficitaire, ou sur des tranches d'âge porteuses de mutualisation.

Ces méthodes de mises à jour, si elles sont potentiellement efficaces, ne sont pas nécessairement justifiées par des arguments techniques robustes et peuvent générer des mises à jour successives contradictoires difficiles à présenter à l'adhérent. Par exemple, le manque d'effectifs pourrait indiquer un P/C favorable à tort conduisant à une mise à jour à la baisse du zonier une année alors que l'année d'après, la souscription par un nombre de PP plus important

conduirait à une mise à jour à la hausse du zonier. Dans un souci de cohérence commerciale, les mises à jour restent assez marginales.

Le but de ce mémoire est dans un premier temps de confronter le zonier actuel à une méthode de construction théorique, et dans un deuxième temps de s'intéresser aux conséquences opérationnelles de la modification d'un zonier de manière globale. Nous utiliserons des modèles couramment utilisés dans la tarification en général et en santé en particulier dont la théorie sera présentée dans la partie suivante.

## PARTIE III : UTILISATION DES MODELES LINEAIRES GENERALISES (GLM) EN SANTE ET METHODE DE CONSTRUCTION DE ZONIER

Nous avons présenté à la partie précédente le zonier actuel. Cependant, afin d'essayer de corriger, au moins partiellement, les limites évoquées, non seulement sur le suivi et la mise à jour du zonier mais aussi sur l'absence de données dans certaines zones, nous allons utiliser des méthodes théoriques permettant la construction d'un zonier. Ce sujet a été traité un certain nombre de fois dans la littérature (comme dans (Beraud-Sudreau, 2017), (Ferrier, 2016), (Fesquet, 2015), (Poubennec, 2017)). Nous allons mettre en œuvre une méthode se basant sur des estimations par modèles linéaires généralisés (GLM). Après avoir présenté dans un premier temps les GLM, nous nous pencherons sur les méthodes courantes de construction de zoniers.

### 1. Présentation théorique des GLM

Cette partie est grandement inspirée de la présentation des modèles linéaires généralisés réalisée dans (Charpentier & Denuit, 2004).

L'objectif de la modélisation est d'identifier l'impact de différentes variables sur notre variable d'intérêt. Aussi, les modèles linéaires généralisés se présentent, de même que le modèle linéaire, comme suit :

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

avec  $N$  la variable aléatoire représentant le nombre de sinistres d'un individu,  $C_i$  le coût unitaire du  $i^{\text{ème}}$  sinistre et  $S$  la prime pure de l'individu.

Il est couramment considéré l'indépendance entre le nombre de sinistres et leur montant permettant ainsi de construire un modèle de fréquence d'une part et un modèle de coût moyen d'autre part. Si cette hypothèse peut aisément se justifier pour un certain nombre de risque, ce n'est pas aussi simple dans le cas de l'assurance santé. Nous reviendrons sur ce point dans la partie relative à l'application en IV.1.1.

Dans le cadre des modèles de régression, de même que pour les GLM, l'objectif est de prédire des espérances conditionnelles. Ainsi, la prime pure à l'intérieur d'une classe de risque  $X$  peut s'écrire :

$$E[S|X] = E\left[\sum_{i=1}^N C_i|X\right]$$

On cherche ici à modéliser l'espérance conditionnelle de la prime pure.

Pour rappel, dans le cadre d'un modèle linéaire (et par extension, dans le cadre des modèles linéaires généralisés qui nous intéressent ici), nous cherchons à modéliser une variable  $S$  à l'aide d'un certain nombre  $p$  de variables explicatives  $X = (X_1, \dots, X_p)^t$ . La régression linéaire revient à supposer que  $S \sim \text{Nor}(\mu, \sigma^2)$  où  $\mu = X^t \beta$ .

Les modèles linéaires imposent un certain nombre de contraintes et hypothèses parfois non respectées avec la réalité des coûts ou des fréquences de sinistres. Ainsi, dans le modèle linéaire, la densité de la variable à expliquer doit être approximativement Gaussienne. Cette hypothèse pourra être relâchée dans les GLM. Les autres hypothèses telle, la propriété d'homoscédasticité des résidus doit être respectée pour les modèles linéaires comme pour les GLM.

Dans le contexte d'un *GLM*, nous considérons que pour une variable aléatoire  $S$  qui correspond à la variable à expliquer, il existe une relation de la forme suivante, en reprenant les notations précédentes :

$$g(E[S|x_1, \dots, x_p]) = \sum_{k=1}^p \beta_k x_k$$

La fonction  $g$  est appelée fonction de lien du modèle et détermine la relation entre le prédicteur linéaire et l'espérance de la variable expliquée. Bien que le choix de cette fonction  $g$  soit relativement libre, il détermine notamment la nature de la relation entre les variables explicatives et la variable à expliquer. Les fonctions, parmi les plus courantes sont l'identité ( $g(x) = x$ ) permettant l'obtention d'un modèle additif :

$$E[S_i] = \alpha_0 + \sum_{j=1}^p \alpha_j x_{i,j} = \sum_{j=0}^p \beta_j$$

$$\text{avec } \forall 1 \leq j \leq p, \quad \beta_j = \begin{cases} \alpha_0 & \text{si } j = 0 \\ \alpha_j x_{i,j} & \text{sinon} \end{cases}$$

Ou encore le logarithme népérien ( $g(x) = \ln(x)$ ) qui permet l'obtention d'un modèle multiplicatif :

$$\ln(E[S_i]) = \alpha_0 + \sum_{j=1}^p \alpha_j x_{i,j} \iff E[S_i] = \exp(\alpha_0 + \sum_{j=1}^p \alpha_j x_{i,j})$$

$$\iff E[S_i] = \prod_{j=0}^p \beta_j$$

$$\text{avec } \forall 1 \leq j \leq p, \quad \beta_j = \begin{cases} \exp(\alpha_0) & \text{si } j = 0 \\ \exp(\alpha_j x_{i,j}) & \text{sinon} \end{cases}$$

Il est à noter qu'un modèle multiplicatif permet d'obtenir une interprétation des coefficients de régression relativement aisée.

Dans un deuxième temps, il faut spécifier la loi de la variable à expliquer  $S$ . Comme évoqué précédemment, la théorie des *GLM* permet de relâcher l'hypothèse sur le caractère normal de la variable à expliquer  $S$ . Ainsi, seule l'appartenance à la famille exponentielle doit être vérifiée. Les lois appartenant à la famille exponentielle sont les lois à deux paramètres  $\theta$  et  $\varphi$  dont la densité peut s'écrire sous la forme :

$$f(s_i|\theta_i, \varphi) = \exp\left(\frac{s_i \theta_i - b(\theta_i)}{\varphi / \omega_i} + c(s_i, \varphi)\right),$$

avec :

- $\theta$  le paramètre de la moyenne,
- $\varphi$  le paramètre de dispersion lié à la variance,
- $b$  une fonction définie sur  $\mathbb{R}$  deux fois dérivable et de dérivée seconde positive,
- $c$  une fonction définie sur  $\mathbb{R}^2$ ,
- $\omega_i$  est un poids strictement positif.



La moyenne et la variance d'une variable aléatoire dont la densité est de la forme exponentielle sont définies de la façon suivante :

$$E[S] = b'(\theta) \quad \text{et} \quad \text{Var}(S) = b''(\theta) \times a(\varphi)$$

Ces lois sont complètement définies par leur moyenne et leur variance, d'une part et, d'autre part, leur variance est une fonction de leur moyenne.

La famille exponentielle contient notamment les lois suivantes :

- loi Normale,
- loi de Poisson,
- loi Binomiale,
- loi Gamma,
- loi Inverse Gaussienne.

Cependant, il est possible de démontrer qu'un certain nombre de lois appartiennent également à la famille des lois exponentielles (notamment des lois composées ou des lois mélangées). C'est le cas notamment de la loi de Tweedie qui sera présentée en partie III.2.

Les coefficients de régression  $\beta_0, \beta_1, \dots, \beta_p$  ainsi que le paramètre de dispersion  $\phi$  étant inconnu doivent être estimés par exemple par la méthode du maximum de vraisemblance. La log-vraisemblance du modèle à maximiser s'écrit :

$$L_S(\theta|s, \varphi) = \sum_{i=1}^p \ln(f_S(s_i | \theta_i, \varphi))$$
$$L_S(\theta|s, \varphi) = \sum_{i=1}^p \frac{s_i \theta_i - b(\theta_i)}{a(\varphi)} + c(s_i, \varphi)$$

Les équations découlant de ce système doivent être résolues à l'aide de méthodes numériques. Par exemple, dans (Charpentier & Denuit, 2004), la méthode de Newton-Raphson est présentée. Pour ce mémoire, le logiciel R, estime les coefficients d'un *GLM* par la méthode des moindres carrés repondérés itérativement (*Iteratively Reweighted Least Squares*)<sup>11</sup> que nous ne détaillerons pas ici, les méthodes numériques n'étant pas l'objet de ce mémoire.

Si dans ce mémoire nous utilisons les modèles *GLM*, il est important de préciser que la modélisation santé se réalise également par les modèles de machine learning comme dans (Aouizerate, 2012), (Paglia, Phéllippé-Guinvarc'h, & Lenca, 2011) et (Phelippe-Guinvarc'h & Paglia, 2011).

## 2. Loi de Tweedie

La plupart des lois usuelles ne permettent pas d'avoir de valeurs nulles. La loi Gamma est une des lois parmi les plus utilisées pour modéliser les coûts des sinistres. Cependant, en anticipation de la construction du modèle, il est important de s'intéresser à notre base de données. En effet, un certain nombre de personnes protégées ne consomme pas sur l'année. Aussi, afin de prendre en compte cette masse non-consommatrice, il sera difficile de considérer une loi Gamma dans la modélisation. Aussi, la loi de Tweedie qui permet la modélisation en incluant une masse de

<sup>11</sup> <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/glm>

probabilité en zéro semble plus adapté. D'autres modèles permettant la modélisation des masses en zéro, tels que les modèles à inflations de zéros, sont évoqués dans (Planchet & Serdeczny, 2014).

Le modèle de Tweedie est présenté et utilisé assez fréquemment dans la tarification, ainsi dans (Miloé, 2011) (Poubennec, 2017) ou (Ferrier, 2016) et est brièvement décrite ci-dessous. Le modèle s'écrit :

$$f(\varphi; \mu, \phi) = a(y, \phi) \exp\left(\frac{1}{\phi} [y\theta(\mu) - \kappa(\theta(\mu))]\right)$$

$$\theta(\mu) = \begin{cases} \frac{\mu^{1-p}}{1-p} & \text{si } p \neq 1 \\ \log \mu & \text{si } p = 1 \end{cases} \quad \text{et } \kappa(\theta(\mu)) = \begin{cases} \frac{\mu^{2-p}}{1-p} & \text{si } p \neq 2 \\ \log \mu & \text{si } p = 2 \end{cases}$$

$E[Y] = \mu$  et  $Var(Y) = \psi\mu^p = \psi E[Y]^p$ , avec  $\psi$  un paramètre de dispersion  $> 0$ .

En reprenant les notations précédentes et en considérant que  $S$  suit une loi de Tweedie d'espérance  $\mu$  et avec  $s_i$ , sa réalisation, nous devons introduire le paramètre de forme  $p$  tel que  $1 < p < 2$ , alors la densité se présente comme suit :

$$f(s_i | \theta_i, \varphi) = \exp\left(\frac{s_i \theta_i - b(\theta_i)}{\varphi / \omega_i} + c(s_i, \varphi)\right),$$

Avec :

$$\begin{cases} b(\theta) = \frac{\mu^{2-p}}{2-p} \\ c(s_i, \varphi) = -\ln\left(\sum_{n=1}^{+\infty} \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} \times \frac{\lambda^n}{n!} \times s_i^{n\alpha-1}\right) \end{cases}$$

en posant,  $\alpha = \frac{2-p}{p-1}$ ,  $\frac{1}{\beta} = \varphi / \omega_i \times (p-1) \times \mu^{p-1}$  et  $\lambda = \frac{\mu^{2-p}}{\varphi / \omega_i \cdot (2-p)}$  :

La variance de  $S$  s'écrit :

$$Var(S) = \frac{\varphi}{\omega_i} \times \mu^p$$

Pour un certain nombre de valeurs de  $p$  nous pouvons retrouver un certain nombre de lois appartenant à la famille exponentielle et couramment utilisées dans le cadre des *GLM*, comme présenté ci-dessous :

$$\begin{cases} \text{pour } p = 0, & \text{ nous retrouvons une loi Normale,} \\ \text{pour } p \rightarrow 1, & \text{ nous nous rapprocherons d'une loi de Poisson,} \\ \text{pour } p \rightarrow 2, & \text{ nous nous rapprocherons d'une loi Gamma.} \end{cases}$$

Pour des valeurs de  $p$  entre 1 et 2, la loi de Tweedie correspondra à une loi composée Poisson-Gamma. En figure 58, ci-dessous, nous avons représenté l'évolution d'une loi de Tweedie de paramètre  $\mu$  et  $\varphi$  équivalents à la moyenne et à l'écart type de notre base d'étude en faisant varier uniquement le paramètre  $p$ .

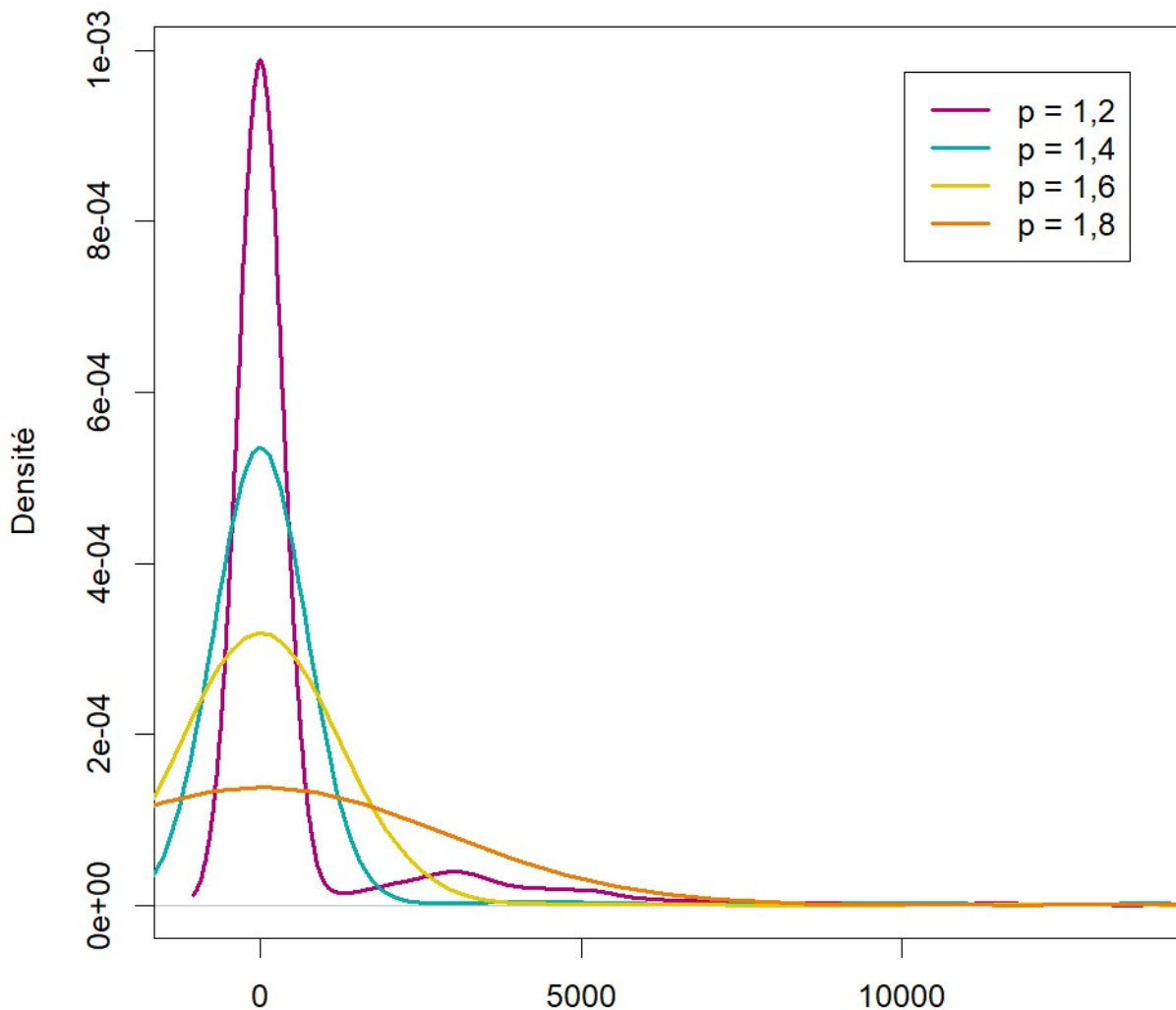


Figure 58 - Loi de Tweedie en fonction de  $p$

### 3. Méthode de construction d'un zonier en utilisant les *GLM*

La théorie des *GLM* permet de construire des modèles d'estimation et de prévision des primes pures associées au risque à l'aide de variables explicatives. Nous pouvons considérer que la modélisation de notre variable à expliquer  $S$  peut être réalisée d'une part par l'intégration de variables explicatives sans lien avec les facteurs géographiques et d'autre part, par des variables explicatives générant directement un impact géographique. L'idée sous-jacente dans la construction d'un zonier est d'isoler la part expliquée par la partie les facteurs géographiques.

Considérons la variable à expliquer  $S_i$  pouvant d'écrire comme suit :

$$\ln(S_i) = \beta_0 + \beta_1 X_1^i + \dots + \beta_p X_p^i + \varepsilon_i$$

L'expression ci-dessus peut s'écrire :

$$S_i = e^{\beta_0} \times e^{\beta_1 X_1^i} \times \dots \times e^{\beta_p X_p^i} \times e^{\varepsilon_i}$$

Suite à une première régression intégrant l'ensemble des variables explicatives, nous obtenons une première estimation de  $S_i$  notée  $\hat{S}_i$  :

$$\hat{S}_i = e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 X_1^i} \times e^{\hat{\beta}_2 X_2^i} \times \dots \times e^{\hat{\beta}_p X_p^i}$$

En considérant que les facteurs géographiques sont portés par la variable  $X_1^i$ , nous obtenons en enlevant cette variable, une deuxième estimation en considérant égaux les coefficients de régression  $\hat{\beta}_0, \hat{\beta}_2, \dots, \hat{\beta}_p$ .

Nous obtenons ainsi une deuxième estimation de notre variable  $S_i$ , notée  $\hat{S}'_i$

$$\hat{S}'_i = e^{\hat{\beta}_0} \times e^{\hat{\beta}_2 X_2^i} \times \dots \times e^{\hat{\beta}_p X_p^i}$$

Nous pouvons alors définir les résidus pour chacune des estimations précédentes et sous l'hypothèse,  $\hat{\beta}_0 = \beta_0, \hat{\beta}_1 = \beta_1, \hat{\beta}_2 = \beta_2, \dots, \hat{\beta}_p = \beta_p$

- On déduit de la première estimation de  $S_i$ , l'expression suivante notée (1) :

$$\frac{S_i}{\hat{S}_i} = \frac{e^{\beta_0} \times e^{\beta_1 X_1^i} \times \dots \times e^{\beta_p X_p^i} \times e^{\varepsilon_i}}{e^{\hat{\beta}_0} \times e^{\hat{\beta}_1 X_1^i} \times e^{\hat{\beta}_2 X_2^i} \times \dots \times e^{\hat{\beta}_p X_p^i}} = e^{\varepsilon_i}$$

- On déduit de la deuxième estimation de  $S_i$ , l'expression suivante notée (2) :

$$\frac{S_i}{\hat{S}'_i} = \frac{e^{\beta_0} \times e^{\beta_1 X_1^i} \times \dots \times e^{\beta_p X_p^i} \times e^{\varepsilon_i}}{e^{\hat{\beta}_0} \times e^{\hat{\beta}_2 X_2^i} \times \dots \times e^{\hat{\beta}_p X_p^i}} = e^{\beta_1 X_1^i} \times e^{\varepsilon_i}$$

Nous pouvons définir des estimations obtenues aux équations (1) et (2) :

$$\begin{cases} S_i = e^{\varepsilon_i} \times \hat{S}_i \\ S_i = e^{\beta_1 X_1^i} \times e^{\varepsilon_i} \times \hat{S}'_i \end{cases}$$

Soit :

$$e^{\varepsilon_i} \times \hat{S}_i = e^{\beta_1 X_1^i} \times e^{\varepsilon_i} \times \hat{S}'_i$$

De l'expression suivante, nous pouvons définir le résidu  $r_i$  pour chaque PP :

$$r_i = \frac{\hat{S}_i}{\hat{S}'_i} = e^{\beta_1 X_1^i}$$

Ainsi,  $r_i$  définissant la part de  $S_i$  expliquée par les facteurs géographiques.

Nous pouvons alors définir par département  $\hat{r}_k^d$ , le résidu d'estimation « moyen » associé à chaque département ou groupe de département  $k$  :

$$\hat{r}_k^d = \frac{\sum_{i=1}^{I^k} e_i r_i}{\sum_{i=1}^{I^k} e_i}$$

en notant  $e_i$  l'exposition associée au PP  $i$ , et  $I^k$  le nombre d'assurés dans le département  $k$ .

Ainsi, nous réalisons dans un premier temps une modélisation globale de la variable à expliquer en intégrant tant les variables explicatives géographiques que les autres variables. Dans un deuxième temps, nous récupérons l'estimation de la variable à expliquer sans les coefficients de régression liés aux départements mais basée sur les mêmes coefficients de régression. Nous pouvons ainsi extraire la part résiduelle intégrant, en plus d'une part non expliquée par le premier modèle, l'effet des variables géographiques.

Une fois cette part résiduelle calculée, nous pouvons constituer une base par département permettant la classification moyenne du résidu. Sur ces résidus moyens par département ou groupe de département, nous appliquerons ensuite des méthodes de classification afin de créer les différentes classes de risque et donc le zonier.

Ainsi, nous pouvons résumer notre méthode avec la figure ci-dessous, inspirée de celle présentée dans (Ferrier, 2016).

