

**Résumé :** Construction d'une table d'expérience prospective pour un portefeuille de rentes de conjoint et impact sur le provisionnement.

**Mots-clés :** table d'expérience, Log-rank, Cox, Kaplan-Meier, Hoem du moment, Hoem générationnel, Méthode du SMR, Méthode du Logit, Provisionnement.

L'objectif de ce mémoire est de construire une table d'expérience prospective d'un portefeuille de rentes de conjoint de Malakoff Humanis. Le but est ainsi de s'assurer de la suffisance du provisionnement de la garantie rente de conjoint grâce à cette table, construite à partir des données disponibles du portefeuille.

Un contrôle de la fiabilité et de la qualité des données a été réalisé avant de passer à l'estimation des taux bruts, puis un ajustement de ces derniers par positionnement a été effectué, suivi d'une fermeture de la table avant d'obtenir la table d'expérience finale.

Les résultats des tests du Log-rank et du modèle de Cox ont montré que la construction d'une table segmentée par sexe était justifiée. Cependant, en raison du faible nombre de données sur la population masculine, seule une table d'expérience générationnelle pour les femmes est construite pour le portefeuille. Le périmètre de l'étude est défini et porte sur une période d'observation de 2000 à 2021 ainsi qu'un intervalle d'âges de 35-100 ans.

Après le contrôle et le nettoyage des données, l'estimation des taux bruts de sorties a été réalisé selon trois méthodes : la méthode de Kaplan-Meier, la méthode des moments de Hoem ainsi que la méthode de Hoem par génération. Après avoir comparé les modèles, la méthode des moments de Hoem a été choisie en raison de son caractère prudent, qui a semblé plus adapté pour la construction de la table.

Après l'estimation des taux bruts et la sélection de la méthode la plus fiable, il convient de vérifier que les taux bruts reflètent de la manière la plus fidèle le phénomène de mortalité observé. Du fait d'une insuffisance de données, les taux peuvent comporter des irrégularités. Il est donc nécessaire de les ajuster. Deux méthodes d'ajustement par positionnement sont utilisées, la méthode du SMR ainsi que la méthode de la régression logistique. Ces méthodes sont basées par le positionnement de la mortalité du portefeuille d'expérience par rapport à une référence de mortalité.

Des tests statistiques ont ensuite été réalisés afin de vérifier la cohérence et la justesse des modèles qui nous ont permis de choisir l'ajustement le plus adapté à notre modèle, qui s'est avéré être l'ajustement par régression logistique.

Après la fermeture de la table par la méthode de Denuit et Goderniaux, une table d'expérience femme prospective est obtenue. Les provisions mathématiques sont alors calculées à partir de cette nouvelle table et sont alors comparées aux provisions de la table réglementaire TGF 05, qui est la table habituellement utilisée pour provisionner les rentes de conjoint.

**Abstract :** Construction of a prospective experience table for a portfolio of spouse annuities and impact on reserving.

**Keywords:** experience table, Log-rank, Cox, Kaplan-Meier, moment Hoem, generational Hoem, SMR method, Logit method, reserving.

The goal of this dissertation is to construct a prospective experience table for a portfolio of spouse annuities of Malakoff Humanis. The aim is to ensure the adequacy of the provision for spouse annuities through this table, built from available data from the portfolio.

A reliability and data quality check was performed before estimating the raw rates. Then, an adjustment of these rates by positioning was carried out, followed by closing the table before obtaining the final experience table.

The results of Log-rank and Cox model tests showed that constructing a segmented table by gender was justified. However, due to the low amount of data on the male population, only a generational experience table for women was constructed for the portfolio. The study's scope is defined and covers an observation period from 2000 to 2021, with an age range of 35-100 years.

After data cleaning and quality check, the estimation of raw mortality rates was performed using three methods: Kaplan-Meier, Hoem's moments, and Hoem's generation method. After comparing the models, Hoem's moments method was chosen due to its conservative nature, which seemed more suitable for building the table.

After estimating the raw rates and selecting the most reliable method, it is necessary to verify that the raw rates reflect the observed mortality phenomenon most faithfully. Due to insufficient data, rates may have irregularities. Therefore, two positioning adjustment methods are used, the SMR method and the logistic regression method. These methods are based on positioning the mortality of the experience portfolio relative to a mortality reference.

Statistical tests were then performed to verify the coherence and accuracy of the models that allowed us to choose the most suitable adjustment for our model, which turned out to be the adjustment by logistic regression.

After closing the table by the Denuit and Goderniaux method, a prospective women's experience table is obtained. Mathematical provisions are then calculated from this new table and compared to the provisions of the regulatory TGF 05 table, which is usually used to provision spouse annuities.

## Note de synthèse :

Les avancées médicales et l'amélioration des conditions de vie au cours du 20ème siècle ont fortement contribué à l'allongement de la durée de vie, créant ainsi un nouveau risque pour les compagnies d'assurance appelé risque de longévité. Les contrats d'assurance sur la vie, et plus particulièrement les rentes viagères, en sont lourdement impactés puisque ces contrats sont un engagement pour l'assureur jusqu'au décès du bénéficiaire. Pour faire face à ce risque, les organismes assureurs utilisent des tables de mortalité prospectives.

Cependant, les tables réglementaires utilisées par les assureurs ne sont pas toujours adaptées aux spécificités des portefeuilles. Ainsi, la construction d'une table d'expérience permet à un organisme assureur de mieux refléter le comportement de sa population assurée.

L'objectif de ce mémoire est ainsi de créer une table d'expérience prospective pour un portefeuille de rentes de conjoint de Malakoff Humanis à partir des données disponibles, afin d'évaluer l'impact d'une nouvelle table sur le provisionnement de la garantie rente de conjoint.

### Méthode mise en place :

Dans cette note de synthèse, les grandes parties de l'étude vont être présentées ainsi que les conclusions obtenues. La méthode mise en place lors de ce mémoire est la suivante :

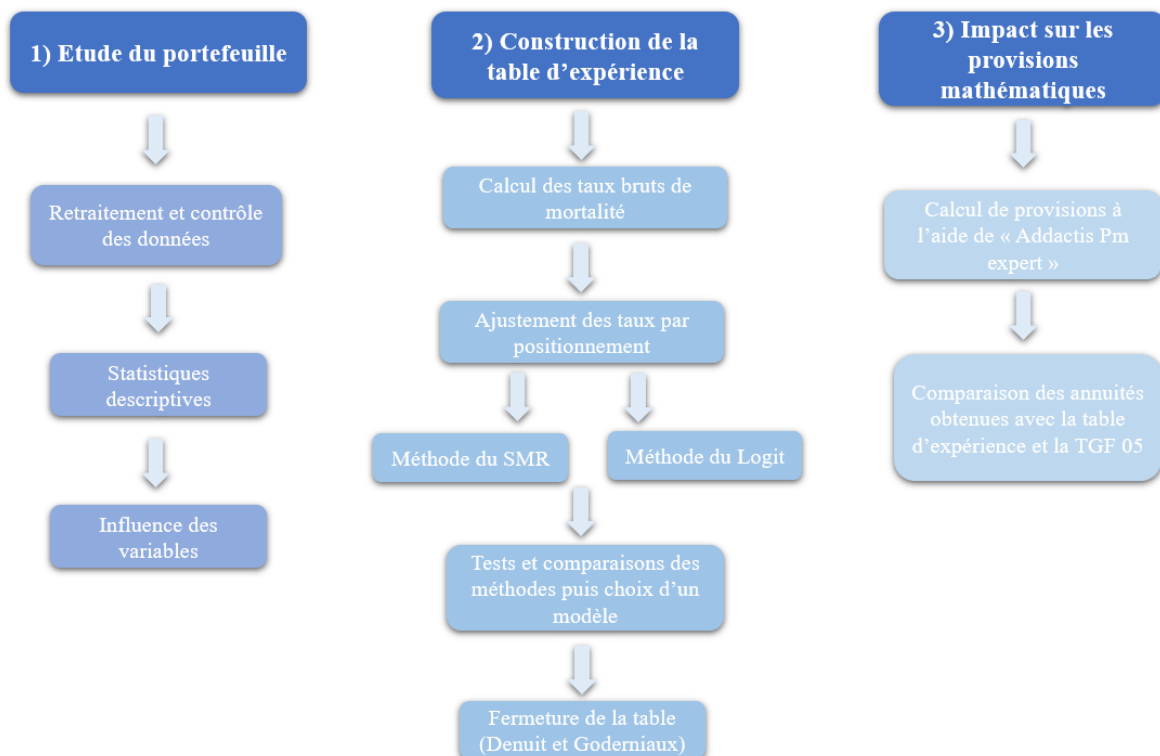


Figure 1 - Méthode appliquée pour la construction de la table

La première partie du mémoire présente le contexte et l'objectif de l'étude, qui porte sur des contrats de rentes de conjoint issus d'un portefeuille de Malakoff Humanis. Des tests de fiabilité et de qualité ont été effectués pour valider la conformité et la cohérence des données, suivis de statistiques descriptives sur le portefeuille.

Les résultats des tests du Log-rank et du modèle de Cox ont montré que la construction d'une table segmentée par sexe était justifiée. Cependant, en raison du faible nombre de données sur la population masculine, seule une table d'expérience générationnelle pour les femmes est construite pour le portefeuille. Réaliser une table générale sans distinction du genre ne serait pas pertinent en raison de la faible proportion d'hommes qui n'est pas représentative d'une population mixte.

Si le nombre de données le permet à l'avenir, la construction d'une table pour les hommes pourrait être envisagée en suivant la même approche que pour la table des femmes.

Enfin, après une présentation des outils requis pour la construction de la table, le périmètre de l'étude a été défini et concernera une période d'observation de 2000 à 2021 et un intervalle d'âges [35,100]ans.

### Estimation des taux bruts de mortalité :

Pour l'estimation des taux bruts, trois méthodes ont été appliquées au portefeuille : la méthode de Kaplan Meier, la méthode des moments de Hoem, puis celle de Hoem par génération. Après avoir comparé les modèles, la méthode des moments de Hoem a été choisie en raison de son caractère prudent, qui a semblé plus adapté pour la construction de la table.

Les graphiques des taux bruts obtenus sont les suivants :

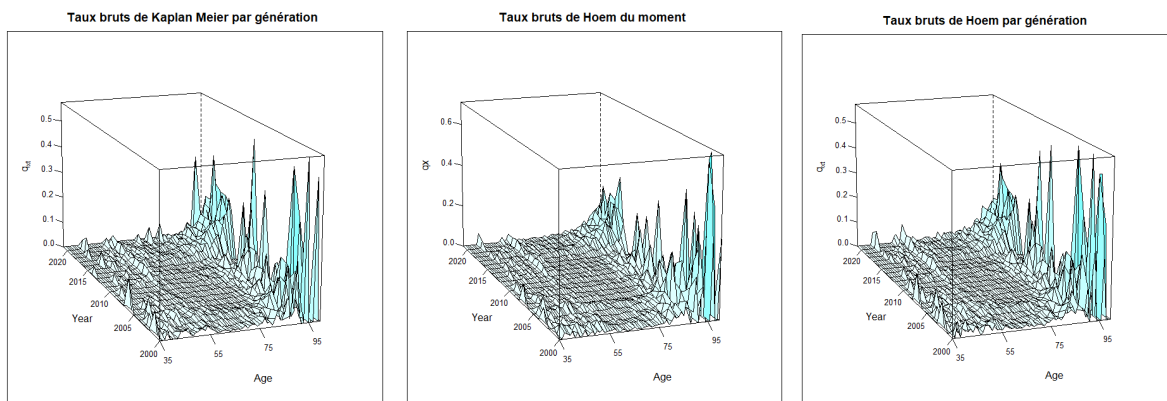


Figure 2 : Estimation des taux bruts par différentes méthodes

Après l'estimation des taux bruts et la sélection de la méthode la plus fiable, il convient de vérifier que les taux bruts reflètent de la manière la plus fidèle le phénomène de mortalité observé. Du fait d'une insuffisance de données, les taux peuvent comporter des irrégularités. Il est donc nécessaire de les ajuster.

### Ajustement des taux bruts par positionnement :

Dans ce mémoire, seules les méthodes d'ajustements par positionnement sont utilisées en raison du faible nombre de données disponibles, qui ne nous permet pas d'avoir une approche directe de modèles intrinsèques. Ces méthodes sont basées sur le positionnement de la mortalité du portefeuille d'expérience par rapport à une référence de mortalité.

La table de référence utilisée, est une table de mortalité prospective proposée par l'Institut des actuaires et représentative de la mortalité de la population générale des rentiers. Cette table fournit des taux de mortalité sur chaque année allant de 2007 à 2060 pour l'intervalle d'âges [30,130]ans.

Les deux méthodes d'ajustement par positionnement ont été appliquées pour ajuster les taux bruts de mortalité :

- La méthode du SMR (*Standardized Mortality Ratio*), qui consiste à calculer un taux d'abattement à partir d'une table de référence ;
- La méthode du Logit qui consiste à positionner les taux à l'aide d'une fonction logistique selon un modèle de régression.

L'intervalle retenu pour l'ajustement est de 2007 à 2021, constitué des années en commun avec notre période d'observation et la plage d'année de la table de référence.

Les taux ajustés obtenus sont les suivants :

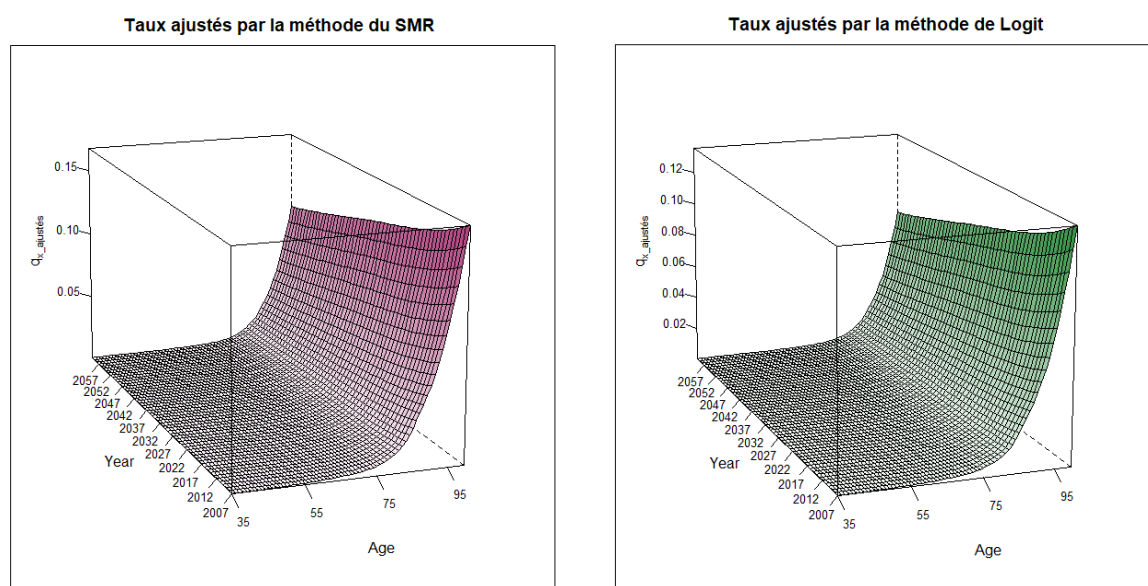


Figure 3 : Taux ajustés par deux méthodes de positionnement

### Choix de la méthode d'ajustement :

Le choix du meilleur ajustement est réalisé selon plusieurs critères de sélection :

- *Le critère de proximité*, qui est vérifié par différents tests statistiques et par des critères quantitatifs.

Les tests du Standardized Mortality Ratio (SMR), du rapport de vraisemblance et de Wilcoxon qui valident l'écart entre les données et les prédictions de la table sont effectués. Des valeurs d'intérêt comme  $\chi^2$ , le *MAPE*, le  $R^2$  et la *déviante* sont également calculées afin de valider les ajustements et de retenir le modèle le plus adapté.

- *Le critère de stabilité*, qui correspond à la régularité des ajustements. Il est vérifié par le test des runs et des signes.
- *Le critère de cohérence* privilégie l'ajustement donnant la mortalité la plus faible en tenant compte du risque de longévité.

La méthode de régression logistique a été retenue car elle permet de valider un plus grand nombre de critères que la méthode du SMR.

### Fermeture de la table :

La méthode utilisée pour la fermeture de la table est la méthode de Denuit et Goderniaux, qui consiste en l'ajustement d'un modèle log-quadratique en imposant deux contraintes aux grands âges.

La première contrainte impose une allure concave aux taux de mortalité pour les âges les plus avancés, ce qui signifie que la courbe doit être incurvée vers le haut pour éviter une décroissance des taux de mortalité à ces âges. La deuxième contrainte impose l'existence d'une tangente horizontale au point  $x = 130 \text{ ans}$ , ce qui signifie que la courbe doit avoir une pente nulle à cet âge.

Une table prospective annuelle complétée s'étalant sur la plage d'âges [35,130] ans et sur les années allant de 2007 à 2060 est obtenue.

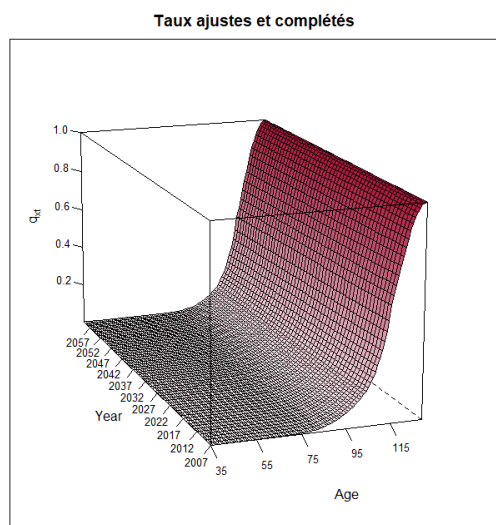


Figure 4 : Taux ajustés et complétés sur la période [2007 ;2060]

### Impact sur les provisions mathématiques :

Dans cette dernière partie du mémoire, les provisions mathématiques calculées à partir de la table réglementaire TGF 05, qui est la table habituellement utilisée pour provisionner les rentes de conjoint, sont comparées avec les provisions obtenues à partir de la table d'expérience construite pour le portefeuille étudié.

L'outil de calcul des provisions mathématiques utilisé pour cette étude est ADDACTIS PM'Expert. La table d'expérience femme construite a été intégrée au logiciel, puis le calcul de provisions mathématiques a été réalisé à la date du 31/12/2021, à iso-paramètres techniques (taux techniques, taux de chargement).

Les provisions mathématiques féminines obtenues ont alors été comparées aux provisions calculées à la clôture 2021 calculées avec la TGF 05.

Les ax moyens par tranche d'âges obtenus sont représentés dans le tableau ci-dessous :

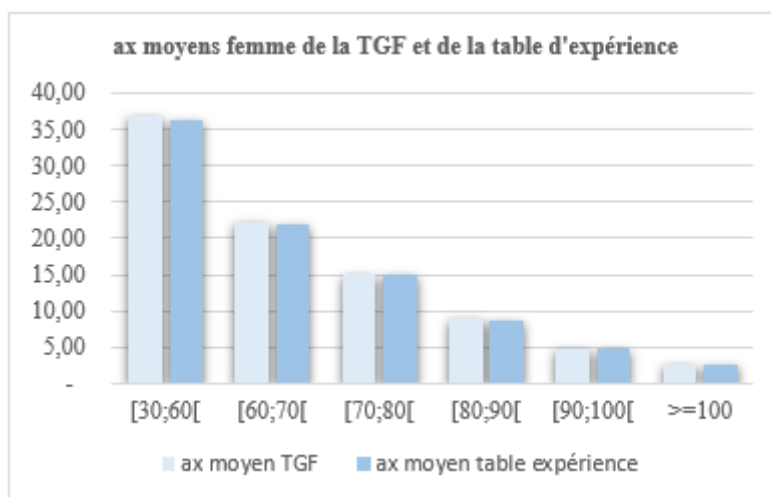


Figure 5 : Histogramme des ax moyens femmes de la TGF et de la table d'expérience

Un tableau des rapports entre les ax de la table d'expérience et les ax de la table réglementaire par tranche d'âges a également été réalisé :

	Tranche d'âges	Rapport ax table d'expérience / ax TGF 05
Tranche 1	[30;60[	-1,62%
Tranche 2	[60;70[	-0,91%
Tranche 3	[70;80[	-0,64%
Tranche 4	[80;90[	-0,53%
Tranche 5	[90;100[	-0,20%
Tranche 6	>=100	0,00%

Tableau 1 : Tableau des rapports entre les ax de la table d'expérience et la TGF 05

Les annuités obtenues avec la loi d'expérience sont légèrement inférieures à celles de la table réglementaire, les écarts étant variables selon les âges.

### Conclusion de l'étude :

La table d'expérience générationnelle construite est plus adaptée à la mortalité de notre portefeuille féminin rente de conjoint que la TGF 05, toutefois le changement de la table aurait peu d'impact sur le provisionnement.

Il est nécessaire d'être conscient des limites inhérentes à la construction d'une table d'expérience. En effet, la quantité d'observations est déterminante pour l'obtention de résultats robustes. Dans notre cas, la taille du portefeuille étant plutôt faible, cela peut se traduire avec des variations brusques de la mortalité. De plus, les données du portefeuille ne représentent la mortalité qu'à un instant précis, ne permettant pas d'avoir de certitudes sur son évolution future.

Dans les années à venir, il serait pertinent de réitérer l'étude sur la population féminine afin de s'assurer du maintien de la qualité du provisionnement, ainsi que de reproduire cette étude sur la population masculine, pour autant que le nombre de données le permette.

Afin d'approfondir cette étude, il serait intéressant de segmenter selon la catégorie socioprofessionnelle, l'ancienneté dans le portefeuille, la zone géographique etc. Cette segmentation permettrait de mieux comprendre les différences dans les tendances d'évolution observées et d'identifier les facteurs qui influencent l'espérance de vie du portefeuille étudié.



## Summary note :

Medical advances and improvements in living conditions during the 20th century have greatly contributed to the increase in life expectancy, creating a new risk for insurance companies known as longevity risk. Life insurance policies, particularly annuities, are heavily impacted by this risk since these contracts are an obligation for the insurer until the death of the beneficiary. To manage this risk, insurance companies use prospective mortality tables.

However, the regulatory tables used by insurers are not always adapted to the specificities of their portfolios. Therefore, constructing an experience table enables an insurance company to better reflect the behavior of its insured population. The objective of this paper is to create a prospective experience table for the spouse annuity portfolio of Malakoff Humanis using available data, in order to evaluate the impact of a new table on the provision of the spouse annuity guarantee.

### Methodology implemented :

In this summary, the major sections of the study will be presented as well as the conclusions obtained. The methodology implemented in this paper is as follows:

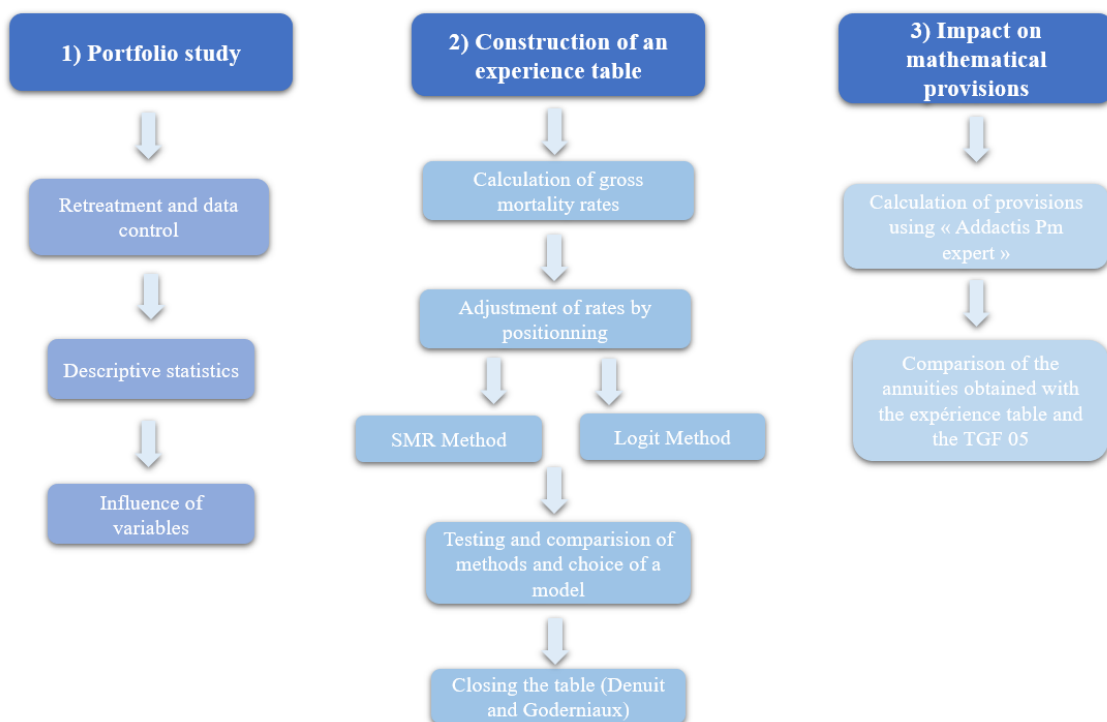


Figure 6 : Method applied for the construction of the table

The first part of the paper presents the context and objective of the study, which focuses on spouse annuity contracts from the Malakoff Humanis portfolio. Reliability and quality tests were performed to validate the conformity and coherence of the data, followed by descriptive statistics on the portfolio.

The results of the Log-rank and Cox model tests showed that the construction of a segmented table by gender was justified, but due to the lack of data on the male population, only a generational experience table for women is constructed for the portfolio. Creating a general table without gender distinction would not be relevant due to the low proportion of men which is not representative of a mixed population.

If the number of data allows in the future, the construction of a table for men could be considered using the same approach as for the women's table.

Finally, after presenting the required tools for the construction of the table, the scope of the study was defined and will cover an observation period from 2000 to 2021 and an age range of [35,100] years.

**Estimation of gross mortality rates :**

For the estimation of gross rates, three methods were applied to the portfolio : the Kaplan Meier method, the Hoem moment method, and the Hoem method by generation. After comparing the models, the Hoem moment method was chosen due to its cautious nature, which seemed more suitable for the construction of the table.

The graphs of the gross rates obtained are as follows:

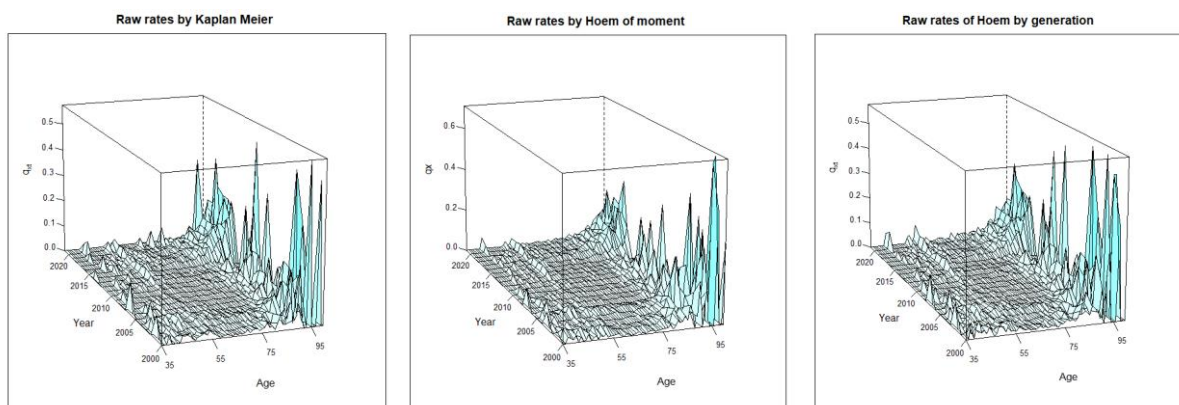


Figure 7 : Estimation of the raw rates

After estimating the gross rates and selecting the most reliable method, it is important to verify that the gross rates accurately reflect the observed mortality phenomenon. Due to insufficient data, the rates may contain irregularities, so they need to be adjusted.

**Taux bruts adjustment by positioning:**

In this paper, only positioning adjustment methods are applied due to the limited number of available data, which does not allow for a direct approach of intrinsic models. These methods are based on positioning the mortality of the experience portfolio relative to a reference mortality.

The reference table used, is a prospective mortality table proposed by the Institute of Actuaries and representative of the mortality of the general population of annuitants. This table provides mortality rates for each year from 2007 to 2060 for the age range [30,130] years.

Both positioning adjustment methods were applied to adjust the crude mortality rates :

- The SMR (Standardized Mortality Ratio) method, which consists of calculating a discount rate from a reference table ;
- The Logit method, which positions the rates using a logistic function according to a regression model.

The interval used for adjustment is 2007 to 2021, which includes the years in common with our observation period and the range of years in the reference table.