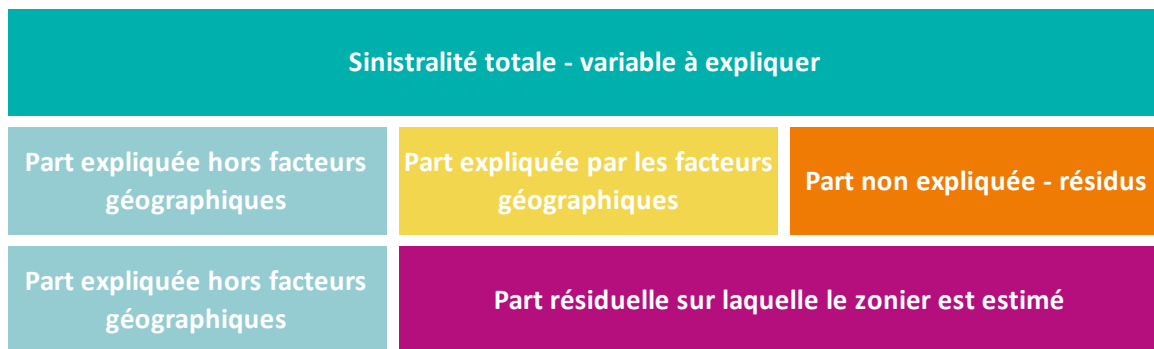


Figure 59 - Méthode de construction d'un zonier

## PARTIE IV : APPLICATION - MODELISATION DU RISQUE SANTE

La théorie et la méthode de travail ayant été présentée à la partie III.1, nous présenterons dans cette partie, leur application sur notre base d'étude présentée en partie II.1.2. Ainsi, après avoir présenté l'application des *GLM* sur un modèle complet, nous appliquerons sur les résidus du modèle excluant les facteurs géographiques, différentes méthodes de classification dans le but de créer plusieurs zoniers. Enfin, nous les comparerons les uns par rapport aux autres afin de sélectionner la meilleure segmentation.

### 1. Modèle complet

#### 1.1. Etapes préliminaires

Nous souhaitons dans un premier temps construire un *GLM* complet intégrant l'ensemble des variables explicatives à disposition. Afin de construire ce modèle, nous détaillons ci-dessous les différentes étapes de modélisation, allant de la sélection des variables à la validation globale du modèle. Comme évoqué dans la partie II.3.1), la variable à modéliser est la prime pure par personne protégée.

#### Distinction fréquence/coût

Dans un certain nombre de mémoires relatifs à la tarification en général, le modèle retenu distingue la fréquence des sinistres et leur coût à travers deux modélisations indépendantes (par exemple : (Fesquet, 2015), (Lagadec, 2009), (Sepulveda, 2016), (Millet, 2016)).

Cependant, comme déjà évoqué dans la partie III.1, retenir un modèle coût / fréquence impose une relation d'indépendance entre le coût et la fréquence des sinistres. En partant du modèle évoqué précédemment :

$$E[S|X] = E\left[\sum_{i=1}^N C_i|X\right]$$

Cette relation, en cas d'indépendance des coûts et des fréquences de sinistre, peut s'écrire :

$$E[S|X] = E[N|X] \times E[C|X]$$

Ce qui revient à modéliser l'espérance de la fréquence des sinistres d'une part et du coût unitaire des sinistres d'autre part.

C'est une hypothèse assez difficile à justifier pour le risque santé. L'augmentation du coût moyen est corrélée directement avec une augmentation du niveau de garantie. Or, avec l'augmentation du niveau de couverture et la diminution du reste à charge pour l'assuré, il est courant de constater une augmentation de la fréquence de consommation. En effet, lorsque nous sommes bien couverts, nous pouvons être plus incités à consommer ce qui constitue une sorte d'aléa moral en santé.

La modélisation peut être réalisée par grand poste de soins comme dans (Poubennec, 2017), (Lagadec, 2009) ou (Sepulveda, 2016). Cependant, la cible de ce mémoire étant la construction d'un zonier, le choix de ne pas faire de

modèle par grand poste, en plus de permettre une modélisation simplifiée et aisément applicable opérationnellement, est plutôt recommandée dans le cadre de la construction d'un zonier. En effet, l'agrégation de zonier est un exercice difficile et non recommandé d'une part et d'autre part, la traduction commerciale et les problématiques liées à la mise en gestion d'un tel zonier par grand poste seraient trop coûteuses pour créer un zonier opérationnellement efficient. C'est pourquoi, nous avons retenu dans ce mémoire une estimation directe sur la prime pure par personne protégée sans distinction fréquence / coût et sans distinction par grand poste de consommation.

#### Traitement des valeurs nulles et choix de la loi de modélisation de la prime pure par pp $S_i$

Comme évoqué dans la partie III.1., la plupart des lois usuelles ne permettent pas d'avoir de valeurs nulles. La loi Gamma est une des lois parmi les plus utilisées pour modéliser les coûts des sinistres. Cependant, un certain nombre de personnes protégées ne consomme pas sur l'année. Aussi, afin de prendre en compte cette masse non-consommatrice, nous ne pouvons considérer que notre variable à expliquer suit une loi Gamma et nous utiliserons la loi de Tweedie présentée en III.2.

Dans notre base, un peu plus de 3000 lignes ressortent sans prestations, soit 0,7% des individus. Cela étant, la proportion de personnes protégées ne présentant aucun sinistre étant relativement faible, nous avons présenté en Annexe 2, les résultats de modélisation avec un modèle Gamma après exclusion des personnes protégées avec une sinistralité nulle.

#### Fonction de lien

Le dernier point à définir sur la variable à expliquer est la fonction de lien. Elle détermine la relation entre le prédicteur linéaire et l'espérance de la variable expliquée (Planchet & Serdeczny, 2014).

Dans le cadre de ce mémoire nous utiliserons la fonction du logarithme népérien comme fonction de lien afin d'identifier aisément l'effet de chacune des modalités des différentes variables explicatives sur la prime pure par personne protégée grâce à l'obtention d'un modèle multiplicatif.

#### Variables explicatives et variables à expliquer

Les *GLM* nécessitent la sélection de variables explicatives. Les variables à notre disposition sont rappelées à la figure 60.

Variable	Définition	Catégories
Tranche d'âge	Décomposition de l'âge en tranche	[0,10] ; (10,20] ; (20,25] ; (25,30] ; (30,35] ; (35,40] ; (40,45] ; (45,50] ; (50,55] ; (55,65] ; (65,75] ; (75,85] ; Plus de 85
Sexe		Homme ; Femme
Produit	Niveau de garantie	bas de gamme ; milieu de gamme ; haut de gamme
Adhésion	Indique si la personne protégée est présente depuis plus d'un an ou non	Oui ; Non
Département	Département ou groupe de départements	1 ; 3 ; 4 ; 5 ; 6 ; 7 ; 11 ; 14 ; 21 ; 25 ; 26 ; 27 ; 34 ; 38 ; 39 ; 42 ; 44 ; 50 ; 53 ; 58 ; 61 ; 66 ; 69 ; 70 ; 71 ; 73 ; 74 ; 75 ; 76 ; 77 ; 78 ; 83 ; 84 ; 89 ; 90 ; 91 ; 92 ; 93 ; 94 ; 95 ; 13-20 ; 15-43-63 ; 16-17-79-86 ; 18-28-36-37-41-45 ; 19-23-87 ; 2-60-80 ; 22-29-35-56 ; 24-33-40-47-64 ; 30-48 ; 54-55-88 ; 59-62 ; 72-49-85 ; 8-10-51-52 ; 9-12-31-32-46-65-81-82
Type PP	Profil d'assuré	Assuré Principal ; Conjoint ; Enfant ; Autres

Figure 60 – Variables explicatives et catégories associées

L'étude du risque santé dans la littérature a souvent été effectuée par le biais de variables relatives à l'âge, au niveau de garantie et au département (plus ponctuellement sur sexe compte tenu de l'impossibilité de discrimination sur ce critère dans la tarification). Nous rajoutons ici la variable précisant la notion d'adhésion supérieur à un an.

Une étape initiale est de tester la corrélation des variables explicatives. Nous choisissons ici de tester l'hypothèse de corrélation entre les variables avec le V de Cramer. Le V de Cramer est une mesure de l'association entre deux variables qualitatives ( $X_1, X_2$ ) à  $K \geq 2$  modalités. Il est basé sur la statistique de test du  $\chi^2$  de Pearson et se calcule comme suit :

$$V = \sqrt{\left(\frac{\chi^2}{N \times \min(C - 1, L - 1)}\right)}$$

avec,

- $N$ , le nombre total d'observations,
- $C$ , le nombre de colonnes du tableau croisé dynamique issu de  $X_1, X_2$  (ou plus simplement le nombre de modalités de  $X_1$ )
- $L$ , le nombre de lignes du tableau croisé dynamique issu de  $X_1, X_2$  (ou plus simplement le nombre de modalités de  $X_2$ )

Le V de Cramér varie dans l'intervalle [0,1]. Le tableau ci-dessous présente les intervalles d'interprétation.



Valeur	Force du lien statistique
0	Absence de relation
Entre 0,05 et 0,10	Très faible
Entre 0,10 et 0,20	Faible
Entre 0,20 et 0,40	Modérée
Entre 0,40 et 0,80	Forte
Entre 0,80 et 1	Colinéarité

Figure 61 - Intervalle d'interprétation V de Cramer<sup>12</sup>

Les résultats sur notre base sont les suivants :

V de Cramer	Département	Produit	Sexe	Tranche d'âge	Adhésion	Type PP
Département		0,12	0,02	0,05	0,05	0,05
Produit	0,12		0,02	0,17	0,06	0,08
Sexe	0,02	0,02		0,09	0,01	0,09
Tranche d'âge	0,05	0,17	0,09		0,09	<b>0,55</b>
Adhésion	0,05	0,06	0,01	0,09		0,05
Type PP	0,05	0,08	0,09	<b>0,55</b>	0,05	

Figure 62 - Test V de Cramer

Comme nous pouvons le voir sur la figure 62, les corrélations entre les variables sont entre faibles et très faibles d'après les éléments d'interprétations fournis à la figure 61, hormis comme déjà évoqué précédemment entre la « Tranche d'âge » et le « Type PP ». En effet, la relation entre l'âge et le type de personne protégée, notamment pour l'enfant est confirmé mathématiquement. Compte tenu de cette corrélation importante, nous faisons le choix de ne pas garder le « Type PP » dans la modélisation. Nous pouvons également noter une corrélation un peu plus élevée entre l'âge et le produit. Cette corrélation reste faible mais peut s'expliquer aisément par l'augmentation avec l'âge, des besoins de couverture d'une part et des moyens à disposition d'autre part pouvant ainsi conduire à réajuster le choix du niveau de garanties.

Pour la modélisation globale, il a été décidé de retenir et de tester l'ensemble des autres variables ci-dessus. Si la segmentation sur le sexe est interdite depuis 2012, notre objectif dans ce présent mémoire n'est pas de faire une tarification segmentant le sexe mais d'élaborer un zonier. Le zonier se basant sur l'exclusion de la part non expliquée relative aux variables autres que celles en lien avec les facteurs géographiques (voir partie III), il semble acceptable de maintenir cette variable sexe dans la modélisation, si elle s'avère significativement explicative sur la prime pure par PP.

Si couramment, l'exposition est utilisée en *offset*, la définition de la prime pure, calculée en fonction de l'exposition, nous a conduit à ne pas le faire dans notre modèle complet. En revanche, cette notion sera utilisée dans la partie de comparaison des zoniers en partie IV.4.

Pour rappel, une *offset* est une variable qui sera traitée comme une constante dans la modélisation. La définition de l'*offset* est présentée de manière assez explicite dans (Planchet & Serdeczny, 2014) et est reprise partiellement ci-dessous.

<sup>12</sup> Source : <https://lemakistatheux.wordpress.com/2013/05/31/le-v-de-cramer>

En partant d'un modèle poissonnien et d'une fonction de lien logarithme pour expliquer, par exemple, le nombre de sinistre  $N$ , nous avons :

$$E[N|X] = \exp\left(\sum_{k=1}^p \beta_k x_k\right) = \exp(\beta'x)$$

Si on veut tenir compte de l'exposition au risque  $d$ , nous savons que l'espérance  $\lambda$  de la loi de poisson devient  $\lambda d$  et le modèle s'écrit alors :

$$E[N|x, d] = d \times \exp\left(\sum_{k=1}^p \beta_k x_k\right) = \exp(\beta'x + \ln(d))$$

Ainsi, cela revient à ajouter une variable explicative pour laquelle le coefficient  $\beta_k$  est connu et égal à 1 et ne doit donc pas être estimé. La variable  $\ln(d)$  est appelée *offset*.

### Définition de l'individu de référence

Les *GLM* sont adaptés pour estimer l'impact d'un facteur d'une variable explicative relativement à un autre facteur. Aussi, l'ensemble de nos variables étant déjà catégorielles, il nous faut définir un individu de référence. L'estimation de la prime pure de cet individu sera contenue dans la constante du modèle et les impacts des autres modalités et facteurs des différentes variables explicatives seront estimées relativement aux différentes caractéristiques de cet individu de référence.

Par simplicité, notre individu de référence correspond pour chaque variable à la modalité la plus représentée. Ainsi, il se caractérise comme suit :

- Femme (56% de la base d'étude),
- entre 55 et 65 ans (20% de la base d'étude),
- provenant du département 38 (12% de la base d'étude),
- ayant souscrit un produit bas de gamme (56% de la base d'étude),
- ayant souscrit depuis plus d'un an (81% de la base d'étude).

### 1.2. Construction du modèle

Une fois les différentes étapes préliminaires réalisées, nous présentons les différentes étapes de construction du modèle.

Dans un premier temps, le *GLM* étant une méthode d'apprentissage, il convient de séparer notre échantillon en deux (voir en trois dans certaines études) :

- un échantillon d'apprentissage pour calibrer le modèle,
- un échantillon de validation : pour tester l'adéquation de notre modèle estimé sur le modèle d'apprentissage,
- éventuellement un échantillon de test : pour les méthodes de data science, l'échantillon de validation permet la consolidation du modèle et l'échantillon de test, la mesure de la qualité de la prédiction (performance du modèle).

Le choix peut se faire de manière aléatoire en prenant par exemple, un tiers de la base comme base de validation, et deux tiers comme base d'apprentissage. Compte tenu de la taille importante de notre base de données et de la segmentation relativement fine par département souhaitée, nous avons retenu pour notre étude 75% des données dans l'échantillon d'apprentissage et 25% dans l'échantillon de validation. Nous n'avons pas utilisé ici de méthode de rééchantillonnage (bagging par exemple) mais cela aurait pu être intéressant compte tenu de la taille des échantillons. Cependant, quelques vérifications sur l'échantillonnage réalisé semblent produire des résultats satisfaisants.

Une vérification empirique sur les moyennes et écarts types de chacun des échantillons a été réalisée et est présentée en figure 63.

	Prime pure par PP	Ecart (%)	Ecart-type	Ecart (%)
Base globale	448,81		753,46	
Base d'apprentissage	448,93	0,0%	752,89	-0,1%
Base de validation	448,14	-0,1%	756,68	0,4%

Figure 63 - Validation empirique de l'échantillonnage

Afin de déterminer quel est le meilleur modèle nous utiliserons l'analyse de la déviance permettant d'estimer la part expliquée par le modèle et donne un bon critère de la qualité globale de l'ajustement. La déviance est l'écart entre la log-vraisemblance obtenue avec notre estimation, et celle obtenue avec un modèle parfait (dit saturé).

Ce critère se définit comme suit :

$$D = 2 \times (\ln L(Y) - \ln L(\hat{\mu}))$$

avec  $\hat{\mu} = g^{-1}(X\hat{\beta})$

$D$  est positif et « petit » pour un modèle de bonne qualité. Cette statistique suit asymptotiquement, du fait de résultats généraux sur les rapports de vraisemblance, une loi du khi-2 à  $n-p-1$  degrés de liberté (son espérance est donc  $n-p-1$ ).

Nous avons présenté plus haut les lois de Tweedie que nous avons convenu d'utiliser dans ce mémoire. Pour rappel, les lois de Tweedie sont dépendant d'un paramètre  $p$  qu'il faudra déterminer (voir partie III.2). Pour sélectionner la valeur optimale pour le paramètre  $p$ , nous utilisons le critère de déviance calculé pour des valeurs de  $p$  allant de 1,10 à 1,90 sur le modèle intégrant l'ensemble des variables explicatives à disposition. Les résultats sont présentés à la figure 64<sup>13</sup>.

Le paramètre  $p$  qui minimise la déviance est égal à 1,90. Nous pouvons donc considérer que notre loi de Tweedie se rapprochera d'un modèle Gamma.

<sup>13</sup> Le choix de cette méthode de sélection est évoqué sur [https://freakonometrics.hypotheses.org/files/2015/12/slides\\_ensae\\_9.pdf](https://freakonometrics.hypotheses.org/files/2015/12/slides_ensae_9.pdf)

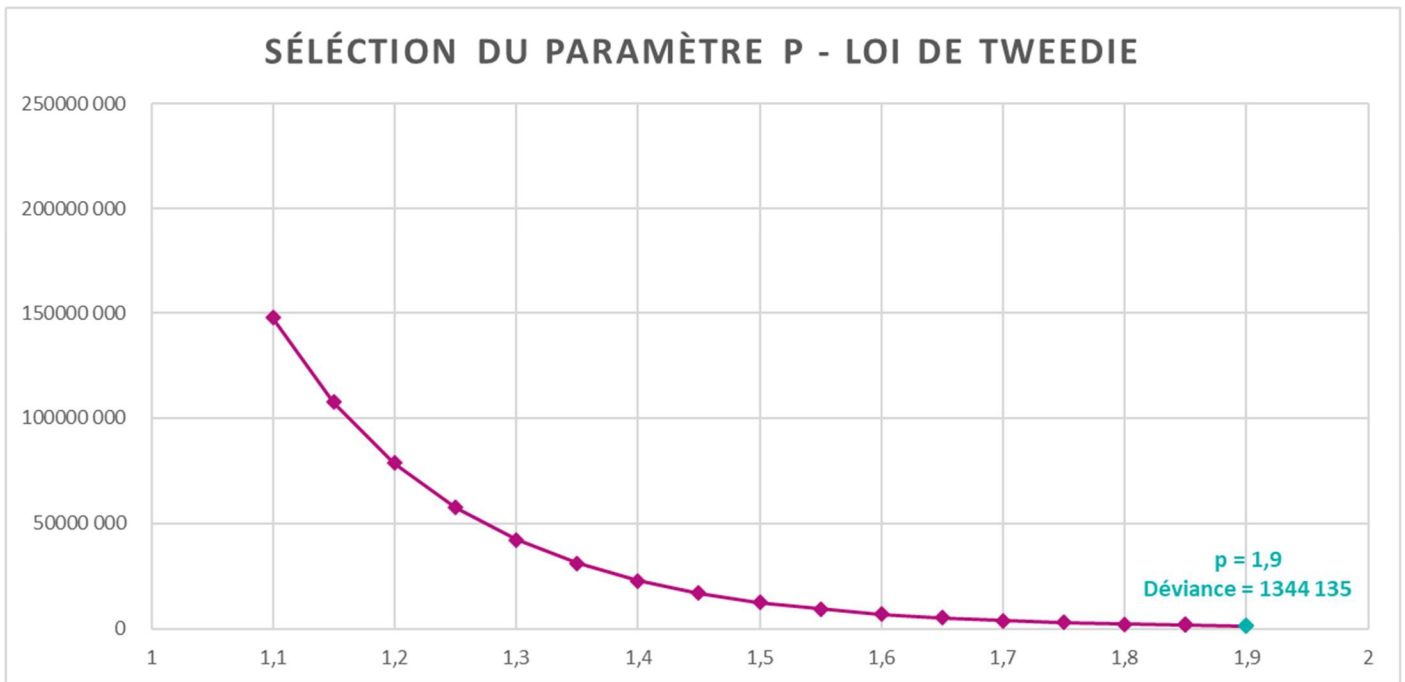


Figure 64 - Choix du paramètre p dans la loi de Tweedie – critère déviance

Afin de tester la significativité de chacune des variables dans le modèle, nous avons choisi d'utiliser une méthode *forward*. Cette méthode consiste à intégrer les variables les unes après les autres et à tester leur apport sur le modèle en utilisant par exemple des critères de comparaison de modèle. Ainsi, une fois le paramètre p sélectionné, nous allons comparer les valeurs de déviance pour sélectionner le meilleur modèle. La méthode *backward*, partant d'un modèle intégrant toutes les variables et les retirant les unes après les autres aurait aussi pu être mis en œuvre. A noter que cette méthode conduirait au même résultat dans notre exemple.

En utilisant ce critère pour sélectionner le meilleur modèle dans notre base de données, nous avons pu réaliser la figure 65 qui présente l'évolution des déviances des différents modèles avec l'ajout d'une variable.

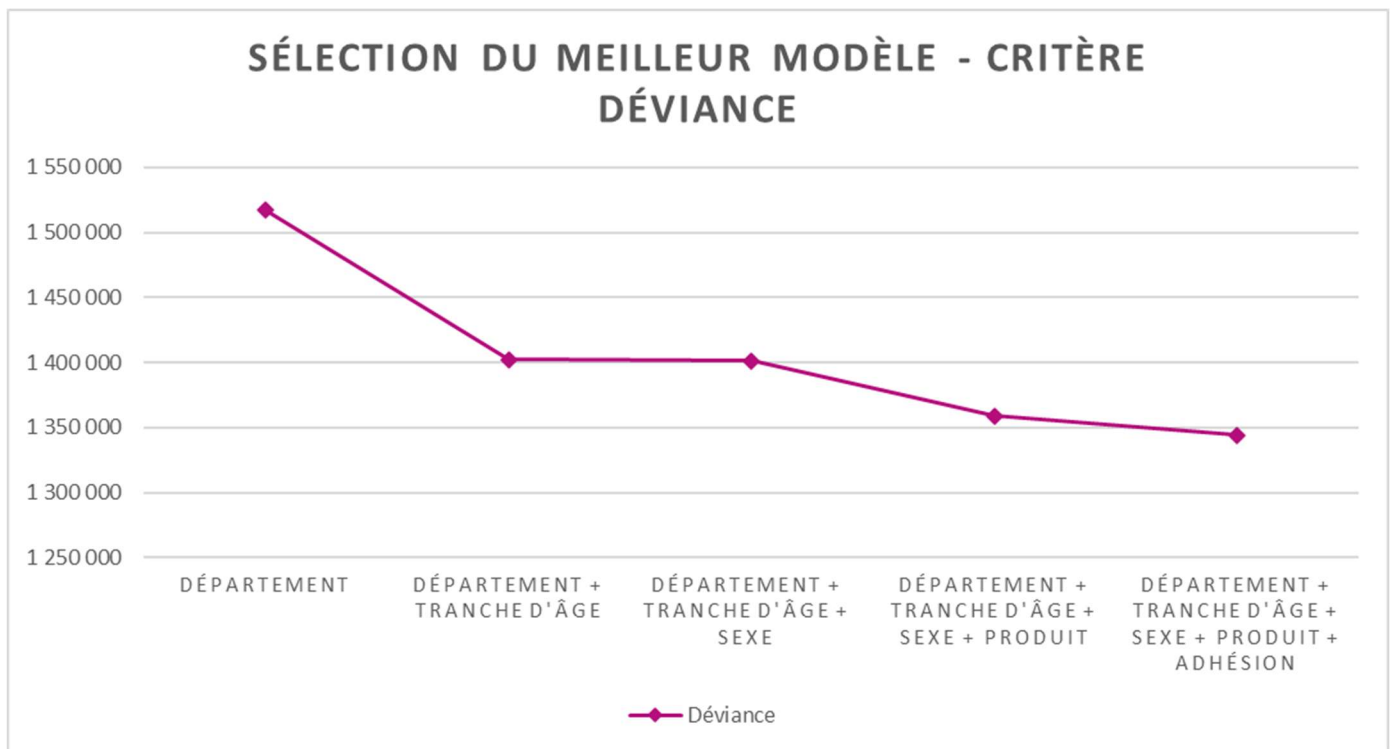


Figure 65 – Sélection du meilleur modèle avec le critère de déviance

La figure 65 semble indiquer que l'ensemble des variables explicatives a un impact significatif sur la prime pure, la courbe de déviance étant décroissante avec l'ajout de variable.

L'utilisation d'autres critères afin de déterminer le meilleur modèle comme l'*AIC (Akaike Information Criterion)* qui prend en compte la complexité du modèle a été réalisé. L'*AIC* permet, outre de tester le pouvoir explicatif du modèle, de prendre en considération la simplicité du modèle. Ainsi, un modèle simple pourra être privilégié par rapport à un modèle peut-être plus explicatif mais plus compliqué. L'*AIC* se définit comme suit :

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

dans lequel  $L$  représente la vraisemblance du modèle et  $k$  le nombre de paramètres estimés du modèle.

Le meilleur modèle sera donné par la valeur la moins élevée. Les résultats sont donnés ci-dessous et les conclusions sont similaires.

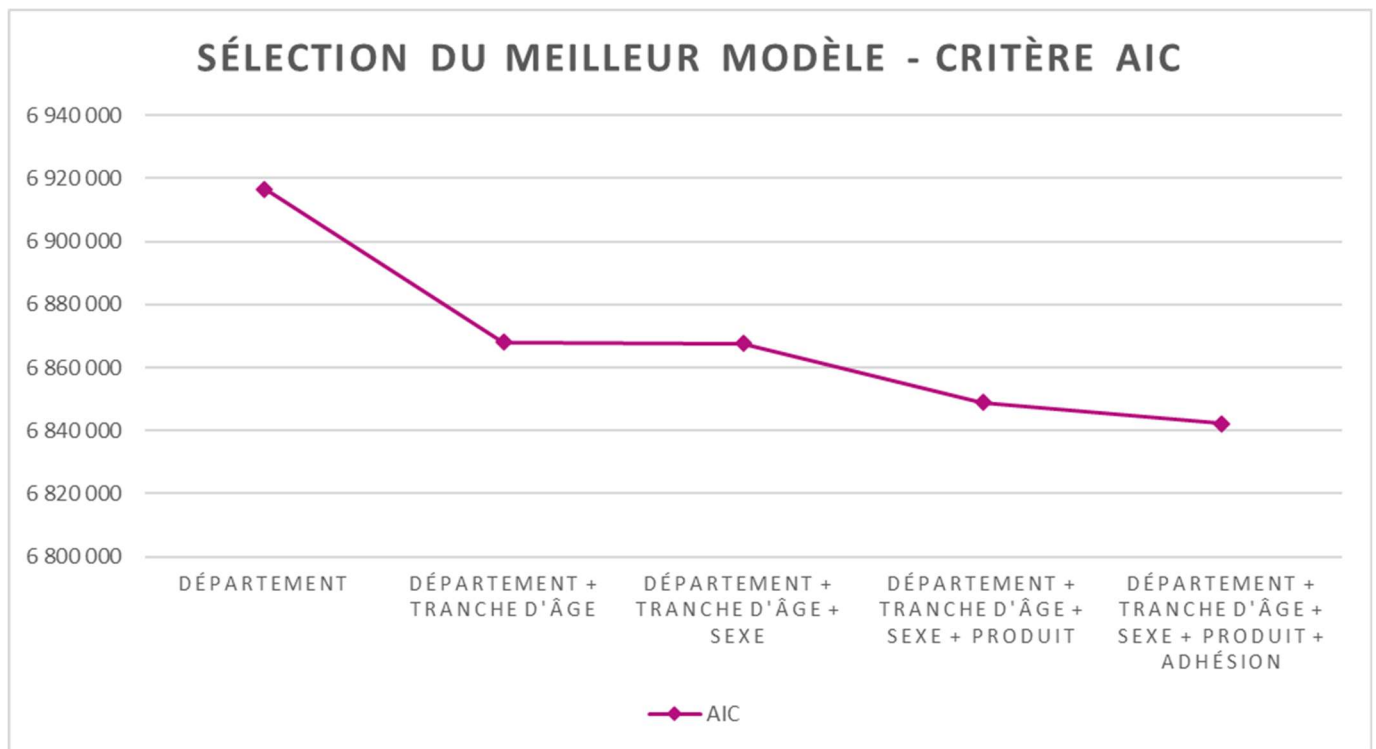


Figure 66 – Sélection du meilleur modèle avec le critère AIC

### 1.3. Validation et analyse du modèle complet

Maintenant que le modèle complet a été implémenté, nous allons pouvoir tester la significativité du modèle.

Le modèle retenu ici est le suivant :

$$\begin{aligned}
 \ln(\text{PRIME PURE}) &\sim \text{constante} \\
 &+ \hat{\beta}_1 * \text{DEPARTEMENT} \\
 &+ \hat{\beta}_2 * \text{TRANCHEAGE} \\
 &+ \hat{\beta}_3 * \text{SEXE} \\
 &+ \hat{\beta}_4 * \text{PRODUIT} \\
 &+ \hat{\beta}_5 * \text{ADHESION}
 \end{aligned}$$

Pour ce faire nous allons d'abord regarder la significativité des différentes modalités des variables explicatives sur la variable à expliquer. Nous avons réalisé un test de Student sur l'ensemble des catégories de nos variables. Les résultats globaux de ce modèle complet sont présentés ci-dessous en figure 67. Un certain nombre de départements n'étant pas significatif et dans un souci de présentation, nous avons regroupé l'ensemble des départements non significatifs dans la catégorie « Autres départements ».

Variable	Coefficients	Écart-type	P-Valeur	Significativité
(Intercept)	6,10	0,01	0,00	***
newdep1	-0,10	0,01	0,00	***
newdep13 20	0,09	0,03	0,00	***
newdep14	-0,10	0,01	0,00	***
newdep18 28 36 37 41 45	-0,10	0,04	0,02	*
newdep21	-0,04	0,01	0,01	**
newdep22 29 35 56	-0,09	0,04	0,01	*
newdep26	-0,15	0,04	0,00	***
newdep27	-0,17	0,04	0,00	***
newdep3	-0,09	0,01	0,00	***
newdep30 48	0,03	0,01	0,02	*
newdep34	0,09	0,01	0,00	***
newdep39	-0,07	0,02	0,00	***
newdep4	-0,13	0,03	0,00	***
newdep44	-0,16	0,07	0,02	*
newdep50	-0,09	0,02	0,00	***
newdep53	-0,20	0,04	0,00	***
newdep54 55 58	-0,15	0,08	0,04	*
newdep61	-0,07	0,02	0,00	***
newdep69	-0,02	0,01	0,04	*
newdep7	-0,10	0,05	0,03	*
newdep71	-0,02	0,01	0,04	*
newdep72 49 85	-0,16	0,02	0,00	***
newdep73	-0,08	0,02	0,00	***
newdep74	-0,13	0,01	0,00	***
newdep75	-0,10	0,04	0,02	*
newdep76	-0,13	0,02	0,00	***
newdep84	-0,07	0,04	0,05	*
newdep89	-0,09	0,03	0,00	**
Autres déps	...	...	...	...
classeage[0,10]	-0,75	0,01	0,00	***
classeage(10,20]	-0,57	0,01	0,00	***
classeage(20,25]	-0,56	0,01	0,00	***
classeage(25,30]	-0,45	0,01	0,00	***
classeage(30,35]	-0,36	0,01	0,00	***
classeage(35,40]	-0,32	0,01	0,00	***
classeage(40,45]	-0,18	0,01	0,00	***
classeage(45,50]	-0,08	0,01	0,00	***
classeage(50,55]	-0,02	0,01	0,05	*
classeage(65,75]	0,23	0,01	0,00	***
classeage(75,85]	0,57	0,01	0,00	***
classeage(85,Inf]	0,85	0,02	0,00	***
SEXEHomme	-0,07	0,00	0,00	***
Produitmilieu de gamme	0,32	0,01	0,00	***
Produithaut de gamme	0,77	0,01	0,00	***
AdhesionNon	0,30	0,01	0,00	***

Figure 67 – Coefficients d'apprentissage et tests de Student sur les variables explicatives

Ainsi, nous pouvons constater sur la figure 67 qu'hormis pour les départements, l'ensemble des variables explicatives intégrées a un impact significatif sur la variable à expliquer. Cela étant, cela ne remet pas en cause le fait que la localisation géographique a un impact mais simplement que cet impact est trop diffus à ce stade, le rendant impossible à utiliser en l'état compte tenu du nombre de modalités.

La tranche d'âge présente bien un impact proche de ceux constatés dans la partie (II.3.1), c'est-à-dire, relativement à notre classe de référence de 55 à 65 ans, les classes d'âge allant de 0 à 55 ans ont un impact négatif sur la prime pure (les prestations moyenne pour les individus de moins de 55 ans sont inférieures à celles de la tranche 55-65) et les classes d'âge supérieures à 65 ans ont un impact positif sur la prime pure par personne protégée.

Nous pouvons aussi constater à la figure 65 que le fait d'être un homme a un impact à la baisse sur la prime pure, conformément à l'attendu même si celui-ci semble assez faible (-0,07).

La logique est également respectée pour la croissance du produit et les coefficients sont en revanche relativement élevés (respectivement 0,32 et 0,77 pour les produits milieu de gamme et haut de gamme).

En outre, nous constatons qu'une adhésion inférieure à 1 an semble engendrer effectivement plus de prestations. Le coefficient est de plus relativement important (0,30).

Enfin, pour les départements, les coefficients sont beaucoup plus compliqués à interpréter. En effet, un certain nombre de coefficient ne sont pas significatifs et parmi ceux qui le sont, les coefficients sont trop nombreux et différents pour être interprété en l'état. Cela justifie à nouveau la mise en œuvre d'un zonier.

Sur le modèle de (Ferrier, 2016), afin de valider le modèle, nous vérifions que l'impact des différentes variables explicatives de notre modèle estimé sur l'échantillon d'apprentissage est similaire à l'estimation réalisée sur l'échantillon de validation. Ainsi, nous appliquons notre modèle sur l'échantillon de validation et nous comparons les effets des différentes variables explicatives.

A la figure 68, nous présentons les résultats pour la variable classe d'âge bien que cette vérification ait été réalisée pour chacune des variables. Ainsi, la croissance des coefficients correspond bien à celle attendu et nous retrouvons en particulier le saut entre 0 et 20 ans expliqué notamment par l'orthodontie tant sur la base d'apprentissage que sur la base de validation.



## REPRÉSENTATION DES COEFFICIENTS GLM PAR TRANCHE D'ÂGE

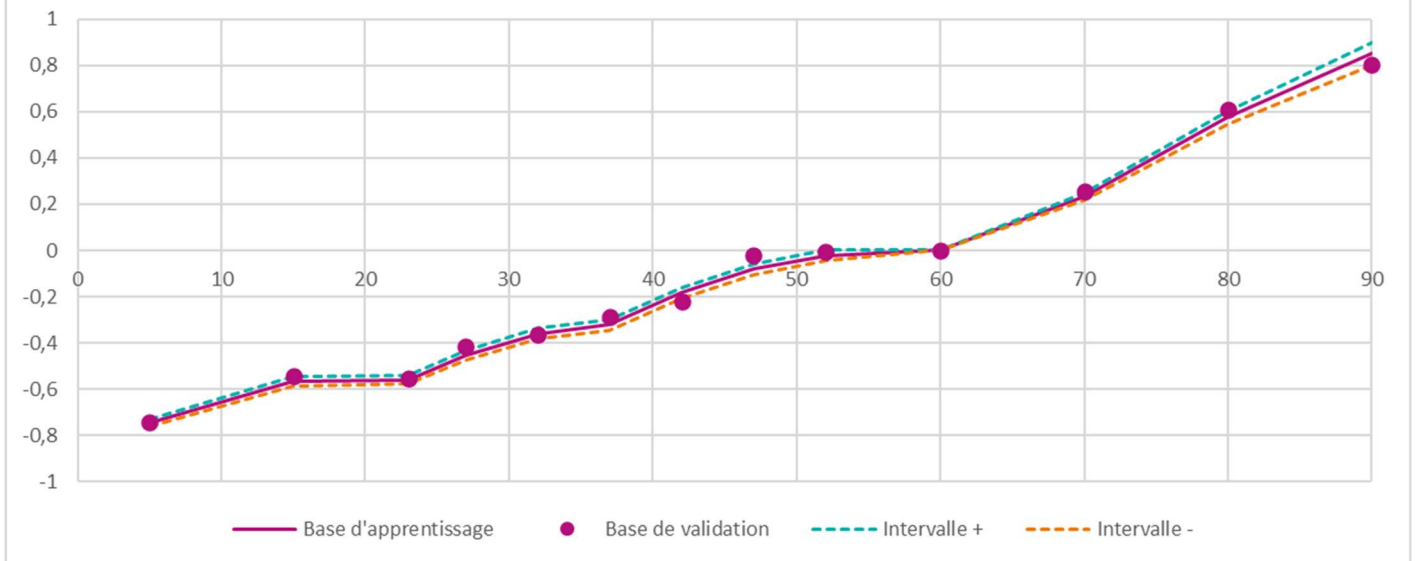


Figure 68 – Coefficients d'apprentissage et intervalle de confiance – Tranche d'âge

Pour chacune des variables, les écarts entre les coefficients calculés sur l'échantillon d'apprentissage et ceux calculés sur l'échantillon de validation sont faibles et permettent de valider empiriquement la qualité du modèle.

En complément, l'analyse des résidus, permet de se donner un aperçu graphique de la qualité du modèle. Ces résidus doivent être aléatoirement répartis autour de 0.

Dans le cadre d'un *GLM*, nous pouvons retenir deux types de résidus. Les résidus de Pearson d'une part qui se définissent comme suit :

$$r_i^p = \sqrt{\omega_i} \times \frac{y_i - \mu_i}{\sqrt{V(\mu_i)}}$$

Le résidu de Pearson se calcule comme le résidu  $(y - \mu)$ , rapporté à l'écart-type estimé de  $\mu$ .

D'autre part, les résidus de déviance qui se définissent comme suit :

$$r_i^p = \pm \sqrt{d_i}$$

avec  $\sum_{i=1}^n d_i = D$ .

Ces résidus mesurent la contribution de chaque observation à la déviance du modèle par rapport au modèle saturé. Pour notre analyse graphique, nous avons retenu l'analyse des résidus de déviance.

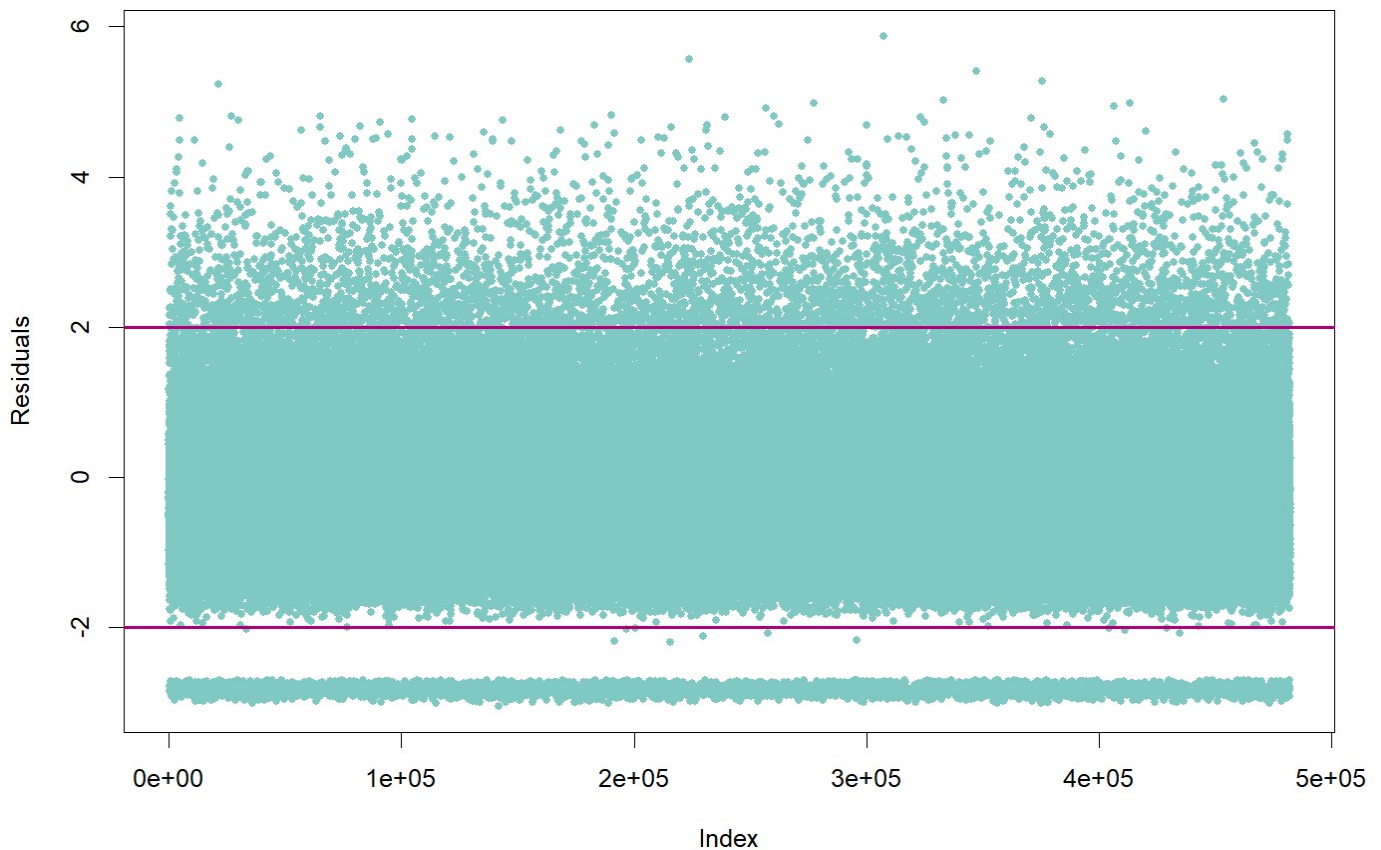


Figure 69 – Représentation des résidus de déviance standardisés

La validation des résidus reste assez subjective. Pour des résidus standardisés, on considère que si l'ensemble des résidus se situe entre -2 et 2 alors l'hypothèse d'homoscédasticité est vérifiée.

A la figure 69, la représentation des résidus n'est pas parfaite. En effet, les résidus ne semblent pas être parfaitement aléatoires. Il semble notamment que l'estimation des prestations nulles n'est pas bien réalisée par le modèle. Cela peut s'expliquer par le choix de  $p$  pour la loi de Tweedie (proche de 2). En effet, l'analyse des minimum et maximum des valeurs prédites sur l'échantillon de validation ne présentent aucune prestation nulle malgré leur présence dans l'échantillon. Cela étant, l'estimation reste globalement assez bien définie compte tenu du grand nombre de résidus entre -2 et 2. En outre, nous n'utilisons pas ce modèle pour obtenir une prime pure mais pour l'obtention d'un zonier se basant sur le différentiel entre deux modèles. Si notre objectif avait été de réaliser une tarification, nous aurions probablement dû retravailler le modèle.

Il est à noter que nous avons également testé un modèle *GLM* en définissant la loi de  $S_i$  comme étant une loi gamma sur nos données (après retraitement des sinistres nulles, non tolérés par le modèle gamma). Les résultats sont présentés en Annexe 2 et les résidus semblent plus aléatoirement réparties. Cela étant, comme évoqué dans les parties précédentes, la suppression des sinistres nulles ne semble pas souhaitable.

## 2. Création d'un zonier

Comme présenté dans la partie III.3, une fois le modèle complet créé, nous pouvons désormais isoler la part expliquée par les facteurs géographiques. Une fois cette part expliquée par les facteurs géographiques uniquement, nous pouvons ensuite, sur la base des résidus, utiliser différentes méthodes de classification et de lissage pour construire les zoniers. Enfin, nous pourrions comparer les différents zoniers obtenus ainsi que celui actuellement en place en les réintégrant dans une régression et en comparant les prédictions obtenues.