The adjusted rates obtained are as follows :

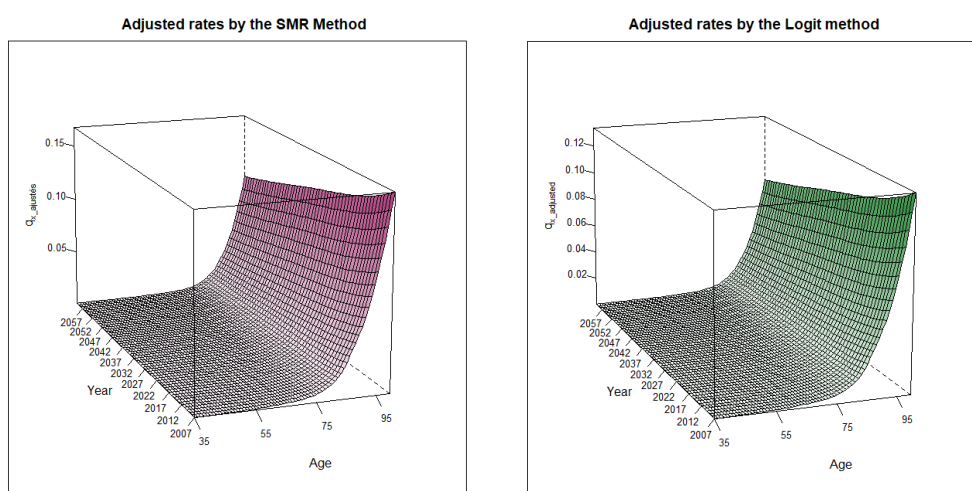


Figure 8 : Adjusted rates using two positioning methods

### **Choice of adjustment method:**

The selection of the best adjustment is made according to several selection criteria:

- *The proximity criterion*, which is verified by various statistical tests and quantitative criteria. The tests of the Standardized Mortality Ratio (SMR), the likelihood ratio, and the Wilcoxon test that validate the difference between the data and the predictions of the table are performed. Interest values such as  $\chi^2$ , *le MAPE*,  $R^2$  and *déviante* sont are also calculated to validate the adjustments and select the most appropriate model. The stability criterion corresponds to the regularity of the adjustments. It is verified by the runs and signs test.
- *The stability criterion* corresponds to the regularity of the adjustments. It is verified by the runs and signs test.
- *The coherence criterion* favors the adjustment giving the lowest mortality while taking into account longevity risk

The logistic regression method was selected because it allows for the validation of a greater number of criteria than the SMR method.

### **Table closure:**

The method used for closing the table is the Denuit and Goderniaux method, which consists of fitting a log-quadratic model while imposing two constraints at the oldest ages.

The first constraint imposes a concave shape to the mortality rates for the oldest ages, which means that the curve must be curved upward to avoid a decrease in mortality rates at these ages. The second constraint imposes the existence of a horizontal tangent at the point  $x=130$  years, which means that the curve must have zero slope at this age.

A completed annual prospective table spanning the age range [35,130] years and the years from 2007 to 2060 is obtained.

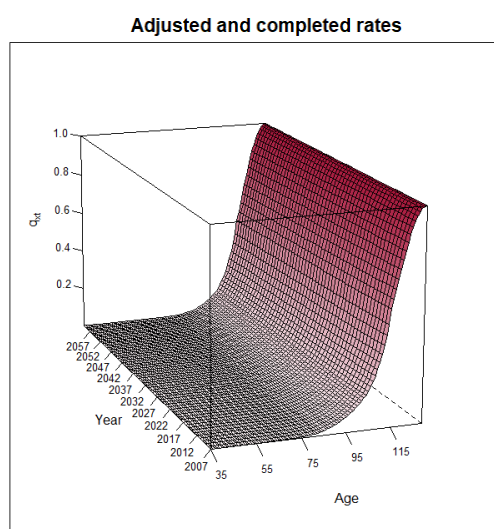


Figure 9: Adjusted and completed rates over the period [2007,2060]

### **Impact on mathematical provisions :**

In this final part of the study, the mathematical provisions calculated from the regulatory table TGF 05, which is the table usually used to provision spousal annuities, are compared with the provisions obtained from the experience table constructed for the portfolio under study.

At Malakoff Humanis, the calculation of mathematical provisions is performed using the "Addactis PM'Expert" software. The constructed female experience table was integrated into the software, and the calculation of mathematical provisions was performed as of 31/12/2021, with a loading rate of 3% and varying technical rates depending on the years of occurrence.

The obtained female mathematical provisions were then compared with the provisions calculated at the closing of 2021 using the TGF 05.

The average annuities by age group obtained are presented in the table below:

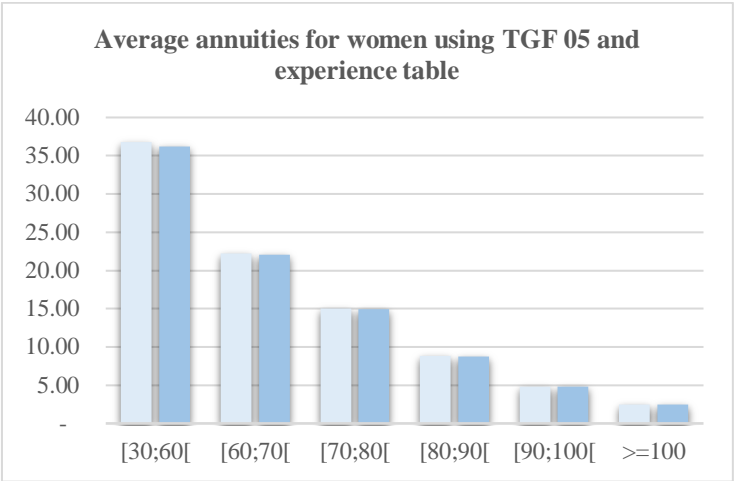


Figure 10 : Average annuities for women using TGF 05 and the experience table

A table of ratios between the ax values of the experience table and the ax values of the regulatory table by age group has also been created :

	Age group	Ratio of experience table ax / TGF 05 ax
Group 1	[30;60[	-1,62%
Group 2	[60;70[	-0,91%
Group 3	[70;80[	-0,64%
Group 4	[80;90[	-0,53%
Group 5	[90;100[	-0,20%
Group 6	>=100	0,00%

Tableau 2 : Table of ratios between the ax of the experience table and the TGF 05

The annuities obtained with the experience law are slightly lower than those of the regulatory table, with variations in the differences depending on the ages.

**Conclusion of the study:**

The generational experience table constructed is better suited to the mortality of our female spouse annuity portfolio than the TGF05, although changing the table would have little impact on provisioning.

It is necessary to be aware of the inherent limitations in constructing an experience table. Indeed, the quantity of observations is crucial for obtaining robust results. In our case, the size of the portfolio being relatively small, this can result in abrupt variations in mortality. In addition, the portfolio data only represent mortality at a specific moment, making it impossible to have certainty about its future evolution.

In the coming years, it would be pertinent to repeat the study on the female population to ensure the quality of the provisioning remains, as well as to replicate this study on the male population, provided that the number of data allows it.

In order to deepen this study, it would be interesting to segment by socioeconomic category, seniority in the portfolio, geographic area, etc. This segmentation would allow us to better understand the differences in observed trends and to identify factors that influence the life expectancy of the studied portfolio.

## ***Remerciements :***

Je souhaite tout d'abord adresser mes remerciements à Amel BOUDARENE, ma tutrice d'alternance, pour sa disponibilité, ses conseils et son expertise ainsi que Alexandre ASTEIX de m'avoir accueillie dans son équipe.

Je remercie également Manon VALDIGUIE et Julie SANCHEZ, collaboratrices de l'équipe MHP pour leurs compétences professionnelles.

Je tiens également à remercier les membres de l'équipe Inventaire de Malakoff Humanis qui m'ont aidée dans l'avancement de mon mémoire ainsi que mon tuteur académique, Monsieur Nicolas Bousquet, pour ses conseils.

Je remercie Garance Genoux de m'avoir accueillie dans son service, pour la confiance qu'elle m'a accordée.

## **Table des matières :**

1	Contexte et objectif de l'étude : étude d'un portefeuille de rentes de conjoint.....	18
1.1	Présentation de la prévoyance .....	18
1.1.1	Le marché de la prévoyance .....	19
1.1.2	Zoom sur risque décès.....	23
1.2	Le groupe Malakoff Humanis .....	25
2	Analyse du portefeuille .....	26
2.1	Etude du risque rente de conjoint sur un portefeuille Malakoff Humanis .....	26
2.1.1	Périmètre de l'étude .....	26
2.1.2	Extraction et description des données brutes.....	26
2.1.3	Retraitement des données.....	28
2.1.4	Contrôle des données .....	29
2.1.5	Statistiques descriptives du portefeuille .....	31
2.2	Influence des variables.....	39
2.2.1	Test du Log-rank.....	40
2.2.2	Méthode de Cox.....	42
2.2.2.1	Présentation de l'estimateur.....	42
2.2.2.2	Application aux données .....	44
3	Construction de la loi d'expérience en rente de conjoint .....	46
3.1	Présentation des tables d'expériences .....	47
3.2	Outils nécessaires au calcul des taux bruts .....	49
3.2.1	Notions de modèles de durée.....	49
3.2.2	Détermination de la période d'exposition au risque des bénéficiaires .....	50
3.2.3	Détermination de la période d'observation de l'étude .....	55
3.2.4	Traitement des censures et troncatures.....	56
3.2.5	Concepts de survie .....	58
3.3	Calcul des taux bruts.....	60
3.3.1	Tranche d'âges retenue pour les taux bruts .....	61
3.3.2	Méthode de Kaplan Meier et application aux données .....	62
3.3.2.1	Présentation de l'estimateur.....	62
3.3.2.2	Application aux données .....	64
3.3.3	Estimateur des moments de Hoem.....	66
3.3.3.1	Présentation de l'estimateur.....	66
3.3.3.2	Application aux données .....	68
3.4	Comparaison des différents modèles .....	69
4	Ajustement des taux bruts.....	71

4.1	Méthode du SMR ( Standardized Mortality Ratio) .....	72
4.2	Méthode de regression linéaire (ou du Logit) .....	73
4.3	Comparaison des ajustements obtenus par les différentes méthodes.....	75
4.3.1	Critere de proximité .....	75
4.3.2	Critère de régularité .....	80
4.3.3	Critère de cohérence .....	81
4.4	Fermeture de la table .....	82
4.4.1	Application de la méthode de Denuit et Goderniaux .....	82
5	Impact de la table sur le provisionnement .....	85
5.1	Les provisions mathématiques .....	85
5.2	Impact de la table construite sur le volume des provisions mathématiques.....	88



## **Introduction :**

En l'espace de quarante ans, la population française a augmenté de près de 12 millions d'habitants. Malgré une démographie dynamique, la population française, comme celle de nombreux pays européens, tend à vieillir.

Cet allongement de la durée de vie ainsi que le vieillissement de la population représentent un défi majeur pour les années à venir, notamment pour les assureurs.

Les rentes viagères étant un engagement pour l'assureur sur une longue période, il est donc important de tenir compte de l'évolution future de la mortalité afin d'estimer de manière adéquate l'évolution de la longévité et l'incidence pour maîtriser ce risque, et ainsi assurer sa solvabilité.

Le rôle d'un contrat de prévoyance est d'assurer les personnes contre la survenance d'éventuels sinistres comme le décès, l'invalidité ou l'incapacité. Ainsi, il est important pour l'assureur de pouvoir faire face à ses sinistres et ainsi respecter les engagements vis-à-vis des assurés, en mettant suffisamment de provisions de côté.

Les tables réglementaires utilisées par les assureurs pour tarifier et provisionner n'étant pas toujours adaptées à la spécificité des portefeuilles, la construction d'une table d'expérience est donc un enjeu majeur pour eux.

Un organisme assureur a la possibilité de construire ses propres tables, ce qui permet de refléter de manière plus juste le comportement de la population assurée. En effet, la table étant construite à partir des propres caractéristiques de la compagnie, l'assureur aura alors la capacité d'établir un tarif plus juste et de prévoir aux mieux les indemnités versées aux sinistrés.

L'assureur peut constituer ses provisions à l'aide des tables réglementaires, si elles sont suffisantes en regard du risque couvert, ou à l'aide de tables d'expérience si elles sont certifiées.

A ce titre, l'objet de ce mémoire consiste à construire une table d'expérience pour les rentes de conjoint en tenant compte de l'évolution de la population Malakoff Humanis. Jusqu'à présent, Malakoff Humanis provisionne cette garantie avec les tables réglementaires.

La problématique de ce mémoire est la suivante : comment construire une table d'expérience sur la garantie rente de conjoint et quel serait l'impact de cette nouvelle table sur le provisionnement ?

La première partie de l'étude présentera le contexte et les objectifs de ce mémoire avec une description du marché de la prévoyance ainsi que de Malakoff Humanis.

Puis dans une seconde partie, des tests de fiabilité et de qualité ont été effectués pour valider la conformité et la cohérence des données, suivis de statistiques descriptives sur le portefeuille.

Dans une troisième partie, les taux bruts seront calculés à l'aide de différents estimateurs statistiques. Puis, dans la partie suivante, des ajustements de taux seront réalisés et des tests statistiques seront effectués afin de déterminer leur cohérence et leur fiabilité et de choisir le meilleur estimateur puis une méthode de fermeture de table sera appliquée.

Enfin, nous présenterons en dernière partie la méthode de calcul des provisions et les impacts des changements de loi sur la variation des provisions.

# 1 Contexte et objectif de l'étude : étude d'un portefeuille de rentes de conjoint

## 1.1 Présentation de la prévoyance

Selon la loi n° 89-1009 du 31 décembre 1989, dite loi EVIN, la prévoyance regroupe « les opérations ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité, des risques d'incapacité de travail ou d'invalidité ou du risque chômage ».

En assurance, la prévoyance désigne de façon générique tous les contrats et garanties qui couvrent les risques sociaux liés à la personne en cas d'arrêt de travail. Ce dernier peut être temporaire ou définitif notamment : accidents de la vie quotidienne ou maladies causant une incapacité de travail, une invalidité, voire un décès. Quel que soit l'événement, l'intérêt des contrats de prévoyance est d'assurer le maintien des revenus de la personne et de sa famille.

Le système de prévoyance en France est regroupé selon plusieurs régimes, obligatoires ou facultatifs.

### La prévoyance en France

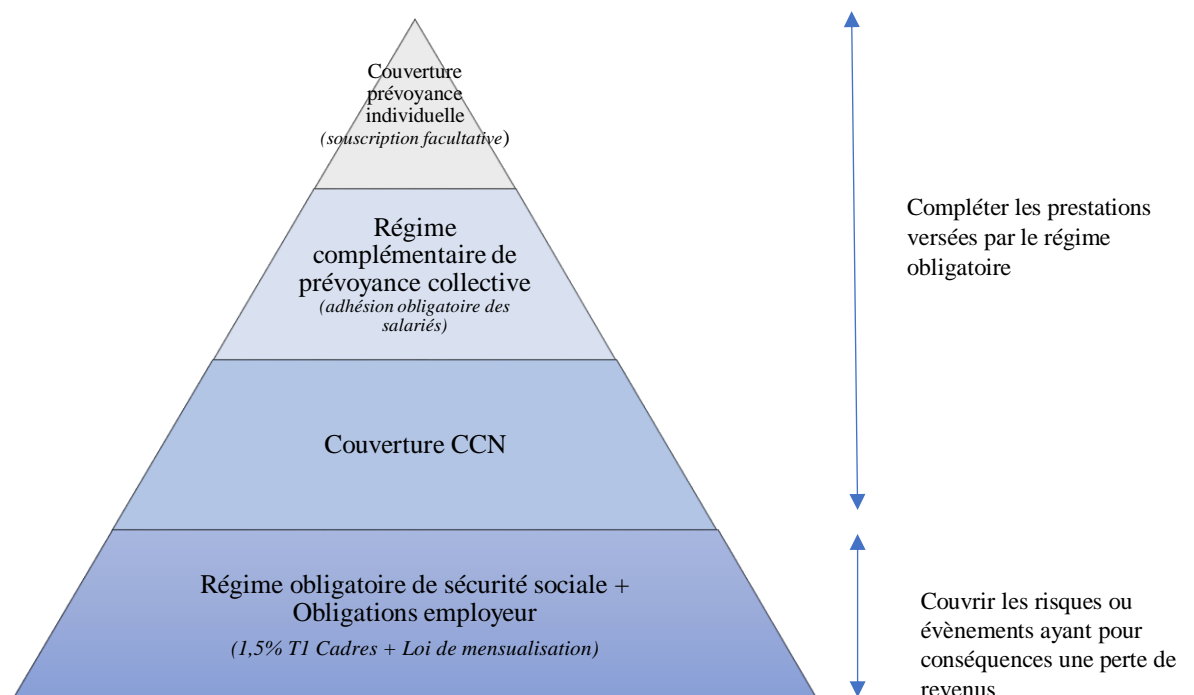


Figure 11: La prévoyance en France

### 1.1.1 Le marché de la prévoyance

Les risques relevant de la prévoyance (incapacité, invalidité, décès) sont couverts :

- Par le régime de base obligatoire de la Sécurité Sociale,
- Par des contrats complémentaires souscrits auprès d'organismes assureurs.

- *Le régime de base de la Sécurité sociale*

Créée en 1945, 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 ». <sup>1</sup>

La sécurité sociale est constituée d'un ensemble d'Institutions ayant pour but de protéger les individus des conséquences de divers événements ou situations, généralement qualifiés de risques sociaux. Elle est destinée à assister financièrement ses bénéficiaires qui rencontrent différents événements coûteux de la vie. Elle repose sur un principe de solidarité, chacun participant à son financement selon ses moyens.

La Sécurité sociale est composée de plusieurs régimes. Ces régimes regroupent les assurés sociaux en fonction de leur catégorie professionnelle. Les principaux sont :

- Le **régime général**, qui concerne les salariés et travailleurs indépendants assimilés et couvre 88% de la population.
- Les **régimes spéciaux**, qui regroupent les fonctionnaires, la SNCF, EDF-GDF, les employés et clercs de notaires, les mines, les cultes, etc... Ces régimes spéciaux couvrent 7% de la population française.
- Le **régime agricole**, qui assure la protection sociale des exploitants et salariés agricoles. Il est géré par la caisse centrale de la Mutualité Sociale Agricole, MSA. Il couvre 5% de la population.
- Les **régimes des non-salariés non agricoles**, qui touchent séparément les artisans, les commerçants ou industriels et les professions libérales pour l'assurance vieillesse. Le risque maladie fait, quant à lui, l'objet d'une gestion commune entre ces différents types de professions.

D'un point de vue fonctionnel, l'organisation actuelle de la Sécurité Sociale résulte de l'ordonnance de 1967 qui instaure la séparation de la Sécurité Sociale en cinq branches autonomes. Chaque branche est alors responsable de ses ressources et de ses dépenses.

- Branche « Maladie » (maternité, paternité, invalidité, décès) ;
- Branche « Accidents du travail et Maladies professionnelles » ;
- Branche « Vieillesse et veuvage » (retraite) ;
- Branche « Famille » (dont handicap, logement, RMI...) ;
- Branche « Autonomie » de la Sécurité sociale a été mise en place au 1 er janvier 2021

---

<sup>1</sup> Source : ordonnance du 4 octobre 1945, texte fondateur

Le graphique ci-dessous synthétise la répartition des recettes par branche pour l'ensemble des régimes obligatoires de base.

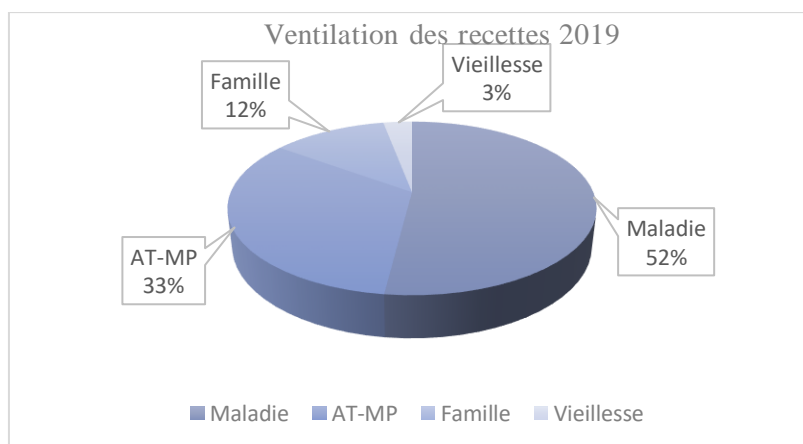


Figure 12: Ventilation des recette 2019<sup>2</sup>

- *Les organismes assureurs*

Les organismes assureurs sont divisés en trois catégories : les mutuelles, les institutions de prévoyance et les sociétés d'assurance. Ils proposent tous au public des contrats d'assurance.

- ✓ Les sociétés d'assurance

Elles sont régies par le Code des Assurances et peuvent prendre différentes formes juridiques : société anonyme, société d'assurance mutuelle ou société mutuelle d'assurance. Contrairement aux autres intervenants, elles ont un but lucratif sur le marché de la prévoyance. Leurs ressources principales sont constituées par des primes d'assurance.

Elles interviennent dans tous les domaines de l'assurance, Prévoyance, Santé, Assurance de Biens ou encore Retraite.

- ✓ Les mutuelles

Régies par le Code de la Mutualité, les mutuelles sont des groupements à but non lucratif qui, essentiellement grâce aux cotisations des membres, mènent des actions de prévoyance.

L'essentiel de leur activité réside dans la couverture des frais de santé, elles interviennent plus rarement sur les garanties de prévoyance lourdes (décès et arrêt de travail). Elles sont contrôlées par les adhérents qui disposent chacun d'une voix dans le cadre des élections en Assemblée Générale.

<sup>2</sup> Source : Annexe 4 du PLFSS pour 2021

✓ Les institutions de prévoyance

Régies par le livre IX du Code de la Sécurité Sociale, les institutions de prévoyance sont des organismes paritaires à but non lucratif. Elles gèrent les garanties de retraite supplémentaire, de prévoyance complémentaire et de complémentaire santé dans le cadre de contrats collectifs pour les salariés. Ils sont dirigés par un conseil d'administration constitué à parts égales de représentants de salariés et de représentants d'entreprises.

En 2020, les cotisations totales des contrats collectifs d'entreprise représentaient 38,2 milliards d'euros d'après l'Urssaf Caisse nationale. (Ce chiffre comprend le périmètre de la prévoyance et de la santé).

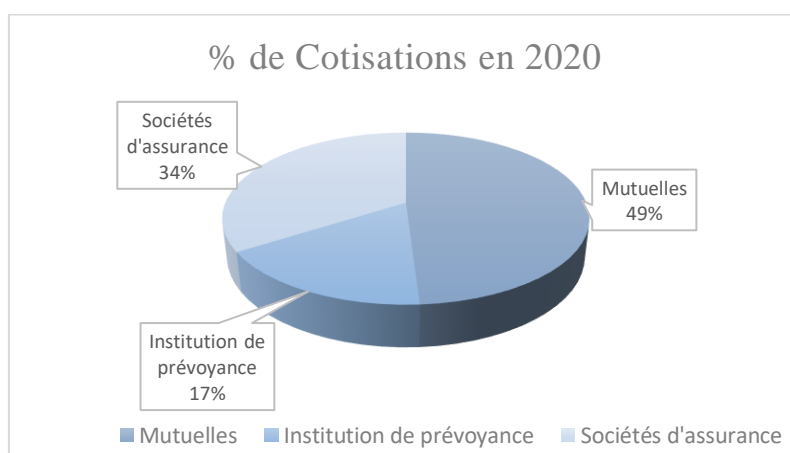


Figure 13 : Pourcentage de cotisations pour les contrats de prévoyance

- *La prévoyance complémentaire*

La couverture de base de la Sécurité Sociale est généralement insuffisante car elle ne comble pas la perte de revenus ou la hausse des dépenses des individus concernés. C'est alors qu'interviennent les régimes complémentaires de prévoyance. En contrepartie de cotisations, ils effectuent des versements supplémentaires qui viennent s'ajouter à ceux de la Sécurité sociale.

Il existe deux types de régimes :

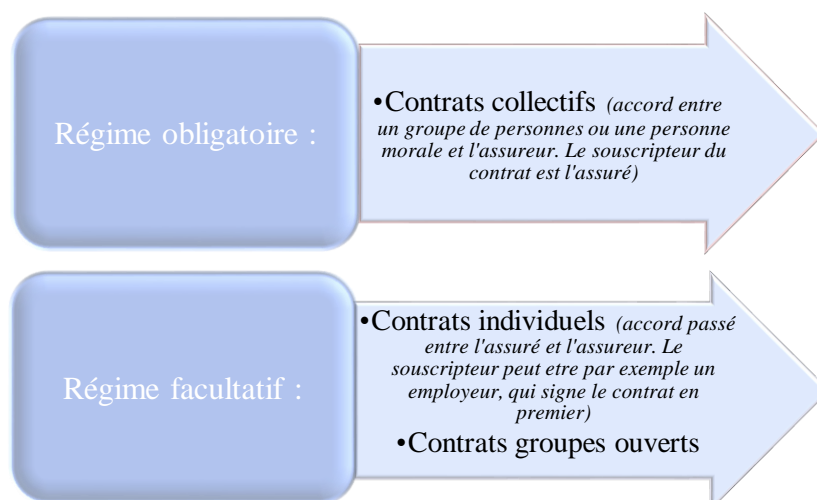


Figure 14: Différents types de régimes

- *La prévoyance collective*

L'objectif d'un contrat de prévoyance collective dans une entreprise est de compléter les prestations des régimes obligatoires de la sécurité sociale afin de mieux protéger les assurés. Un contrat de prévoyance collective propose des garanties couvrant les risques liés au décès, à l'arrêt de travail et à l'invalidité.

Contrairement à la complémentaire santé, la complémentaire prévoyance n'est pas encore légalement obligatoire pour tous les salariés. Seuls les cadres désignés par les articles 4 et 4bis de la Convention Collective Nationale (CCN) de 1947 sont concernés. Leurs employeurs doivent mettre en place un régime de prévoyance et le financer à hauteur minimale de 1,50% de la tranche A de leur salaire.

Pour les autres catégories de salariés, cela dépendra de si la convention collective nationale impose un régime de prévoyance. De ce fait, sans obligation conventionnelle de mise en place, cette prévoyance collective est facultative

Un contrat collectif fait intervenir trois parties ; l'entreprise, le salarié et l'assureur ou l'organisme en charge de la gestion de l'assurance, le fonctionnement et les engagements pris sont les suivants :

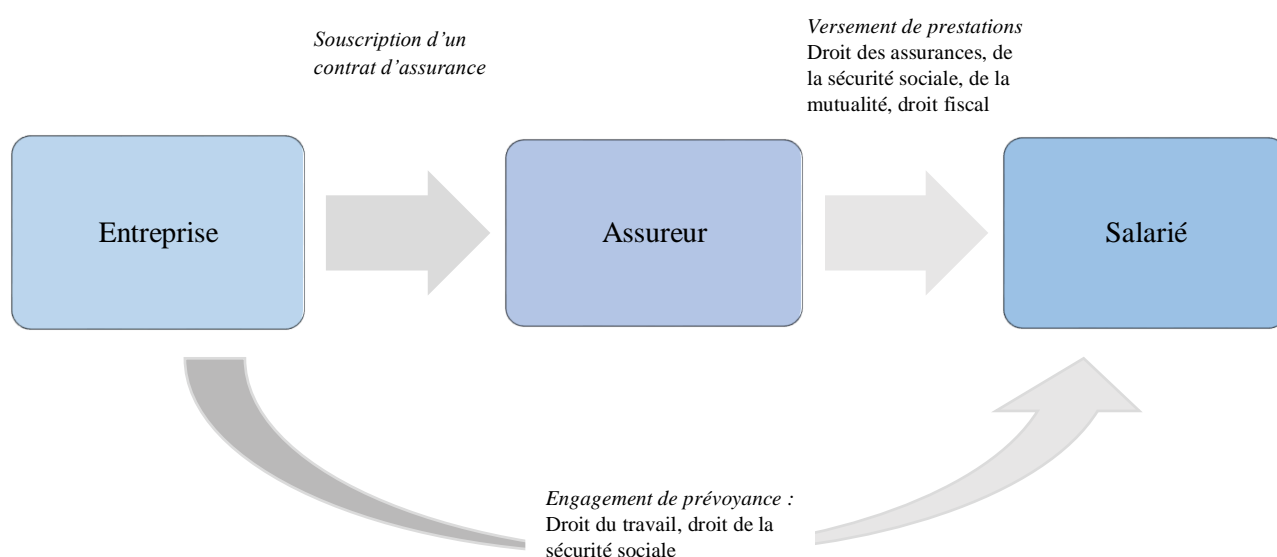


Figure 15: Schéma de fonctionnement d'un contrat collectif

- *La prévoyance individuelle*

La prévoyance peut s'envisager de manière collective dans une entreprise mais également de manière individuelle. Un contrat de prévoyance individuelle peut être souscrit en complément d'un contrat de prévoyance collective, en absence d'un contrat de prévoyance collective ou bien lors du départ en retraite.