### 2.1. Présentation des méthodes de mise en place d'un zonier

Dans toutes les techniques de zonage, il existe une étape préliminaire permettant d'isoler l'effet du risque géographique. La stratégie pour construire un zonier est donc de ne pas intégrer le facteur de risque géographique dans la calibration du modèle, puis travailler sur les résidus pour faire ressortir cet effet.

Afin de réaliser un zonier applicable commercialement et paramétrable opérationnellement nous avons décidé d'utiliser une segmentation basée sur les départements. Ainsi, le seul facteur géographique que nous isolerons de notre modèle complet est la variable département.

Sur la base des éléments présentés à la partie IV.1 sur la validation théorique de la méthode nous pouvons déduire notre modèle sans facteurs géographiques de notre modèle précédent. Ainsi, notre modèle est le suivant :

$$\begin{aligned} \ln(\text{PRIME PURE}) &\sim \text{constante} \\ &+ \hat{\beta}_2 * \text{TRANCHEAGE} \\ &+ \hat{\beta}_3 * \text{SEXE} \\ &+ \hat{\beta}_4 * \text{PRODUIT} \\ &+ \hat{\beta}_5 * \text{ADHESION} \end{aligned}$$

Comme nous pouvons le voir, seule la variable relative à la segmentation départementale a été retraitée. Il est important de noter que l'ensemble des coefficients  $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$  et  $\hat{\beta}_5$  mais surtout la *constante* sont bien strictement égaux à ceux calculés dans le modèle précédent.

### 2.2. Construction d'une base de résidus par département

Conformément à la présentation réalisée à la partie III.2, nous pouvons constituer les résidus  $r_i$  comme suit :

$$r_i = \frac{\hat{Y}_i}{\hat{Y}'_i} = e^{\beta_1 x_1^i}$$

Nous pouvons alors définir par département,

$$\hat{r}_k^d = \frac{\sum_{i=1}^{I^k} e_i r_i}{\sum_{i=1}^{I^k} e_i}$$

Avec,  $e_i$  l'exposition de l'individu  $i$ ,  $\sum_{i=1}^{I^k} e_i$  l'exposition constatée par département,  $I^k$  le nombre d'assurés dans le département  $k$ .

Les données obtenues peuvent être lissées avant classification. L'objectif du lissage est avant tout d'éviter les sauts entre les différents départements ainsi que de corriger quelques expériences exceptionnelles qui pourraient nuire à la qualité du zonier. Aussi, ce lissage n'est pas un lissage géographique.

Il est important de noter que si ce lissage aurait été particulièrement indiqué dans le cas d'une conservation de l'information par départements<sup>14</sup>, son intérêt, compte tenu de la répartition des départements telle que définie à la partie II.3.1, reste assez limité, les résidus constatés allant de 0,82 à 1,13. Cela étant, la méthode de lissage de Whittaker Henderson est présentée en Annexe 3.

Nous pouvons reconstituer à l'issue de cette étape une nouvelle base de travail présentant une ligne par département et en colonne, l'exposition, le résidu lissé, la prime pure prédite, la prime pure observée. Une fois ces éléments déterminés, nous pouvons utiliser des méthodes de classification afin de regrouper les départements entre eux en fonction du résidu observé.

### 3. Méthodes de classification

Le but de ce mémoire étant d'optimiser la création d'un zonier nous utiliserons deux méthodes de classification différentes. La classification sera réalisée sur la moyenne des résidus pondérées par département soit  $\hat{\rho}_k^d$ . Nous réaliserons une première classification avec une méthode de Classification Ascendante Hiérarchique (CAH) et une deuxième avec une méthode de K-moyennes. Il existe de nombreuses techniques statistiques visant à partitionner une population en différentes classes ou sous-groupes. Notre choix s'est porté sur des méthodes aussi simple à comprendre qu'à mettre en œuvre opérationnellement.

#### 3.1. Classification ascendante hiérarchique (CAH)

Pour la CAH, nous cherchons à ce que les individus regroupés au sein d'une même classe (homogénéité intra-classe) soient le plus semblables possible tandis que les classes soient le plus dissemblables (hétérogénéité inter-classe).

Il est à noter que notre classification se fera sur un vecteur (de dimension 1 donc). Si le choix d'appliquer ces méthodes de segmentation sur un vecteur peut être discutable, les méthodes présentées n'en sont pas moins applicables et présenteront des résultats tout à fait exploitables.

Le principe de la CAH est de rassembler des individus selon un critère de ressemblance défini au préalable qui s'exprimera sous la forme d'une matrice de distances, exprimant la distance existante entre chaque individu pris deux à deux. Deux observations identiques auront une distance nulle. Plus les deux observations seront dissemblables, plus la distance sera importante. La CAH va ensuite rassembler les individus de manière itérative afin de produire un dendrogramme ou arbre de classification. La classification est ascendante car elle part des observations individuelles ; elle est hiérarchique car elle produit des classes ou groupes de plus en plus vastes, incluant des sous-groupes en leur sein. En découpant cet arbre à une certaine hauteur choisie, on produira la partition désirée.

La notion de ressemblance entre observations est évaluée par une distance entre individus. De nombreuses solutions et choix de distances existent (saut minimum, distance maximum, moyenne, Ward...). Chacune d'elles produira un

---

<sup>14</sup> Les modèles intégrant l'ensemble des départements s'ils ont été réalisés ne sont pas présentés dans ce mémoire.

dendrogramme différent. Nous ne détaillerons pas ici ces différentes techniques. Cependant, la méthode de Ward semble être la plus couramment utilisée. De manière simplifiée, cette méthode cherche à minimiser l'inertie intra-classe et à maximiser l'inertie inter-classe afin d'obtenir des classes les plus homogènes possibles. Cette méthode vise l'augmentation minimum de la somme des carrés des variances<sup>15</sup>. Dans notre cas, nous allons opter pour la méthode de Ward appliquée au carré des distances.

Le choix du nombre de partition est un exercice délicat. Pour obtenir une partition de la population, il suffit de découper le dendrogramme obtenu à une certaine hauteur. En premier lieu, une analyse de la forme du dendrogramme pourra nous donner une indication sur le nombre de classes à retenir. La figure 70 représente le dendrogramme sur notre base.

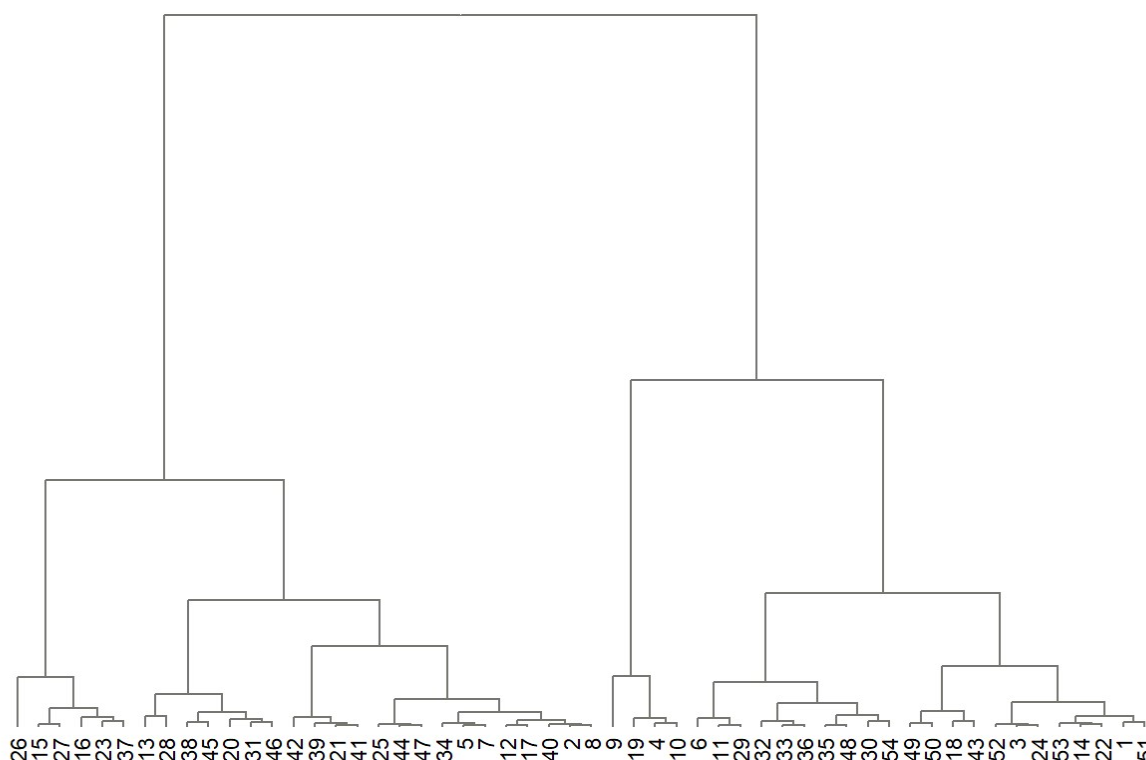


Figure 70 - Dendrogramme – résidus par département - Distance de Ward

Comme nous pouvons le voir à la figure 71, il semblerait envisageable, graphiquement, de découper la série en 3 ou 4 classes a priori. Cependant, pour nous aider, nous pouvons représenter les sauts d'inertie du dendrogramme selon le nombre de classes retenues.

<sup>15</sup> Source : <https://larmarange.github.io/analyse-R/classification-ascendante-hierarchique.html>

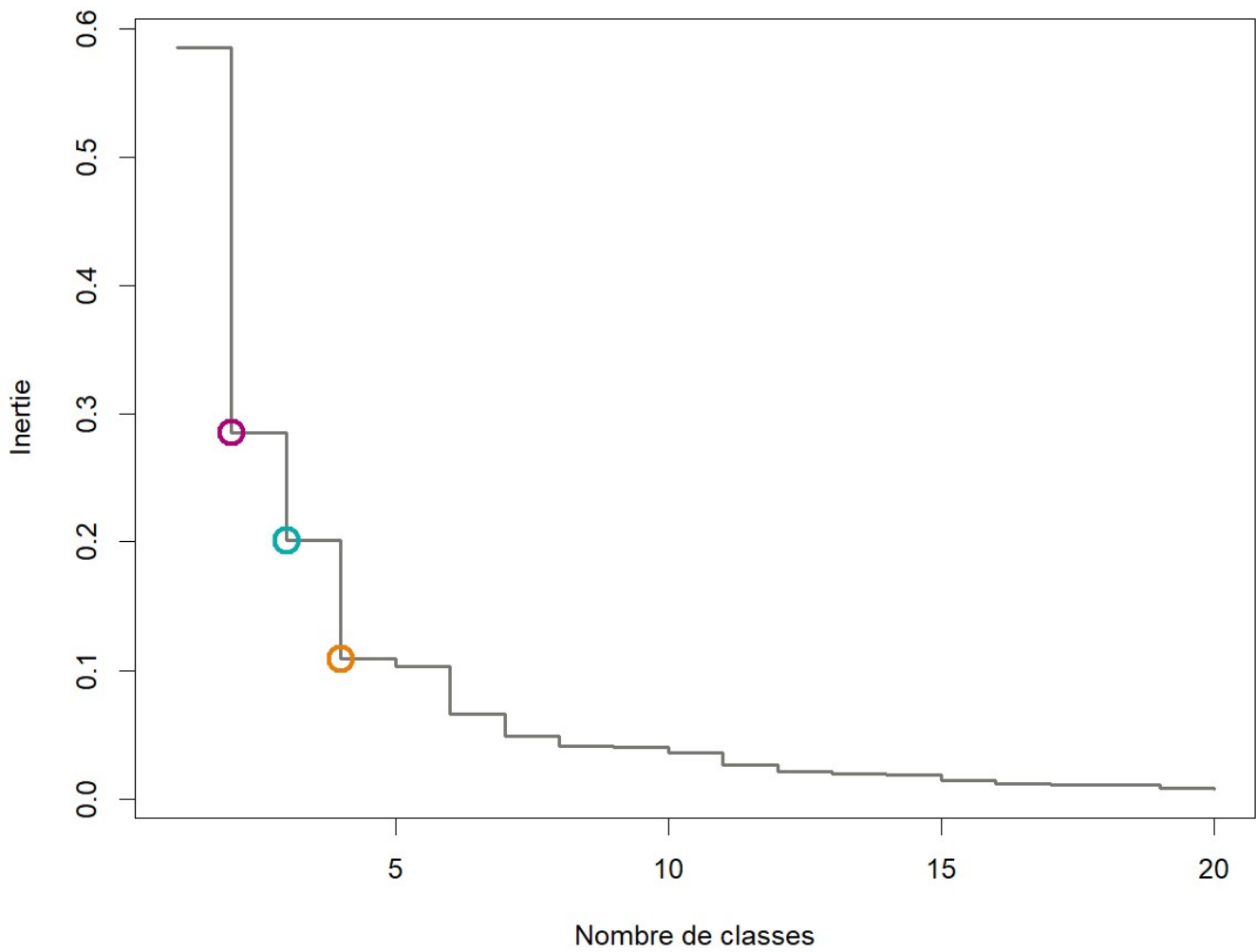


Figure 71 - Sauts d'inertie du dendrogramme

Nous pouvons remarquer à la figure ci-dessus, trois sauts assez nets à 2, 3 et 4 classes, que nous avons représentés ci-dessus respectivement en violet, en bleu et en orange.

Nous pouvons également représenter le graphique des pertes relatives d'inertie. La meilleure partition selon ce critère est représentée par un point noir et la deuxième par un point gris.

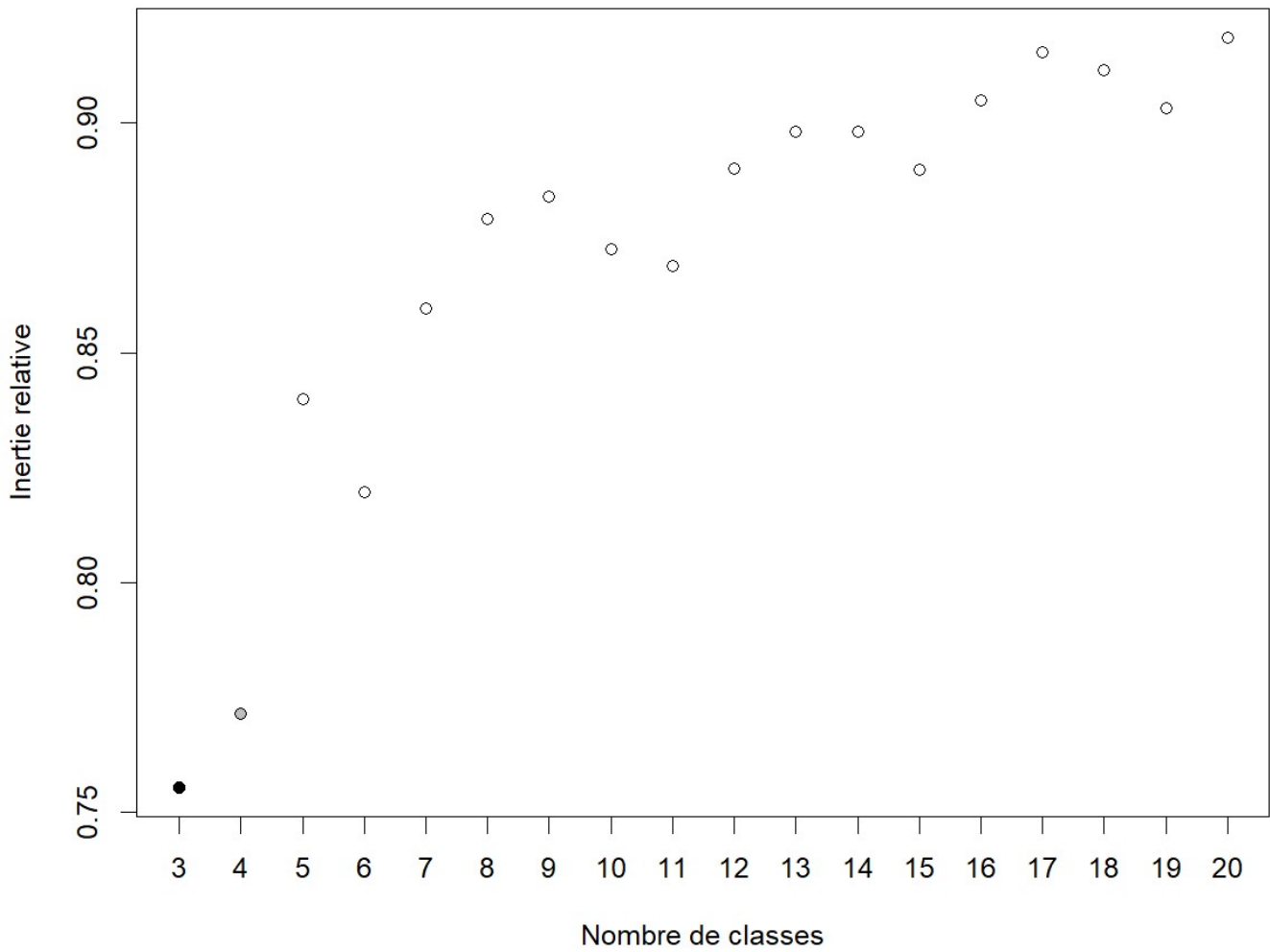


Figure 72 - Perte relative d'inertie selon le nombre de classes

En l'occurrence, la meilleure répartition semble être de 3 classes et la deuxième de 4 classes. La classification à 4 classes est plus proche des segmentations appliquées sur le marché et correspond à celle qui avait été mise en place au niveau technique lors de la création du précédent zonier ADREA Mutuelle. La segmentation en 4 classes sera celle que nous retiendrons ici. Cela étant, les résultats pour 3 classes sont également présentés en annexe 4.

Le dendrogramme avec les classes sélectionnées est repris à la figure 73.

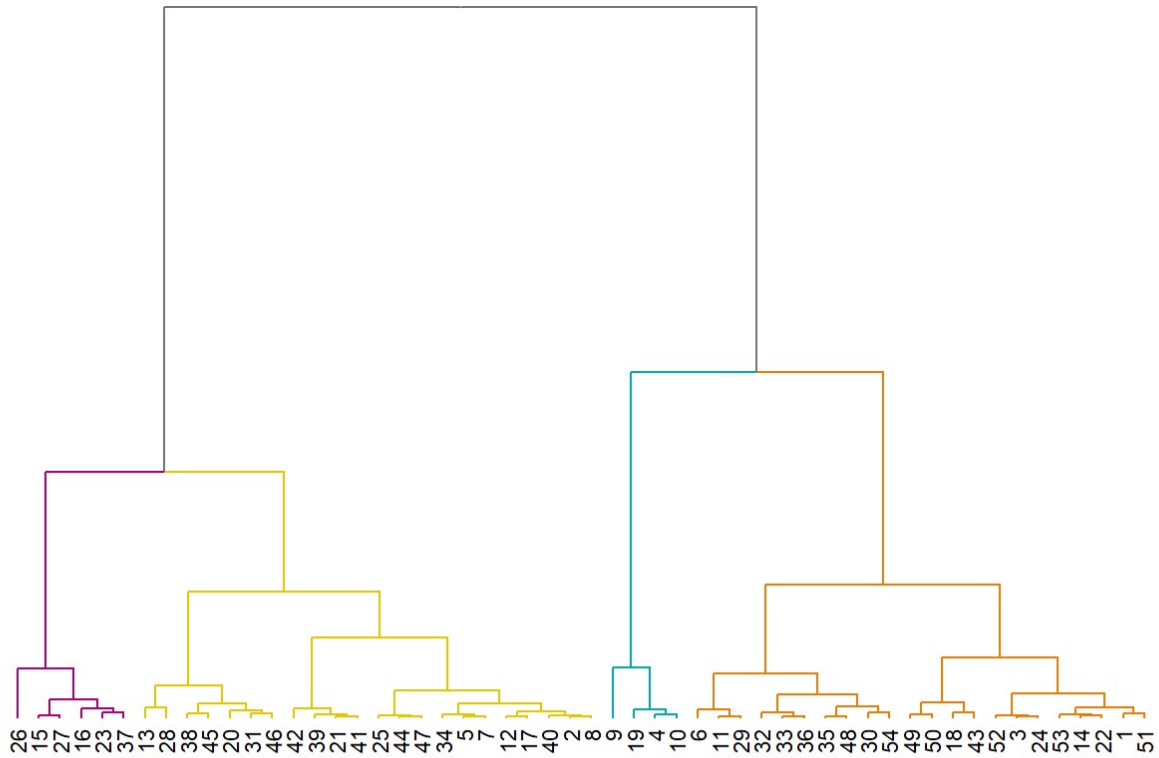


Figure 73 - Dendrogramme avec choix de décomposition en trois classes

### 3.2. Classification avec la méthode des K-moyennes

Afin de construire les classes, nous utiliserons également l'algorithme des K-moyennes (ou *k-means*). Cependant, le nombre de classe doit être fixé a priori.

L'algorithme des K-moyennes est expliqué ci-dessous (Truffery, 2017) :

- Choisir K éléments initiaux "centres" des K groupes  $\{C_1, C_2, \dots, C_k\}$
- Placer les objets dans le groupe de centre le plus proche soit :

$$x_j \in c_k \text{ si } \forall j, |x_i - \mu_k| = \min |x_i - \mu_j|$$

- Recalculer le centre de gravité de chaque groupe :

$$\mu_k = \frac{1}{N} \sum_{x \in C_k} x_i$$

- Itérer l'algorithme jusqu'à ce que les objets ne changent plus de groupe.

Aussi, dans un premier temps, nous devons déterminer le nombre optimal de zone à définir. Pour mémoire, le zonier actuel ne comporte que quatre zones distinctes d'un point de vue strictement technique.

S'il existe certains critères statistiques externes, ces méthodes restent empiriques et sont basées sur la taille de l'échantillon (Ferrier, 2016).



Ainsi, avec n représentant le nombre de départements ou groupe de départements dans notre étude, soit 54 :

- Critère de Wong :  $n^{0,3}$ , soit 3 classes,
- Formule de Brooks-Carruthers :  $5 * \log(n)$ , soit 8 classes,
- Formule de Huntsberger :  $1 + 3,332 * \log(n)$  soit 7 classes,
- Formule de Sturges :  $\log_2(n+1)$ , soit 6 classes.

Aucune de ces formules ne permet de déterminer le nombre de classe exacte. En complément, une méthode graphique de sélection des classes basée sur la méthode des K-moyennes existe.

Il existe une méthode basée sur les K-moyennes qui aide à déterminer un nombre de classe optimal. Cette méthode consiste à utiliser les K-moyennes avec différents nombres de classes et de calculer la variance intra-classe. C'est la méthode *Elbow* (« coude » en anglais). La variance intra-classe est la somme des distances entre chaque centre d'une classe et les différentes observations incluses dans cette même classe. Ainsi, nous cherchons à trouver le nombre de classes permettant la minimisation de la distance entre leurs centres et les observations dans la même classe.

La variance intra-classe se calcule comme suit :

$$V_{intraclasse} = \sum_j \sum_{x_i \rightarrow c_j} D(c_j, x_i)^2$$

avec :

- $c_j$  le centre de la classe,
- $x_i$ , la  $i^{\text{ème}}$  observation dans la classe ayant pour centre  $c_j$ ,
- $D(c_j, x_i)$ , la distance (euclidienne ou autre) entre le centre de la classe et le point  $x_i$ .

Le graphique obtenu à la forme d'un bras où le point le plus haut représente l'épaule et le point associé au nombre k le plus grand représente la main. Le nombre optimal de classe est au niveau du coude. Le coude représente en théorie le nombre de classes à partir duquel la variance ne se réduit plus significativement.

En appliquant cette méthode sur nos données nous obtenons la figure ci-dessous :

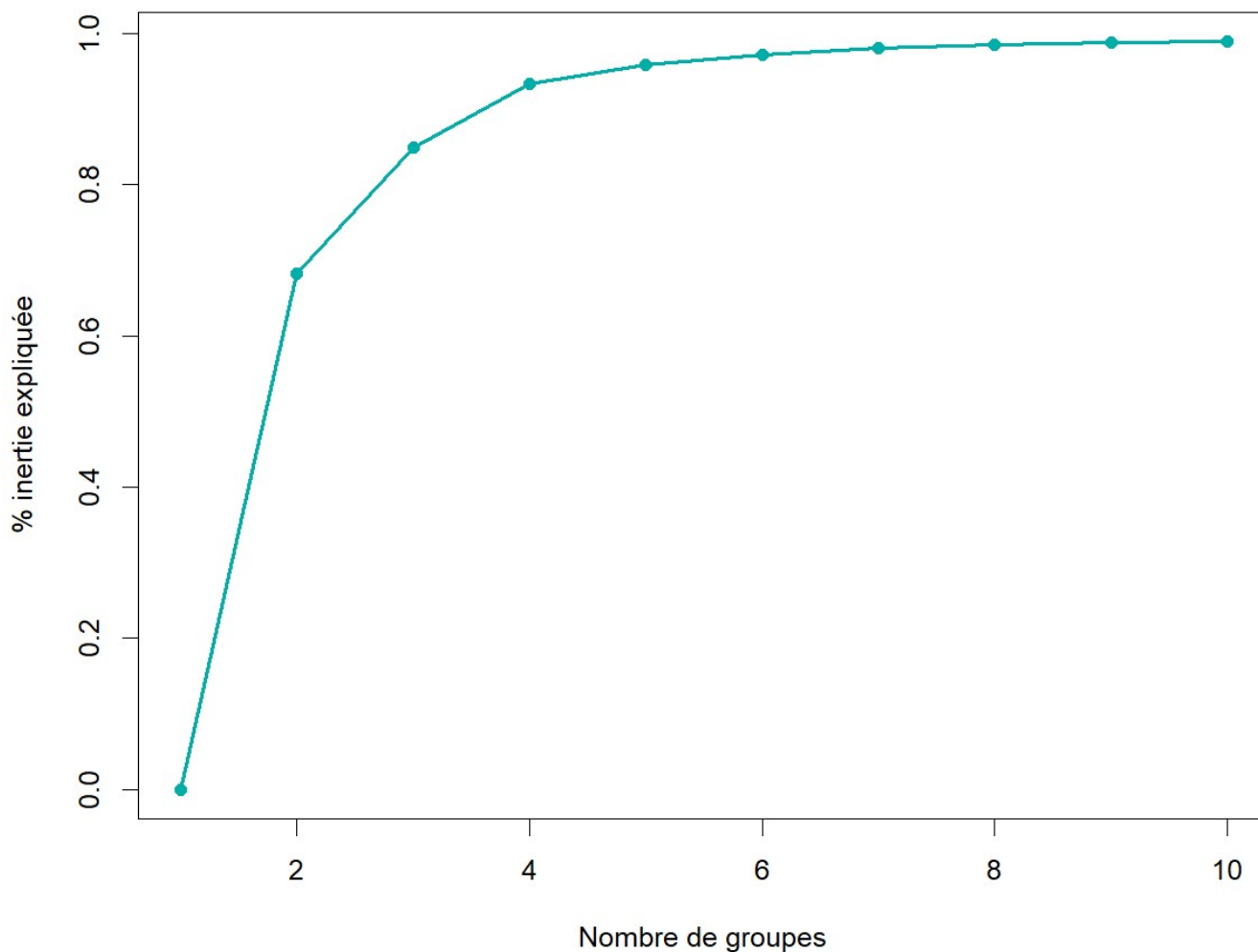


Figure 74 - Application de la méthode Elbow sur les données

A la figure 74, nous pouvons constater que le coude semble indiquer une segmentation optimale entre 3 et 4 classes. Compte tenu du choix optimal de 4 classes évoquée dans la segmentation par CAH, nous retiendrons également ici le choix de 4 classes.

#### 4. Comparaison des zoniers obtenus

Une fois ces deux classifications obtenues, nous allons les comparer l'une par rapport à l'autre ainsi qu'avec le zonier en place afin de déterminer lequel semble fournir le meilleur zonage d'une part et s'ils sont plus adaptés que le zonier actuel d'autre part. La comparaison des zoniers est un sujet fréquemment traité, notamment dans les mémoires d'actuariat comme dans (Bellina, 2014), (Sepulveda, 2016) ou encore (Ferrier, 2016).

En considérant la moyenne des résidus par zone, nous obtenons les coefficients correcteurs à appliquer dans le cadre d'un zonier. Les coefficients associés à chacune des zones sont les suivants :

Coefficients de zonage	CAH	K-moyennes
1	0,84	0,87
2	0,91	0,91
3	0,99	0,99
4	1,09	1,09

Figure 75 - Coefficients associés aux zones - CAH et K-moyennes

Cependant, la CAH et la classification K-moyennes ne présentent que très peu d'écart. En effet, seuls les départements 4, 74, 76, 77 présentent un écart de zone (zone 2 pour la CAH et zone 1 pour les k-moyennes).

Dans le cadre d'une tarification basée sur la prime pure, nous pourrions appliquer ces coefficients directement afin d'obtenir un tarif différencié par zone. En outre, nous pourrions déduire par exemple, qu'avec la classification proposée par les k-moyennes, une personne résidant dans la zone 1 devrait payer de l'ordre de 25% plus cher que celui en zone 4 (contre 30% pour la CAH).

Pour rappel, le zonier en place est le suivant :

Zone	Coefficient
1	0,95
2	1
3	1,05
4	1,15

Figure 76 - Zonier en place (vision technique)

Nous pouvons constater que les zoniers obtenus par CAH et K-moyennes sont plus discriminant, la variation entre la zone la moins chère et la plus chère n'étant que de 21% d'écart sur le coefficient technique du zonier en place.

Nous avons repris ci-dessous les départements avec la différence de rattachement la plus significative entre le zonier en place et le zonage proposé par CAH ou K-moyennes.

A la figure 77, les départements avec une zone en place significativement (d'au moins deux niveaux) moins importante que la classification CAH ou K-moyennes sont indiqués :

Départements	Zonier en place	CAH	K-moyennes
2	2	4	4
12	1	3	3
15	1	3	3
19	2	4	4
21	1	3	3
23	1	4	4
25	1	3	3
43	1	3	3
60	2	4	4
63	1	3	3
70	1	3	3
71	1	3	3
80	2	4	4
87	2	4	4
90	1	3	3

Figure 77 – Différence de rattachement significatif – zonier en place < zoniers théoriques

Il est important de noter que dans cette liste, dix ont fait l'objet d'un regroupement compte tenu du manque de représentativité du département.

La figure 78 représente la liste les départements avec une zone en place moins importante d'un niveau que la classification CAH ou K-moyennes :

Départements	Zonier en place	CAH	K-moyennes
1	1	2	2
3	1	2	2
4	1	2	1
5	2	3	3
7	1	2	2
9	2	3	3
10	1	2	2
11	2	3	3
14	1	2	2
18	1	2	2
22	1	2	2
29	1	2	2
32	2	3	3
34	3	4	4
35	1	2	2
36	1	2	2
37	1	2	2
38	2	3	3
39	1	2	2
46	2	3	3
48	2	3	3
50	1	2	2
56	1	2	2
61	1	2	2
65	2	3	3
66	2	3	3
69	2	3	3
79	1	2	2
82	2	3	3
86	1	2	2

Figure 78 – Différence de rattachement (1 niveau) – zonier en place < zoniers théoriques

D'autre part, les départements avec une zone en place plus importante que la CAH ou les K-moyennes (d'un niveau ou plus) sont représentés ci-dessous :

Départements	Zonier en place	CAH	K-moyennes
77	3	2	1
83	4	2	2
6	4	3	3
8	3	2	2
33	3	2	2
40	3	2	2
54	2	1	1
55	2	1	1
64	3	2	2
74	2	2	1
75	3	2	2
76	2	2	1

Figure 79 - Différence de rattachement significatif – zonier en place > zoniers théoriques

Sur la figure 79, nous pouvons constater que seuls les départements 77 et 83 présentent un écart de zone supérieur à deux niveaux pour la classification avec k-moyennes. En outre, il est à noter que le département 75 (Ville de Paris) a été classé en zone de consommation moyenne (zone 2) pour les classifications théoriques alors qu'il était plutôt classé en zone de consommation élevée (zone 3) actuellement et est souvent considéré comme un département avec une consommation assez forte, au même titre que le reste de l'Île de France. Aussi, soit la consommation dans ce département est particulière sur le portefeuille d'ADREA Mutuelle du fait potentiellement d'une exposition pas suffisamment significative, soit l'explication de la surconsommation attendue dans la ville de Paris est expliquée par d'autres facteurs déjà embarqués dans le prix en assurance individuelle (par exemple l'âge moyen des personnes ayant recours à des soins).

Sur les 92 départements intégrés à notre analyse (hors Alsace-Moselle et DOM-TOM), 36 présentent une classification équivalente à celle en place.

Après l'analyse des coefficients obtenus pour chacun des zoniers, nous allons appliquer des méthodes théoriques présentées en partie III afin de comparer les zoniers et valider le choix d'un des zoniers parmi les trois proposés. Ainsi, afin d'utiliser l'expérience des *GLM*, nous allons réaliser une régression avec chacun des zoniers estimés et comparer la déviance et les *AIC* pour savoir si la modification de zonier est effectivement statistiquement pertinente.

Comme évoqué dans (Planchet & Serdeczny, 2014), afin de tester la qualité de notre zonier, nous intégrons dans notre modèle *GLM* les coefficients calculés avec chacun des zoniers obtenus en *offset*, c'est-à-dire s'apparentant à une variable traitée comme une constante dans la modélisation (voir partie IV.1).

Nous comparons ici deux indicateurs déjà évoqués précédemment :

- Le critère de déviance,
- Le critère *AIC*.

<b>Comparaisons</b>	<b>Deviance</b>	<b>AIC</b>
<i>CAH</i>	1 581 695	8 049 690
<i>K-moyennes</i>	1 581 615	8 049 653
<i>En place</i>	1 583 669	8 050 584
<b>Plus adapté</b>	Kmeans	Kmeans
<b>Moins adapté</b>	En place	En place

Figure 80 - Comparaisons des critères de modèle pour choix de zonier

Comme nous pouvions le prévoir compte tenu des faibles écarts entre la CAH et la classification des k-moyennes, les indicateurs présentés à la figure 80 sont très proches pour ces deux classifications. Cependant, le zonier obtenu avec les K-moyennes semble être un peu meilleur selon ces critères. En revanche, le zonier en place présente pour chaque indicateur une estimation de moins bonne qualité. Il est important de noter que les écarts de déviance et d'*AIC* sont relativement faibles, laissant penser que le zonier en place permet une bonne estimation.

D'autres critères de comparaison sont présentés ci-dessous. Ce sont des critères qui auraient pu être utilisés pour déterminer le nombre de classes optimal. Nous les utiliserons ici pour comparer les trois zoniers obtenus.

Les deux critères qui seront présentés ici sont :

- Le critère Silhouette<sup>16</sup>,
- Le critère de Calinski-Harabasz<sup>17</sup>.

Le critère Silhouette, noté  $S(K)$ , est déterminé en calculant un indice propre à chaque individu  $i$ , noté  $s(i)$ , variant entre -1 et 1.  $s(i)$  prend en compte la moyenne notée  $c(i)$  des distances entre  $i$  et tous les individus de sa classe, mais aussi la moyenne  $v(i)$  avec ceux de la classe la plus proche.

$$s(i) = \frac{v(i) - c(i)}{\max(c(i); v(i))}$$

Plus  $s(i)$  est proche de 1, plus on peut affirmer que l'individu  $i$  se trouve dans la bonne classe. A contrario, s'il est trop proche de -1, l'individu est mal classé. Le critère Silhouette  $S(K)$  d'une partition à  $K$  fixé, correspond à la moyenne des  $s(i)$ . C'est pourquoi nous cherchons à le maximiser.

Le critère Calinski-Harabasz est défini comme suit :

$$\text{Critère}_{\text{Calinski-Har}} = \frac{SS_B}{SS_W} \times \frac{(N - k)}{(k - 1)}$$

où  $SS_B$  est la somme des variances inter-classes,  $SS_W$  la somme des variances intra-classe,  $k$  le nombre de classes et  $N$  le nombre d'observations.

Plus la valeur de ce rapport est élevée, plus les classes seront cohérentes (faible variance intra-classe) et plus les classes individuelles seront distincts / séparées (variance élevée inter-classes). Les résultats obtenus sont les suivants :

<b>Comparaisons</b>	<b>Silhouette</b>	<b>Calinski-Harabasz</b>
<i>CAH</i>	0,46	138,00
<i>K-moyennes</i>	0,50	158,00
<i>En place</i>	-0,10	6,30
<b>Plus adapté</b>	Kmeans	Kmeans
<b>Moins adapté</b>	En place	En place

Figure 81 - Critère de sélection de zonier – Silhouette et Calinski-Harabasz

Nous constatons à la figure 81, que selon les critères Silhouette et Calinski-Harabasz la classification par les K-moyennes semble être la plus indiquée. En revanche, le zonier en place semble une fois de plus moins efficient que les deux autres.

Compte tenu de l'ensemble de ces critères nous retiendrons le zonier obtenu par k-moyennes.

<sup>16</sup> Définition issue du site : <https://www.quantmetry.com/blog/initiation-au-clustering/>

<sup>17</sup> Définition issue du site : <https://help.tableau.com/current/pro/desktop/fr-fr/clustering.htm#:~:text=Le%20crit%C3%A8re%20Calinski-Harabasz%20est%20d%C3%A9fini%20comme%20o%C3%B9%20SS,nombre%20de%20clusters%20et%20N%20le%20nombre%20d%27observations.>

Dans la dernière partie de ce mémoire, les études complémentaires réalisées seront donc basées sur le zonier suivant :

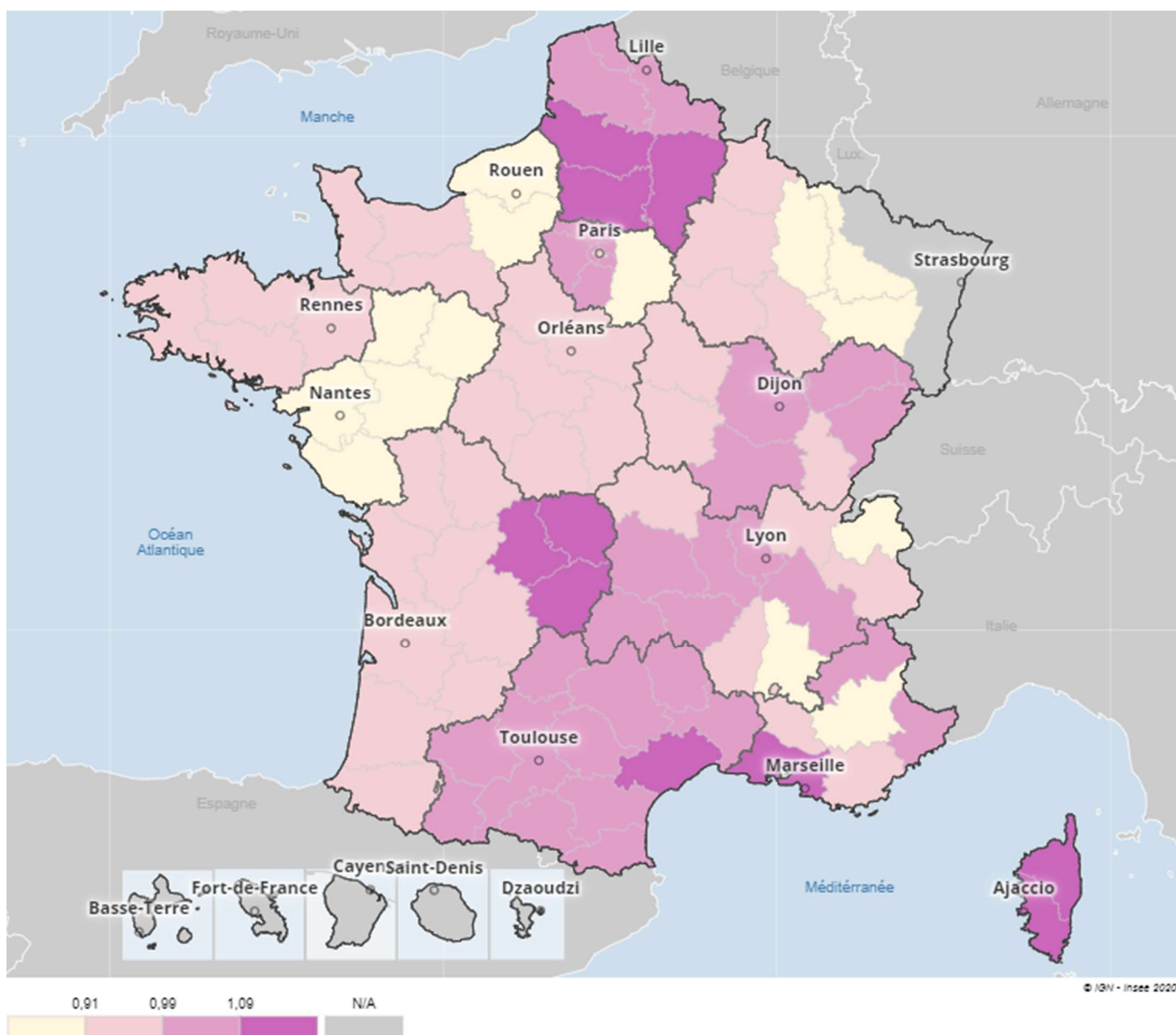


Figure 82 – Proposition de mise à jour du zonier technique

A noter sur la figure 82 que les départements « N/A » correspondent aux DOM ainsi qu’aux départements d’Alsace-Moselle pour lesquels le zonier n’a pas été estimé. La correction que nous avons proposée ici concerne le coefficient du zonier technique. Les modulations commerciales réalisées par ailleurs sont maintenues. Cela étant, les ajustements commerciaux avaient été déterminés sur la base du précédent zonier. Aussi, pour affiner l’exercice, ce zonier technique devrait être soumis à la direction commerciale afin de recenser leur demande d’ajustements éventuels. En effet, en cas de volonté commerciale importante sur certaines zones de conquête sur lesquelles la représentativité actuelle est faible, nous pourrions proposer des ajustements marginaux.

En outre, l’application de certaines réductions commerciales sur le zonier technique, conduit à la création de zone avec des coefficients de zonage extrêmement proche (voir figure 83).



<b>Zone</b>	<b>Coefficient</b>	<b>Exposition</b>
<b>1</b>	78%	6 230
<b>2</b>	82%	2 298
<b>3</b>	87%	9 428
<b>4</b>	89%	12 714
<b>5</b>	91%	48 953
<b>6</b>	98%	1 194
<b>7</b>	99%	61 052
<b>8</b>	109%	5 433

*Figure 83 - Nouveau zonier y compris coefficients commerciaux – avant rapprochement*

Ainsi, par cohérence opérationnelle et dans un souci de simplification et d'adhésion commerciale, nous avons fait le choix de regrouper les zones les plus proches sur un coefficient moyen. Cela étant, nous aurions aussi pu maintenir le zonier réalisé. C'est le cas pour les zones 6 et 7 qui présentent respectivement un coefficient de 98% et 99% et donc un écart de moins d'un point. En outre, l'exposition associée à la zone 6, relativement faible, constitue un argument supplémentaire pour procéder à ce rapprochement.

La figure 84 présente les coefficients commerciaux associé au zonier final.

<b>Zone</b>	<b>Coefficient</b>	<b>Exposition</b>
<b>1</b>	78%	6 230
<b>2</b>	82%	2 298
<b>3</b>	87%	9 428
<b>4</b>	89%	12 714
<b>5</b>	91%	48 953
<b>6</b>	99%	62 246
<b>7</b>	109%	5 433

*Figure 84 – Proposition finale de zonier*

## PARTIE V : CONSEQUENCES OPERATIONNELLES DE LA MODIFICATION D'UN ZONIER

La mise à jour d'un zonier existant, lorsqu'elle est de grande ampleur comme cela pourrait être le cas en cas d'adoption de ce zonier, serait lourde de conséquences sur de nombreux points.

Dans cette partie, nous présenterons les conséquences d'un changement de zonier à travers la projection du compte de résultats en se basant sur des hypothèses de production et de résiliation commerciales sensible au changement de zonier.

Nous présentons à la figure 85, le compte de résultats en survenance 2019. Pour rappel, le P/C brut correspond au rapport entre les prestations et les cotisations hors taxes (HT) et la marge brute à la différence entre les cotisations HT et les prestations.