Il vient pallier les manques ou l'absence du contrat collectif et de la sécurité sociale lorsque les indemnités versées ne sont pas suffisantes pour faire face aux aléas de la vie.

Les contrats de prévoyance proposent des garanties qui couvrent plusieurs risques :

- Le décès ;
- L'invalidité ;
- Incapacité.

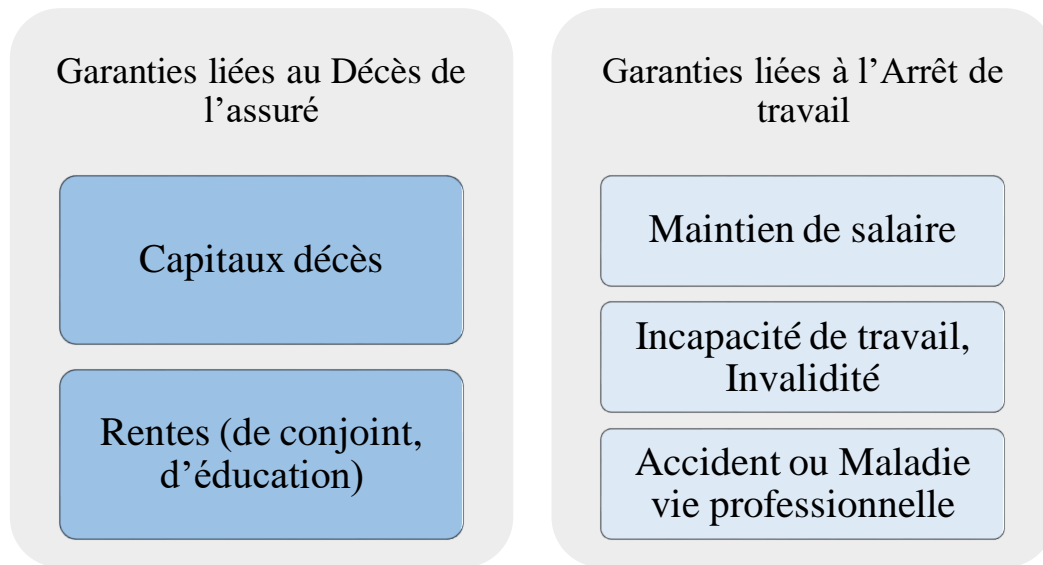


Figure 16: Les différentes garanties

Lorsque survient un événement couvert par le contrat de prévoyance individuelle, la prise en charge se traduit par le versement d'un capital, d'une rente ou d'indemnités journalières selon la nature du sinistre.

### 1.1.2 [Zoom sur risque décès](#)

La couverture du risque décès consiste à verser des prestations aux bénéficiaires du contrat en cas de décès de l'assuré. Il existe deux formes de prestations : les prestations en capital (capital décès) et les prestations en rentes (rentes de conjoint et rentes éducation).

- *Les garanties en capital*

Elles correspondent à la définition d'un montant versé aux bénéficiaires de l'assuré sous forme d'un versement unique au moment de son décès. Le paiement du capital est libératoire pour l'assureur.

En prévoyance collective, le capital décès est généralement calculé en fonction de la rémunération annuelle brute du salarié assuré. Il est le plus souvent exprimé en pourcentage de la totalité de cette rémunération ou d'une ou plusieurs tranches de celle-ci. Dans les contrats de prévoyance individuels ou à adhésion facultative, le montant forfaitaire est défini contractuellement lors de la souscription du contrat.

- *Les garanties en rente*

Elles correspondent à des prestations « échelonnées » versées aux bénéficiaires sous forme de revenus réguliers, généralement selon un fractionnement trimestriel (mais il peut être mensuel, semestriel ou annuel).

Les contrats d'assurance complémentaires distinguent deux types de rentes selon les garanties :

#### ✓ **La rente de conjoint**

La rente de conjoint est définie comme le versement d'une somme déterminée lors de la souscription au contrat au conjoint survivant non remarié. Elle est versée régulièrement par l'assureur.

Elle peut être versée au conjoint marié non séparé de corps par un jugement définitif, au partenaire de PACS ou au concubin notoire (concubin non marié avec l'assuré tout en vivant avec lui de manière stable et continue).

En cas de remariage, ou de conclusion d'un PACS par le conjoint survivant, certains contrats prévoient la cessation du versement de cette rente au conjoint survivant, sauf si des dispositions contractuelles prévoient le maintien de la rente, notamment en cas d'enfant(s) à charge.

Cette rente peut être versée en complément du capital décès, sous forme viagère et/ou temporaire.

Une rente de conjoint peut être temporaire ou viagère.

- *Temporaire* : La rente « temporaire » est une rente versée pendant une période préalablement définie au contrat. Elle cesse donc d'être versée à une date donnée, même si son bénéficiaire est encore en vie.
- *Viagère* : La rente « viagère » est une rente qui prend fin au décès de son bénéficiaire, autrement dit qui est versée tant que ce dernier est en vie.

Dans les contrats de prévoyance collectifs à adhésion obligatoire, le calcul du montant de la rente au conjoint survivant peut être basé sur divers éléments. Parmi eux :

- Un pourcentage du dernier salaire perçu par l'assuré
- La moyenne des salaires des 12 derniers mois
- Les points AGIRC-ARRCO accumulés par le salarié

Dans les contrats de prévoyance individuelle ou à adhésion facultative, la rente est fixée en montant, défini contractuellement.

#### ✓ **La rente d'éducation**

La rente éducation est une rente temporaire versée aux enfants à charge de l'assuré décédé. Cette rente est généralement calculée au prorata du dernier salaire de l'assuré. A partir d'un certain âge, elle peut dépendre de la poursuite d'études du bénéficiaire. Le montant de la rente peut être constant ou exprimé par paliers en fonction de l'âge du bénéficiaire.



Par exemple, la rente versée peut être égale à :

- 12% du salaire jusqu'à 11 ans ;
- 18% du salaire de 12 à 17 ans ;
- 24% du salaire de 18 à 21 ans ou 26 ans (si poursuite d'études).

## 1.2 Le groupe Malakoff Humanis

Le groupe Malakoff Humanis intervient dans les domaines de la protection sociale pour les entreprises et les particuliers

Malakoff Humanis est né de la fusion des groupes Malakoff Médéric et Humanis le 01<sup>er</sup> janvier 2019.

Ce nouveau groupe est issu de plusieurs rapprochements successifs. Pour le groupe Malakoff Médéric, un rapprochement entre Malakoff et Médéric avait été effectué en 2008 tandis que le groupe Humanis est né du rapprochement des groupes Vauban Humanis et Aprionis en janvier 2011.

Aujourd'hui le groupe Malakoff Humanis propose des contrats pour couvrir l'ensemble des besoins de protections de personne en retraite complémentaire, santé, prévoyance et épargne, pour les entreprises et salariés, les indépendants et particuliers. Il est un acteur de la retraite complémentaire et compte deux caisses de retraite complémentaire : Malakoff Humanis Agirc-Arrco et une caisse de retraite complémentaire est dédiée à l'Outre-mer et aux expatriés : Malakoff Humanis International Agirc-Arrco. C'est également le premier assureur en santé collective en prévoyance collective.

Le groupe est composé de plusieurs institutions de prévoyance et de mutuelles.

Le groupe Malakoff Humanis est un groupe de protection sociale paritaire et mutualiste à but non lucratif français. C'est un groupe « Paritaire », administré à parts égales par des représentants des confédérations syndicales de salariés et des représentants des organisations patronales et « Mutualiste », administré par des représentants élus par les adhérents.

L'ensemble des bénéfices sont réinvestis au profit des clients, en services, en accompagnement social ou pour soutenir des causes d'intérêt général.

## 2 Analyse du portefeuille

L'objectif est d'établir une table d'expérience sur l'un des portefeuilles du Groupe Malakoff Humanis afin de mesurer la suffisance des tables réglementaires en regard du risque.

### 2.1 Etude du risque rente de conjoint sur un portefeuille Malakoff Humanis

L'objet de cette étude est de créer une table d'expérience prospective pour les bénéficiaires des rentes de conjoint à partir des données d'expérience.

Chez Malakoff Humanis, le calcul des provisions mathématiques des rentes de conjoint est réalisé à l'aide des tables réglementaires TGH/TGF-05.

Dans le cas où la table serait certifiée, le provisionnement des rentes de conjoint pourrait être établi à partir de cette nouvelle table.

Dans ce mémoire, nous nous intéressons aux conjoints survivants qui sont les bénéficiaires de la rente à la suite du décès de l'assuré. Le bénéficiaire est observable à partir du moment où la rente lui est octroyée jusqu'à sa sortie du portefeuille dans le cas d'une rente viagère, et jusqu'à la date limite de la rente dans le cas d'une rente temporaire. Les adhérents qui sortent en capital ou en rente unique ne sont pas considérés dans l'étude.

#### 2.1.1 Périmètre de l'étude

Pour construire une table d'expérience, il est nécessaire d'exploiter de la meilleure manière les données issues de notre échantillon. La pertinence et la fiabilité des données conditionnent la qualité de la table construite, il faut bien s'assurer qu'une bonne compréhension des données a été faite pour ne pas fausser les résultats.

Il convient de présenter les données telles qu'elles ont été extraites et de décrire le cheminement effectué pour arriver à une base sinistre exploitable.

Deux bases de données ont été utilisées pour cette étude : une base de données extraite du SIP, qui contient les données ex Malakoff Médéric, et une base de données extraite de Pléiade, qui contient les données ex Humanis.

#### 2.1.2 Extraction et description des données brutes

Pour le portefeuille Malakoff Médéric, un premier code Sas a été réalisé permettant d'extraire les données souhaitées pour l'étude à partir d'une base de données des prestations de prévoyance. Les données depuis 1990 y étaient disponibles. Cette base de données a dû être fusionnée avec une ancienne base de données qui contenait les données de prestations de prévoyance depuis la création (historiques), afin d'être complétée.

Une nouvelle table a été recréée à partir des deux bases de données, avec les noms des champs nécessaires pour l'étude et en faisant un filtre sur le risque Rentes de conjoint.

Les variables utiles à l'étude sont sélectionnées. Puis, une clé sera définie afin d'éviter les doublons d'individus. Pour finir, les variables qui semblent influencer le modèle seront extraites afin d'effectuer des tests de significativité.

De nombreuses variables étaient présentes dans la base brute. Une sélection de variables a dû être effectuée pour permettre la construction de la loi de survie.

La première sélection de variables faite est la suivante :

VARIABLES	
Liées au bénéficiaire et à l'assuré	Liées au contrat
<ul style="list-style-type: none"> <li>• Sexe du bénéficiaire (H, <i>Homme</i> ou F, <i>Femme</i>)</li> <li>• Date de naissance du bénéficiaire</li> <li>• <i>Date de naissance de l'assuré</i></li> <li>• <i>Date de décès de l'assuré</i></li> <li>• <i>Catégorie socio-professionnelle de l'assuré</i></li> <li>• Date de survenance, <i>date décès assuré (date de survenance du décès qui est l'élément déclencheur de la rente)</i></li> </ul>	<ul style="list-style-type: none"> <li>• Numéro de dossier (<i>numéro de dossier de rente</i>)</li> <li>• Date d'octroi de rente</li> <li>• Date de début de période</li> <li>• Date de fin de période</li> <li>• Type de rente (<i>temporaire ou viagère</i>)</li> <li>• Age au terme du contrat (50, 55, 60 pour les rentes temporaires et 120 pour les rentes viagères)</li> <li>• Date de l'état <i>date de l'état du contrat (voir après pour le détail)</i></li> <li>• Code motif, (<i>décès, âge atteint, remariage, transfert assureur...</i>)</li> <li>• <i>Montant de base de la rente</i></li> </ul>

Les paiements relatifs à un seul individu font souvent l'objet de plusieurs lignes dans la base de données. En effet, l'individu peut détenir plusieurs contrats de prévoyance, percevoir des prestations de plusieurs contrats différents ou bien avoir des dates qui correspondent à des périodes de paiement de prestations. Le but est de n'avoir qu'une seule ligne par individu et par type de rente afin de ne pas biaiser les estimations. Un conjoint rentier peut bénéficier de 2 rentes, une temporaire et une viagère.

Les clés uniques pour la base sont la date de naissance du bénéficiaire, son sexe, son numéro de dossier. Une procédure de tri a été établie dans un code Sas par les trois clés uniques citées et les doublons ont été retirés. Puis, une autre procédure de tri a été réalisée de manière décroissante par la date état afin de retenir le dernier état en date du dossier, ainsi qu'une nouvelle suppression des doublons.

Une nouvelle base est donc créée. Après l'obtention de cette nouvelle base, une extraction a été réalisée sous Excel de la table afin de pouvoir l'exploiter.

La seconde base de données utilisée est une base de données Humanis de rente de conjoint avec une période observée de 2000 à 2022, depuis l'origine également.

Pour le portefeuille Humanis, un code Sas a également été créé permettant d'extraire les données de l'étude. Les mêmes champs que pour Malakoff Médéric ont été sélectionnés, un filtre a été appliqué pour sélectionner l'entité « Humanis Prévoyance », et les différents assureurs correspondant aux rentes de conjoint.

Puis une procédure de tri a été faite selon le numéro de dossier, l'identifiant du bénéficiaire et le type de rente de conjoint. Ensuite, une élimination des doublons a été réalisée.

Pour finir, une base contenant le code socio-professionnel a été fusionnée avec celle obtenue précédemment afin de récupérer les catégories socio professionnelles des bénéficiaires pour la suite de l'étude. La base finale obtenue a alors été exportée vers Excel.

A la suite de l'obtention des deux extractions sur Excel, un retraitement des bases de données a été effectué, qui sera évoqué dans la partie suivante.

### 2.1.3 Retraitement des données

Une analyse de la base de données doit être réalisée avant tout afin de vérifier la cohérence des données. Il peut exister des anomalies ou des erreurs. Des valeurs manquantes peuvent également être remarquées et il faut donc les retraiter. Le but de cette étape est de récupérer le maximum d'informations et de conserver la majeure partie des données afin de ne pas biaiser notre étude.

Le retraitement des données se découpe en plusieurs étapes.

Pour notre base de données, les retraitements ont été faits à l'aide de logiciels Excel et sas.

Pour commencer, les bases de données ex Malakoff et ex Humanis, aussi conséquentes qu'elles soient, ont dû être réduites afin de ne garder que les informations nécessaires à l'étude.

Tout d'abord, un nouveau tri a été effectué pour enlever les rentes qui ne rentrent pas dans le cadre de l'étude. Les rentes ayant une date de sinistre inférieure à 1980 ont été retirées car elles ne reflètent plus la réalité du portefeuille actuel.

Des nouvelles variables ont été créées pour retraiter certains champs afin de pouvoir analyser au mieux le portefeuille et construire la loi ensuite.

Les champs retraités sont les suivants :

- Age du bénéficiaire à la date de survenance
- Type de rente, *V(Viagère)* ou *T(Temporaire)*

#### Retraitements sur les âges :

- Age du bénéficiaire à la date de décès de l'assuré pour dossiers en cours, en attente de pièces ou clos. Pour obtenir l'âge à la date du décès qui n'est pas disponible.

On applique le même type de procédure pour les calculs des âges ci-dessous.

- Age bénéficiaire à la date de son décès,
- Age bénéficiaire à la date de sa fin de rente pour âge atteint,
- Age bénéficiaire à la date de sa fin de rente pour remariage, pour transfert assureur et pour motif de clôture non renseigné ou autre motif,
- Âge de l'assuré à son décès : si la date de survenance et la date de naissance de l'assuré ne sont pas vides, alors on fait la différence entre la date de survenance et la date de naissance de l'assuré.

#### Sur les dates :

- **Année de survenance**
- **Année de clôture** : si le code état est clos, alors on met l'année du code état.
- Ecart d'âge entre les conjoints à la date de décès de l'assuré

### Sur les codes motifs :

Un retraitement sur les motifs de clôture a été effectué et ils ont été regroupés par catégorie :

- Décès du bénéficiaire
- Remariage
- Age atteint
- Autre motif (transfert assureur, pas de motif renseigné, fin à charge, en attente de pièces...)

#### 2.1.4 Contrôle des données

Afin de préserver la qualité des données, des contrôles ont dû être réalisés sur la base de données afin de vérifier leur cohérence. Pour commencer, certaines variables ont été corrigées ou supprimées, dans le cas où certaines étaient manquantes ou bien aberrantes.

##### ✓ Corrections de variables

Une vérification de la cohérence sur le sexe du bénéficiaire et la date de naissance du bénéficiaire a été réalisée, ainsi que sur d'autres variables qui sont des variables nécessaires à l'étude.

Les règles suivantes ont été appliquées pour l'étude pour compléter les valeurs manquantes ou traiter des anomalies :

- *Date de naissance du bénéficiaire*

Les dates de naissance des bénéficiaires sont indispensables à la réalisation de l'étude.

Les dates manquantes ont été recherchées à la main dans le SIP, système qui contient toutes les informations sur les assurés, puis ont été extraites sur un fichier Excel.

Les lignes pour lesquelles la date de naissance n'a pas pu être complétée ont été supprimées de la base de données.

- *Age bénéficiaire à la date de survenance,*

Une vérification a été réalisée afin de s'assurer que les dates de naissance des bénéficiaires n'étaient pas incohérentes. Les bénéficiaires ayant un âge inférieur à 18 ans à la date de survenance du sinistre sont supprimés de la base.

- *Sexe du bénéficiaire,*

Si le sexe du bénéficiaire est vide ou à « Inconnu » et que le sexe de l'assuré est mis à « F » (*Femme*), alors on considère que le bénéficiaire est un homme le sexe est forcé à « M » (*Homme*). Inversement si le sexe de l'assuré est à « M », alors on force le sexe du bénéficiaire à « F ».

On contrôle également les anomalies suivantes au cas par cas. Dans le cas où les sexes du bénéficiaire et de l'assuré sont mis à « Inconnu » ou ne sont pas renseignés, alors les sexes ont été cherchés dans le

SI de Gestion afin de récupérer les valeurs manquantes. Pour les valeurs introuvables, les bénéficiaires concernés ont été retirés de l'étude.

- *Date de survenance,*

Si elle est vide, alors on met la date de décès de l'assuré. Un contrôle a également été fait dans le cas où la date de survenance est différente de la date de décès du bénéficiaire. Dans ce cas, on prend la date qui est la plus ancienne entre les 2. Cela représente un faible nombre dans la base de données.

- *Age limite de la rente,*

Une vérification a été réalisée pour les âges limites des rentes viagères, à « 120 » ans pour les rentes viagères et à « 50,55,60 » pour les rentes temporaires.

- *Type de rente,*

Lorsque le type de rente n'est pas renseigné, l'âge limite de la rente est regardé, s'il est à 120, alors le type de rente est rempli à « rente viagère » et dans le cas où les limites de rente sont à 50,55,60,65 alors il est mis à « rente temporaire ».

#### ✓ Suppression d'incohérences :

Les lignes incohérentes sont alors supprimées de la base de données pour ne pas fausser l'étude.

- Plusieurs vérifications ont été faites sur la date de naissance de bénéficiaire,
  - Elle doit être inférieure à la date de décès de l'assuré ;
  - Elle doit également être inférieure à la date d'effet de contrat.
- Certaines anomalies non consécutives sur la date d'effet de rente vide alors que la date du bénéficiaire était renseignée ont été remarquées sur la base
- La date de début de période doit également être inférieure à la date de fin d'observation
- La date d'effet de rente (date début période) doit être inférieure à la date de fin de période

Après tous les retraitements et les contrôles de qualité de données faits sur le fichier Excel, une colonne a été créée afin d'éliminer des lignes qui ne rentreront pas dans notre étude. Les lignes éliminées sont les suivantes :

- Les lignes des bénéficiaires qui ont moins de 18 ans ;
- Les dates de naissance des bénéficiaires qui sont égales à « 01/01/1900 » ;
- Les lignes où la date de survenance et la date de décès de l'assuré ne sont pas renseignées ainsi que les lignes où l'année de survenance et l'année de décès de l'assuré sont différentes et dont une information est mal renseignée ;
- Les cas d'état de sinistre mis à « radié » ;

- Lorsque la date de fin de période est vide et que la date de décès du bénéficiaire n'est pas renseignée (nous ne pouvons ainsi pas déterminer la sortie du portefeuille) ;
- Lorsque la date de fin de période n'est pas renseignée ainsi que la date de décès du bénéficiaire.

### ✓ Concaténation des données

Le but de cette étude étant d'obtenir une table d'expérience Malakoff Humanis, fusionner les données des portefeuilles Malakoff Médéric et Humanis était une étape essentielle.

Après obtention de deux extractions retraitées pour les deux portefeuilles, une sélection, un regroupement et un renommage de variables essentielles pour la suite de l'étude ont été réalisés. Puis, une concaténation des données a été réalisée sur R avec la fonction rbind. Une extraction de cette nouvelle base est faite sous Excel.

### 2.1.5 Statistiques descriptives du portefeuille

Après le retraitement des données, des statistiques descriptives ont été réalisées sur les données retraitées afin de décrire le portefeuille.

Premièrement, des graphiques en secteurs ont été tracés afin de visualiser certaines proportions du portefeuille :

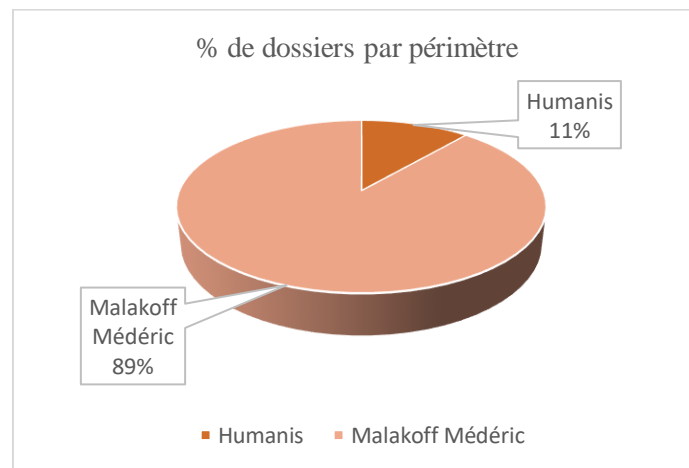


Figure 17 : Pourcentage de dossiers par périmètre

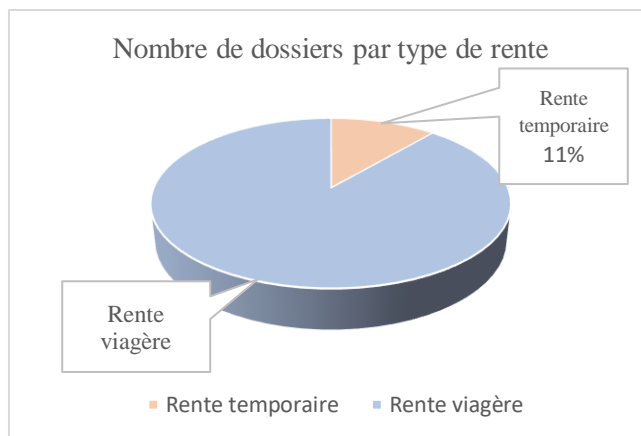


Figure 18 : Nombre de dossiers par type de rente

Le premier graphique représente le pourcentage de dossiers par périmètre. On remarque que le portefeuille étudié est majoritairement composé de dossiers de Malakoff Médéric.

Le second graphique réalisé représente le nombre de dossiers par type de rente. Nous avons un portefeuille contenant 31 637 rentes en tout dont 3575 rentes temporaires et 28062 rentes viagères, ce qui représente 11% contre 89%. Notre portefeuille est majoritairement constitué de rentes viagères.

Deux autres graphiques en secteurs ont été créés représentant le nombre de dossiers par sexe pour les 2 types de rentes :

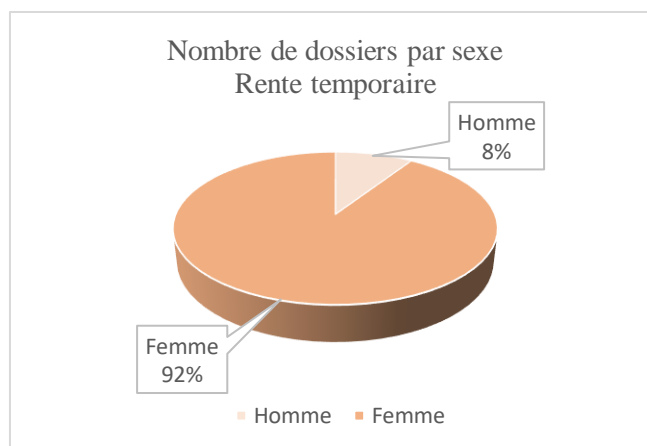


Figure 19 : Nombre de dossiers par sexe pour les rentes temporaires



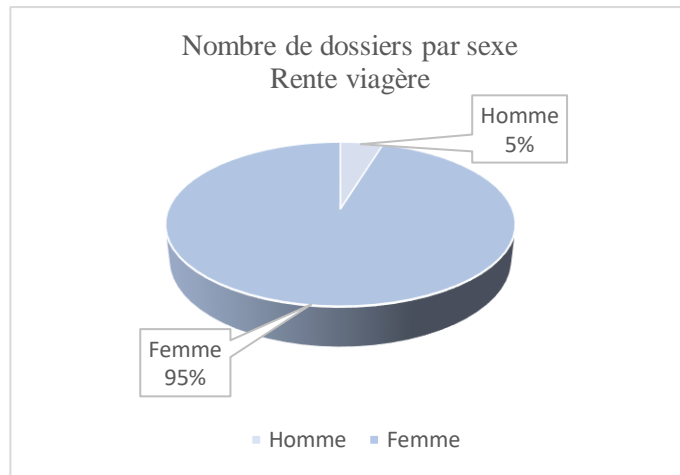


Figure 20 : Nombre de dossiers par sexe pour les rentes viagères

On peut remarquer une part importante de femmes pour les rentes viagères 95% (27 136 rentes) contre 5% (1339 rentes hommes), Une part importante de femmes est également remarquée pour les rentes temporaires 91% (3270 rentes) contre 9% rentes hommes (332).