Année	Exposition	Cotisations HT	Prestations	P/C Brut	Marge brute
<b>2019</b>	147 301	84 362 804 €	64 730 068 €	76,7%	19 632 736 €

Figure 85 - Compte de résultats - survenance 2019

Il est à noter que dans ces projections, la contribution forfait médecin traitant, égale à 0,8% des cotisations hors taxes a été intégrée en charge de prestations. En outre, les départements liés au régime Alsace-Moselle ont été exclus de la projection.

L'objectif de cette partie va être de proposer, par un exemple théorique, les conséquences possibles d'une modification de zonier sur la projection de notre compte de résultats à un an. La projection pluriannuelle n'a pas été retenue pour plusieurs raisons :

- Les hypothèses utilisées pour la modification d'un zonier seront difficiles à justifier du fait de l'absence de données et de recul pour les valider,
  - ⇒ La projection pluriannuelle, elle aussi basée sur des hypothèses de projection, aurait présenté des résultats encore plus difficiles à interpréter et à exploiter.
- Nous avons, faute de recul sur les événements récents lié au COVID-19, décidé de ne pas retenir dans la projection théorique 2020 les éléments récents constatés soit :
  - La baisse de prestations attendue sur 2020,
  - La contribution COVID fixée à 2,6% des cotisations hors taxes.
  - ⇒ De même, l'absence de prise en compte de ces hypothèses rend la projection pluriannuelle encore moins significative que la projection 2020 que nous saurons n'être que théorique.
- La mise à jour des cotisations, suite à la modification du zonier aura de fortes conséquences la première année et moins les années suivantes, à moins qu'elle entraîne une modification profonde de la structure de portefeuille.
  - ⇒ Si c'est effectivement une possibilité, la sensibilité de notre projection aux hypothèses retenues sera trop importante pour rendre une projection pluriannuelle de qualité.

Ainsi, nous allons réaliser une projection de ces résultats 2019 en 2020 en considérant que la modification du zonier aurait été réalisée au 1<sup>er</sup> janvier 2020.

La première conséquence de la modification d'un zonier est évidemment tarifaire. En effet, pour l'adhérent, la modification du coefficient d'abattement associé à sa zone va engendrer un changement de cotisations à la hausse ou à la baisse.

Cependant, il est important, en prérequis, de ramener l'application du zonier à une cotisation équivalente. Ainsi, si le zonier proposé est plus discriminant, il est important que son application conduise à une cotisation quasi équivalente à celle en place. L'objectif de la mise à jour du zonier n'est pas d'augmenter globalement la cotisation mais bien d'avoir une discrimination plus adaptée à chaque zone. C'est pourquoi, nous avons, sur la population en place par département calculé un coefficient de correction à appliquer sur la cotisation sur laquelle le zonier sera appliqué.

Pour se faire, pour chaque zonier, nous avons calculé un coefficient de correction moyen selon la formule suivante :

$$Coeff\_corr_{en\ place} = \frac{\sum_{i=1}^N expo_i \times coeff\_zoneenplace_i}{\sum_{i=1}^N expo_i}$$

$$Coeff\_corr_{propo} = \frac{\sum_{i=1}^N expo_i \times coeff\_zonepropo_i}{\sum_{i=1}^N expo_i}$$

*i* représentant le numéro d'un département.

Le coefficient correcteur appliqué sur la cotisation permettant un raisonnement à cotisations équivalente est donc :

$$Correction_{zonier} = \frac{Coeff\_corr_{en\ place}}{Coeff\_corr_{propo}} - 1$$

Les coefficients que nous comparons sont ceux des zoniers commerciaux. Ainsi, la proposition de modification que nous réalisons ne concerne que la partie technique du zonier, les modulations commerciales ayant été maintenues tels qu'elles existaient (voir partie II.2-2.2 et IV.3-3.2).

Dans notre exemple, les résultats sont les suivants :

$$Coeff\_corr_{en\ place} = 96,6\%$$

$$Coeff\_corr_{propo} = 94,3\%$$

Soit un coefficient de correction sur le tarif avant application du zonier de :

$$Correction_{zonier} = 2,4\%$$

Ainsi, nous allons augmenter le tarif moyen de 2,4% afin de lui appliquer un zonier plus discriminant que celui en place.

Une fois cette étape réalisée, nous pouvons déterminer les départements pour lesquels le tarif va augmenter ou diminuer et dans quelles proportions. La figure 86 présente ces résultats sous forme de tableau.

Impact sur les cotisations	Nombre de Départements	Exposition
inférieur à - 10%	12	12 813
supérieur à -10% et inférieur à -5%	34	22 925
supérieur à -5% et inférieur à 0%	30	74 170
supérieur à 0% et inférieur à 5%	10	36 996
supérieur à 5%	6	397

Figure 86 – Impact de la modification du zonier sur la cotisation

Les autres hypothèses de projections à retenir sont les suivantes :

- Taux d'inflation : il représente l'augmentation prévisionnelle des prestations entre 2019 et 2020. Ce taux prend à la fois en compte une augmentation structurelle des soins mais aussi d'éventuelles modifications réglementaires (par exemple le 100% Santé). Comme déjà évoqué, nous ne prenons pas en compte les évolutions de prestations liées au COVID.  
⇒ Pour notre projection, ce taux d'inflation a été fixé à 1,5%.
- Le taux directeur : il représente le taux d'augmentation qui sera appliqué sur les cotisations entre 2019 et 2020. Il est en général assez corrélé avec le taux d'inflation. Cependant, il peut aussi permettre de redresser un portefeuille déficitaire ou de financer des investissements structurels d'une organisation en cas de fonds propres insuffisants par exemple.  
⇒ Pour notre projection, ce taux sera fixé à 1,5%, le portefeuille ne présentant pas de nécessité de redressement particulier.
- Taux de résiliation / adhésion : la projection d'un portefeuille nécessite de considérer les mouvements de portefeuille. Ainsi, afin de réaliser une projection au plus près de la réalité, nous avons utilisé les taux d'adhésion et de résiliation constatés sur les années précédentes. Si nous avons pris le taux moyen des résiliations 2017 et 2018, nous n'avons pris pour l'adhésion, que le taux constaté en 2018. En effet, le portefeuille étant encore en croissance forte entre 2017 et 2018, il est important de ne considérer que l'année de souscription la plus récente compte tenu du poids du stock. Ainsi, si la courbe de croissance d'un portefeuille est croissante jusqu'à atteindre son rythme de croisière, sa dérivée est décroissante compte tenu notamment du poids du stock et de l'impact de plus en plus limité des affaires nouvelles. La croissance nette des résiliations constatée entre 2018 et 2019 n'étant que de moins de 3% (contre 9% entre 2017 et 2018), il semble que l'application d'une croissance brute de résiliation stable soit acceptable pour 2020.  
⇒ Les taux de résiliation / adhésion ont été calculés propre aux départements sur lesquels l'exposition 2019 dépassait 1000 personnes protégées afin de prendre en compte les spécificités de distribution / radiation dans les zones à fortes présence commerciale et en moyenne sur l'ensemble des autres départements.

Il est important de préciser que, dans une première projection, nous ne modifierons pas les taux d'adhésion / résiliation et nous baserons sur les constats ci-dessus malgré la mise à jour du zonier avant d'impacter ces taux dans une seconde projection. En outre, pour ces deux projections, l'absence de modification sur certains points implique des hypothèses sous-jacentes comme l'absence de modification de l'âge moyen du groupe. Ainsi, nous acceptons que les adhésions / résiliations n'impactent pas de fait l'âge moyen de la population.

Dans un premier temps, nous avons donc réalisé la projection globale ci-dessous pour 2020.

Année	Exposition	Cotisations HT	Prestations	P/C Brut	Marge brute
<b>2020v1</b>	150 779	87 708 297 €	67 244 410 €	76,7%	20 463 887 €
<b>Ecart 2020v1/2019 - 1</b>	2,4%	4,0%	3,9%	-0,1%	4,2%

Figure 87 - Première projection 2020 - pas de modification des hypothèses d'évolution d'effectifs 2020

Comme nous pouvons le constater à la figure 87, selon nos hypothèses, nous constatons une augmentation des effectifs de 2,4%, qui, associée à un taux directeur et un taux d'inflation de 1,5%, entraîne une augmentation des cotisations et des prestations de 3,9%. Comme attendu, le P/C brut global n'est que très légèrement modifié (-0,1%) et la marge brute progresse de l'effet conjoint de l'augmentation des cotisations et de la faible amélioration du P/C brut. En revanche, le zonier ayant été modifié, il est important de comprendre que des variations de P/C brut par département importantes sont constatées. Ainsi, 17 départements présentent un P/C brut qui se dégrade de plus de 10% en valeur absolue et jusqu'à un maximum de +22,5% pour le département 83. C'est la nouvelle répartition des personnes protégées par département qui entraîne une légère modification du P/C. Cela étant, l'estimation réalisée n'intégrant pas d'impact propre au zonier sur les effectifs, l'analyse ne sera pas plus détaillée.

Nous prenons maintenant la position de modifier de manière structurante les hypothèses d'adhésion / résiliation en prenant en compte l'impact potentiel des modifications du zonier sur les hypothèses de production et de résiliation.

Afin de trouver une règle simple mais cohérente, nous nous baserons sur un indicateur important qui sera l'impact adhérent. L'impact adhérent correspond à l'évolution de la cotisation pour une personne protégée donnée. Il se décompose, hors modification d'un zonier, des paramètres suivants en assurance individuelle :

- le taux directeur (déjà évoqué ci-dessus),
- le pas d'évolution par âge.

En individuel, la cotisation étant dépendante de l'âge pour l'adhérent, le fait de vieillir d'un an, engendrera une indexation automatique de la cotisation. Ce pas d'âge est estimé sur Flex'ADREA en moyenne à 2,7%. Ainsi, compte tenu de la modification du zonier, nous pouvons estimer un impact adhérent qui correspondra au produit des trois paramètres suivants :

- taux directeur,
- pas d'âge,
- modification zonier.

Sur ces bases, nous proposons la règle ci-dessous pour prendre en compte l'impact de la modification du zonier.

Impact cotisation adhérent	Impact adhésion	Impact radiation	Nombre de département	Exposition
<b>inférieur à -7%</b>	25%	-50%	12	12 813
<b>supérieur à -5% et inférieur à 0%</b>	15%	-25%	22	10 968
<b>supérieur à +1% et inférieur à +3%</b>	0%	0%	31	56 613
<b>supérieur à +6% et inférieur à +8%</b>	-25%	50%	11	29 515
<b>supérieur à +10%</b>	-50%	50%	16	37 393

Figure 88 - Proposition de règle de projection des effectifs

Ces coefficients viendront impacter les coefficients d'adhésion et de résiliation évoqués dans la projection précédente. Les justifications de ces règles restent qualitatives. Il faut tout d'abord préciser que, compte tenu des modifications de zonier et des impacts harmonisés de taux directeurs et de pas d'âge, les impacts adhérents ne sont pas illimités puisque nous en obtenons douze différents, ce qui justifie les intervalles choisis.

Ainsi, pour chaque impact adhérent retenu :

- Au-delà ou égal à -7% :
  - Résiliation : nous avons considéré que la sensibilité à la résiliation était majoritairement liée au prix et qu'une diminution supérieure à 7%, quand la majoration standard attendue par l'adhérent est plutôt autour de +4%, devait limiter de manière très importante les résiliations. Une part des résiliations étant liée à des décès, il semble important de maintenir des résiliations structurelles.
  - Adhésion : un des freins principaux au développement est le prix. Ainsi, des prix attractifs pourraient engendrer une « sur-adhésion » significative compte tenu de l'impact sur le prix du changement de zonier. Cela étant, cette hypothèse n'est pas différenciée par département alors qu'il semble probable que l'impact pourrait être très différent en fonction des départements (entre ceux pour lesquels la cotisation est déjà bien positionnée et ceux où elle est aujourd'hui mal positionnée). Cela étant, l'impact sur les ventes a été estimé moins important que sur les résiliations. En effet, nous avons considéré que les impacts principaux se feraient sur des départements sur lesquels les équipes de ventes étaient déjà assez représentées et finalement un peu moins dans des départements dans lesquels la mutuelle est moins présente malgré le fait que nous pourrions avoir dans ces départements des augmentations de vente de plus de 100 ou 200%.
- Entre -5% et 0% :
  - Résiliation : Le raisonnement est sensiblement le même que dans le premier cas et évolue en cohérence avec l'impact associé.
  - Adhésion : Le raisonnement est sensiblement le même que dans le premier cas et évolue en cohérence avec l'impact associé.
- Entre +1% et +3% :
  - Résiliation : un impact adhérent entre +1% et +3% correspond peu ou prou aux standards des renouvellements appliqués chaque année. C'est pourquoi, nous avons considéré que ces modifications ne viendraient pas modifier les taux de résiliation appliqués dans la projection précédente.
  - Adhésion : De même que pour les résiliations, un impact adhérent entre +1% et +3% correspondant peu ou prou aux standards des renouvellements appliqués chaque année, nous avons considéré que ces modifications ne viendraient pas modifier les taux d'adhésion appliqués dans la projection précédente.
- Entre +6% et +8% :
  - Résiliation : la sensibilité aux prix étant la cause principale de résiliation, nous considérons que l'impact d'une telle augmentation de prix sera très impactante sur le portefeuille. En effet, à la suite d'échange avec les commerciaux, un impact adhérent supérieur à +6% est estimé lourd de conséquence sur le maintien ou non dans une gamme.
  - Adhésion : dans le même sens, l'impact sur les adhésions pourrait être très pénalisant, le risque de se « déposer » étant très important avec des augmentations de cotisations aussi importantes. Au contraire, avec les baisses importantes envisagées dans les départements pour lesquels nous sommes déjà « hors marché », la conséquence sera finalement très faible car il n'y aura pas plus de vente. En revanche, dans un département dans lequel il existe un gros portefeuille historique, le potentiel de vente pourrait être annihilé.
- Supérieur à +10% :
  - Résiliation : l'impact sur la résiliation a été estimé identique à celle d'un taux supérieur à +6%.
  - Adhésion : sur l'adhésion, nous avons considéré qu'une augmentation supérieure à +10% engendrerait un impact plus important encore et freinerait les ventes de moitié.

Si l'ensemble de ces coefficients sont effectivement sujet à discussion, l'absence de données rend leur validation délicate. C'est pourquoi, nous avons réalisé des tests de sensibilité, non pas sur le choix des intervalles mais bien sur les conséquences associées à chaque intervalle.

Le résultat de la projection obtenue au global est le suivant :

Année	Exposition	Cotisations HT	Prestations	P/C Brut	Marge brute
<b>2020v2</b>	144 960	83 830 185 €	64 339 085 €	76,7%	19 491 099 €
<b>Ecart 2020v2/2019 - 1</b>	-1,6%	-0,6%	-0,6%	0,0%	-0,7%
<b>Ecart 2020v2/2020v1- 1</b>	-3,9%	-4,4%	-4,3%	0,1%	-4,8%

Figure 89 - Projection suite à modification des hypothèses d'adhésion / résiliation

Comme nous pouvons le voir à la figure 89, l'impact de la modification des hypothèses d'adhésion / résiliation entraîne une baisse globale de production et de fait de cotisations mais ne modifie que très marginalement le P/C brut et l'évolution de la marge brute est directement corrélée avec l'évolution des cotisations. En revanche, entre 2019 et 2020, comme évoqué précédemment, les modifications par département sont significatives.

L'analyse détaillée sur les départements les plus impactés par les modifications, présentée en figure 90 permettra de mesurer l'impact d'un département à l'autre. A noter que le total de la figure 90 correspond à l'impact sur l'ensemble des départements et pas uniquement ceux avec un impact important.

Départements	Exposition	Cotisations HT	Prestations	P/C Brut	Marge brute	2020	2020	2019
						Impact PP	Impact Cot (€)	P/C (en points)
4	879	422 182 €	372 917 €	88,3%	49 265 €	7,2%	28 457 €	5,3%
5	1 893	1 054 834 €	803 381 €	76,2%	251 453 €	-5,5%	-61 118 €	-1,3%
6	331	188 009 €	139 519 €	74,2%	48 491 €	13,5%	22 344 €	8,5%
21	3 321	2 184 248 €	1 611 987 €	73,8%	572 261 €	-18,2%	-484 550 €	-5,2%
25	13 473	8 006 493 €	6 162 267 €	77,0%	1 844 227 €	-16,4%	-1 566 630 €	-5,4%
27	697	322 382 €	272 344 €	84,5%	50 037 €	7,2%	21 730 €	5,0%
34	3 882	2 503 760 €	1 822 030 €	72,8%	681 730 €	-19,3%	-596 908 €	-4,8%
38	17 657	11 372 187 €	8 567 691 €	75,3%	2 804 496 €	-4,9%	-590 497 €	-1,3%
53	749	362 937 €	319 089 €	87,9%	43 848 €	7,2%	24 463 €	5,2%
58	823	424 773 €	355 028 €	83,6%	69 745 €	7,2%	28 631 €	5,3%
63	1 386	759 908 €	616 914 €	81,2%	142 994 €	-18,1%	-168 435 €	-5,7%
69	8 621	5 094 310 €	4 027 458 €	79,1%	1 066 851 €	-5,9%	-320 092 €	-1,4%
70	1 304	720 397 €	622 920 €	86,5%	97 477 €	-18,1%	-159 002 €	-6,1%
71	7 153	4 504 898 €	3 401 138 €	75,5%	1 103 759 €	-18,4%	-1 012 797 €	-5,3%
72	1 490	766 166 €	648 681 €	84,7%	117 485 €	4,6%	33 517 €	5,1%
73	4 078	2 224 868 €	1 743 870 €	78,4%	480 998 €	4,1%	87 556 €	4,9%
74	9 620	4 553 211 €	3 585 967 €	78,8%	967 244 €	14,0%	559 136 €	8,4%
75	600	309 905 €	260 144 €	83,9%	49 761 €	17,4%	45 937 €	9,0%
76	3 269	1 624 501 €	1 363 236 €	83,9%	261 265 €	15,0%	211 840 €	8,9%
83	332	196 075 €	135 485 €	69,1%	60 591 €	13,5%	23 302 €	12,7%
84	708	345 153 €	308 601 €	89,4%	36 552 €	7,2%	23 265 €	5,6%
89	985	496 746 €	413 632 €	83,3%	83 115 €	20,9%	85 945 €	5,2%
90	514	272 056 €	225 400 €	82,9%	46 656 €	-18,3%	-61 046 €	-5,8%
<b>Total</b>	<b>144 960</b>	<b>83 830 185 €</b>	<b>64 339 085 €</b>	<b>76,7%</b>	<b>19 491 099 €</b>	<b>-3,9%</b>	<b>-3 878 112 €</b>	<b>0,0%</b>

Figure 90 - Résultats de la modification de zonier – zoom par département

Comme nous pouvons le constater à la figure 90, les modifications engendrées par la baisse de zonier semblent être défavorables en matière de production, les PP diminuant de -3,9% et les cotisations de près de 4 millions d'Euros par rapport à la projection 2020 avant impact sur les adhésions / résiliations.

Les départements les plus impactés (impact CA supérieur à -150 K€) sont les départements 21, 25, 34, 38, 63, 69, 70 et 71. Si certains sont associés à un nombre de personne protégée important couvert par ADREA Mutuelle (25, 38, 69 et 71), d'autres le sont beaucoup moins comme les départements 63 et 70. Il est à noter que les quatre départements les plus exposés représentent un impact sur les cotisations de -3,5 M€.

Comme nous pouvons le voir à l'inverse, seuls les départements 74, 76 et dans une moindre mesure, 89 et 73 pourraient bénéficier de la modification de zonier de manière significative en matière de personnes protégées et de CA puisque ce dernier pourrait augmenter d'un total de 944 K€ par rapport à la projection 2020 sur ces trois départements.

L'analyse d'évolution des P/C entre 2019 et 2020 peut conduire à des situations par département très différentes. Ainsi, en considérant des frais de fonctionnement de l'ordre de 20%, les P/C inférieurs à 80% peuvent être estimés bénéficiaires et ceux supérieurs à 80% peuvent être estimés déficitaires.

- Les P/C aujourd'hui bénéficiaires qui se dégradent :
  - o La dégradation rend les P/C déficitaires : départements 27, 58, 72, 75, 76 et 89.  
Pour ces départements, la problématique de renouvellement sera importante. En effet, la dégradation des P/C n'est pas liée à la problématique du zonier mais bien à une répartition plus importante sur certains produits plus déficitaires que d'autres ou à une adéquation entre le tarif et la consommation par âge moins adaptée. En outre, ces P/C sont associés à des départements en croissance de portefeuille. Aussi, en dehors de la problématique de renouvellement, la souscription sur ces départements doit nécessiter une attention particulière pour ne pas nuire globalement à l'équilibre de la gamme.
  - o Les P/C dégradés qui restent bénéficiaires : départements 6, 73, 74 et 83.  
Pour ces départements, si le changement de zonier va venir diminuer de manière importante la marge, les P/C resteront favorables tant à un renouvellement amoindri qu'à une souscription importante.
- Les P/C déficitaires qui s'améliorent :
  - o L'amélioration permet aux P/C de devenir bénéficiaire : départements 25, 69 et 71.  
Pour ces départements, le zonier présente un intérêt de renouvellement très important techniquement. Cependant, comme nous avons pu le voir pour les départements 25 et 71, ces mises à jour peuvent être très impactantes sur le stock et la production (-2,6M€ de CA entre 2019 et 2020).
  - o Malgré une amélioration, les P/C restent déficitaires : départements 63, 70, 90.  
Pour ces départements, la mise à jour du zonier a également un impact bénéfique sur le suivi technique. La mise à jour du zonier a ainsi un impact significatif sur le P/C (amélioration supérieure à 5,5%) et concerne des départements avec une exposition limitée (entre 500 et 1500 personnes protégées).
- Les P/C bénéficiaires qui s'améliorent : département 5, 21, 34, 38  
Hormis pour le département 5, ce sont des départements avec une exposition importante (en particulier pour le département 38 avec une exposition de plus de 18 000 personnes protégées en 2019). Sur ces départements, la mise à jour du zonier va engendrer, dans notre projection, une érosion du CA importante malgré des P/C aujourd'hui bénéficiaire (-1,7M€ de CA). Cette mise à jour va nécessairement engendrer une incompréhension importante au niveau commercial. Il sera probablement nécessaire en cas de mise à jour de prévoir une transition en douceur pour les adhérents et les chargés de développement par le biais par exemple d'action commerciales ponctuelles (mois gratuits par exemple). Si cela engendrera un coût supplémentaire pour



l'entreprise, la fidélisation des adhérents reste une problématique essentielle d'autant plus compte tenu de la mise en place de la résiliation infra-annuelle depuis le 1<sup>er</sup> décembre 2020.

- Les P/C déficitaires dégradés par la mise à jour du zonier : départements 4, 53, 84.  
Ces départements présentent une exposition relativement faible (inférieure à 900 personnes protégées). Cela étant, préconiser le développement sur ce type de département doit être traité avec beaucoup de précaution. La dégradation du P/C alors qu'il est déjà déficitaire n'est pas parfaitement intuitive. Aussi, s'il est possible que le fait que le P/C soit dégradé pourrait être lié à la souscription de produits particulièrement dégradés, ou encore de souscription sur des tranches d'âge pour lesquels le P/C est dégradé, nous ne pouvons exclure que le zonage ne soit pas parfait. En outre, la classification sera plus importante pour les éléments dont l'écart à la moyenne de la classe sera le plus élevé. Il faudra analyser et comprendre sa source avant toute action commerciale d'ampleur.

La figure 91 présente une synthèse des éléments explicités ci-dessus.

Evolution	Départements	Impact Cotisations HT	P/C Brut 2020 moyen	Impact P/C Brut (2020-2019)	Impact Marge Brute
P/C 2019 < 80% => P/C 2020 > 80%	27, 39, 72, 75, 76, 89, 92	192 049,94 €	84%	7%	204 841,77 €
P/C 2019 < 80% en dégradation => P/C 2020 < 80%	1, 21, 71, 73, 74	-127 765,64 €	76%	1%	208 961,24 €
P/C 2019 > 80% => P/C 2020 < 80%	25, 26, 69	-353 589,92 €	78%	-4%	-430 646,28 €
P/C 2019 > 80% en amélioration => P/C 2020 > 80%	4, 58, 63, 70, 90	-145 755,67 €	84%	-3%	-46 476,68 €
P/C 2019 < 80% en amélioration => P/C 2020 < 80%	34, 38	-430 676,20 €	75%	-2%	-171 577,02 €
P/C 2019 > 80% en dégradation => P/C 2020 > 80%	53, 84	28 857,26 €	89%	5%	33 658,97 €

Figure 91 – Synthèse des évolutions de P/C suite à la mise à jour du zonier

Comme évoqué précédemment, l'exemple ci-dessus est basé sur des hypothèses structurante d'évolution d'effectifs. Pour analyser ce point, nous proposons trois sensibilités basées sur les hypothèses suivantes :

- Scénario 1 :
  - o sensibilité à la résiliation plus marquée en cas de hausse et de baisse de prix (+ ou – 10 points),
  - o sensibilité à la souscription plus marquée en cas de hausse et de baisse de prix (+ ou – 10 points).
- Scénario 2 :
  - o sensibilité à la résiliation plus marquée en cas de hausse de prix et moins en cas de baisse (+ ou – 10 points),
  - o sensibilité à la souscription moins marquée en cas de baisse de prix et plus en cas de hausse (+ ou – 10 points).
- Scénario 3 : taux directeur de 5% et taux d'inflation de 3%.

Pour plus de précisions, les tableaux d'impacts associés aux scénarios 1 et 2 sont présentés en annexe 5.

Le scénario de référence, permettant le calcul des écarts en figure 92 correspond au scénario présenté ci-dessus y compris hypothèses de modifications d'adhésions / résiliations.

Impact	Effectifs (%)	CA (%)	Prestations (%)	P/C (point)	Marge (€)
<b>Scénario 1</b>	-0,7%	-0,9%	-0,9%	0,0%	197 227 €
<b>Scénario 2</b>	-2,0%	-2,0%	-2,0%	0,0%	367 796 €
<b>Scénario 3</b>	-5,1%	-1,9%	-3,8%	1,4%	808 942 €

Figure 92 – Synthèse des scénarios de sensibilité

Les scénarios 1 et 2 ne semblent pas montrer une sensibilité trop importante au choix de nos hypothèses. Le troisième scénario en revanche engendre une baisse d'effectifs couverts importante.

En synthèse, la mise à jour d'un zonier est opérationnellement fortement impactant. Si elle permet l'identification potentielle de sources de développement sur des départements peu ou pas représentés en portefeuille, son impact sur le stock peut être significatif en matière de maintien en portefeuille.

En outre, le suivi du zonier tel qu'il est réalisé en analysant les P/C par département semble contre indiqué du fait des éléments externes contenus dans les P/C (tranches d'âge souscrites, produits déficitaires sur représentés, etc.). Aussi, s'il semble plus indiqué, pour faire le suivi du zonier au cours du temps de le faire en mettant à jour les données et en réestimant le zonier systématiquement (par exemple, déplacer les départements classés dans des zones différentes d'une analyse à l'autre, ou revaloriser ou diminuer marginalement le coefficient associé à une zone), nous sommes aussi conscients des limites de l'exercice :

- la mise à jour et le suivi est lourd,
- la mise à jour systématique du zonier même marginalement ne serait pas bien comprise par les commerciaux d'une part et par les adhérents d'autre part, pour lesquels les évolutions de cotisations ne seraient pas aussi standardisées d'une année sur l'autre,
- le zonage théorique peut conduire à des conclusions contre-intuitives, comme l'a montré l'exemple ci-dessus qui seront difficile à mettre en œuvre,
- comme pour beaucoup de méthodes théoriques appliquées opérationnellement, l'erreur d'analyse existe, en particulier dans ce genre de méthodes « itératives » ou une anomalie d'analyse en première étape peut conduire à des incohérences majeures.

Enfin, comme nous avons pu le constater dans la sensibilité au prix intégrant le taux directeur majoré, il semble important, en cas de choix de modification ou de création de zonier sur une gamme existante de le faire une année où le taux directeur sera relativement faible (voir scénario 3).

Il est à noter que les conséquences déjà évoquées plus haut qu'elles soient techniques ou commerciales doivent être complétées des conséquences pour la gestion et le paramétrage, ainsi que les modifications contractuelles et juridiques.

En effet, le paramétrage d'un zonier peut être assez conséquent puisqu'il est généralement associé à plusieurs offres individuelles. Aussi, la modification peut engendrer un nombre important d'ajustements nécessaires dans les outils de gestion et générer des sources d'erreur.

Il est également important de noter que, chez ADREA Mutuelle la modification du zonier a proprement parlé n'engendre pas de modification du règlement mutualiste car les éléments relatifs au zonier sont strictement internes et ne sont pas définis dans le règlement. Cela étant, en fonction de la définition du règlement mutualiste, la conséquence de modification du zonier peut être plus impactante.

En outre, la modification des cotisations pour les adhérents nécessite une communication soignée et justifiée ce qui pourra engendrer des réclamations massives qui devront être traitées en agence. Or, comme nous l'avons vu plus haut, la modification d'un tarif lié à la modification du zonier peut être un message difficile à porter pour les services commerciaux.

## Conclusion

La réalisation d'un zonier basée sur les données d'expérience si elle est désormais relativement maîtrisée chez les différents assureurs n'est pas pour autant chose aisée. En effet, comme nous avons pu le voir, les données d'expérience parfois inexistantes sur certaines zones ne permettent pas la création d'un zonage de qualité sur ces régions. S'il aurait effectivement été possible d'utiliser des variables exogènes issues des données nationales (population, bassin de santé, salaire moyen, proportion urbain/rural, couverture médicale, etc.), ce n'était pas l'objet de ce mémoire.

En outre, l'utilisation des *GLM* pour la réalisation de la régression globale qui impose la spécification de la distribution de la variable à expliquer, conduit à des résultats d'estimation de la prime globale plus ou moins satisfaisants. Ainsi, le choix de la loi dans notre mémoire ne permet pas de bien estimer les prestations importantes et les prestations nulles malgré le choix d'une modélisation basée sur la loi Tweedie. Les méthodes de zonage se basant sur la comparaison des résidus avec et sans l'analyse d'une variable de zonage, il devrait être possible d'utiliser le même raisonnement en utilisant des méthodes de machine learning potentiellement plus efficaces bien qu'avec des résultats qui pourraient s'avérer plus difficiles à analyser.

L'objectif de ce mémoire était de savoir si le suivi du zonier réalisé aujourd'hui de manière empirique en analysant les P/C par département ne pouvait être remplacé par une mise à jour un peu plus globale du zonier sur les zones avec une certaine présence commerciale. La création d'un nouveau zonier qui semblait être statistiquement plus adapté a donc été réalisée. Il a notamment conduit à valider le nombre de zones de différenciation de 4 mais tend à justifier que la méthode de suivi du zonier par les P/C, si elle a le mérite d'être simple et comprise, ne semble pas parfaitement adaptée et peut conduire à des erreurs de raisonnement. En effet, le P/C par département peut s'expliquer par des critères de population spécifiques comme la répartition par âge ou par produit conduisant à des déficits plus importants.

Cependant, nous avons pu constater dans la dernière partie que la mise à jour d'un zonier est lourde de conséquences opérationnellement parlant. Ainsi, si elle permet d'identifier des opportunités de développement commerciales dans des zones sur lesquelles l'ancien zonier pouvait constituer un frein, il semble difficile d'impacter les adhérents en portefeuille de la modification globale du zonier. Aussi, la proposition de mise à jour du zonier, si elle ne semble pas pouvoir s'appliquer sur un produit existant, semble tout à fait adaptée pour participer à la réflexion non seulement d'un développement commercial sur des zones aux tarifs inadaptés dans le zonage en place mais aussi pour sa mise en place sur un nouveau produit.

Comme déjà évoqué précédemment, la limitation du zonage par l'absence de données sur certaines zones, pourra être comblée par la fusion à venir avec les portefeuilles de deux acteurs mutualistes, présent sur des zones sur lesquelles ADREA Mutuelle n'est pas présente et notamment le Nord de la France.

## Bibliographie

- ADREA. (2018). *Rapport sur la solvabilité et la situation financière (SFCR) - ADREA MUTUELLE*. ACPR.
- Aouizerate, J.-M. (2012). Alternative neuronale en tarification santé. *BULLETIN FRANCAIS D'ACTUARIAT*, vol. 12, n°23, 97-127.
- Bellina, R. (2014). *Méthodes d'apprentissage appliquées à la tarification non-vie*. Mémoire Actuariat ISFA.
- Beraud-Sudreau, G. (2017). *Construction d'un zonier en MRH à l'aide d'outils de DATA SCIENCE*. Mémoire Actuariat CNAM.
- Charpentier, A., & Denuit, M. (2004). *Mathématiques de l'assurance non vie Tome I et II*. ECONOMICA.
- Ferrier, C. (2016). *Le zonier en tarification IARD : approche comparative de deux techniques de segmentation géographique en assurance habitation*. Mémoire Actuariat ISFA.
- Fesquet, C. (2015). *Zoniers et utilisation de facteurs exogènes en tarification MRH*. Mémoire Actuariat ISFA.
- Lagadec, F. (2009). *Tarification d'un contrat complémentaire santé par un modèle Linéaire généralisé*. Mémoire Actuariat EURIA.
- Millet, F. (2016). *Analyse de la sinistralité et tarification d'un produit d'assurance complémentaire santé*. Mémoire Actuariat ISFA.
- Miloé, Y. (2011). *Tarification d'un produit MRH à l'aide de la méthode des marges*. Mémoire d'actuariat ISFA.
- Mutuelle, A. (2018). *Rapport sur la solvabilité et la situation financière*.
- Paglia, A., Phélippe-Guinvarc'h, M., & Lenca, P. (2011). *Adaptation de l'algorithme CART pour la tarification des risques en assurance non-vie*. Consulté le 5 9, 2019, sur <https://hal.archives-ouvertes.fr/hal-00631121>
- Phelippe-Guinvarc'h, M., & Paglia, A. (2011). Tarification des risques en assurance non-vie, une approche par modèle d'apprentissage statistique. *BULLETIN FRANCAIS D'ACTUARIAT*, Vol. 11, n°22,, 49-81.
- Planchet, F., & Serdeczny, G. (2014). *Modèles fréquence - coût : Quelles perspectives d'évolution*. Journée IARD de l'Institut des Actuaire - Niort.
- Poubennec, F. (2017). *Refonte de la tarification d'un produit assurance santé modulaire avec implémentation s'un lissage spatial*. Mémoire Actuariat EURIA.
- Senné, T. (2015). *Apport de l'analyse prospective au pilotage d'un portefeuille santé individuel fermé à la souscription au sein d'une mutuelle*. Mémoire Actuariat CEA.
- Sepulveda, C. (2016). Modélisation du risque géographique en Santé, pour la création d'un nouveau Zonier. Comparaison de deux méthodes de lissage spatial. *Mémoire actuariat ISFA*.
- Truffery, S. (2017). *Data Mining et statistique décisionnelle*. TECHNIP.

## Annexes

### Annexe 1 : Droit et mutualité

Cette annexe est inspirée de (Senné, 2015).

Le schéma d'adhésion en assurance individuelle est présenté en figure 93. Ainsi, en individuel, la personne physique adhère au règlement mutualiste (RM) Individuel. Le signataire du bulletin d'adhésion devient membre participant de la mutuelle (ou encore adhérent).

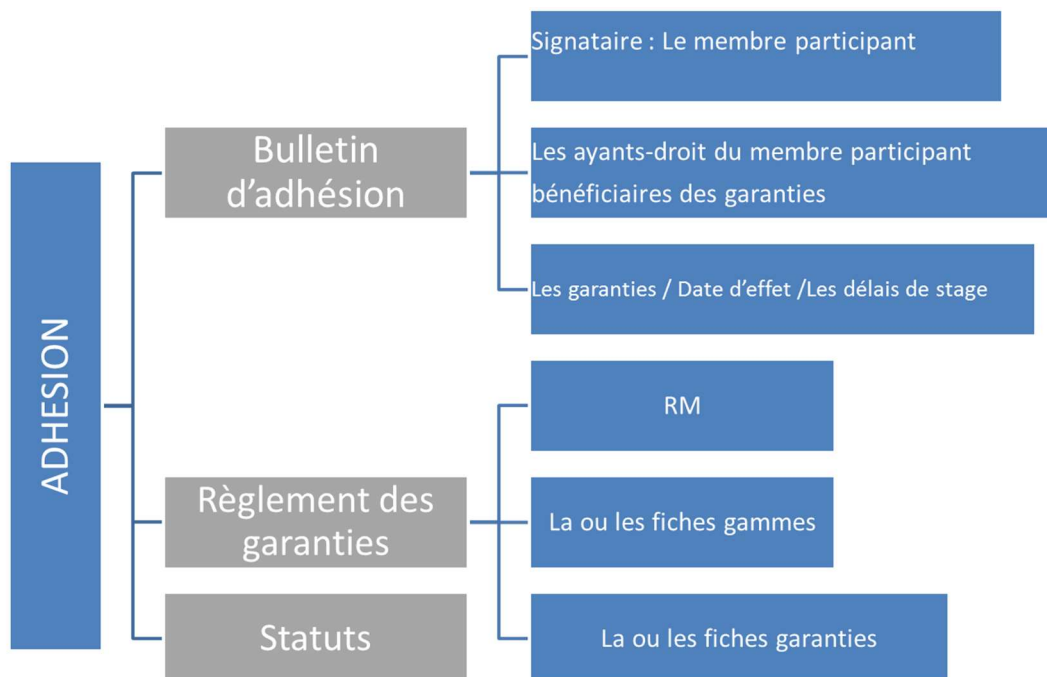


Figure 93 : le contrat d'adhésion dans les opérations individuelles

La mutuelle peut modifier le contrat d'adhésion, via une modification du règlement mutualiste, au niveau des règles de gestion, la définition des prestations et le niveau des cotisations. L'adhérent peut modifier le contrat par l'ajout d'un bénéficiaire, un changement de garantie ou encore les modalités de paiement. La résiliation du contrat n'est possible pour ADREA mutuelle que pour non-paiement des cotisations par le membre participant.

L'adhérent peut, quant à lui, résilier à l'échéance, en cas de modification du risque sous certaines conditions, en cas de mise en place d'un régime obligatoire par l'employeur, en cas de décès, en cas de bénéfice de la Couverture Médicale Universelle-Complémentaire ou de l'Aide à la Complémentaire Santé. On parle dans les autres cas de figure d'une possibilité de résiliation unilatérale à l'initiative de l'adhérent.

Dans le cadre de l'assurance collective, comme présenté en figure 94 et 95, la personne morale, pouvant être une association ou plus souvent une entreprise, peut :

- 1- adhérer au règlement mutualiste collectif
- 2- Souscrire un contrat collectif

Dans les deux cas, la personne morale devient membre honoraire de la mutuelle. On constate que les opérations collectives entraînent une relation tripartite entre la mutuelle, le souscripteur (l'entreprise ou l'association) et le

bénéficiaire du régime (salariés pour les entreprises ou adhérents pour une association). Ce dernier devient membre participant de la mutuelle comme dans le cadre des opérations individuelles.

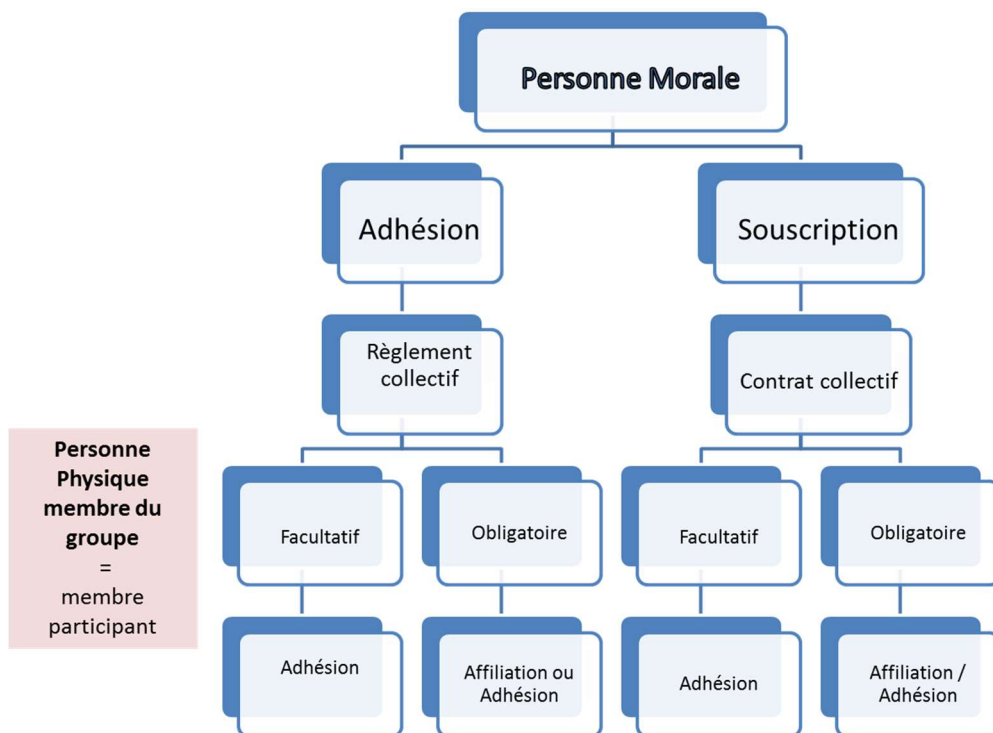


Figure 94 : l'adhésion ou la souscription dans les opérations collectives

Le contrat lie la personne morale à la mutuelle mais les prestations sont versées par la mutuelle aux bénéficiaires directement. En revanche, le contrat est annuel et se renouvelle par tacite reconduction et la possibilité et la résiliation est possible tant par la mutuelle que par la personne morale, sous réserve du respect des clauses contractuelles (délai de préavis de résiliation notamment).

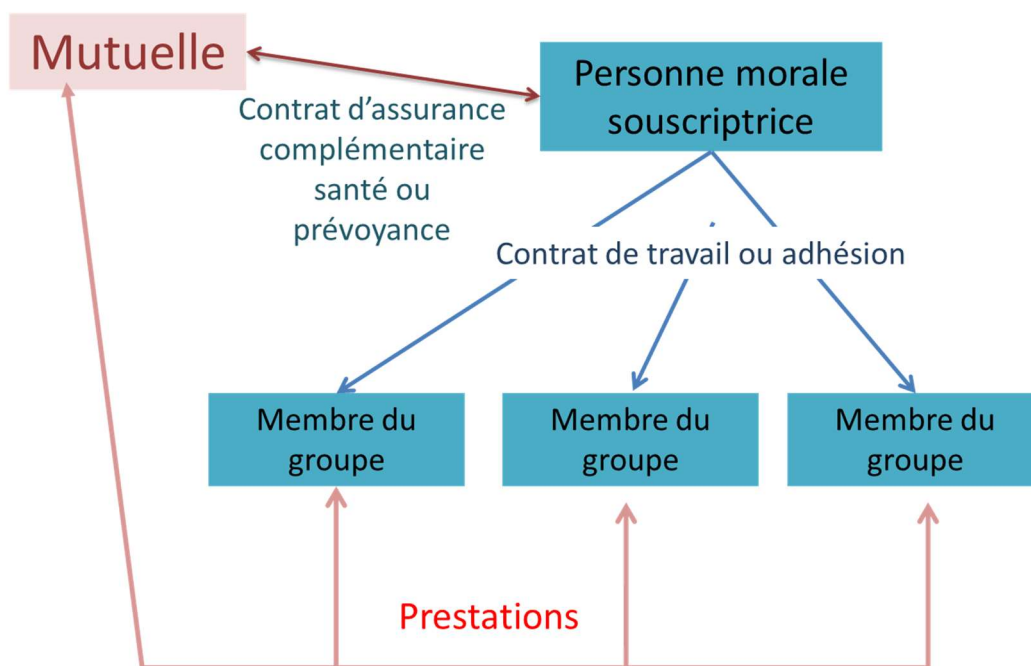


Figure 95 : La relation tripartite des contrats collectifs

## Annexe 2 : Modélisation Gamma

La part de prestations nulles étant relativement faible, nous avons tester l'hypothèse d'une fonction de distribution gamma pour la prime pure, variable à expliquer du modèle, sur une base de travail retraitées des prestations nulles. En effet, la modélisation Gamma ne permet pas de modéliser des sinistres nuls.