Les âges moyens ont été déterminés pour l'ensemble du portefeuille et ont été répertoriés dans le tableau suivant :

	Hommes	Femmes	Population globale
Age moyen du bénéficiaire à la date de survenance	50	54	54
Age moyen de l'assuré à la date de son décès	51	49	49
Age moyen du bénéficiaire à la date de sortie	66	73	73
Age moyen du bénéficiaire à son décès	70	79	79

Tableau 3 : Ages moyens par genre

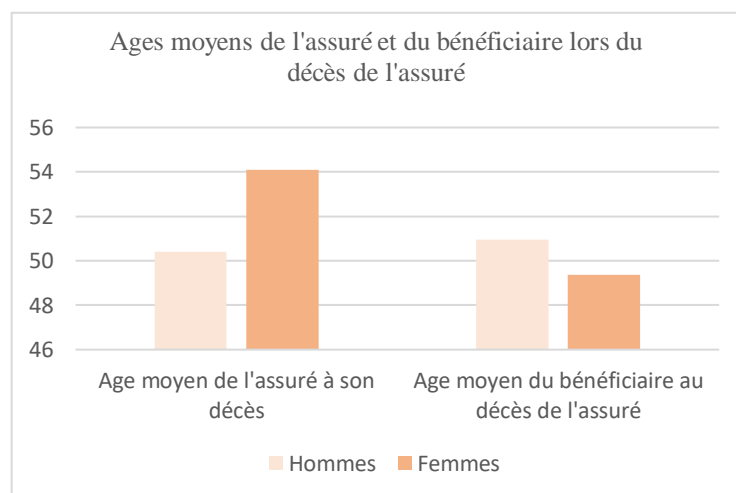


Figure 21 : Ages moyens de l'assuré et du bénéficiaire lors du décès de l'assuré

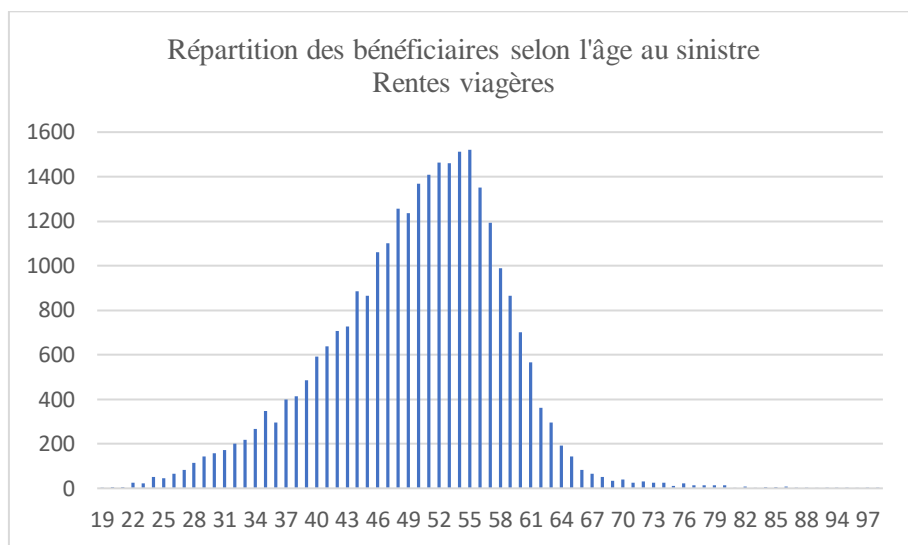


Figure 22 : Répartition des bénéficiaires selon l'âge au sinistre pour les rentes viagères

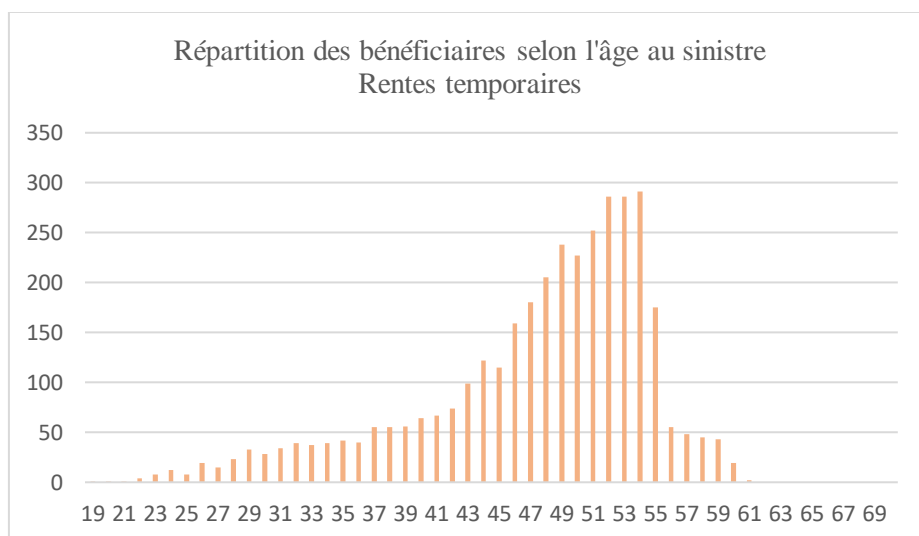


Figure 23 : Répartition des bénéficiaires selon l'âge au sinistre selon le type de rente

Des histogrammes empilés ont ensuite été réalisés :

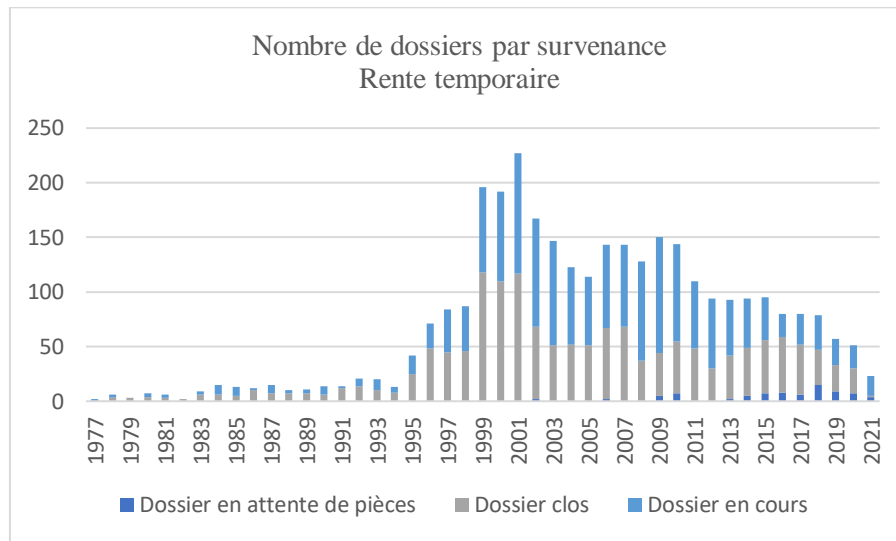


Figure 24 : Nombre de dossiers par survenance pour les rentes temporaires

L’histogramme ci-dessus représente le nombre de dossiers par survenance pour les rentes temporaires en fonction de l’état des dossiers depuis la création des rentes de conjoint. On peut constater qu’un nombre important de dossiers sont clos au long des années.

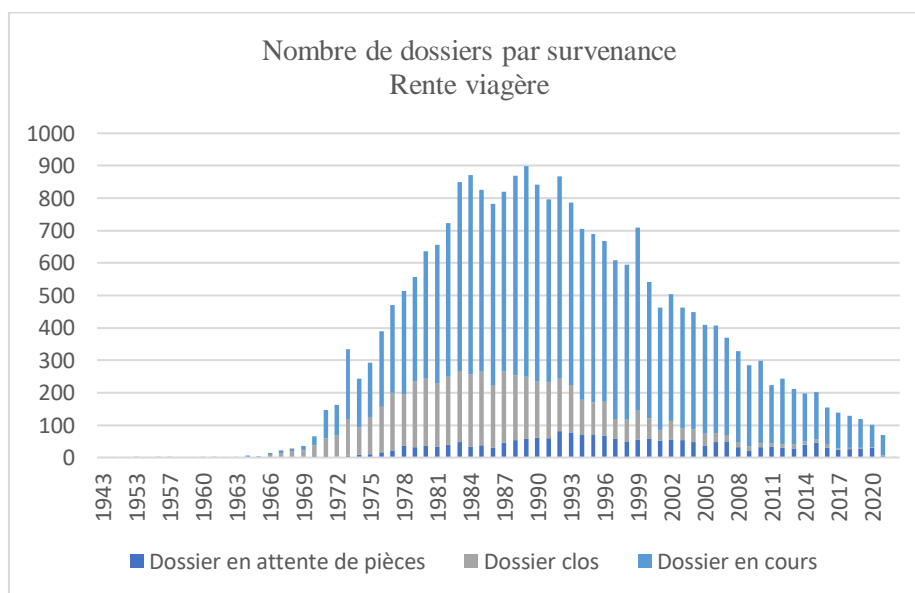


Figure 25 : Nombre de dossiers par survenance pour les rentes viagères

Le nombre de dossiers par âge à la survenance a maintenant été tracé pour les deux types de rentes :

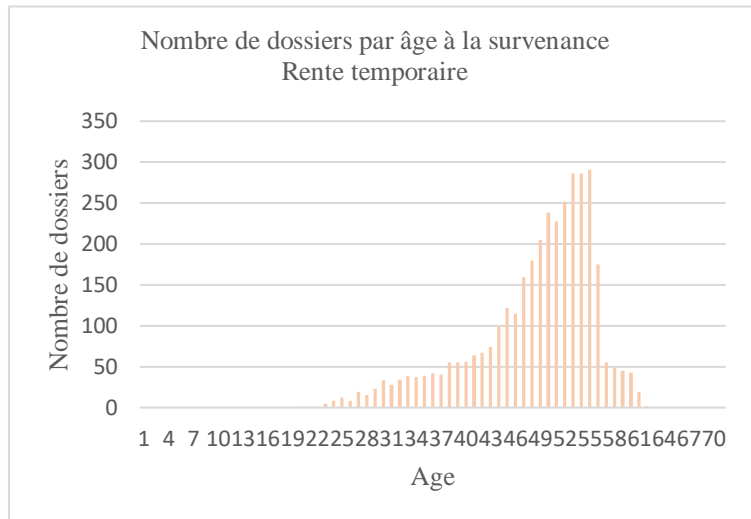


Figure 26 : Nombre de dossiers par âge à la survenance pour les rentes temporaires

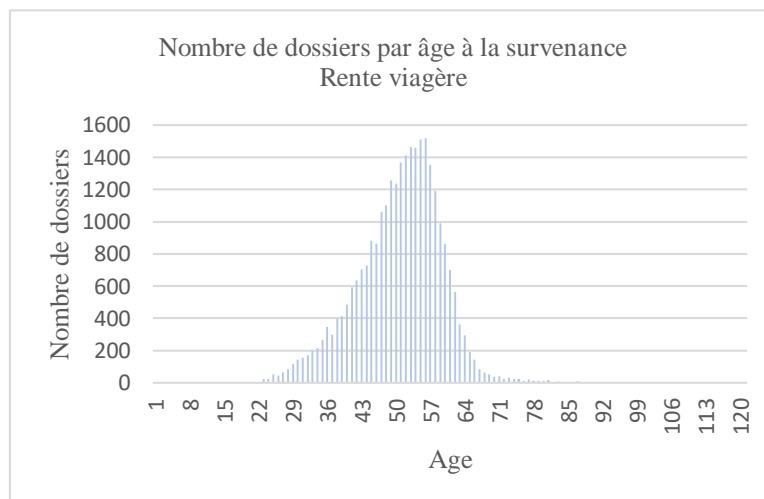


Figure 27 : Nombre de dossiers par âge à la survenance les rentes viagères

On remarque un pic de dossiers autour de 52 ans pour les rentes temporaires, tandis que pour les rentes viagères, un pic est observé autour de 56 ans soit un peu plus tard. La majorité de dossiers est concentrée autour de 48 à 55 ans pour les rentes temporaires et de 50 à 65 pour les rentes viagères.

Les 2 graphiques ci-dessous représentent le nombre de dossier par âge à l'inventaire pour les deux types de rentes.

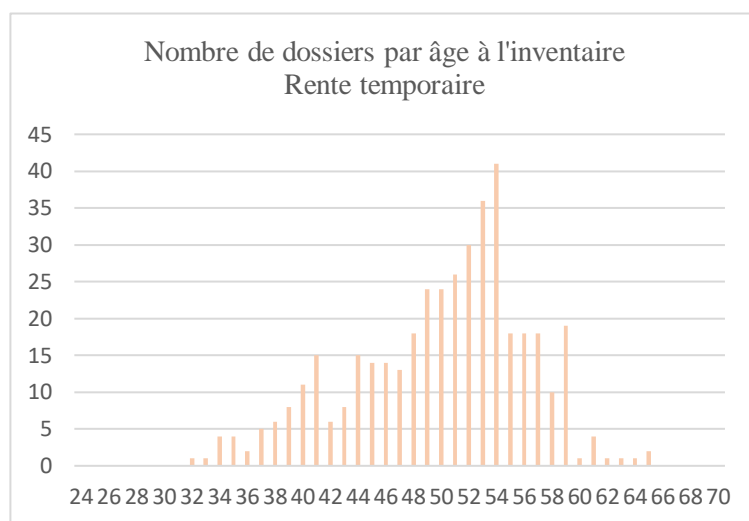


Figure 28 : Nombre de dossiers par âge à l'inventaire pour les rentes temporaires

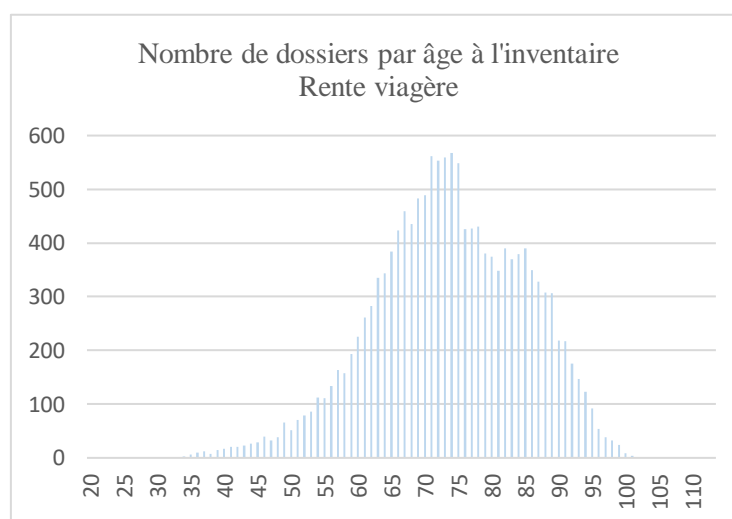


Figure 29 : Nombre de dossiers par âge à l'inventaire pour les rentes viagères

De nombreux contrats sont remarquables entre 65 et 90 ans avec un pic à [70,75] ans pour les rentes viagères. Pour les rentes temporaires, les contrats sont majoritairement concentrés autour de [50,55] ans.

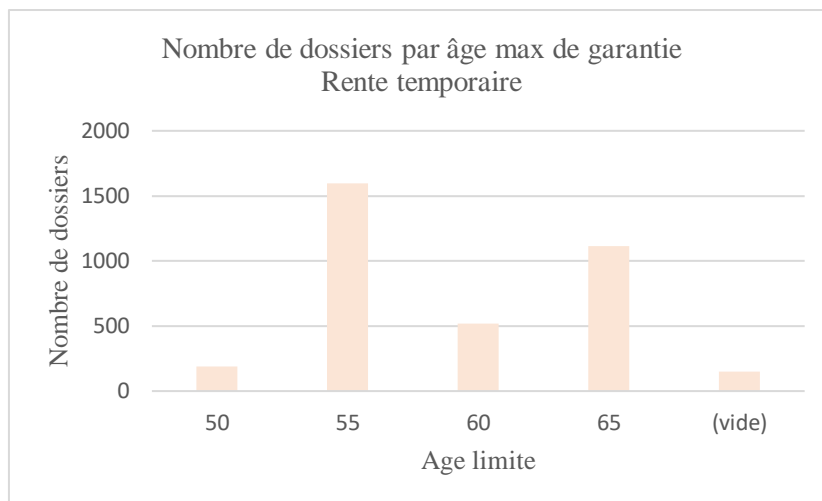


Figure 30 : Nombre de dossiers par âge max de garantie pour les rentes temporaires

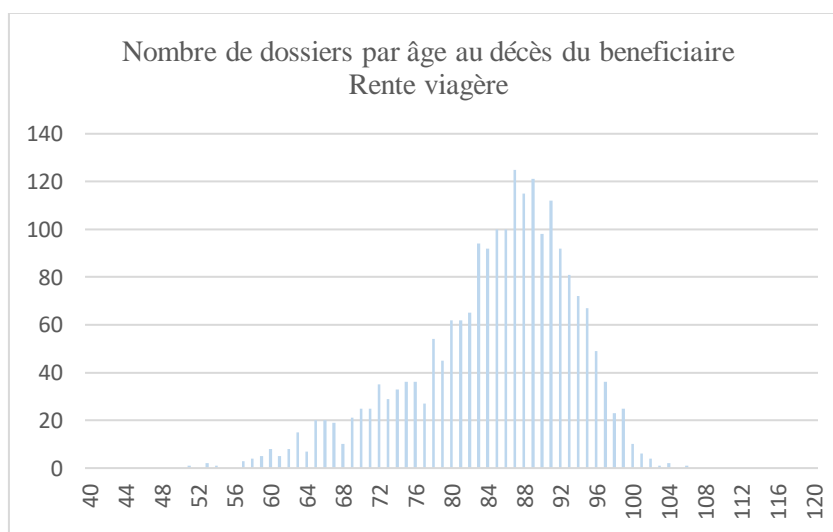


Figure 31 : Nombre de dossiers par âge au décès du bénéficiaire pour les rentes viagères

## 2.2 Influence des variables

Nous cherchons à déterminer l'influence des variables pour la construction de la loi. Pour cela, une étude descriptive a été réalisée afin de faire un tri parmi les variables explicatives.

Les variables qui segmentent la table sont principalement :

- La variable AgeEntree qui correspond à l'âge d'entrée dans le portefeuille.
- La variable DureeExposition qui correspond à la durée d'un individu dans l'étude.

Les variables ci-dessous peuvent aussi influencer la durée dans le portefeuille :

- La variable Sexe qui correspond au genre de l'assuré,
- Le type de rente (temporaire ou viagère),
- Âge limite de la rente,
- Le code socio professionnel (ou profession ou montant de base),
- L'année de survenance du sinistre, ou date de décès du conjoint (ou date entrée),
- Périodicité de paiement,
- Ancienneté du contrat,
- Le motif de clôture (remariage ou non)

En raison de la volumétrie de la base de données, la loi ne peut pas être segmentée selon chacune de ces variables.

La corrélation entre les variables sélectionnées a été testée afin d'optimiser les résultats du modèle. Elles seront tout de même analysées à l'aide de tests statistiques dans le chapitre suivant pour évaluer leur influence ainsi que leur significativité sur les taux de sortie.

Nous avons tout d'abord commencé par observer la répartition de l'exposition par âge des bénéficiaires sur le portefeuille global.

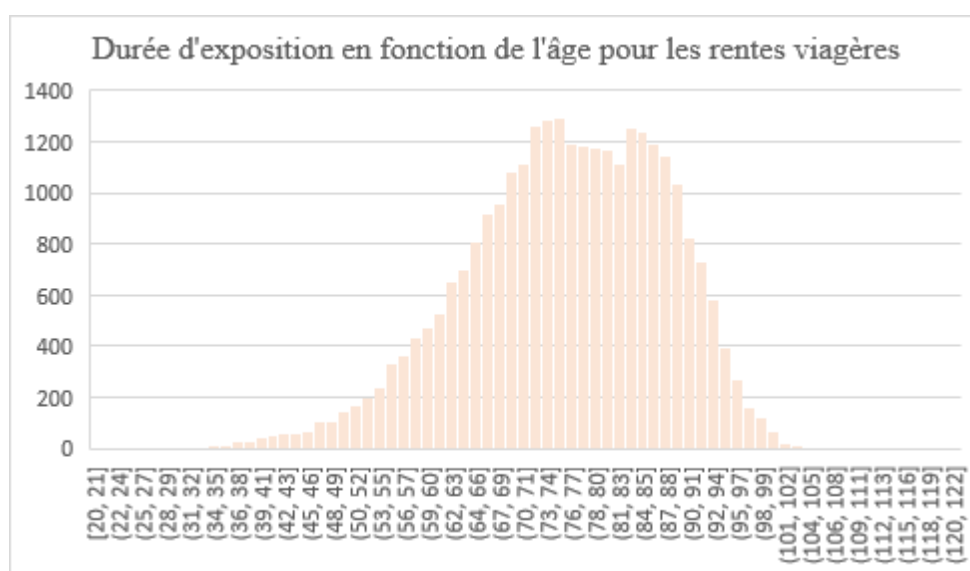


Figure 32 : Durée d'exposition en fonction âge pour les rentes viagères

On remarque qu'une grande partie de l'exposition se situe pour les âges entre 65 et 90 ans. La durée moyenne d'exposition d'un bénéficiaire a été calculée et est de 23 ans.

Nous allons maintenant réaliser des tests statistiques afin de voir quelles variables sont influentes et voir s'il est intéressant de les segmenter dans notre étude.

### 2.2.1 Test du Log-rank

La période d'observation choisie étant volontairement longue (de 2000 à 2021), il est important de s'assurer que des évolutions trop importantes n'ont pas été remarquées sur cette période. C'est pour cela que nous avons utilisé le test du Log-rank. C'est un test qui permet de vérifier la stabilité dans le temps.

Le test du Log-rank est une approche non paramétrique qui permet de comparer les distributions de deux courbes de survie en prenant en compte les données censurées.

La survie entre deux populations peut être comparée sans faire d'hypothèse au préalable sur leur distribution. Pour comparer les fonctions de survie, les deux populations doivent être indépendantes et les observations doivent appartenir à des périodes disjointes.

L'utilisation du test des Log Rank permet de comparer les courbes de survie mais ne permet pas de quantifier l'impact de la variable explicative sur la mortalité contrairement au modèle de Cox que nous verrons plus loin.

L'objectif est de tester les hypothèses suivantes :

*Hypothèse nulle*  $H_0$  : « les probabilités de survie entre les 2 populations sont identiques ainsi que les allures des fonctions de survie. »

*Hypothèse alternative*  $H_1$  : « les probabilités de survie entre les 2 populations sont différentes, les allures des fonctions de survie diffèrent. »

#### ✓ Paramètre et loi statistique du test

Le test statistique est basé sur la statistique du Khi-deux. Sous l'hypothèse  $H_0$  on calculera le nombre de sorties « attendues » dans chacun des groupes (noté E).

On comparera ensuite le nombre de sorties observées (O) aux nombres de sorties attendues (E).

#### ✓ Description du test pour notre portefeuille

Nous souhaitons ici comparer la durée de vie des bénéficiaires dans le portefeuille en distinguant les hommes et les femmes. Nous disposons de 2 échantillons.

Nous supposons que les deux échantillons (A et B) sont indépendants et censurés.

Nous souhaitons tester l'hypothèse nulle d'égalité des fonctions de survie des deux échantillons :

- Hypothèse  $H_0$  :  $S_A = S_B$  les fonctions de survie des deux échantillons sont égales
- Hypothèse  $H_1$  :  $S_A \neq S_B$  les fonctions de survie des deux échantillons sont différentes.

On compare alors le nombre de sorties observées dans chaque groupe aux sorties théoriques (ou attendues).

Sous l'hypothèse  $H_0$ , le taux de sorties attendu est identique dans les échantillons A et B.



Soient :

- $d_{i,j}^{obs}$  le nombre de sorties observées dans le groupe j au temps i.
- $d_{i,j}^{theo}$  le nombre de sorties attendues dans le groupe j au temps i. On calcule pour chaque temps d'événement le nombre de décès observés et ceux attendus.

On calcule pour chaque temps d'événement le nombre de décès observés et ceux attendus.

- $D_j^{obs} = \sum_{i=1}^N d_{i,j}^{obs}$
- $D_j^{theo} = \sum_{i=1}^N d_{i,j}^{theo}$

La statistique du Log-Rank s'écrit :

$$\phi = \sum_{j=1,2} \frac{(D_j^{obs} - D_j^{theo})^2}{D_j^{theo}}$$

Cette statistique suit asymptotiquement une loi de khi-deux à k-1 degrés de liberté, k étant le nombre de groupes à comparer (k=2).

Soit :

$$\phi = \sum_{j=1,2} \frac{(D_1^{obs} - D_1^{theo})^2}{D_1^{theo}} + \frac{(D_2^{obs} - D_2^{theo})^2}{D_2^{theo}} \rightarrow \chi(1)$$

Sous l'hypothèse nulle, cette quantité suit approximativement une loi du  $\chi^2$  à 1 ddl sous  $H_0$ .

On calcule également la p-value de niveau  $1 - \alpha = 95\%$ .

- Si  $p \leq \alpha$  : on rejette l'hypothèse  $H_0$ .
- Sinon on ne rejette pas  $H_0$ .

### ✓ Résultats

Le test du log-rank a été appliqué dans R avec la fonction `survdiff` du package `survival` pour tester les différentes variables.

On obtient les résultats suivants :

	Segmentation	Khi-2	p-value
<i>Test du Log rank</i>	Sexe	350	p= <2e-16
	Type de rente	2909	p= <2e-16

Tableau 4 : Résultats du test du Log-Rank

### Pour le sexe et le type de rente :

La p-value étant inférieure au seuil de 5%, cela conduit à rejeter l'hypothèse  $H_0$  selon laquelle les deux courbes de survie des populations (masculine et féminine) ainsi que (rente temporaire rente viagère) sont égales.

On peut alors conclure que la distinction suivant le sexe et le type de rente concernant la survie des bénéficiaires est justifiée statistiquement.

Le test de Cox est alors appliqué sur le portefeuille afin de confirmer ces hypothèses.

## 2.2.2 Méthode de Cox

Le modèle de Cox est un modèle qui permet de voir si certaines variables exercent une influence sur le modèle.

### 2.2.2.1 Présentation de l'estimateur

Le modèle de Cox est un outil semi-paramétrique d'estimation de la fonction de survie. C'est un modèle de régression qui a pour but d'évaluer simultanément l'effet de plusieurs facteurs sur la survie, en d'autres termes cela permet d'examiner comment des facteurs spécifiques influencent le taux d'occurrence d'un événement particulier (dans notre cas la sortie du portefeuille des bénéficiaires), à un moment donné. Il cherche donc à évaluer l'effet de variables explicatives sur la fonction de hasard de base.

Dans notre cas, le but de l'utilisation d'un modèle de Cox est ici de tester la significativité des variables explicatives en vérifiant la pertinence du modèle.

#### ✓ Présentation du modèle

Le modèle de Cox permet d'exprimer le risque instantané de survenue de l'événement en fonction de l'instant  $t$  et des variables explicatives  $X_j$ . Le risque instantané de survenue de l'événement représente la probabilité d'apparition de l'événement à l'instant  $t$  sachant qu'il ne s'est pas encore réalisé juste avant l'instant  $t$ .

Soit  $\theta_0 = (\theta_1, \dots, \theta_p)$ .

Le modèle de Cox s'écrit :

$$h(t|X, \theta) = \exp(-X^t \theta) * h_0(t)$$

Où :

- $h_0(t)$  la fonction de hasard de base commune à tous les individus.
- $\theta_0 = (\theta_1, \dots, \theta_p) \in \mathbb{R}$  les paramètres du modèle, ils ne dépendent pas du temps
- $t$  L'ancienneté du contrat en jours
- $X = (X_1, \dots, X_n)^t$  les variables explicatives

✓ Hypothèses du modèle

Le modèle de Cox repose sur deux hypothèses de base :

- l'hypothèse de proportionnalité des fonctions de risques instantanés au cours du temps entre les différents profils, entre le profil de base 0 et le profil  $i$  par exemple :

$$\ln\left(\frac{h_i(t)}{h_0(t)}\right) = X^t \beta$$

- l'hypothèse de linéarité, le rapport des fonctions de risque instantané pour deux individus et de caractéristiques et ne dépend pas du temps :

$$\frac{h_i(t, X_i)}{h_j(t, X_j)} = \frac{\exp(X_i^t \beta)}{\exp(X_j^t \beta)} = \text{constante}$$

C'est-à-dire que les fonctions de risques instantanés des deux individus  $i$  et  $j$  sont proportionnelles, et que leur rapport de proportionnalité ne dépend pas du temps  $t$ .

✓ Estimation des paramètres :

Les estimateurs sont obtenus par la méthode du maximum de vraisemblance. Plus exactement, seule la partie de la vraisemblance comportant de l'information sur les coefficients est retenue pour les calculs. Il s'agit de la "vraisemblance partielle" ou de la "vraisemblance de Cox" :

$$L = \prod_{i=1}^n \left( \frac{\exp(-x_i^t \theta)}{\sum_{j=1}^{n_i} \exp(-x_j^t \theta)} \right)^{d_i}$$

Où :

- $d_i$  est l'indicateur de censure du  $i$  ème contrat, égal à 1 si le contrat est censuré et 0 sinon,
- $n_i$  le nombre de contrats exposés au risque avant le temps  $t_i$ , temps de chute du contrat  $i$ ,
- $x_i$  sont les caractéristiques du contrat qui chute à  $t_i$ ,
- $x_j$  sont les caractéristiques du  $i$ ème contrat des  $n_i$  contrats.

Le paramètre de nuisance  $h_0$  n'intervient pas dans la vraisemblance partielle. Le principe pour le modèle de Cox est de n'estimer que les coefficients  $\theta_i$ . Nous ne cherchons pas à estimer  $h_0$ .

Les valeurs de  $\theta$  qui maximisent la fonction de vraisemblance partielle sont aussi celles qui maximisent le logarithme de cette fonction.

On peut estimer le risque cumulé de base par l'estimateur de Breslow :

$$\hat{h}_0(t) = \sum \frac{d_i}{\sum_{i=1}^{n_i} \exp(-x_j^t \theta)}$$

On peut alors en déduire un estimateur de la fonction de survie pour un vecteur de covariable X :

$$S(t|X, \theta) = [S_0(t)]^{\exp(-X^t\theta)}$$

Avec

- $S_0$  la fonction de survie du groupe de référence.
- $\theta_0 = (\theta_1, \dots, \theta_p) \in \mathbb{R}$  les paramètres du modèle
- $t$  L'ancienneté du contrat en jours
- $X = (X_1, \dots, X_n)^t$  les variables explicatives

✓ Effet des covariables :

Soit  $X_i$  une variable qualitative tel que :  $X_i = 1$  si l'individu  $i$  est un homme et 0 sinon

$$X_i = \begin{cases} 1 & \text{si l'individu } i \text{ est un homme} \\ 0 & \text{si l'individu } i \text{ est une femme} \end{cases}$$

$$\frac{h_0(t) \exp(\theta_1 * 1 + \dots + \theta_p X_{ip})}{h_0(t) \exp(\theta_1 * 0 + \dots + \theta_p X_{ip})} = \exp(\theta_1)$$

$\exp(\theta_1)$  est alors le rapport de risque entre un homme et une femme toutes choses étant égales par ailleurs (c'est à dire après avoir pris en compte la valeur des autres covariables).

- Si  $\theta_1 > 0$ ,  $\exp(\theta_1) > 1$  et le risque que l'évènement d'intérêt se produise est plus élevé chez les hommes que chez les femmes.
- Si  $\theta_1 < 0$ ,  $\exp(\theta_1) < 1$  et le risque que l'évènement d'intérêt se produise est plus faible chez les hommes que chez les femmes.
- Si  $\theta_1 = 0$ ,  $\exp(\theta_1) = 1$  et le risque instantané est le même chez les hommes et chez les femmes.

#### 2.2.2.2 Application aux données

✓ Validation des hypothèses

Nous allons essayer de vérifier les hypothèses du modèle de Cox :

L'hypothèse des risques proportionnels est l'hypothèse majeure du modèle.

- Principe de proportionnalité :

La première hypothèse à valider est le principe de proportionnalité. Cette hypothèse suppose que les résidus de Schoenfeld ne dépendent pas du temps. Elle peut être vérifiée en testant l'indépendance entre la variable relative au temps et chacune ou toutes les variables explicatives du modèle. Ce test est appelé test des résidus de Schoenfeld. Une valeur de p-value inférieure à 5% indique que l'hypothèse n'est pas

vérifiée. Notre objectif est donc d'avoir pour chacune des variables explicatives ainsi que pour l'ensemble du modèle une p-value  $> 5\%$ .

- *Principe de log linéarité*

L'hypothèse de log-linéarité (difficile à vérifier en pratique).

L'écriture du modèle suppose l'existence d'une relation Log linéaire entre la fonction de risque instantanée et les covariables.

#### ✓ Résultats

Pour illustrer la mise en œuvre du modèle de Cox, la fonction `coxph` du package `survival` de R a été utilisée, qui calcule les coefficients de régression pour tester la significativité de la variable « Sexe » sur la loi de maintien.

Les coefficients  $\theta$  obtenus par le modèle de Cox sont de -0.25 pour la variable « Sexe » et de -0.57 pour le « type de rente » avec deux p value significativement faibles et inférieures à  $2^{-16}$ .

L'hypothèse de proportionnalité n'est donc pas vérifiée, l'hypothèse  $H_0$  est rejetée. Il est donc nécessaire de segmenter selon le sexe et le type de rente pour notre étude. De plus, Le provisionnement de la garantie rente de conjoint est réalisé par deux tables générationnelles : la TGH05 pour les hommes et la TGF05 pour les femmes, tous types de rentes confondues. Il serait ainsi judicieux de créer deux tables d'expérience : une table générationnelle homme ainsi qu'une table générationnelle femme, les deux populations ayant une tendance de mortalité légèrement différente.

#### Conclusion de cette partie :

L'analyse que nous venons de présenter avec le test du Log-rank et le modèle de Cox montre que certaines variables telles que le sexe ou le type de rente peuvent être intéressantes pour l'étude afin qu'elle soit plus fine et segmentée.

Cependant, en raison du faible nombre de données sur la population masculine, seule une table d'expérience générationnelle pour les femmes sera construite pour le portefeuille, qui est principalement constitué de femmes. Réaliser une table générale sans distinction du genre ne serait pas pertinent en raison du faible nombre d'hommes dans le portefeuille, qui ne représenterait pas un poids assez conséquent pour les considérer dans une table mixte.

La table d'expérience ne sera pas segmentée selon le type de rente en raison du manque de données à disposition mais également car les tables générationnelles distinguent seulement le genre et non le type de rente.

En conséquence, seule une table femme sera construite dans la suite de l'étude en ne distinguant pas le type de rente.

A l'avenir, si le nombre de données nous le permet, une table homme pourra être envisagée avec la même démarche que dans cette étude avec la table femme.

### **3 Construction de la loi d'expérience en rente de conjoint**

Les tables de mortalité servent à tarifer et à provisionner. La réglementation actuelle impose aux assureurs d'utiliser des tables de mortalité « prudentes » (Art. A335-1 C. Ass.) ce qui conduit à construire des tables majorant les taux de décès pour les risques en cas de vie et majorant les espérances de survie pour les risques en cas de vie (y compris les rentes).

Les tables d'expérience sont construites sur la base de données du portefeuille de l'organisme assureur alors que les tables réglementaires sont construites sur la base de données de la population française.

Utiliser sa propre table d'expérience permet aux assureurs de provisionner et de tarifer de la façon la plus adéquate. En effet, la base brute de départ de l'étude est constituée des données issues de la propre gamme de produits de l'assureur, ainsi que de sa propre clientèle. Les probabilités de décès sont alors construites à partir de l'historique du portefeuille ce qui permet de mieux prendre en compte les risques qu'elle supporte.

Du point de vue de l'assureur, on distinguera les tables réglementaires et les tables d'expérience. On distinguera également les tables du moment et les tables prospectives (appelées aussi tables par génération), intégrant l'aspect dynamique de la mortalité.

### 3.1 Présentation des tables d'expériences

Dans cette partie nous allons expliciter les caractéristiques des tables de mortalité. Une table de mortalité est une table qui permet d'étudier le nombre de décès, les probabilités de décès ou de survie et l'espérance de vie selon l'âge et le sexe d'une population donnée.

Il existe deux types de tables de mortalité, les tables de mortalité réglementaires et les tables de mortalité d'expérience.

#### ✓ Tables réglementaires

Les tables de mortalité de la population générale sont établies à partir des données de recensement d'une population, comme la date de naissance, celle du décès .... En France, ces tables sont calculées par l'INSEE (Institut National de la Statistique et des Etudes Economiques). Elles sont généralement établies à l'aide de recensements opérés sur la population et peuvent être utilisées par les assureurs.

#### ✓ Tables d'expériences

Les tables d'expérience sont établies à partir des données historiques d'un portefeuille de prévoyance d'une ou plusieurs sociétés d'assurance (date d'entrée en portefeuille, date de sortie du portefeuille par décès ou résiliation, date de naissance de l'assuré...) ; les données utilisées correspondent donc à un échantillon de la population générale composé uniquement d'assurés.

Les tables d'expériences ont été instaurées en raison de l'écart souvent constaté entre la mortalité de la population assurée et les tables de mortalité de la population générale.

Elles permettent de refléter davantage le risque réel porté par l'assureur puisqu'elles sont construites à partir des données du portefeuille.

La réalisation d'une table d'expérience se fait en 3 étapes :

1. La construction de la table
2. La certification de la table par un actuare agréé et indépendant de la compagnie
3. Le suivi annuel de la table, destiné à prolonger le droit d'utiliser la table établie.

Les deux types de tables rencontrées le plus souvent sont définies ci-dessous. Elles peuvent être par génération ou du moment.

- *Les tables par génération ou tables prospectives*

Les tables de mortalité prospectives sont des tables qui représentent la mortalité d'une population en fonction de l'âge et de l'année de naissance qui représente leur génération : elles permettent d'établir des projections de la mortalité dans le futur.

Elles permettent d'observer les probabilités de décès de personnes de même génération, donc nées la même année : on fixe une année de naissance, et on regarde les vivants, nés cette année-là, à différentes années. Elles sont établies à partir de modèles statistiques estimant les tendances de la mortalité et les extrapolant dans le futur.

Elle possède l'avantage d'être plus représentative de la réalité mais ne peut être construite qu'une fois la génération entièrement éteinte.

- *Les tables du moment*

Les tables de mortalité du moment consistent à estimer les taux de mortalité à un moment donné, pour chaque âge : on fixe un moment, et on regarde les vivants d'âges différents. Elles sont élaborées à l'aide de données relatives à la mortalité du portefeuille des assurés sur une courte période d'observation (1 à 5 ans) sans prendre en compte l'année de naissance des assurés. Les tables de mortalité du moment ne permettent pas de prendre en compte l'évolution de la mortalité des assurés en raison de la période de suivi trop courte.

Les grandes étapes de la construction d'une table peuvent être schématisées comme ceci :

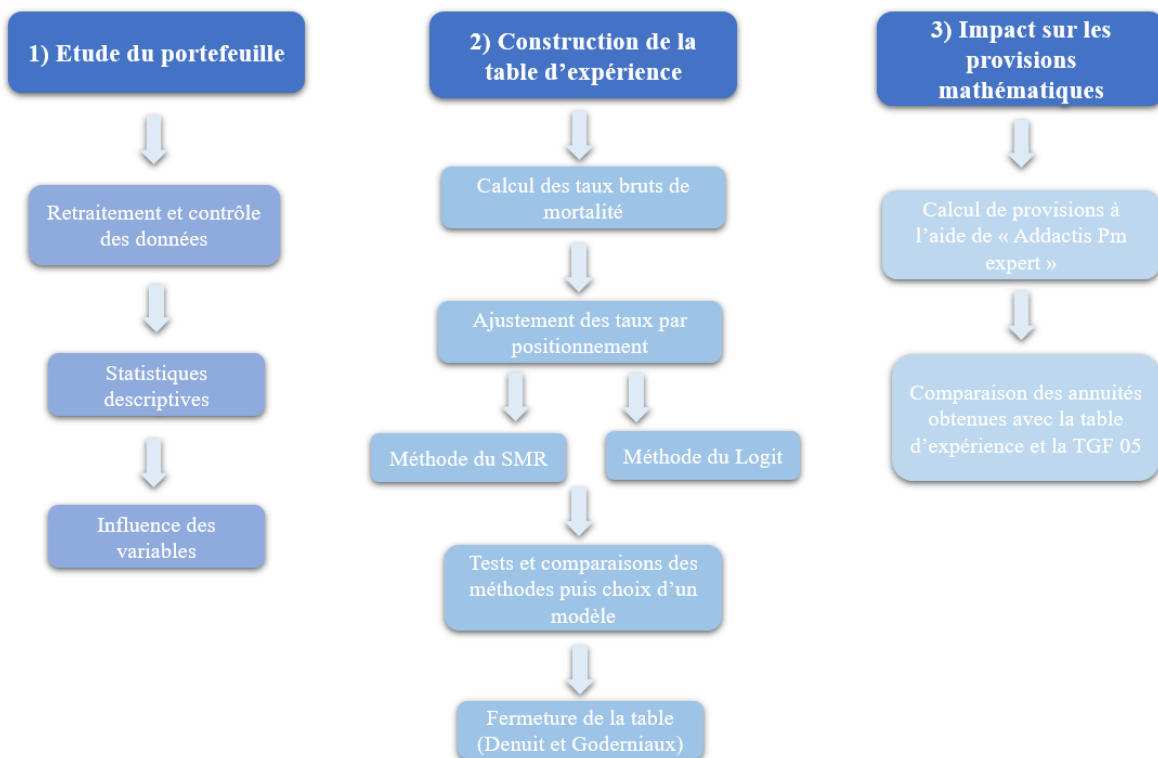


Figure 33 : Etapes de construction d'une table



## 3.2 Outils nécessaires au calcul des taux bruts

L'objectif de l'étude est de modéliser la durée de vie d'un bénéficiaire dans le portefeuille à partir du moment où il commence à percevoir sa rente jusqu'à sa sortie du portefeuille, qui correspond à son décès.

Dans un premier temps, plusieurs modèles seront testés et ensuite le modèle le mieux adapté à notre étude sera choisi. Puis, nous verrons comment passer d'un modèle de durée à une table d'expérience directement exploitable par l'assureur ainsi que les grandes étapes d'une construction de table.

### 3.2.1 Notions de modèles de durée

Le but de cette étude est de construire un modèle exploitable, la partie qui suit permettra de déterminer si le portefeuille rentes de conjoint étudié contient tous les éléments nécessaires à la construction d'une table d'expérience.

#### Les notations :

Le rôle d'une table de mortalité est de fournir, en fonction de l'âge d'une personne, sa probabilité de décéder avant d'atteindre l'âge suivant.

$T_x$  est la variable aléatoire positive qui représente la durée de vie restante d'un individu de la population. Nous définissons une suite de variables aléatoires  $\{T_x = 1, 2, \dots\}$ , où  $T_x$  est la durée de vie restante d'un individu ayant atteint l'âge  $x$ .

$$P[T_x > t] = P[T > x + t | T > x]$$

Lorsque  $x < y$ ,  $P[x + T_x > t] \leq P[y + T_y > t]$  quel que soit  $t > x$ .

$$\text{Car } P[x + T_x > t] = P[T > t | T > x] = \frac{P[T > t]}{P[T > x]} \leq \frac{P[T > t]}{P[T > y]} = P[y + T_y > t]$$

La probabilité pour un individu de vivre  $t$  années baisse avec l'âge.

On note  ${}_tL_x$  le nombre d'individus âgés  $x$  vivant au 1er janvier de l'année calendaire  $t$ .

${}_tD_x$  est le nombre de décès enregistré à l'âge  $x$  durant l'année calendaire  $t$  ou les individus sont exposés au risque de décès. Il s'agit de la durée totale vécue par ces individus durant la période d'observation.

On peut définir les probabilités de survie et de décès par âge d'un assuré. La probabilité de survie d'un assuré notée  ${}_tP_x$  est la probabilité qu'un assuré survive entre l'âge  $x$  et l'âge  $x + t$  alors que la

probabilité de décès notée  ${}_tq_x = 1 - {}_tP_x$  désigne l'évènement contraire c'est-à-dire la probabilité qu'un assuré décède entre l'âge  $x$  et l'âge  $x + t$ .

La probabilité de survie  ${}_tP_x$  est définie par :

$${}_tP_x = P[T_x > t] = P[T_x > x + t | T < x]$$

La probabilité de décès  ${}_tq_x$  est définie par :

$${}_tq_x = 1 - {}_tP_x = P[T_x \leq t] = P[T \leq x + t | T > x]$$

Pour un assuré d'âge  $x$ , l'espérance de vie résiduelle représente le temps moyen à vivre pour ce bénéficiaire. On a :

$$e_x = E(T_x) = \int_{t \geq 0} {}_tP_x dt$$

### 3.2.2 Détermination de la période d'exposition au risque des bénéficiaires

Une étude de survie tient compte :

- Du nombre de bénéficiaires qui entrent dans l'étude (quel que soit le moment de leur entrée),
- Du nombre de sujets qui décèdent pendant la durée de l'étude,
- Du nombre de sujets qui disparaissent de l'étude parce qu'ils sont perdus de vue : on parle de sujets "Censurés".

Il faut donc connaître pour chaque bénéficiaire :

- Ses dates d'entrée et de sortie de l'étude, ou la durée pendant laquelle le sujet a été maintenu dans l'étude,
- La raison ou l'évènement qui a justifié sa sortie de l'étude (décès, censure, sortie pour remariage...) ou s'il est toujours en vie à la fin de l'étude.

Les variables gardées et renommées indispensables à la construction de la loi sont les suivantes :

- Le sexe du bénéficiaire
- La date de naissance du bénéficiaire
- L'âge au terme du contrat (50, 55, 60 pour les rentes temporaires et 120 pour les rentes viagères)
- Le type de rente (temporaire ou viagère)
- Etat du contrat (clos, en cours, ...)
- Motif de clôture
- Date d'entrée
- Date de sortie

✓ Période d'exposition du bénéficiaire

La sélection des dates pour la période d'exposition a été faite afin que la période d'exposition de l'individu soit la plus longue possible.

D'une part, déterminer la date de début de période et la date de fin de période de la rente était indispensable à la création de la table. En effet, la date de début de période que nous souhaitions obtenir correspond à la date de premier versement de rente du bénéficiaire, et la date de fin de période correspondant à la date de fin de versement de la rente.

Or, dans notre base, chaque individu possède plusieurs lignes pour chaque trimestre, semestre... de versement de sa rente. Afin d'obtenir la période d'exposition la plus juste du bénéficiaire, il faut récupérer sa première date de versement de la rente, et sa dernière date d'indemnisation dans le but de ne garder qu'une seule ligne par individu et par rente.

C'est pour cela que dans le code Sas, nous avons sélectionné le MIN de début de période et le MAX de fin de période et donc les bonnes dates souhaitées, le tout en les convertissant au bon format de dates dans SAS. Une vérification a été faite par la suite.

Les dates correspondant à « date de début de période », et à « date de fin de période » correspondent dans la base ex-Humanis à « date de premier règlement » et « date de dernier règlement ». La même méthode que pour la base ex-Malakoff Médéric est utilisée, le Min de « date de premier règlement » est sélectionné, converti en date et nommé « date de début de période » ainsi que le Max de « date de dernier règlement », qui est également converti et nommé « date de fin de période ».

La date d'entrée correspond dans notre étude à la date où le bénéficiaire a commencé à percevoir sa rente tandis que la date de sortie correspond à la sortie du bénéficiaire du portefeuille pour différentes raisons (décès, remariage...), et donc correspond à la date de fin de versement de la rente de conjoint.

Nous cherchons à déterminer la durée d'observation de chaque bénéficiaire dans le portefeuille pour la construction de la loi. Pour cela, il est nécessaire de déterminer de la manière la plus judicieuse la date d'entrée du portefeuille du conjoint survivant ainsi que sa sortie du portefeuille.

✓ Règles appliquées pour le choix des dates d'entrée et de sortie :

• *Date d'entrée :*

- Si la date de début de période et la date d'effet de rente ne sont pas vides, alors on prend le minimum entre les 2. Dans le cas où l'une des deux dates est vide, alors on met la date non vide. (1)

Un graphique a été réalisé afin de voir le nombre d'entrées par date pour voir si des pics n'étaient pas dus à des migrations de données.

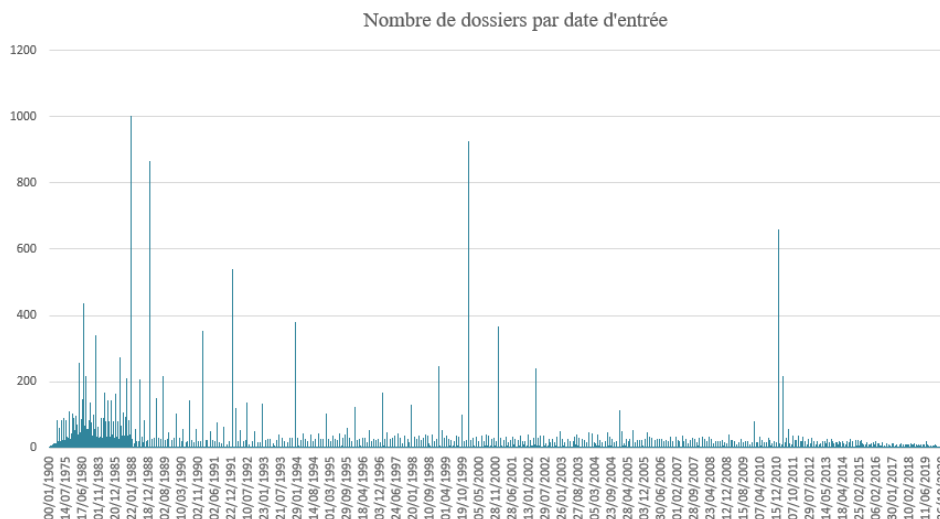


Figure 34 : Nombre de dossiers par date d'entrée

Quatre gros points d'accumulation ont été remarqués sur le graphique :

- Le 01/01/2000
- Le 01/01/1988
- Le 01/01/1989
- Le 01/01/2011

Dans le cas de ces quatre points d'accumulation, la date d'entrée est mise à la date de survenance + 5 jours (une moyenne de 5 jours a été remarquée dans la base de données entre la date de survenance et la date d'effet de rente). Cette règle a alors été appliquée pour les quatre cas de pics de dossiers. Dans le cas où la date d'effet de rente ne ferait pas partie de ces quatre dates, alors on applique la règle énoncée plus haut (1).

Après application de la règle précédente sur la correction des points d'accumulation, nous obtenons le graphique suivant :

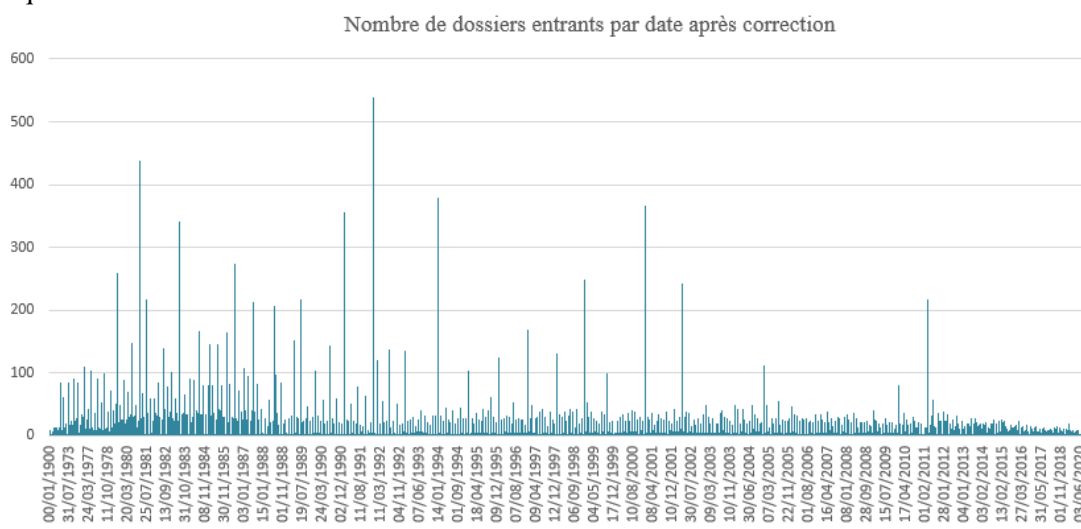


Figure 35 : Nombre de dossiers entrants par date d'entrée après correction

- *Date de sortie* :

Dans le cas où le bénéficiaire est décédé et que sa date de décès est remplie, alors on met la date de son décès comme date de sortie. Sinon, si le motif de clôture est mis à « Décès du bénéficiaire », alors on renseigne la date de clôture du décès si cette dernière est remplie. Dans le cas inverse, on met la date de fin de période comme date de sortie.

✓ Exposition au risque pour chaque individu

L'exposition au risque à l'âge  $x$  représente le temps qu'un bénéficiaire passe dans le portefeuille. Ainsi l'exposition est exprimée en jours par la formule suivante :

$$expo_x^i = \max[0; \min(x, \hat{age}_{sortie}) - \max(x, \hat{age}_{entrée}) \forall x \in [0; 120]$$

Pour chaque âge, l'exposition totale est calculée en sommant l'exposition de chaque individu :

$$expo_x = \sum_i (expo_x^i)$$

Le calcul de l'exposition par âge selon le sexe a été effectué avec l'outil R. Le graphique obtenu est le suivant :

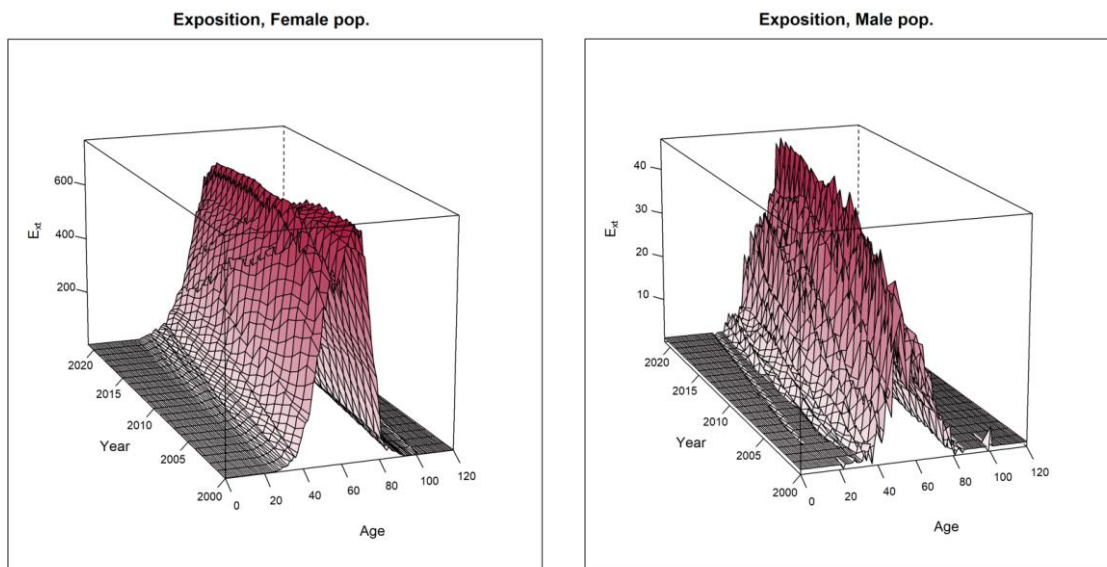


Figure 36 : Exposition par âge des bénéficiaires selon le sexe

La distribution est concentrée entre 30 et 95 ans, avant nous avons très peu d'informations. L'exposition des femmes est bien plus élevée sur tous les âges que celle des hommes, ainsi que beaucoup moins irrégulière (surement en raison du plus grand nombre de femmes dans le portefeuille). Nous nous concentrons donc sur ces âges pour la suite de l'étude (la méthode de Cochran sera appliquée par la suite afin de s'assurer de la validité statistique de cette tranche d'âges).

Le nombre de décès observés sont également tracés sur R :

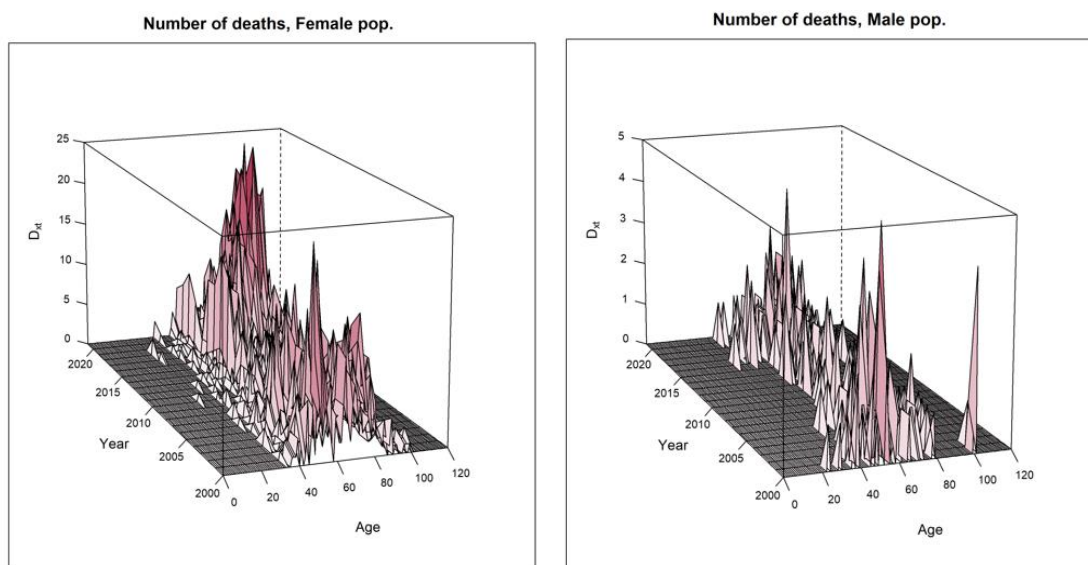


Figure 37 : Nombre de décès par genre

La plupart des sorties pour les femmes sont observées pour la tranche d'âge [40 :90 ans]. L'âge moyen de sortie de la population est de 73 ans. La moyenne d'âge de sortie pour les hommes est de 66 ans contre 74 ans pour les femmes de notre portefeuille. On observe de très grosses irrégularités de taux ainsi qu'un très faible nombre de décès pour les hommes en raison de leur très faible proportion.

Des retraitements ont également été effectués dans R pour la suite de la construction de la loi :

- Les variables *date de naissance du bénéficiaire*, *date d'entrée* et *date de sortie* ont été converties au format date.
- La période d'observation a été définie dans R du 01/01/2000 au 31/12/2021 (le choix de la période d'observation est expliqué dans la partie 3.2.3 de l'étude).
- Les âges d'entrée et de sortie ont été calculés et des nouvelles colonnes ont été rajoutées dans le dataframe de la manière suivante :

$$\begin{cases} \text{âge d'entrée} = \frac{\text{date d'entrée} - \text{date de naissance du bénéficiaire}}{365,25} \\ \text{âge de sortie} = \frac{\text{date de sortie} - \text{date de naissance du bénéficiaire}}{365,25} \end{cases}$$

- Après avoir calculé les âges, des vérifications ont été réalisées comme la positivité des âges d'entrée et de sortie, ainsi que l'âge de sortie supérieur à l'âge d'entrée.

### 3.2.3 Détermination de la période d'observation de l'étude

Plusieurs conditions doivent être réunies afin de déterminer la période d'observation optimale. En effet, un nombre suffisant de données est nécessaire pour que les résultats soient robustes, tout en ayant des générations aux comportements stables dans le temps afin que les résultats soient représentatifs du portefeuille.

Le premier décès de bénéficiaire de la base de données date de 1946. Toutefois, nous ne conservons pas les survenances antérieures à 2000 afin d'éviter de fausser la mortalité sur des générations trop lointaines. Il convient cependant de conserver une période d'observation assez importante pour avoir un nombre suffisant de données afin d'avoir un échantillon représentatif.

Un histogramme est réalisé afin de visualiser si le nombre de dossiers est homogène et afin de nous aider à déterminer une période d'observation judicieuse.