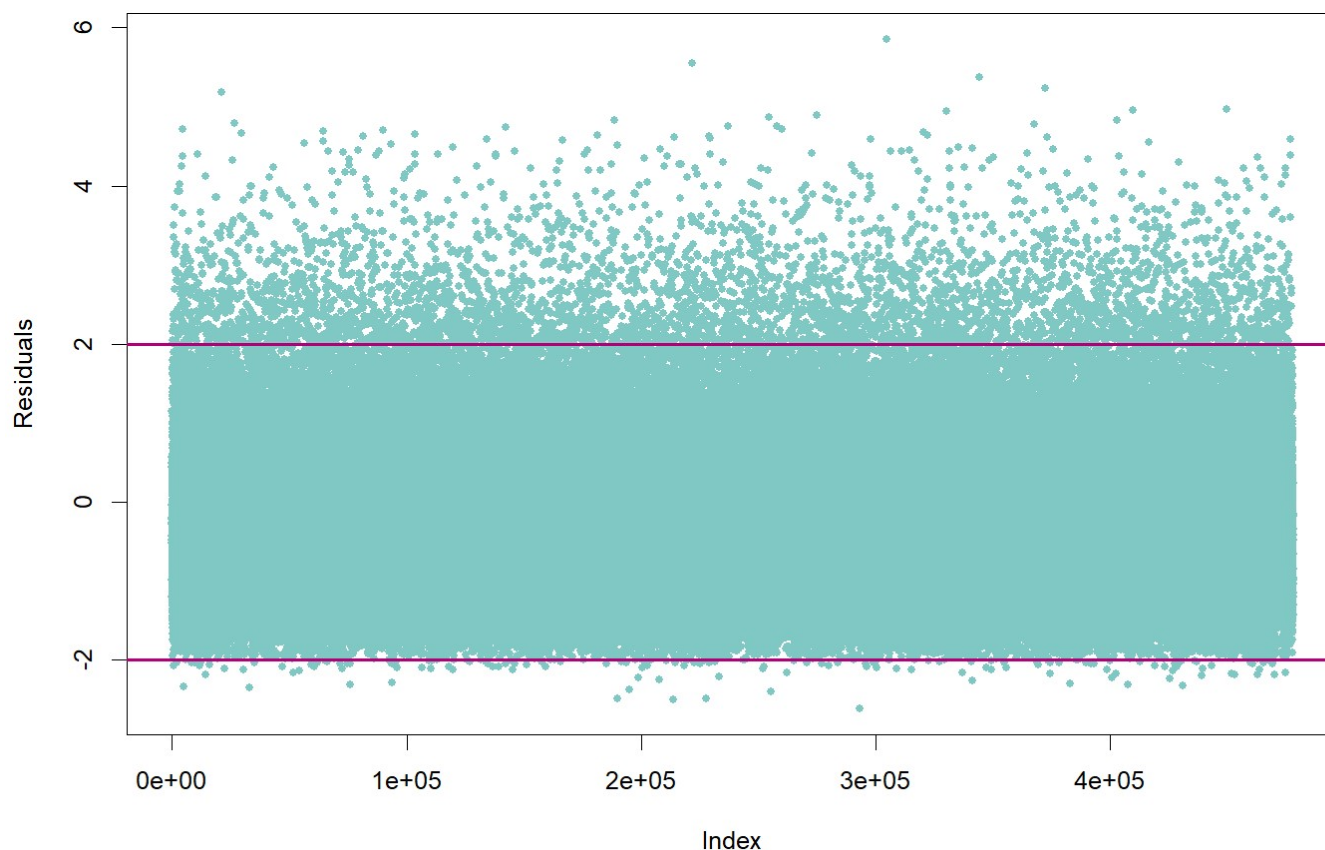


Figure 96 – Résidus standardisées – modèle Gamma

Les résidus obtenus semblent graphiquement permettre une validation de l'hypothèse d'homoscédasticité des résidus. En outre, si les critères AIC et de Déviance sont également inférieurs à ceux obtenus avec le modèle de Tweedie et pourraient conduire à retenir ce type de modélisation, la suppression des lignes de prestations sans sinistres, n'est pas une hypothèse à retenir et les comparaisons sont rendus délicates.

## Annexe 3 : Lissage Whittaker Henderson

Nous présentons ici la méthode de Whittaker Henderson. C'est une méthode qui combine un critère de fidélité aux données à un critère de régularité dans la forme de la courbe lissée.

Cette méthode non paramétrique est utilisée, entre autres, par le Bureau Commun des Assurances Collectives pour le lissage des barèmes de provisions mathématiques en incapacité/invalidité.

Soit U une matrice colonne de dimension n contenant les valeurs à lisser :  $U = (u_1, \dots, u_n)$ .

Soit V une matrice colonne de dimension n contenant les valeurs lissées :  $V = (v_1, \dots, v_n)$ .

L'objectif de cette méthode consiste à réaliser un compromis entre la fidélité aux données brutes et la régularité des données lissées, grâce à deux paramètres h et z entiers positifs.

Plus ces paramètres h et z sont grands, et plus le lissage est important, au détriment de la fidélité aux données brutes.

Les critères utilisés pour mesurer la fidélité et la régularité sont les suivants :

- Critère de fidélité :  $F = \sum_{i=1}^n (v_i - u_i)^2$
- Critère de régularité :  $s = \sum_{i=1}^n [\Delta_z(v_i)]$

Où  $\Delta_z$  est la différence d'ordre z entre la valeur de V.

Pour comprendre les raisons qui rendent S critère de régularité, il faut mettre en relation les différences d'ordre z et la dérivation d'ordre z d'une fonction continue v de la variable x.

En approchant  $v'(x)$  par  $v(x+1) - v(x)$ , on obtient :

$$\begin{aligned} v'_{(x)} &\approx v(x+1) - v(x) = \Delta_1(x), \\ v''_{(x)} &\approx v(x+2) - 2v(x+1) + v(x) = \Delta_2(x), \\ v'''_{(x)} &\approx v(x+3) - 3v(x+2) + 3v(x+1) - v(x) = \Delta_3(x), \dots \\ v^{z}_{(x)} &= \Delta_z(x) \end{aligned}$$

La minimisation de S dans le cadre discret revient donc à une minimisation, en valeur absolue, des dérivées d'ordre z, en chaque du vecteur lissé. Choisir z égal à 1 n'autorise comme régularité la dérivée première, c'est-à-dire la constance entre  $v_i$  et  $v_{i+1}$ . Le choix de z=3 conduit à annuler une dérivée d'ordre 3, c'est-à-dire que la régularité autorisée par ce choix est celle d'un polynôme d'ordre 2. Ainsi se vérifie la croissance de la régularité avec le degré du polynôme et l'emploi de S comme critère de régularité.

On définit la matrice  $K_z$ , telle que  $K_z \cdot V = \Delta_z(v_i)$ , i allant de 1 à n-z. On peut montrer que  $K_z$  est la matrice à n-z lignes et n colonnes contenant les coefficients binomiaux d'ordre z affectés de leur signe.

Cela provient de la définition  $\Delta_z(v_i)$  :

$$\Delta_z(v_i) = \sum_{k=0}^z c_z^k \cdot (-1)^{z-k} \cdot v_{i+k}$$

Le lissage résulte de la minimisation d'une combinaison linéaire entre :

- une mesure de la précision entre valeurs brutes et les valeurs lissées
- une mesure de la force de lissage

On cherche à minimiser  $M = F + h \cdot S$ , c'est-à-dire on cherche les valeurs de  $V^*$  qui minimisent M.

On peut écrire que M vaut :

$$M = {}^t_i(V - U) + h \cdot {}^t_i V \cdot ({}^t_i K_z K_z) \cdot V$$

L'idée est de compléter le carré tel que :

$$M = {}^t_i(V - X) + (Id + h \cdot ({}^t_i K_z K_z))^{-1} \cdot (V - X)$$

Si on développe les deux expressions de M, qu'on supprime tout ce qui est indépendant de V et qu'on égalise les 2 expressions obtenues, on obtient :

$$U = (Id + h \cdot ({}^t_i K_z K_z)) \text{ soit } X = (Id + h \cdot ({}^t_i K_z K_z))^{-1} \cdot U$$

Or M est minimal si  $V-X=0$ .



On obtient donc la solution du problème de lissage :

$$V^* = (Id + h \cdot ({}^t K_z K_z))^{-1} \cdot U$$

Il reste à choisir les valeurs optimales de paramètres h et z.

#### **Le paramètre z**

Les essais réalisés par les auteurs de la méthode ont abouti aux observations suivantes : les paramètres z=1 et z=2 ne lissent pas suffisamment les irrégularités, tandis que z=4 tend à représenter une droite.

#### **Le paramètre h**

Ce paramètre est variable selon l'intensité du lissage souhaité :

- h < 30 : lissage faible
- h entre 30 et 100 : lissage habituel
- h > 100 : lissage prononcé

Des tests visuels sont effectués en fonction de la courbe à lisser et la valeur de h retenue dépend de ces tests visuels.

### Annexe 4 : Zonier avec 3 classes

Si la classification retenue dans notre application est celle en 4 classes, nous avons présenté en annexe les résultats de segmentation en 3 classes en figure 94.

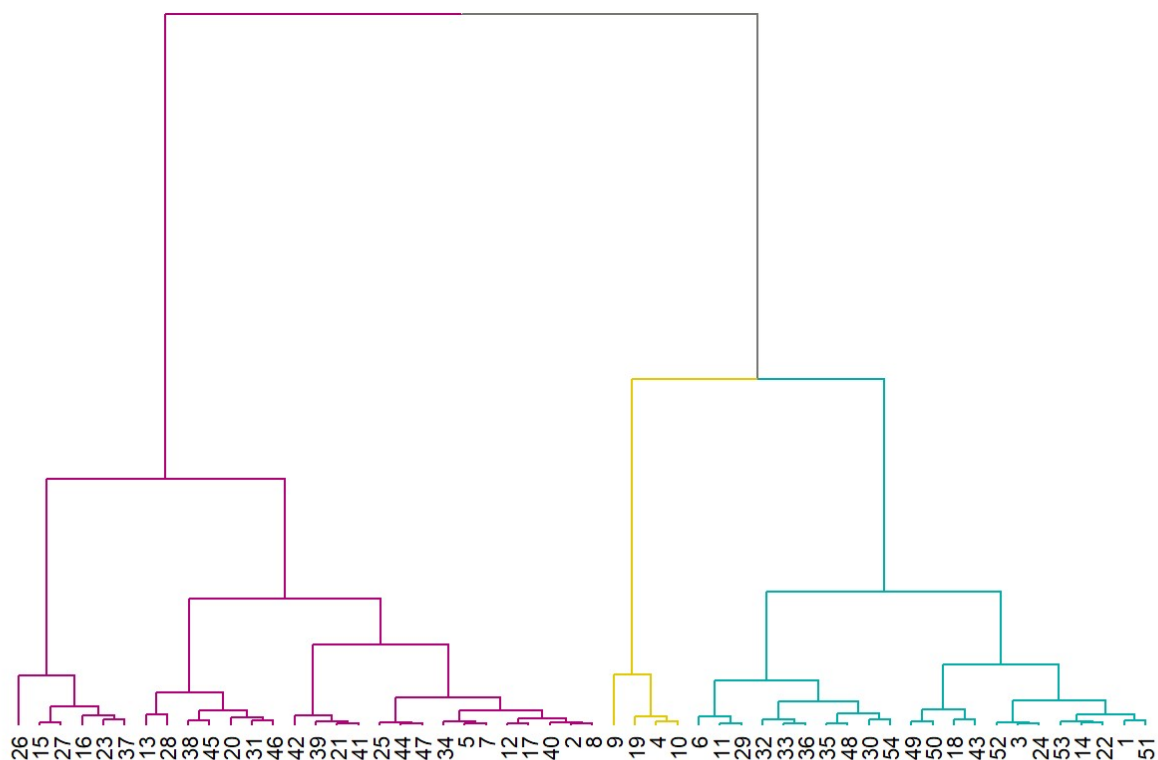


Figure 97 – Dendrogramme avec segmentation en 3 classes

Les coefficients résiduels associés à cette segmentation sont les suivants :

Coefficients de zonage	CAH
1	0,90
2	0,99
3	1,09

Figure 98 – Coefficient de zone – 3 classes

Comme nous pouvons le constater, la zone 1 représente un grand nombre de départements et le zonier est moins discriminant que la classification en 4 classes (30% contre 21% pour le zonier à 3 classes).

## Annexe 5 : Hypothèse des scénarios de sensibilités

Les hypothèses associées aux scénarios de tests relatifs aux hypothèses d'adhésion et de résiliation sont les suivants :

### Scénario 1 :

- sensibilité à la résiliation plus marquée en cas de hausse et de baisse de prix (+ ou – 10 points),
- sensibilité à la souscription plus marquée en cas de hausse et de baisse de prix (+ ou – 10 points).

Scénario 1	Impact adhésion	Impact Radiation	Nombre de département
Impact adhérent < -7%	35%	-60%	12
Impact adhérent > -5% < 0%	25%	-35%	22
Impact adhérent > 1% et < 3%	0%	0%	31
Impact adhérent > 3% < 8%	-35%	60%	11
Impact adhérent > 10%	-60%	60%	16

Figure 99 - Hypothèses adhésion / résiliation - Scénario 1

#### Scénario 2 :

- sensibilité à la résiliation plus marquée en cas de hausse de prix et moins en cas de baisse (+ ou – 10 points),
- sensibilité à la souscription moins marquée en cas de baisse de prix et plus en cas de hausse (+ ou – 10 points).

Scénario 2	Impact adhésion	Impact Radiation	Nombre de département
Impact adhérent < -7%	15%	-40%	12
Impact adhérent > -5% < 0%	5%	-15%	22
Impact adhérent > 1% et < 3%	0%	0%	31
Impact adhérent > 3% < 8%	-35%	60%	11
Impact adhérent > 10%	-60%	60%	16

Figure 100 - Hypothèses adhésion / résiliation - Scénario 2

## Table des figures

Figure 1 – Structure de financement de la CSBM – source DREES .....	8
Figure 2 - Exemple de bases et taux de remboursements pour médecins généralistes et spécialistes .....	12
Figure 3 - Le parcours de soins.....	13
Figure 4 - Principe de remboursement de la Sécurité Sociale .....	14
Figure 5 - Exemple de remboursement.....	14
Figure 6 - Structure de la CSBM et de la dépense de chacun de ses financeurs en 2019.....	15
Figure 7 - Structure du financement des grands postes de la CSBM en 2019 .....	15
Figure 8 - Part des organismes complémentaires dans les principaux postes de 2010 à 2019 .....	16
Figure 9 - Part de marché par type d'organismes entre 2010 et 2019.....	17
Figure 10 - Evolution du nombre d'organismes sur le marché de l'assurance entre 2006 et 2017 .....	17
Figure 11 - Prestations versées selon le type de soins et de contrat en 2019 .....	19
Figure 12 - Calendrier de mise en place de la réforme du 100% Santé .....	22
Figure 13 - Répartition des personnes protégées (PP) en 2018.....	25
Figure 14 - Pourcentage de pénétration d'ADREA Mutuelle par département - Nombre de personnes protégées rapportées à la population du département .....	26
Figure 15 - Evolution des indicateurs de solvabilité 2 – ADREA Mutuelle .....	27
Figure 16 – Ventilation des cotisations et des prestations 2018 entre Vie et Non-Vie.....	27
Figure 17 – Evolution de la part des contrats collectifs et individuels en santé .....	28
Figure 18 - Ventilation de la marge technique économique brute (en millions d'euros (M€)).....	29
Figure 19 – Zoom sur la marge économique brute en individuel (en M€) .....	30
Figure 20 - Premiers retraitements sur la base d'étude.....	32
Figure 21 - Coefficient d'extrapolation à appliquer sur les prestations constatées au 31.07.2020 pour 2018 et 2019 .....	33
Figure 22 - Modularité de la gamme Flex'ADREA .....	35
Figure 23 - Les seize combinaisons de la Gamme Flex'ADREA.....	36
Figure 24 - Pyramide des âges - Offre Flex'ADREA .....	37
Figure 25 - Suivi des personnes protégées – Offre Flex'ADREA .....	38
Figure 26 - Suivi du chiffre d'affaires – Offre Flex'ADREA .....	38
Figure 27 - Suivi de la marge - Offre Flex'ADREA .....	39
Figure 28 - Suivi Prestations / Cotisations nettes – Offre Flex'ADREA .....	39
Figure 29 - CA des produits les plus souscrits – Offre FlexADREA .....	40
Figure 30 - Marge des produits les plus souscrits – Offre Flex'ADREA.....	40
Figure 31 – Zoom sur les produits les plus déficitaires – Offre Flex'ADREA .....	41
Figure 32 - Zoom sur les produits les plus excédentaires – Offre Flex'ADREA .....	41
Figure 33 – Boite à moustache de la prime pure – avant retraitement.....	43
Figure 34 – Boite à moustache – Prime pure .....	44
Figure 35 - Deuxième retraitement sur la base d'étude .....	44
Figure 36 – Distribution de la prime pure .....	45
Figure 37 - Répartition des effectifs par produit avant regroupement des modalités .....	46
Figure 38 - Evolution de la prime pure en fonction du produit .....	47
Figure 39 - Dernier retraitement sur la base d'étude .....	47
Figure 40 - Evolution de la prime pure en fonction du produit après exclusion des données Alsace-Moselle .....	48
Figure 41 - Evolution de la prime pure moyenne selon le choix final de regroupement des produits.....	49
Figure 42 - Répartition des effectifs par produit – Regroupement retenu .....	49
Figure 43 - Répartition Femme / Homme .....	50
Figure 44 – Prime pure par sexe .....	51
Figure 45 - Répartition des PP par tranche d'âge.....	51
Figure 46 - Evolution de la prime pure en fonction de l'âge.....	52
Figure 47 - Répartition du personnes protégées par cohorte - année de soins 2019.....	53
Figure 48 - Coût technique par cohorte - année de soins 2019.....	53
Figure 49 – Répartition Adhésion > à 1 an / ≤ à 1 an .....	54

Figure 50 – Prime pure en fonction de la durée d’adhésion.....	54
Figure 51 - Répartition par type de personnes protégées .....	55
Figure 52 – Evolution de la prime pure en fonction du type de personne protégées .....	56
Figure 53 – Evolution de la prime pure en fonction du département ou groupe de département.....	57
Figure 54 – Evolution de la prime pure en fonction du département ou groupe de département (département avec Exposition > 3000 PP).....	57
Figure 55 - Présentation du zonier actuel – Zonier technique .....	59
Figure 56 - Présentation du zonier actuel – Zonier commercial .....	60
Figure 57 - Extrait de décision : Mise à jour du zonier.....	61
Figure 58 - Loi de Tweedie en fonction de p .....	67
Figure 59 - Méthode de construction d'un zonier .....	69
Figure 60 – Variables explicatives et catégories associées.....	72
Figure 61 - Intervalle d'interprétation V de Cramer.....	73
Figure 62 - Test V de Cramer .....	73
Figure 63 - Validation empirique de l'échantillonnage .....	75
Figure 64 - Choix du paramètre p dans la loi de Tweedie – critère déviance .....	76
Figure 65 – Sélection du meilleur modèle avec le critère de déviance.....	77
Figure 66 – Sélection du meilleur modèle avec le critère AIC.....	78
Figure 67 – Coefficients d’apprentissage et tests de Student sur les variables explicatives.....	79
Figure 68 – Coefficients d’apprentissage et intervalle de confiance – Tranche d’âge.....	81
Figure 69 – Représentation des résidus de déviance standardisés.....	82
Figure 70 - Dendrogramme – résidus par département - Distance de Ward.....	85
Figure 71 - Sauts d’inertie du dendrogramme .....	86
Figure 72 - Perte relative d’inertie selon le nombre de classes .....	87
Figure 73 - Dendrogramme avec choix de décomposition en trois classes .....	88
Figure 74 - Application de la méthode Elbow sur les données.....	90
Figure 75 - Coefficients associés aux zones - CAH et K-moyennes.....	91
Figure 76 - Zonier en place (vision technique).....	91
Figure 77 – Différence de rattachement significatif – zonier en place < zoniers théoriques.....	92
Figure 78 – Différence de rattachement (1 niveau) – zonier en place < zoniers théoriques .....	93
Figure 79 - Différence de rattachement significatif – zonier en place > zoniers théoriques.....	93
Figure 80 - Comparaisons des critères de modèle pour choix de zonier .....	94
Figure 81 - Critère de sélection de zonier – Silhouette et Calinski-Harabasz.....	95
Figure 82 – Proposition de mise à jour du zonier technique .....	96
Figure 83 - Nouveau zonier y compris coefficients commerciaux – avant rapprochement.....	97
Figure 84 – Proposition finale de zonier .....	97
Figure 85 - Compte de résultats - survenance 2019 .....	98
Figure 86 – Impact de la modification du zonier sur la cotisation .....	100
Figure 87 - Première projection 2020 - pas de modification des hypothèses d'évolution d'effectifs 2020.....	101
Figure 88 - Proposition de règle de projection des effectifs .....	101
Figure 89 - Projection suite à modification des hypothèses d'adhésion / résiliation .....	103
Figure 90 - Résultats de la modification de zonier – zoom par département.....	103
Figure 91 – Synthèse des évolutions de P/C suite à la mise à jour du zonier .....	105
Figure 92 – Synthèse des scénarios de sensibilité .....	105
Figure 93 – Résidus standardisées – modèle Gamma.....	111
Figure 94 – Dendrogramme avec segmentation en 3 classes.....	114
Figure 95 – Coefficient de zone – 3 classes.....	114
Figure 96 - Hypothèses adhésion / résiliation - Scénario 1.....	115
Figure 97 - Hypothèses adhésion / résiliation - Scénario 2.....	115