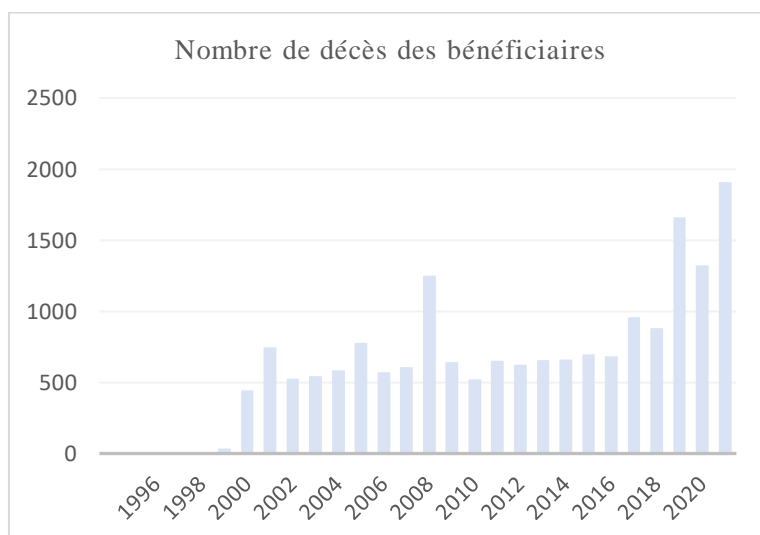


Figure 38 : Nombre de décès des bénéficiaires par année de décès

Sur cet histogramme, un nombre élevé de contrats est remarqué entre les années 2000 à 2021 tandis que le nombre de contrats avant 2000 est très faible.

La volumétrie importante et la représentativité du portefeuille étant indispensables à l'établissement d'un modèle robuste, les années d'observations allant de 2000 à 2021 sont retenues pour la suite de l'étude.

La date de fin de période est choisie au 31 décembre 2021 afin d'avoir des années complètes et de perdre le moins de dossiers possible. La période d'observation retenue est donc de 21 ans.

### 3.2.4 Traitement des censures et troncatures

Le traitement des censures et des troncatures est un sujet important lors de la construction d'une loi de survie. En effet, les données sont souvent incomplètes car nous n'avons pas accès à toute l'information. C'est pour cela que nous avons souvent des données censurées et tronquées. La censure est le phénomène le plus couramment rencontré lors du recueil de données de survie.

Estimer une loi de survie nécessite d'observer la durée de vie des individus d'un échantillon. Cependant, nous ne pouvons pas toujours observer la durée de vie complète d'un individu.

En effet, dans une table d'expérience, nous observons la durée de présence de l'individu dans le portefeuille, ce qui correspond parfois à observer la durée de vie entière d'un assuré, mais aussi quelques fois à observer un minorant de cette durée. La durée exacte de vie n'est donc pas connue. Dans le dernier cas, on parlera de censure ; on dira que l'observation est censurée.

Il existe deux types de censure, la censure à gauche (avant le début de l'observation) et la censure à droite (après la fin de l'observation). Une observation peut être soit censurée à gauche, auquel cas on parlera de troncature, soit censurée à droite, auquel cas on parlera de censure. Le cas mixte peut aussi se présenter, à savoir qu'une observation peut tout à fait être à la fois tronquée et censurée.

Dans notre cas, la censure à droite (censure) correspond aux rentiers sortis durant la période d'observation par une autre cause que le décès ou bien ceux qui sont encore vivants à la fin de cette période d'observation.

La censure à gauche (troncature) signifie que le décès du rentier est survenu avant la période d'observation. On ne les considère donc pas dans l'échantillon.

Dans notre cas, les contrats sont observés entre janvier 2000 et décembre 2021, nous sommes donc confrontés à des censures droites et des troncatures gauches.

Une rente qui a commencé à être versée avant janvier 2000 correspond à une troncature à gauche.

Une rente qui n'a pas fini d'être versée à la fin de la période d'observation soit au 31 décembre 2021 correspond à une censure à droite.



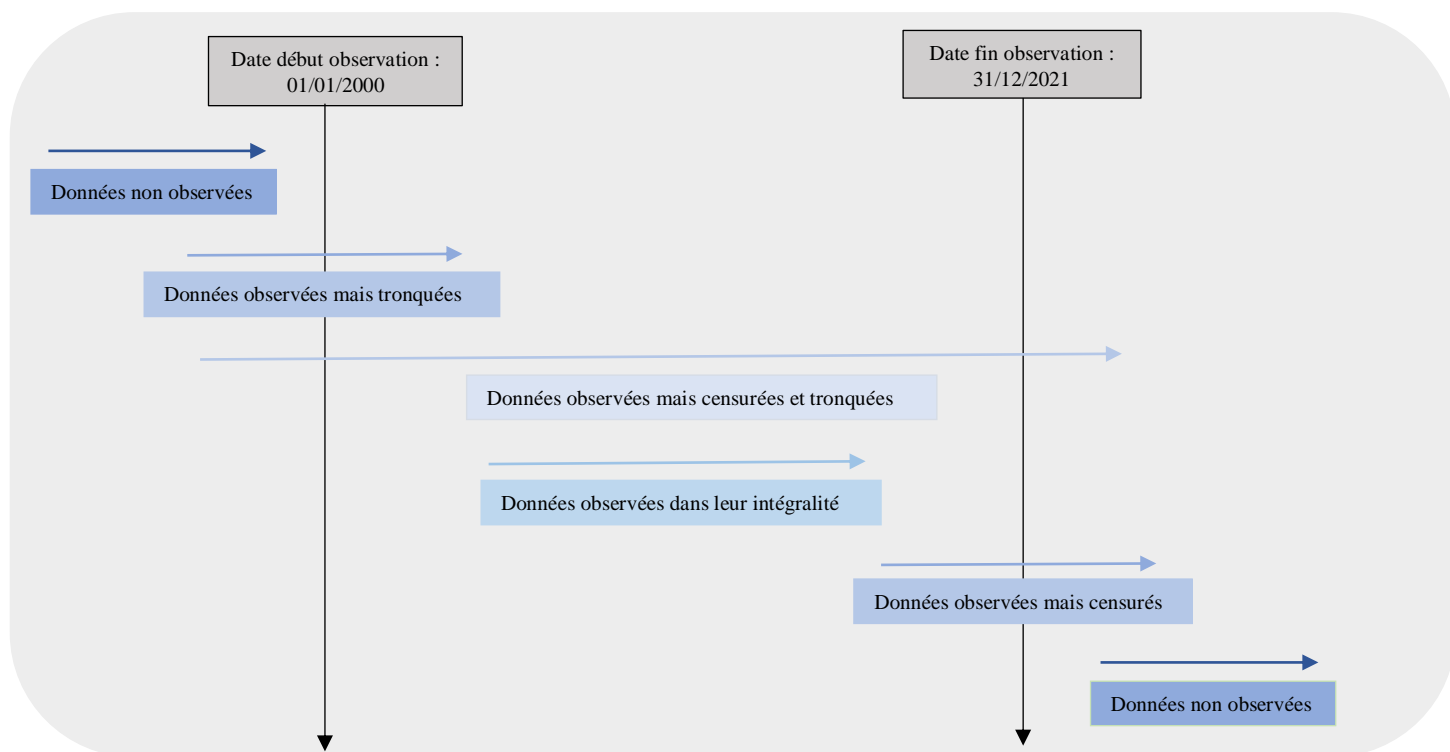


Figure 39 : Schéma des censures et troncatures

### 3.2.5 Concepts de survie

On considère une variable aléatoire  $T$  à valeurs dans  $[0, +\infty[$ .

La durée écoulée  $t$  avant la survenue d'un évènement correspond aux réalisations d'une variable aléatoire  $T$  continue et positive, de fonction de répartition  $F(t)$  et de densité  $f(t)$ . Par définition, la fonction de répartition correspond à :

$$F(t) = \int_0^t f(u) du = P(T \leq t)$$

Et s'interprète comme la probabilité que l'évènement se réalise avant la date  $t$  considérée. La densité  $f(t)$  est définie par

$$f(t) = \lim_{dt \rightarrow 0^+} \frac{P(t \leq T < t + dt)}{dt}$$

Et correspond à la probabilité dite instantanée que l'évènement survienne dans l'intervalle de temps infinitésimal  $[t, t + dt]$ .

#### ✓ Fonction de survie

La fonction de survie donne la probabilité que l'évènement survienne après un instant  $t$ , soit

$$S(t) = 1 - F(t) = P(T > t)$$

Cette fonction est continue décroissante tel que  $S(0) = 1$ .

Pour survivre en  $t_i$ , il faut avoir survécu en  $t_{i-1}, t_{i-2} \dots t_1$ .

#### ✓ La fonction de survie conditionnelle :

La fonction de survie conditionnelle est la survie d'un élément après un instant, sachant qu'il a déjà survécu jusqu'en :  $S_u(t) = P(T > u + t | T > u)$

La survie conditionnelle s'exprime simplement à l'aide de la fonction de survie.

#### ✓ Fonction de hasard

La fonction de hasard évalue le risque que l'évènement survienne à un instant précis, sachant qu'il n'a jamais été observé auparavant.

La fonction de hasard (ou taux de panne, taux de défaillance, taux de décès, risque instantané, etc.) est par définition :

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t)$$

✓ Fonction de hasard cumulé :

Elle est définie par :

$$H(t) = \int_0^t h(s)ds$$

Ainsi, on peut exprimer la fonction de survie par la fonction de hasard de la façon suivante :

Il en résulte directement que la fonction de hasard détermine entièrement la loi de T la relation suivante est obtenue :

$$S(t) = \exp(-H(t))$$

### 3.3 Calcul des taux bruts

Pour construire des tables d'expérience, nous devons d'abord estimer les taux bruts de mortalité de la population à partir de la base de données construite précédemment. Pour cela nous avons plusieurs estimateurs à notre disposition. Une des approches consiste à utiliser une méthode non paramétrique, nous la traiterons dans ce mémoire avec la méthode de Kaplan Meier, ensuite nous testerons un modèle paramétrique, la méthode de Hoem et nous finirons par la méthode de Hoem par génération.

Dans la suite de l'étude, une table sur le risque rente de conjoint sera construite et l'évènement de sortie qui nous intéresse est la sortie du bénéficiaire du portefeuille pour son décès.

Une distribution peut être caractérisée par plusieurs fonctions comme la fonction de hasard, la fonction densité, la fonction de survie.

#### Types de méthodes (paramétriques ou non paramétriques)

Il existe trois catégories de méthodes :

✓ Les méthodes non paramétriques

Les tests non paramétriques ne font aucune hypothèse sur le type de loi de distribution des données. Ils se basent uniquement sur les propriétés numériques des échantillons. Ils peuvent donc être utilisés même si les conditions de validité des tests paramétriques ne sont pas vérifiées.

Ils sont très peu sensibles aux valeurs aberrantes et sont donc conseillés dans ce cas.

Les tests non-paramétriques sont plus robustes que les tests paramétriques. Ils peuvent être utilisés dans un plus grand nombre de situations. De plus, ils sont très faciles à appliquer et à comprendre également, ils ont une complexité plus faible. Cependant, ce type d'approche nécessite une base de données suffisamment conséquente.

✓ Les méthodes paramétriques

Les méthodes paramétriques fonctionnent en supposant que les données que l'on a à disposition suivent un type de loi de distribution connu. Par conséquent, certaines conditions de validité doivent être vérifiées pour que le résultat d'un test paramétrique soit fiable.

L'estimation se réduit alors à estimer les paramètres de ce modèle.

Ces tests sont en général très fins, mais ils nécessitent que les données suivent effectivement la loi de distribution supposée. Ils sont en particulier très sensibles aux valeurs aberrantes et ne sont pas conseillés si des valeurs aberrantes sont détectées.

Les tests paramétriques sont, eux, plus puissants en général que leurs équivalents non-paramétriques. Autrement dit, un test paramétrique sera plus apte à aboutir à un rejet de  $H_0$ , si ce rejet est justifié. La plupart du temps, la p-value calculée par un test paramétrique sera inférieure à la p-value calculée par un équivalent non-paramétrique exécuté sur les mêmes données.

✓ Les méthodes semi-paramétriques

Pour modéliser la fonction de survie, on souhaite faire appel à l'estimation de paramètres mais aussi à une estimation dite non paramétrique, ce qui permet un ajustement propre aux données. Les modèles «

à hasard proportionnel » sont des modèles semi-paramétriques, ils sont appelés ainsi car la fonction de hasard estimée est proportionnelle à une fonction de hasard « de référence ».

### **Segmentation Homme/ Femme :**

Précédemment, des tests d'homogénéité des variables ont été réalisés ainsi que l'évaluation de leur influence. La segmentation homme femme a été testée par le test du Log-rank. Cependant, au vu de la faible proportion d'hommes dans le portefeuille, seulement une table femme sera réalisée dans ce mémoire.

#### 3.3.1 Tranche d'âges retenue pour les taux bruts

Précédemment, les taux bruts ont été calculés pour les âges compris entre 20 et 120 ans afin d'avoir une vision globale. Nous allons maintenant utiliser le critère de Cochran afin de déterminer la taille de la population qui est statistiquement la plus significative pour le calcul des taux bruts.

- Critère de Cochran :

- **Critère 1 :**

Le nombre des décès en l'âge  $x$  doit être une valeur théorique non nulle :

$$E_x \hat{q}_x \geq 5$$

- **Critère 2 :**

Et le nombre des décès doit avoir une valeur théorique supérieure ou égale à 5 :

$$E_x(1 - \hat{q}_x) \geq 5$$

- Critère d'exposition

L'exposition doit être supérieure à 2000. **Critère 3 :**  $E_x \geq 2000$

Avec  $E_x$  l'exposition au risque de l'individu à l'âge  $x$  et  $\hat{q}_x$  l'estimation du taux de mortalité empirique  
 $\hat{q}_x = \frac{d_x}{N_x}$ .

Les critères ont été appliqués à notre portefeuille (voir Annexe A).

Les intervalles d'âges validés sont :

- Par le critère 1 : [35,100]
- Par le critère 2 : [20,110]
- Par le critère 3 : [20 ;108]

On retient donc la tranche d'âge : [35, 100] ans qui est la tranche validée par les trois critères.

La mise en place des notions et du périmètre de travail était nécessaire avant d'entrer dans la pratique. À présent, la première étape de la construction d'une table de mortalité va être réalisée : l'estimation des taux de mortalité.

### 3.3.2 Méthode de Kaplan Meier et application aux données

#### 3.3.2.1 *Présentation de l'estimateur*

Publié en 1958 par Edward L. Kaplan et Paul Meier, l'estimateur de Kaplan-Meier, également connu sous le nom d'estimateur produit-limite, est un estimateur du maximum de vraisemblance visant à estimer la fonction de survie. L'avantage de cet estimateur est qu'il permet de tenir compte des données censurées. Le volume des censures étant conséquent, ignorer ces contrats représenterait une grande perte d'informations.

L'estimateur de Kaplan Meier est un estimateur non paramétrique, ce qui signifie qu'il n'est pas nécessaire d'adopter une quelconque spécification de loi pour estimer la fonction de survie.

Aucune hypothèse n'est émise à priori sur la distribution de la loi de survie, cette loi sera entièrement déterminée à partir de nos données. Cela nécessite une base de données suffisamment robuste.

Une distribution peut être caractérisée par plusieurs fonctions comme la fonction de hasard, la fonction densité, la fonction de survie.

C'est un estimateur pertinent lorsque l'intervalle de temps pris en compte est petit relativement à la vitesse de variation de la fonction de survie. Pour notre portefeuille, le nombre important de contrats permet d'avoir un calcul précis de la fonction de survie, ainsi que notre connaissance des âges d'entrée et de sortie sur la période d'observation choisie (2000-2021).

La limite de l'estimateur de Kaplan Meier est qu'il ne prend pas en compte les variables qualitatives de l'étude qui peuvent être significatives. Cette limite n'est pas vraiment une entrave à notre étude car nous étudions la population globale de notre portefeuille pour cette partie de l'étude.

#### ✓ Principe de l'estimateur

On peut définir le terme "durée de survie" comme le temps qui s'écoule jusqu'à la survenance d'un événement particulier, ici le décès du bénéficiaire, ou plus particulièrement l'arrêt de versement de la rente de conjoint.

Si  $T$ , variable aléatoire, représente la durée écoulée depuis un instant  $t_0$  pour chaque individu avant qu'il n'ait connu l'échéance de l'événement alors :

$$S(t) = P(T > t)$$

D'après la formule de Bayes, pour tout  $t_m > t_{m-1} > \dots > t_2 > t_1$  la fonction de survie peut alors s'écrire :

$$\begin{aligned} S_x(t_m) &= P[T > t_m | T > t_{m-1}] S(t_{m-1}) \\ &= P[T > t_m | T > t_{m-1}] \cdot P(T > t_{m-1} | T > t_{m-2}) \cdot S(t_{m-2}) \\ &= \dots \\ &= [T > t_m | T > t_{m-1}] \dots P(T > t_2 | T > t_1) \cdot S(t_1) \end{aligned}$$

Avec  $T$  l'âge à la survenance du sinistre.

En choisissant comme instants de conditionnement les âges  $t_j \in [x, x + 1[$  où se produisent des évènements, avec  $t(0) = x$ , cela revient alors à estimer les probabilités :

$$p_j = P(T \geq t_{(j+1)} | T > t_{(j)}) \text{ et } i_j = P(T < t_{(j+1)} | T > t_{(j)})$$

Avec  $p_j$  la probabilité pour un individu de rester vivant sur l'intervalle d'âge  $[t_j; t_{j+1}]$  sachant qu'il est vivant à l'âge  $t_j$ . Et  $i_j = 1 - p_j$  la probabilité de sortir du portefeuille sur l'intervalle d'âge  $[t_j; t_{j+1}]$  sachant qu'on est valide à l'âge  $t_j$ .

La fonction de survie au point  $t$  pour un âge  $x$  :

$$S_x(t) = p_0 p_1 \dots p_{t-1} = \prod_{j | t_{(j)} < x} (1 - i_j)$$

#### ✓ Expression de l'estimateur

Les  $i_j = 1 - p_j$  sont alors estimés par la méthode du maximum de vraisemblance à partir de l'observation d'un  $n$ -échantillon.

En notant,  $d_j$  le nombre d'évènements à la date  $t_j$  et  $n_j$  le nombre de personnes encore à risque juste avant  $t_j$  (i.e. qui sont toujours suivies, et donc notamment toujours non censurées),

L'estimateur de Kaplan-Meier (parfois appelé estimateur produit limite) est donné par :

$$\hat{S}(t) = \prod_{j: t_{(j)} \leq t} \left(1 - \frac{d_j}{n_j}\right)$$

Il est possible de déterminer la variance de l'estimateur  $\hat{S}(t)$  pour préciser l'estimation  $S(t)$ . La formule la plus utilisée pour déterminer la variance de l'estimateur est la variance de Greenwood. Ce dernier est un estimateur consistant de la variance. Il se calcule comme suit :

$$\hat{\sigma}_{KM}^2(t) = \hat{S}_{KM}^2(t) \cdot \sum_{j=1}^t \frac{d_j}{n_j(n_j - d_j)}$$

Où  $d_j$  et  $n_j$  représentent respectivement le nombre de décès et le nombre d'individus sous risque à l'âge  $j$ .

L'estimateur de Kaplan-Meier est asymptotiquement gaussien et centré sur  $S(t)$ . On peut donc déterminer, en reprenant les notations précédentes un intervalle de confiance asymptotique de la survie par la formule suivante :

$$S_t \left( 1 \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\sum \frac{d_j}{n_j(n_j - d_j)}} \right)$$

Où  $z_{\frac{\alpha}{2}}$  représente le quantile d'ordre  $\frac{\alpha}{2}$  d'une gaussienne centrée réduite.

### 3.3.2.2 Application aux données

#### ✓ Calcul de l'estimateur

Pour le calcul de cet estimateur, le package Survival a été utilisé sur R. La fonction *survfit()* permet une analyse par génération de la survie des individus présents dans le portefeuille grâce à la fonction *Surv()* qui dépend du « temps » et du « statut » qui jouent un rôle important dans l'analyse.

La sortie du portefeuille est due au décès du bénéficiaire. Les censures sont également à déterminer au préalable. Les patients encore dans le portefeuille à la date de fin d'observation, c'est-à-dire ici le 31/12/2021, sont alors mis au statut de censure.

L'estimateur de Kaplan Meier nécessitant des intervalles de temps petits, nous avons donc estimé la fonction de survie par ancienneté avec un pas en jour.

La fonction *survfit()* permet de créer des courbes de survie et d'imprimer le nombre de valeurs, le nombre d'évènement, le temps médian et l'intervalle de confiance à 95%.

Une fois la fonction de survie estimée, les taux bruts de mortalité à chaque âge  $x$  peuvent être déterminés :

$$\hat{q}(x) = 1 - \hat{S}(x) = 1 - \prod_{T_i \leq x} \left( 1 - \frac{d_i}{L_i} \right)$$



Pour chaque âge  $x$  nous calculons en pratique les taux bruts de mortalité par la formule :

$$\hat{q}(x) = P[T \leq x + 1 | T > x] = 1 - \frac{S(x + 1)}{S(x)}$$

A la vue des statistiques descriptives établies précédemment sur la période d'observation, les taux bruts seront estimés seulement pour les âges compris entre 35 ans et 100 ans en raison du manque de données sur la sinistralité du portefeuille pour les âges inférieurs à 35 ans et supérieurs à 100 ans.

De plus, les effectifs sinistrés sont insuffisants pour calculer directement les taux bruts avant 35 ans et après 100 ans. Le choix d'estimer les taux bruts seulement entre 35 et 100 ans a donc été fait.

Après l'étude précédente de l'influence de nos variables, la courbe des taux bruts de Kaplan Meier n'est tracée que pour les femmes en ne distinguant pas le type de rente.

La courbe des taux bruts estimés a été tracée ci-dessous. Ces taux sont déterminés pour la période allant du 01/01/2000 au 31/12/2021 pour les âges compris entre 35 ans et 100 ans.

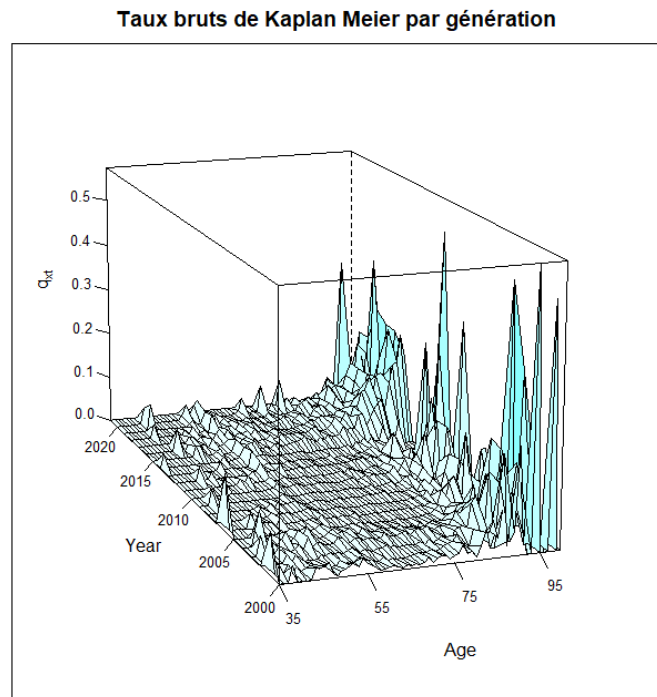


Figure 40 : Taux bruts par âge estimés par Kaplan-Meier

Graphiquement, on remarque une croissance des taux avec l'âge : plus l'assuré vieillit, plus la probabilité qu'il décède et donc qu'il sorte du portefeuille augmente. Les taux sont relativement bas jusqu'à 80 ans. De nombreux pics avec de nombreuses irrégularités sont alors remarqués entre 80 et 100 ans pour toutes les générations observées.

On remarque également des fluctuations qui surviennent à partir de 90 ans, ce qui est expliqué par le volume assez faible de données disponibles pour les âges élevés.

### 3.3.3 Estimateur des moments de Hoem

#### 3.3.3.1 *Présentation de l'estimateur*

Contrairement à l'estimateur de Kaplan Meier, l'estimateur des moments de Hoem est un estimateur paramétrique qui généralise l'estimateur binomial. C'est un estimateur qui prend en compte les censures et les troncatures.

Cet estimateur se base sur l'exposition d'une personne au risque à l'âge  $x$  sur un intervalle de temps qui correspond à la durée d'observation d'une personne, de sa date de début d'observation à sa date de fin d'observation.

L'exposition au risque à l'âge  $x$  pour l'année  $N$  représente donc le temps qu'un assuré passe dans le portefeuille à l'âge  $x$  durant l'année  $N$ . L'exposition est donc une variable qui s'exprime en années.

#### 1. Principe de l'estimateur :

On note :

- $n_x$  : le nombre d'individus dans le portefeuille d'âge  $x$
- $D_x$  : la variable aléatoire représentant le nombre de décès observés sur  $][x, x + 1]$
- $X_k$  variables aléatoires de Bernoulli
- $[a_k, b_k]$  représente l'intervalle de temps inclus dans  $[x, x + 1]$  pour lequel le bénéficiaire  $k$  est sous observation ; avec  $a_k$  la date de début d'observation du bénéficiaire à l'âge  $x$  et  $b_k$  la date de fin d'observation du bénéficiaire pour l'âge  $x$ .

L'estimateur des moments de Hoem se construit en considérant que chaque bénéficiaire est exposé au risque de décès à l'âge  $x$  sur l'intervalle  $[a_k, b_k] \subset [x, x + 1]$

On fait les hypothèses suivantes :

- La probabilité d'un individu vivant à l'âge  $x$  de sortir du portefeuille dans l'intervalle  $[x, x+1]$  est  $(b_k - a_k)q_x$
- Soit  $D_x$ , une variable aléatoire comptabilisant le nombre de sorties sur l'intervalle  $[x, x+1]$  et  $X_1 \dots X_{n_x}$  de loi de Bernoulli de paramètre  $(b_k - a_k)q_{x+a_k}$ .
- La probabilité de sortie du portefeuille est une fonction linéaire du temps, ainsi il en découle que :

$$E(X_k) = (b_k - a_k)q_x$$

Et donc :

$$D_x = \sum_{k=1}^{n_x} X_k$$

## 2. Construction de l'estimateur :

On suppose les  $X_1 \dots X_{n_x}$  indépendants.  $D_x$  suit une loi Binomiale  $\beta(n_x, b_k - a_k i_{x+a_k})$ .

Finalement :

$$E(D_x) = E\left(\sum_{k=1}^{n_x} X_k\right) = \sum_{k=1}^{n_x} E(X_k) = \sum_{k=1}^{n_x} (b_k - a_k)q_k$$

Donc :

$$q_x = \frac{E(D_x)}{\sum_{k=1}^{n_x} (b_k - a_k)}$$

En notant  $d_x$  une réalisation de  $D_x$ , et en appliquant la loi des grands nombres, on obtient :

$$\hat{q}_x = \frac{d_x}{\sum_{k=1}^{n_x} (b_k - a_k)}$$

## 3. Estimateur de Hoem :

L'estimateur de Hoem est un estimateur sans biais :

$$E(\hat{q}_x) = \frac{E(D_x)}{\sum_{k=1}^{n_x} (b_k - a_k)} = q_x$$

Si les hypothèses paramétriques adoptées sont proches de la loi modélisée, alors les estimations paramétriques obtenues avec l'estimateur de HOEM sont les mieux adaptées.

## 4. Détermination des intervalles de confiance :

D'après le Théorème Central Limite,

$$n_x \frac{\hat{q}_x - q_x}{\sqrt{n_x q_x (1 - q_x)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Soit, d'après le théorème de Slutsky :

$$\sqrt{n_x} \frac{\hat{q}_x - q_x}{\sqrt{\hat{q}_x (1 - \hat{q}_x)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

On obtient donc :

$$\hat{q}_x \xrightarrow{L} \mathcal{N} \left( q_x, \sqrt{\frac{q_x(1-q_x)}{n_x}} \right)$$

Les bornes de l'intervalle de confiance asymptotique de niveau  $1 - \alpha$  sont donc données par :

$$\hat{q}_x \pm u_{1-\frac{\alpha}{2}} \sqrt{\frac{q_x(1-q_x)}{n_x}}$$

Seulement l'écart type  $\sqrt{\frac{q_x(1-q_x)}{n_x}}$  est inconnu. Il est donc remplacé par l'écart type estimé.

Si l'on suppose que  $n_x$  est suffisamment grand, et en désignant  $u_{\frac{\alpha}{2}}$ , le quantile d'ordre  $\alpha$  d'une loi normale centrée réduite, nous en déduisons alors l'intervalle de confiance asymptotique de niveau  $\alpha$  suivant :

$$IC_{\alpha} = \left[ \hat{q}_x - u_{\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{n_x}}, \hat{q}_x + u_{\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{n_x}} \right]$$

### 3.3.3.2 Application aux données

La courbe des taux bruts a été obtenue en implémentant sous le logiciel R une fonction qui calcule l'exposition par année d'exercice à l'aide de la formule détaillée plus haut.

La figure ci-dessous présente les taux bruts estimés par la méthode des moments de Hoem :

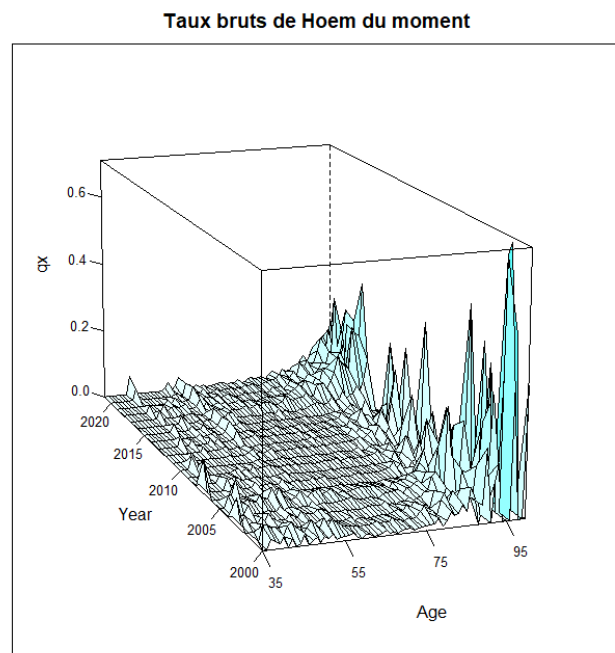


Figure 41 : Taux bruts par âge estimés par la méthode d'Hoem du moment

Tout comme l'estimateur de Kaplan-Meier, on remarque une croissance globale des taux avec l'âge.

On retrouve également une forte irrégularité des taux de mortalité pour les âges élevés. Ces irrégularités sont probablement dues à la volatilité des décès due au faible nombre de données pour les âges élevés.

Il est également intéressant d'utiliser l'estimateur de Hoem par génération car il prend en compte la génération dans le calcul des taux bruts. Il est basé sur la même approche que l'estimateur de Hoem des moments, mais le principe de calcul est basé sur une stratification des données par génération.

Les taux de mortalité par âge sont estimés en rapportant les décès sur les expositions pour chaque génération présente dans la base de données.

Le graphique obtenu est le suivant :

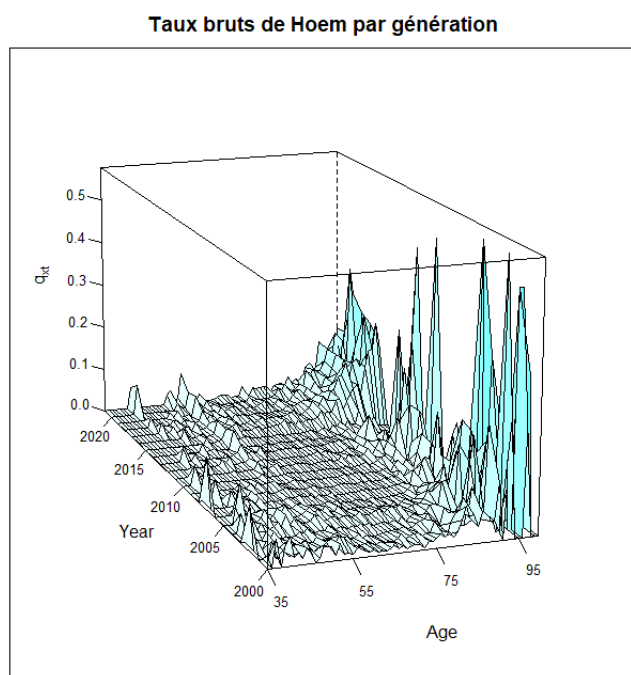


Figure 42 : Taux bruts de Hoem par âge par génération

On constate toujours une forte volatilité des taux bruts aux âges élevés liée à l'insuffisance de données et la forte variabilité du nombre de décès. La tendance globale reste inchangée.

### 3.4 Comparaison des différents modèles

Les taux bruts obtenus présentent des irrégularités importantes qui ne permettent pas d'avoir une tendance de la mortalité cohérente. Avant l'étape de l'ajustement des taux, il est nécessaire de comparer les différents modèles établis et de choisir celui qui illustre le mieux la mortalité de la manière la plus prudente possible.

Graphiquement, les estimateurs de Kaplan Meier et de Hoem par génération ont l'air très proches. Même pour les âges élevés, les mêmes pics de mortalités sont observés. Le graphique des taux bruts de Hoem des moments présente des taux légèrement plus bas, c'est-à-dire une mortalité plus faible, plus principalement pour les âges élevés.

À partir de ces observations, il serait plus prudent de retenir la méthode des moments de Hoem, car en comparaison des autres méthodes, elle nous donne un taux de mortalité plus faible.

Une comparaison des taux a ainsi été réalisée sur les âges [60,100] où on estime que les données sont les plus significatives afin de vérifier cette hypothèse.

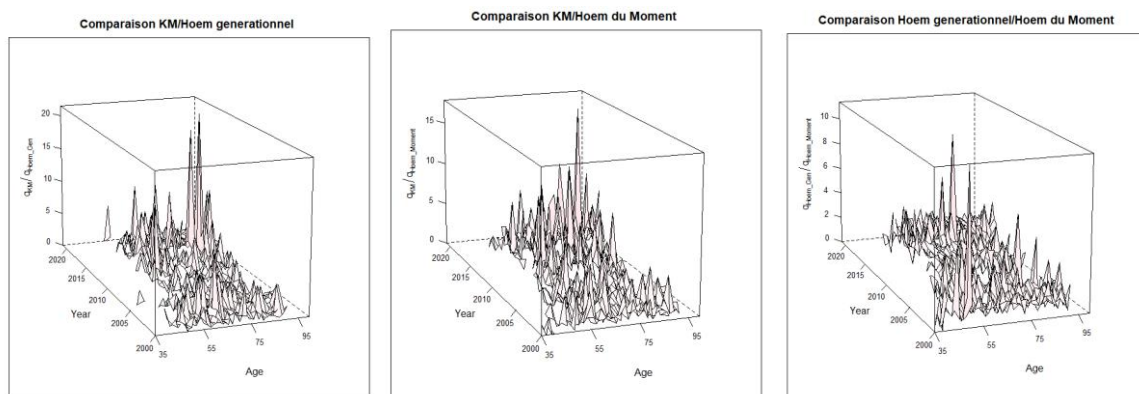


Figure 43 : Comparaison des taux bruts de mortalité

Les ratios moyens annuels sur une moyenne des années d'observations entre les différents modèles ont été rapportés :

	$\frac{KM}{Hoem\ Générationnels}$	$\frac{KM}{Hoem\ du\ Moment}$	$\frac{Hoem\ générationnels}{Hoem\ du\ Moment}$
Ratio moyen :	1.19	1.20	1.08

Tableau 5 : Ratios moyens annuels sur une moyenne des années d'observations des différents modèles

A l'issue de ces comparaisons, le calcul des ratios moyens annuels donne des résultats supérieurs à 1 ce qui signifie que les taux de Kaplan Meier sont en moyenne plus élevés que ceux de Hoem générationnels et que ceux des moments de Hoem. Le troisième ratio moyen annuel montre également que les taux bruts de Hoem générationnels sont en moyenne plus élevés que ceux des moments de Hoem.

Compte tenu du contexte de risque de longévité, on retiendra donc l'estimateur de Hoem des moments qui s'avère plus prudent que les deux autres car les taux bruts estimés sont plus faibles en moyenne que ceux obtenus par les deux autres méthodes.

## 4 Ajustement des taux bruts

Après l'estimation des taux bruts et la sélection de la méthode la plus fiable, il est important de vérifier que les taux bruts reflètent de la manière la plus fidèle le phénomène de mortalité observé. Les taux peuvent représenter de grandes irrégularités, en partie dues à une insuffisance de données. Il est ainsi nécessaire d'ajuster les taux obtenus.

Il existe deux types d'approches pour l'ajustement des taux bruts, l'approche intrinsèque et l'approche par positionnement. Dans ce mémoire, seules les méthodes d'ajustements par positionnement seront utilisées en raison du manque de données disponibles en termes de volume, ce qui ne nous permet pas d'avoir une approche directe de modèles intrinsèques.

L'utilisation d'une référence externe consiste à rechercher un « positionnement » de la table d'expérience par rapport à une table de référence donnée. Dans ce chapitre seront présentées deux méthodes qui permettent d'ajuster la table afin de se rapprocher au plus de la réalité. Des indicateurs de validation seront ensuite utilisés afin de comparer les différentes méthodes et de choisir la plus adaptée, c'est-à-dire l'ajustement le plus prudent et le plus fidèle à nos données.

En raison du faible nombre de données dans le portefeuille étudié, les méthodes de lissages classiques ne peuvent pas être appliquées. Il est ainsi nécessaire de positionner la mortalité du portefeuille par rapport à une table de référence.

La table de référence utilisée est une table de mortalité prospective proposée par l'Institut des actuaires et représentative de la mortalité de la population générale des rentiers. Cette table fournit des taux de mortalité sur chaque année allant de 2007 à 2060 pour l'intervalle d'âges [30-130]ans. On peut donc supposer que la mortalité du groupe étudié dans ce mémoire sera proche de la mortalité représentée par la table réglementaire.

Le taux de référence sur lesquels nous allons positionner la mortalité du portefeuille est défini ci-dessous :

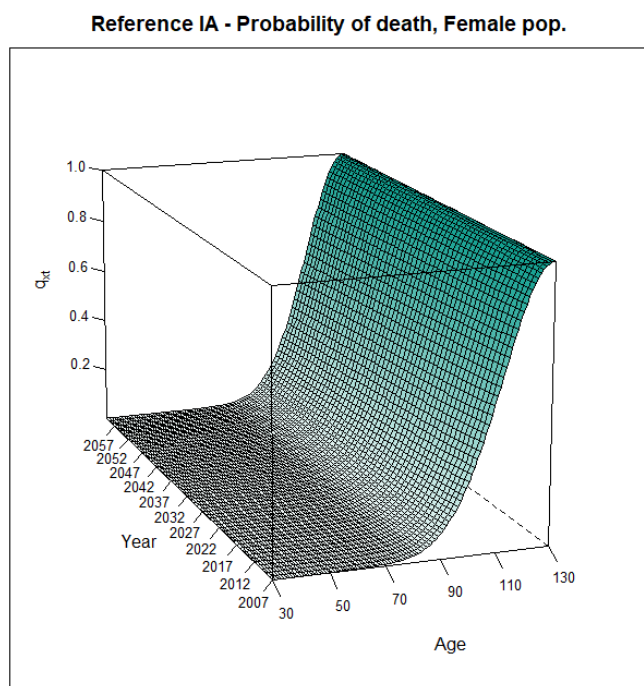


Figure 44 : Taux bruts de mortalité de la table de référence

Ce graphique représente les taux lissés de 30 à 130 ans sur les années 2007 à 2060. Ces taux de mortalité ont été estimés sur la base d'une population de rentiers issus de plusieurs organismes assureurs.

L'intervalle retenu pour l'ajustement est [2007 ;2021], constitué des années en commun avec notre période d'observation et la plage d'année de la table de référence.

La table prospective finale contiendra donc les taux ajustés sur l'intervalle 2007-2021 puis extrapolés sur l'horizon 2022-2060. Nous utilisons les fonctions du package ELT de R pour appliquer les méthodes d'ajustement par positionnement.

Par la suite seront présentées deux méthodes par positionnement, la méthode du SMR ainsi que la méthode du Logit qui seront ensuite comparées à l'aide de plusieurs critères afin de déterminer quelle méthode est la plus adaptée à notre portefeuille.

#### 4.1 Méthode du SMR ( Standardized Mortality Ratio)

L'approche à un paramètre du SMR est la méthode la plus simple envisagée. Elle consiste à appliquer un seul coefficient de réduction/majoration aux probabilités conditionnelles de décès de la référence, de manière multiplicative aux probabilités conditionnelles de décès de la référence. Ce coefficient est appelé le SMR pour la population considérée.

Ainsi, on obtient les probabilités de décès d'expérience par :

$$q_x = SMR * q_{x*}^{ref} \text{ avec } SMR = \frac{D}{E} = \frac{\sum_{x=1}^p D_x}{\sum_{x=1}^p E_x q_{x*}^{ref}}$$

où  $q_{x*}^{ref}$ ,  $D_x$  et  $E_x$  dénotent respectivement la probabilité de décès, le nombre de décès et l'exposition au risque à l'âge  $x$  durant l'année calendaire  $t$ .

La fonction Method1() a été mise en oeuvre sur R afin d'effectuer un positionnement de nos données d'expérience par la méthode du SMR.

La plage d'âges est un choix très important car la table de mortalité qui sera construite par positionnement n'aura de validité que sur la plage d'âges que nous aurons choisie.

C'est pourquoi nous appliquons le calcul du SMR à différentes tranches d'âges afin de sélectionner celle qui semble la plus pertinente pour la suite :

Le SMR obtenu est le suivant:

Plage d'âges	35-100	40-100	35-90	40-95
SMR	0.9985	0.9972	1.03	1.004

Tableau 6 : Résultats obtenus par tranche d'âges pour la méthode du SMR

Lorsque le SMR est <1, la mortalité du portefeuille est inférieure à celle de la population générale des tables de référence. Les tables de références sur estiment donc la mortalité du portefeuille. Dans le cas inverse, la mortalité du portefeuille est supérieure à celle des tables de référence.



Plus la valeur du SMR se rapproche de 1, plus la mortalité de la population étudiée se rapproche de celle prédite par la table de référence. Du fait d'une exposition plus large ainsi qu'un SMR proche de 1, nous retenons la plage d'âges [35-100]ans.

Les coefficients obtenus nous permettent d'extrapoler les taux de mortalité aux années postérieures à 2019 jusqu'en 2060 grâce à la relation  $q_x = SMR * q_{x^*}^{ref}$ . La courbe obtenue a l'allure suivante :

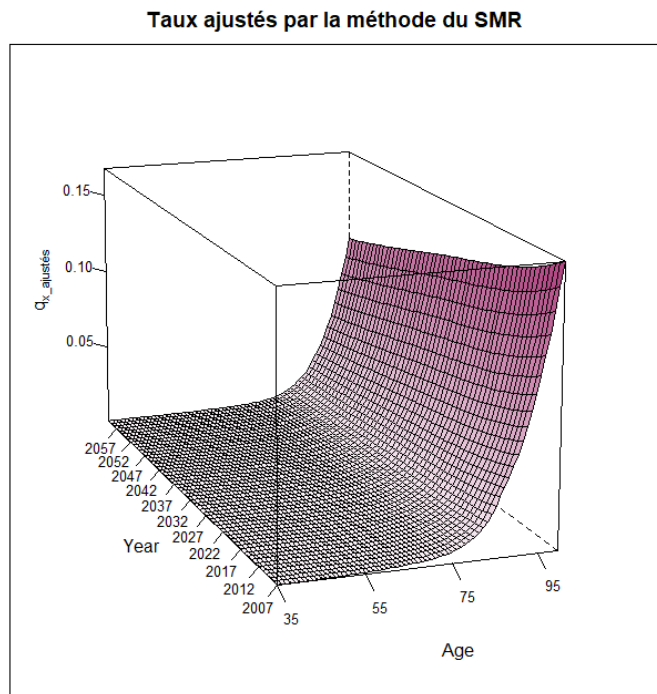


Figure 45 : Taux ajustés et extrapolés par la méthode du SMR

Cette courbe correspond aux taux estimés sur la plage [35 ;100]. Les taux sont quasi nuls jusqu'à 75 ans puis une hausse brutale de la mortalité jusqu'à 100 ans. On observe également une tendance à la baisse de la mortalité avec le temps.

## 4.2 Méthode de regression linéaire (ou du Logit)

La deuxième méthode est une variante semi-paramétrique du modèle de Brass (1971). Le positionnement est effectué à l'aide de la fonction logistique selon le modèle de regression suivant :

$$\text{logit}(q_x) = \alpha + \beta \cdot \text{logit}(q_{x^*}^{ref}) + \varepsilon_{x^*}$$

où  $x^*$  correspond à la plage d'âges commune avec la référence de mortalité,  $\varepsilon_{x^*}$ , est le terme d'erreur et  $q_{x^*}^{ref}$  est la référence de mortalité. Le paramètre  $\alpha$  est un indicateur de mortalité qui affecte tous les âges alors que le paramètre  $\beta$  modifie cet effet avec l'âge.

Les logits des probabilités conditionnelles de décès sont ajustées sur les logits des probabilités de décès de la table de référence.

L'estimation des paramètres est effectuée par une minimisation de la distance pondérée par l'exposition entre les forces de mortalité observées et estimées :

$$\sum |E_{x^*} \times (\hat{q}_{x^*} - \tilde{q}_{x^*})|$$

Enfin, on obtient les probabilités conditionnelles de décès par la formule suivante :

$$\tilde{q}_x = \frac{\exp(\alpha + \beta \cdot \text{logit}(q_x^{ref}))}{1 + \exp(\alpha + \beta \cdot \text{logit}(q_x^{ref}))}$$

Pour l'exécution de cette méthode, la fonction Method2() a été utilisée sur R, en spécifiant la tranche d'âge concernée [35 ;100].

L'application de cette méthode sur la plage d'âges [35-100] ans comme déterminé à la méthode précédente donne les coefficients suivants :

Coefficients	$\alpha$	$\beta$
<b>Population</b>	-0.91	0.92

Tableau 7 : Résultats des coefficients obtenus pour la méthode du Logit

Les taux ajustés et extrapolés obtenus par la méthode des Logits sont représentés sur le graphique suivant :

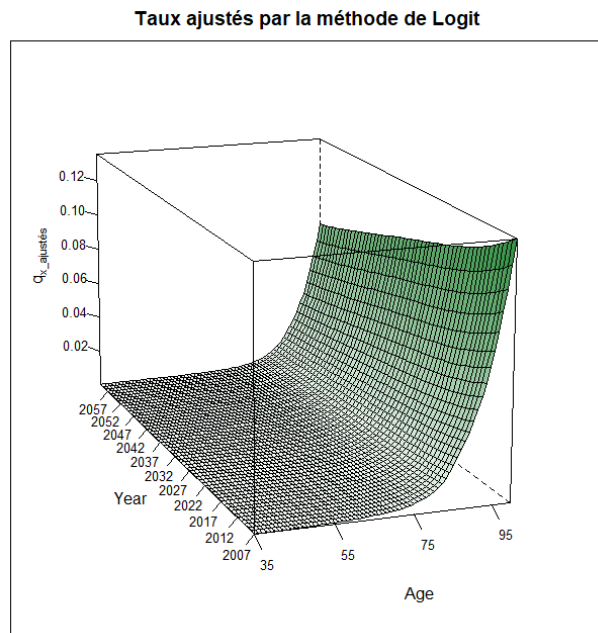


Figure 46 : Taux ajustés et extrapolés par la méthode de la régression logistique

Comme pour la méthode du SMR, on observe une baisse progressive de la mortalité avec le temps.

### 4.3 Comparaison des ajustements obtenus par les différentes méthodes

Après avoir calibré différents modèles, il est important de vérifier la validité des ajustements par des indicateurs de validation qui permettront de vérifier que les taux ajustés ne sont pas trop loin des taux bruts.

Il est également nécessaire de comparer les différents modèles et de déterminer l'ajustement le plus pertinent pour notre étude en se basant sur la cohérence de la table ainsi que de sa future utilisation tout en respectant la fidélité des données du portefeuille. Le tout afin de valider un ajustement pour la construction de la table.

Une analyse graphique n'est pas suffisante, et il est important d'utiliser un nombre d'indicateurs et de tests statistiques pour retenir la meilleure méthode d'ajustement.

Plusieurs critères sont utilisés afin de sélectionner le meilleur ajustement :

- Critère de proximité ;
- Critère de régularité ;
- Critère de cohérence, qui privilégie l'ajustement donnant la mortalité la plus faible en tenant compte du risque de longévité.

#### 4.3.1 Critère de proximité

Ce critère repose sur le degré de justesse de la mortalité ajustée par rapport à la mortalité observée. Il peut être évalué selon l'aspect graphique ainsi que l'aspect statistique.

##### ✓ Les résidus

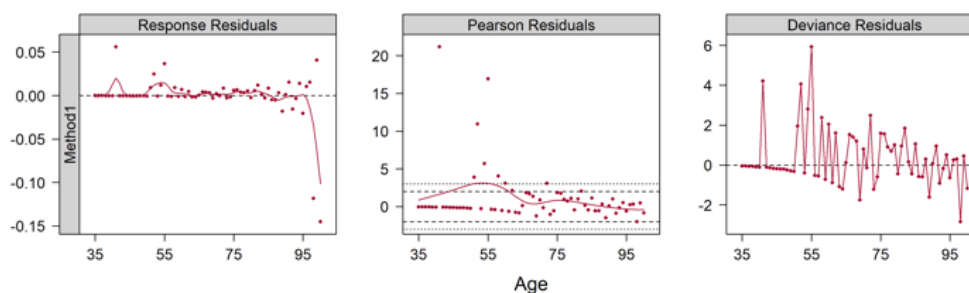
Les différents résidus présentés ci-dessous vont nous permettre d'évaluer la qualité de l'ajustement des modèles établis :

- ✓ *Résidus de Réponse* :  $r_x = \hat{q}_x - \tilde{q}_x$ , sont définis comme l'écart entre les quotients de mortalité ajustés et ceux observés. Ils sont utiles pour évaluer l'adéquation globale du modèle aux données, en particulier pour les modèles de régression linéaire.
- ✓ *Résidus de Pearson* :  $r_x = \frac{(D_x - E_x \tilde{q}_x)}{\sqrt{\text{Var}[E_x \tilde{q}_x]}}$ , sont calculés en divisant les résidus de Pearson par l'écart type résiduel. Ils sont utilisés pour évaluer l'ajustement local du modèle aux données, c'est-à-dire pour identifier les points de données qui sont particulièrement mal ajustés par le modèle.
- ✓ *Résidus de la Déviance* :  $r_x = \text{signe}(D_x - E_x \tilde{q}_x) \times \sqrt{\text{Deviance}_x}$ , ils sont utilisés pour évaluer la qualité de l'ajustement d'un modèle de régression logistique ou de Poisson. Ils sont calculés en soustrayant la déviance résiduelle du modèle ajusté de la déviance nulle, qui est la déviance d'un modèle n'ayant aucun prédicteur. Les résidus de déviance peuvent aider à identifier les observations qui ont une influence importante sur l'ajustement du modèle.

Les graphique des résidus de la réponse, des résidus Pearson et des résidus de la déviance avec les deux méthodes pour les plages [35,100] :

Méthode 1 du SMR :

### Method1 Female - Plot of the residuals - year: 2021



Méthode 2 du Logit :

### Method2 Female - Plot of the residuals - year: 2021

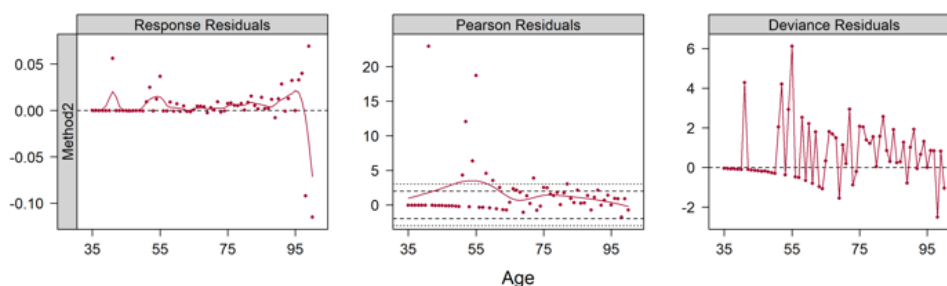


Figure 47 : Graphiques des résidus pour l'année 2021

Pour les deux modèles, les résidus de Pearson sont globalement tous compris dans l'intervalle  $[-2,2]$ , ce qui signifie que l'approche modélise correctement la variabilité des données.

Une succession de signes négatifs puis positifs est observée pour les deux modèles sur l'intervalle d'âges  $[50 ; 75]$  pour les résidus de la déviance, ce qui indique un ajustement légèrement sur-lissé.

#### ✓ Les intervalles de confiance des décès prédits et observés

En partant de l'hypothèse  $D_{x,t} \sim N(E_{x,t} \cdot q_x(t)(1 - q_x(t)))$ , le nombre de décès est prédit avec l'intervalle de confiance ci-dessous :

$$IC = [\hat{D}_{x,t} - z_{1-2\alpha} \cdot \sqrt{E_{x,t} \cdot q_x(t)(1 - q_x(t))}, [\hat{D}_{x,t} - z_{1-2\alpha} \cdot \sqrt{E_{x,t} \cdot q_x(t)(1 - q_x(t))}]$$

Où  $\hat{D}_{x,t} = \hat{q}_x(t) \cdot E_{x,t}$ ,  $z_{1-2\alpha}$  étant le quantile de la loi normale avec  $\alpha = 0.05$ .

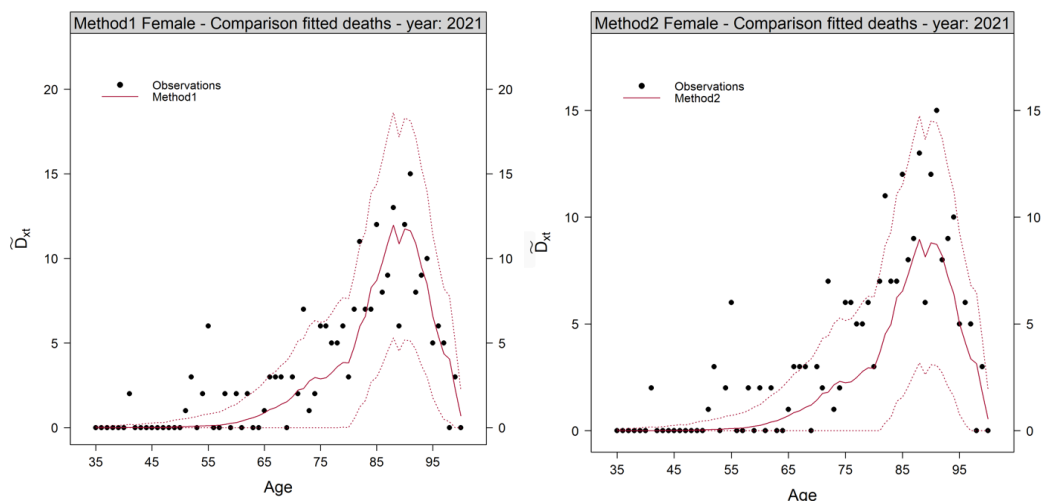


Figure 48 : Graphique des décès théoriques et observés pour l'année 2021 pour les deux méthodes

La même tendance est remarquée pour les deux approches. Malgré une forte volatilité, les décès observés sont quasi tous observés dans la bande de confiance sur l'ensemble de la période d'observation, ce qui témoigne d'une bonne représentation de la réalité.

Ces graphiques nous donnent une première indication de la justesse des modèles cependant il est indispensable d'étudier l'aspect statistique des différents ajustements.

#### ✓ Les tests statistiques

On rappelle en ce qui concerne les tests les définitions suivantes :

- ✓ **Le test du rapport de vraisemblance** permet de déterminer si l'ajustement par le modèle correspond à la loi de mortalité sous-jacente. En ce qui concerne la règle de décision, nous retiendrons l'hypothèse nulle  $H_0$  si la statistique du test suit une loi de  $\chi^2$  avec un nombre de degré de liberté égal au nombre d'observations.

Les hypothèses suivantes sont testées :

- $H_0$  : l'ajustement effectué correspond à la loi de mortalité sous-jacente ;
- $H_1$  : l'ajustement effectué ne correspond pas à la loi de mortalité sous-jacente

La statistique du test du rapport de vraisemblance :

$$\xi = \sum_{x,t} D_{x,t} \ln\left(\frac{D_{x,t}}{E_{x,t} \tilde{q}_{x,t}}\right) - (D_{x,t} - E_{x,t} \tilde{q}_{x,t})$$

On retient l'hypothèse  $H_0$  si  $\xi \sim \chi^2(n)$  avec  $n$  le nombre d'observations et on la rejette dans le cas où  $\xi > \chi^2_{(1-\alpha)}(n)$  avec  $\chi^2_{(1-\alpha)}(n)$  correspondant au quantile  $(1 - \alpha)$  de la distribution du  $\chi^2$  à  $n$  degrés de liberté.

- ✓ **Le SMR** est défini comme le rapport du nombre de décès observés au nombre de décès prédits dans une population de référence.

$$SMR = \frac{\sum_{x,t} D_{x,t}}{\sum_{x,t} E_{x,t} \tilde{q}_{x,t}}$$

Avec les  $\tilde{q}_{x,t}$  taux de mortalité ajustés et  $D_{x,t} \sim P(E_{x,t} q_{x,t})$

Les hypothèses que nous testons sont donc :

- $H_0$  : le SMR n'est pas significativement différent de 1 ;
- $H_1$  : le SMR est significativement différent de 1.

Si le  $SMR > 1$ , les décès ajustés sont sous-estimés et réciproquement si  $SMR < 1$ . Plus ce ratio est proche de 1, meilleure est l'estimation.

La statistique du test est calculée :

$$\begin{cases} \xi = 3 \cdot \sqrt{D} \cdot \left( 1 - (9D)^{-1} - \sqrt{\frac{D}{E}} \right), \text{ si } SMR > 1 \\ \xi = 3 \cdot \sqrt{D^*} \cdot \left( \sqrt{\frac{D^*}{E}} + (9D^*)^{-1} - 1 \right), \text{ si } SMR < 1 \end{cases}$$

Où  $D = \sum_{x,t} D_{x,t}^{obs}$ ,  $D^* = \sum_{x,t} D_{x,t}^{obs} + 1$  et  $E = \sum_{x,t} E_{x,t} \tilde{q}_{x,t}$

L'hypothèse nulle  $H_0$  est rejetée si  $\xi > N_{1-\alpha}(0,1)$ .

- ✓ **Le test de Wilcoxon** permet de tester l'hypothèse nulle que la médiane entre chaque paire est nulle. Les paires étant constituées des probabilités de décès observées d'une part et prédites par le modèle d'autre part. L'hypothèse nulle  $H_0$  est vraie si les probabilités de décès observées et prédites sont équivalentes.

Les hypothèses que nous testons sont donc :

- $H_0$  : la médiane de la différence entre chacune des paires est nulle ;
- $H_1$  : la médiane de la différence entre chacune des paires n'est pas nulle.

$$\xi = \frac{(w - \frac{1}{2} - n(n+1)/4)}{\sqrt{n(n+1)(2n+1)/24}}$$

Ainsi,  $H_0$  est retenue si  $\xi \sim N(0,1)$  et rejetée si  $|\xi| > N_{1-\frac{\alpha}{2}}(0,1)$  avec  $N_{1-\frac{\alpha}{2}}(0,1)$  correspondant au quantile  $(1 - \frac{\alpha}{2})$  de la distribution normale centrée réduite.

- ✓ **Le test du  $\chi^2$**  permet de calculer la distance  $d^2$  entre les taux bruts et les taux lissés et entraînera un rejet si cet écart est trop grand.

Le  $\chi^2$  est un indicateur permettant d'évaluer la qualité de l'ajustement du modèle effectué. Sa valeur est ici déterminée par :

$$\chi^2 = \sum_{(x,t)} \frac{(d_{x,t} - E_{x,t} \tilde{q}_{x,t})^2}{E_{x,t} \tilde{q}_{x,t} (1 - \tilde{q}_{x,t})}$$

Avec les  $\tilde{q}_{x,t}$  taux de mortalité ajustés.

La valeur du  $\chi^2$  est recherchée de sorte à être le plus faible possible.

- ✓ **Le  $R^2$** , qui correspond au coefficient de détermination et donne une évaluation de l'adéquation du modèle d'ajustement aux données observées, est calculé par la formule suivante :

$$R^2 = 1 - \frac{\sum_{x,t} (\widehat{q}_{x,t} - \tilde{q}_{x,t})}{\sum_{x,t} (\widehat{q}_{x,t} - \frac{\sum \widehat{q}_{x,t}}{n})^2}$$

Une valeur du  $R^2$  proche de 1 permet de conclure à un bon modèle.

Avec  $n$  correspondant au nombre d'observations et les  $\tilde{q}_{x,t}$  taux de mortalité ajustés. La valeur du  $R^2$  est recherchée de sorte à être la plus proche possible de 1.

- ✓ **Le MAPE** ou encore Mean Average Pourcentage Error est un indicateur de l'exactitude entre les observations et l'ajustement. Il est déterminé par la moyenne des écarts en valeur absolue par rapport aux données observées. Plus le MAPE est faible, meilleure est la qualité de l'ajustement.

$$MAPE = \frac{\sum_{x,t} |(\widehat{q}_{x,t} - \tilde{q}_{x,t}) / \widehat{q}_{x,t}|}{\sum_{x,t} d_{x,t}}$$

Avec les  $\tilde{q}_{x,t}$  taux de mortalité ajustés et  $\widehat{q}_{x,t}$  les taux de mortalité observés.

Un tableau avec les résultats des différents tests pour les deux méthodes d'ajustements a été réalisé :

Tests		Méthode du SMR	Méthode du Logit
Test du rapport de vraisemblance	$\xi$	1399.05	1521.09
	Hypothèse acceptée	$H_1$	$H_1$
	p_value	0	0.003
Test du SMR	SMR	1	1.21
	$\xi$	0.0063	4.21
	Hypothèse acceptée	$H_0$	$H_1$
	p_value	0.50	0.001
Test de Wilcoxon	$\omega$	23475	27975
	$\xi$	3.13	0.03
	Hypothèse acceptée	$H_1$	$H_0$
	p_value	0.017	0.76
Grandeurs statistiques	$R^2$	0.72	0.86
	$\chi^2$	556.76	351.67
	MAPE	54.96	42.88
	Déviante	832.54	968.98
Résidus standardisés	>2	54	47
	>3	15	11

Tableau 8 : Résultats des tests statistiques pour comparer la méthode du SMR et la méthode du Logit

Pour le test du rapport de vraisemblance, aucune méthode n'est acceptée tandis que la méthode de Wilcoxon est validée par la méthode du Logit avec une  $p\_value$  assez proche de 1 et la méthode du SMR est validée par le test du SMR avec une  $p\_value$  à 0.50.

Pour les grandeurs statistiques, rappelons que le  $\chi^2$ ,  $MAPE$  et la *déviante* doivent être minimisés tandis que le  $R^2$  doit être maximisé et le plus proche de 1.

D'après ces résultats, les valeurs du  $\chi^2$  et du  $MAPE$  sont les plus faibles pour la méthode du Logit. Le  $R^2$  qui est le plus proche de 1 est également celui du Logit. De plus, les résidus standardisés étant plus faibles pour cette méthode, on favorise également la méthode du Logit.

A première vue, les tests précédents préconisent l'utilisation de la méthode de régression logistique. Les tests de régularité de l'ajustement vont être effectués afin de vérifier ces résultats.

#### 4.3.2 Critère de régularité

Nous allons à présent tenter de valider la régularité des ajustements grâce aux tests des runs et des signes.

##### ✓ Test des runs

Ce test détermine si les éléments d'une séquence sont mutuellement indépendants.

En notant  $n$  le nombre d'éléments d'une séquence,  $n_+$  le nombre de signes positifs et  $n_-$  le nombre de négatifs, avec  $n = n_+ + n_-$ , on a :

$$\mu = \frac{2n_+n_-}{n} - 1, \text{ et}$$

$$\sigma^2 = \frac{2n_+n_-(2n_+n_- - n)}{n^2(n-1)}$$

Si  $m$  désigne le nombre de runs d'une séquence, alors la statistique du test des runs,  $\xi^{RUN}$ , s'écrit :

$$\xi^{RUN} = \frac{m - \mu}{\sigma}$$

L'hypothèse nulle sera rejetée si :

$$|\xi^{RUN}| > N_{1-\frac{\alpha}{2}}(0,1),$$

##### ✓ Test des signes

Ce test examine la fréquence des changements de signe de la différence entre les probabilités de décès observées et ajustées. Sous l'hypothèse nulle  $H_0$ , la médiane entre les signes positifs et négatifs de cette différence est nulle. Si on note  $n_+$  le nombre de signes positifs et  $n_-$  le nombre de signes négatifs, avec  $n = n_+ + n_-$ , la statistique du test des signes,  $\xi^{SIG}$ , s'écrit :

$$\xi^{SIG} = \frac{|n_+ + n_-| - 1}{\sqrt{n}}$$

L'hypothèse nulle sera rejetée si :

$$|\xi^{SIG}| > N_{1-\frac{\alpha}{2}}(0,1),$$

Où  $N_{1-\frac{\alpha}{2}}(0,1)$  est le  $\left(1 - \frac{\alpha}{2}\right)$  quantile de distribution Normale centrée réduite avec  $\alpha = 5\%$ .



Pour le test des signes nous obtenons les résultats suivants :

**TEST DES RUNS :**

	<i>Méthode du SMR</i>	<i>Méthode du Logit</i>
<i>Nombre de runs</i>	70	83
<i>Signes (-)</i>	94	86
<i>Signes (+)</i>	86	94
$ \xi $	3.11	1.17
<i>Seuil</i>	1.96	1.96
<i>Hypothèse</i>	$H_1$	$H_0$
<i>p-value</i>	0.01	0.55

Tableau 9 : Résultats obtenus pour le test des runs

**TEST DES SIGNES :**

	<i>Méthode du SMR</i>	<i>Méthode du Logit</i>
<i>Signes (+)</i>	86	94
<i>Signes (-)</i>	94	86
$ \xi $	0.58	0.52
<i>Seuil</i>	1.96	1.96
<i>Hypothèse</i>	$H_1$	$H_0$
<i>p-value</i>	0.04	0.61

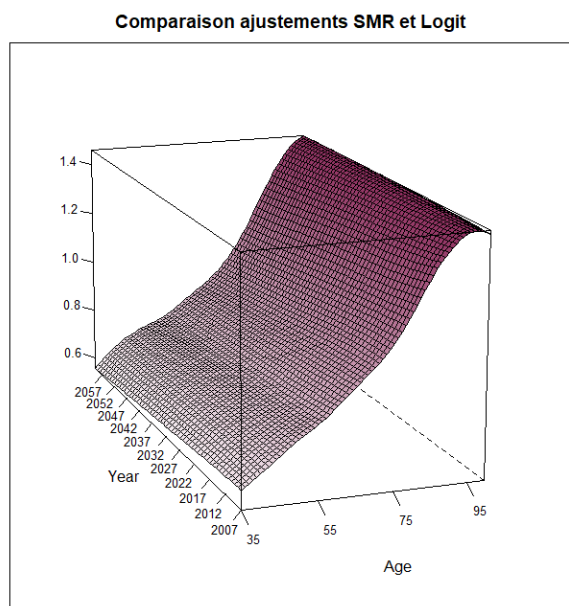
Tableau 10 : Résultats obtenus pour le test des signes

Les *p-value* ne sont pas acceptables avec des valeurs trop proches de 0 pour les deux tests pour le modèle du SMR. L'hypothèse  $H_0$  est donc rejetée pour la méthode du SMR. Cependant, les *p-values* sont acceptables pour la méthode du Logit, qui semble plus adaptée.

#### 4.3.3 Critère de cohérence

Ce critère privilégie l'ajustement donnant la mortalité la plus faible en tenant compte du risque de longévité.

Le rapport des taux ajustés par la méthode du Logit sur les taux ajustés par la méthode du SMR est par conséquent étudié :



*Figure 49 : Rapport des taux ajustés par la méthode du Logit sur la méthode du SMR*

Une augmentation linéaire est observée du rapport des taux en fonction des âges, ce qui signifie que la mortalité modélisée par méthode du Logit croît plus rapidement que celle donnée par la méthode du SMR. Le rapport moyen global des deux modèles est de 0.98. L'ajustement par la méthode de régression linéaire est légèrement plus prudent que celui de la méthode du SMR.

Les critères de prudence et de cohérence étant satisfaisants pour le modèle du Logit, il est ainsi retenu pour la suite de l'étude, nous pouvons à présent procéder à la fermeture de la table. La méthode utilisée ici dans le cadre de la fermeture de la table de mortalité est la méthode de Denuit et Goderniaux (2005).

#### 4.4 Fermeture de la table

Les taux ajustés obtenus par la méthode de la régression logistique ne sont calibrés que sur l'intervalle d'âge [35 :100]. L'ajustement des taux pour les âges supérieurs à 100 ans n'a pas été réalisé en raison de l'insuffisance des données pour les âges élevés. Il est ainsi nécessaire d'envisager une procédure pour la fermeture de la table.

##### 4.4.1 Application de la méthode de Denuit et Goderniaux

La fermeture d'une table s'effectue en deux étapes, la première étant de définir la tranche d'âge dans laquelle l'âge optimal de départ est déterminé pour le calcul des coefficients du modèle et ensuite de déterminer la régularité de la fermeture visuellement.

Les tables obtenues ne sont valables que sur les âges de 35 à 100 ans car le modèle a été appliqué sur cet intervalle d'âges. Compte tenu de l'insuffisance de données au-delà de 100 ans, il serait incohérent d'utiliser une méthode d'ajustement sur ses âges.

Une méthode de fermeture de table est ainsi appliquée, la méthode de Denuit et Goderniaux, afin d'extrapoler la mortalité aux âges élevés (pour notre cas sur les âges [100 :130]).

Cette méthode consiste à ajuster par moindres carrés le modèle log-quadratique :

$$\ln(\hat{q}_x) = a + bx + cx^2 + \varepsilon_x$$

Avec  $\varepsilon_x$ , une variable aléatoire iid suivant une loi normale  $N(0, \sigma^2)$ .

Deux contraintes sont intégrées au modèle :

- ✓  $q_{x=130} = 1$ , l'âge limite de fermeture est fixé à 130 ans. Cette hypothèse semble adaptée car elle est prudente, même s'il n'existe pas d'âge limite dans la réalité.
- ✓  $\frac{\delta q_x}{\delta x} |_{x=130} = 0$ , la deuxième contrainte est une contrainte d'inflexion.

Ces contraintes imposent à la courbe des taux de mortalité une allure concave aux grands âges et l'existence d'une tangente horizontale au point  $x = 130$  ans. Cette contrainte a pour conséquence d'empêcher une décroissance éventuelle des  $q_x$  aux âges élevés. Cela conduit à :

$$a + bx + cx^2 = c(130^2 - 260x + x^2)$$

Le paramètre  $c$  est alors déterminé par un critère de moindres carrés.

Il convient maintenant de déterminer l'âge optimal de départ à partir duquel les taux bruts sont remplacés par les taux ajustés en se basant sur la maximisation du coefficient de détermination  $R^2$  en effectuant le test à plusieurs âges. L'âge optimal retenu est de 95 ans.

Ainsi, nous conservons les  $\hat{q}_x$  pour les âges [35,100] ans et nous remplaçons les probabilités annuelles de décès au-delà de 100 ans par les valeurs ajustées provenant du modèle de fermeture.

Le paramètre  $c$  obtenu reste négatif et décroît de manière linéaire plus les années avancent.

Les taux initialement ajustés par le modèle du SMR sont alors remplacés par les taux obtenus par le modèle de fermeture pour les âges supérieurs à 100 ans avec la formule  $\hat{q}_x(t) = \exp(\hat{c}_t(130 - x)^2)$ .

Une table prospective annuelle complétée est obtenue s'étalant sur la plage d'âges [35 ans ; 130 ans] et sur les années allant de 2007 à 2060. Le graphique suivant est obtenu :

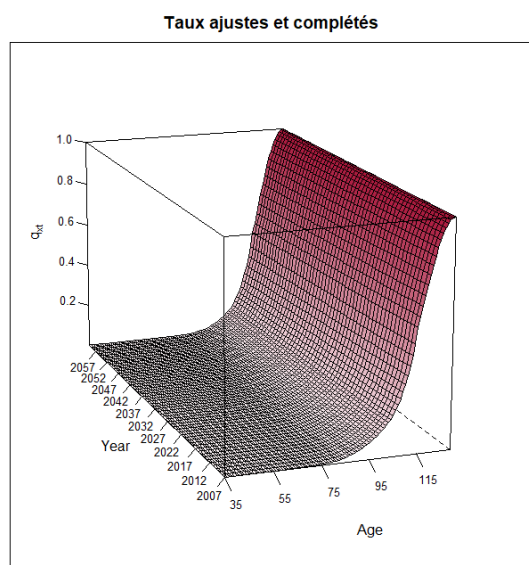


Figure 50 : Taux bruts de mortalité ajustés et complétés

La jonction entre les probabilités de décès ajustées et les valeurs obtenues par le modèle de fermeture est relativement lisse. Aucune cassure n'est identifiée dans la tendance de mortalité aux âges élevés. De plus, les taux de mortalité aux âges élevés semblent cohérents, ce qui traduit bien le phénomène de mortalité réel avec une augmentation de la mortalité en fonction de l'âge. L'approche de Denuit et Goderniaux semble alors satisfaisante.

Après l'obtention d'une table d'expérience prospective pour les femmes, les provisions mathématiques vont être calculées avec cette nouvelle table et comparée avec la table réglementaire afin de voir l'impact de celle-ci sur le provisionnement.

## 5 Impact de la table sur le provisionnement

Afin de mieux appréhender l'impact de la table d'expérience obtenue sur le résultat à court et moyen terme, les provisions mathématiques obtenues avec la table réglementaire et avec la table d'expérience vont être comparées.

### 5.1 Les provisions mathématiques

Le provisionnement est une étape importante pour un assureur : afin de faire face à ses engagements futurs, il doit provisionner le mieux possible les risques auxquels il est confronté, c'est-à-dire évaluer au plus juste le montant des engagements pris envers les assurés.

Il est intéressant de mesurer l'impact que peut avoir l'utilisation de tables d'expériences sur le provisionnement.

Les provisions associées aux rentiers vont être calculées avec la loi d'expérience construite pour voir si des écarts importants de provisions sont constatés avec le calcul de provisions fait avec la loi réglementaire.

#### ✓ Définition :

La provision mathématique est une réserve permettant à l'assureur, en contrepartie du versement d'une prime par le bénéficiaire, de lui garantir le versement d'une prestation en cas de survenance du risque couvert.

Elle est définie dans l'article 343-3 du Code des Assurances comme étant « la différence entre les valeurs actuelles des engagements respectivement pris par l'assureur et par les assurés ».

La provision mathématique est calculée d'après des taux d'intérêt et des tables de mortalité déterminés dans des conditions et limites fixées par arrêté du ministre chargé de l'économie.

A la souscription, il existe par construction un équilibre entre les engagements respectifs de l'assureur et de l'assuré :

$$VAP (Assureur)_k - VAP(assuré)_k = 0$$

A toute date  $t > 0$ ,  $PM_k = VAP (Assureur)_k - VAP(Assuré)_k$

Pour calculer donc les PM, il faut évaluer les VAP de chaque partie à la date de d'évaluation.

✓ Formules des PM en fonction des types de rentes :

**Notions de commutations :**

$l_x$  : nombre de survivants à l'âge  $x$

$d_x = l_x - l_{x+1}$  : nombre de décès entre l'âge  $x$  et  $x + 1$

$v = \frac{1}{1+i}$  : taux actualisation

Une fois l'assuré décédé, il ne reste plus que l'engagement de l'assureur à remplir.

Dans le portefeuille étudié dans le cadre des rentes de conjoint, les rentes sont payables à terme échu ou à échoir. On a alors les provisions mathématiques ci-dessous :

**Rente temporaire :**

La provision mathématique pour le versement d'une rente de conjoint viagère de 1 euro par an est :

✓ Terme à échoir :

$$\ddot{a}_{x:n}^{(m)} = \left[ \sum_{k=0}^{n-1} \frac{l_{x+k}}{l_x} * \frac{1}{(1+i)^k} \right] - \frac{m-1}{2m} * \left[ 1 - \frac{1}{(1+i)^n} * \frac{l_{x+n}}{l_x} \right]$$

✓ Terme échu :

$$a_{x:n}^{(m)} = \left[ \sum_{k=0}^{n-1} \frac{l_{x+k}}{l_x} * \frac{1}{(1+i)^k} \right] - \frac{m+1}{2m} * \left[ 1 - \frac{1}{(1+i)^n} * \frac{l_{x+n}}{l_x} \right]$$

**Rente viagère :**

La provision mathématique pour le versement d'une rente de conjoint viagère de 1 euro par an est :

✓ Terme à échoir :

$$\ddot{a}_x^{(m)} = \left[ \sum_{k=0}^{\sup-x} \frac{l_{x+k}}{l_x} * \frac{1}{(1+i)^k} \right] - \frac{m-1}{2m}$$

✓ Terme échu :

$$a_x^{(m)} = \left[ \sum_{k=0}^{\sup-x} \frac{l_{x+k}}{l_x} * \frac{1}{(1+i)^k} \right] - \frac{m+1}{2m}$$

- A partir de l'âge  $x$  du conjoint à la date de calcul, jusqu'à sa mort
- Avec  $m$  versements par an
- Avec sa loi de mortalité  $l_x$
- Avec un taux technique  $i$
- $\sup$  : âge maximal de la table de mortalité utilisée.

✓ Taux technique :

Le taux technique  $i$  est choisi selon l'année de survenance du sinistre. Le taux technique est plafonné pour les années antérieures à 2009 afin d'avoir un provisionnement plus prudent.

✓ Interpolation linéaire :

Une interpolation linéaire est réalisée :

$$M = (Age2 - Age) * a_{Age1}^{(m)} + (Age - Age1) * a_{Age2}^{(m)}$$

Avec :

- $Age1$  la valeur entière inférieure à l'âge de la date de calcul.
- $Age2$  la valeur entière supérieure à l'âge à la date de calcul.

## 5.2 Impact de la table construite sur le volume des provisions mathématiques

Nous avons intégré la table construite précédemment dans le logiciel de calcul de provisions mathématiques « Addactis PM expert ».

Ensuite, le calcul de provisions a été fait sur la base de cette nouvelle loi et les résultats obtenus ont été comparés avec le montant des provisions calculées sur les tables réglementaires, à iso-paramètres techniques (chronique de taux techniques par année de survenance et taux de chargement).

Pour l'étude, les provisions mathématiques de l'année 2021 vont être calculées avec la table d'expérience créée précédemment. Les provisions mathématiques obtenues sont alors comparées aux provisions calculées sur l'inventaire 2021 à la date du 31/12/2021.

Les résultats des calculs des ax obtenus sont les suivants :

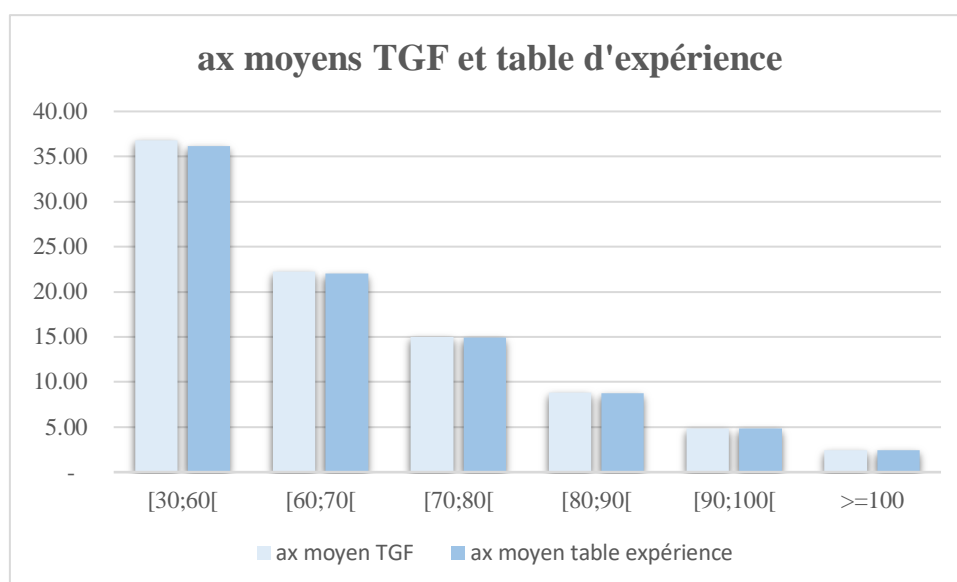


Figure 51 : Histogramme des ax moyens de la TGF et de la table d'expérience

Les rapports par tranche d'âges entre les ax de la table d'expérience et la TGF05 ont également été tracés ci-dessous :

	Tranche d'âges	Rapport ax table d'expérience / ax TGF
Tranche 1	[30;60[	-1,62%
Tranche 2	[60;70[	-0,91%
Tranche 3	[70;80[	-0,64%
Tranche 4	[80;90[	-0,53%
Tranche 5	[90;100[	-0,20%
Tranche 6	>=100	0,00%

Tableau 11 : Rapports entre les ax de la table d'expérience et la TGF 05



Les annuités viagères obtenues sont très légèrement inférieures pour la loi d'expérience que pour la table réglementaire, les écarts étant variables selon les différentes tranches d'âges.

Les provisions sont majorées par la table TGF 05 d'environ 0.6%. Ainsi, les provisions mathématiques pourraient être réduites de 0.6% si la table d'expérience était appliquée à la place de la table réglementaire.

Nous pouvons donc conclure sur le fait que les rentes de conjoint du portefeuille étudié présentent un risque moins élevé que celui de la population générale.

## Conclusion :

Au cours de ce mémoire, une table d'expérience générationnelle féminine a été construite. L'objectif de l'étude était d'analyser si les tables réglementaires utilisées pour provisionner les rentes de conjoint reflétaient correctement la mortalité d'un portefeuille, et de déterminer si le provisionnement actuel était adéquat ou s'il devait être ajusté en utilisant une table d'expérience.

Le provisionnement de la garantie rente de conjoint est réalisé par deux tables générationnelles : la TGH05 pour les hommes et la TGF05 pour les femmes, tous types de rentes confondues. Cependant, le portefeuille étudié étant constitué à 91% de femmes, il n'était pas judicieux de considérer les hommes dans notre étude. A l'avenir, si le nombre de données nous le permet, la construction d'une table homme pourra être envisagée en utilisant la même démarche que pour la table femme.

Lors de la construction de ce mémoire, un retraitement et un contrôle des données a été réalisé afin de construire une base de données solide. Par la suite, des statistiques descriptives ont été réalisées sur le portefeuille. En effet, il est essentiel d'avoir une connaissance approfondie des données brutes utilisées avant de construire la table.

Plusieurs méthodes d'estimation des taux bruts de mortalité ont été réalisées, à la suite de la sélection d'une période d'observation pertinente (2000 à 2021) et d'une tranche d'âges [35 :100]ans grâce à la méthode de Cochran. L'estimateur des moments de Hoem a été retenu en raison de son caractère prudent. Les taux bruts ont ensuite été ajustés par deux méthodes de positionnement, la méthode du SMR ainsi que la méthode du Logit. Après une comparaison des ajustements obtenus selon différents critères de validation, la méthode du Logit a été sélectionnée pour la construction de la table. Enfin, après la fermeture de la table par la méthode de Denuit et Goderniaux, la table a été établie.

La table d'expérience créée a été intégrée dans le logiciel de calcul des provisions mathématiques « Addactis PM expert », dans le but de comparer les provisions obtenues avec la nouvelle table et avec la table réglementaire. Puisque seulement une table femme a été construite, seulement les provisions des femmes ont été comparées avec les provisions de la TGF05.

La mortalité du portefeuille observée est très légèrement plus élevée que celle qui est observée sur la population générale et par la réglementation. Les provisions mathématiques obtenues sont alors légèrement en dessous des provisions obtenues avec la TGF05.

La table d'expérience générationnelle construite est légèrement plus adaptée à la mortalité du portefeuille rente de conjoint étudié de Malakoff Humanis que la TGF05. Toutefois, le changement d'utilisation de la table aurait peu d'impact sur le provisionnement.

Il est cependant nécessaire d'être conscient des limites que l'on peut rencontrer dans la construction d'une table d'expérience. En effet, la quantité d'observations est une variable qui peut fortement influencer les résultats. Dans notre cas, la taille du portefeuille étant plutôt faible, cela peut se traduire par des variations brusques de la mortalité. De plus, les données du portefeuille ne représentent la mortalité qu'à un instant précis et ne permettent pas d'avoir de certitudes sur son évolution future.

Dans les années à venir, il serait judicieux de réitérer l'analyse sur la population masculine, pour autant que le nombre de données le permette, ainsi que de refaire l'étude sur la population féminine afin de pallier le risque lié à la longévité et de maintenir la qualité du provisionnement.

Afin d'approfondir cette étude, il a été montré dans ce mémoire qu'il était intéressant de segmenter selon le sexe, le type de rente ou encore la catégorie socioprofessionnelle dans le portefeuille, la zone géographique etc. Cette segmentation permettrait de mieux comprendre les différences dans les tendances d'évolution observées et d'identifier les facteurs qui influencent l'espérance de vie du portefeuille étudié.

## Annexes :

### Annexe A : Critères de Cochran

Age	Critère 1	Critère2	Critère3	Age	Critère1	Critère2	Critère3	Age	Critère1	Critère2	Critère3
20	0	6	2121	54	142	17131	6304881	88	478	4446	1797208
21	0	9	3366	55	218	18591	6865184	89	465	3672	1509871
22	0	15	5574	56	1607	18327	7275844	90	391	2998	1237013
23	0	36	13286	57	156	19712	7251720	91	350	2420	1011076
24	1	63	23161	58	179	20648	7602086	92	340	1856	801325
25	0	108	39538	59	207	21383	7880424	93	261	1425	615516
26	0	151	55221	60	251	21929	8095669	94	252	1011	460680
27	1	212	77630	61	785	21706	8209362	95	177	731	331099
28	2	279	102647	62	181	22019	8103000	96	126	517	234833
29	2	368	135008	63	177	22070	8120338	97	109	350	167653
30	3	492	182103	64	196	21959	8086797	98	72	234	111604
31	2	642	235161	65	211	21730	8008486	99	58	145	74187
32	4	816	299520	66	228	21384	7888455	100	41	81	44566
33	4	1012	370919	67	267	20904	7727186	101	4	42	25717
34	4	1237	453530	68	212	20411	7527326	102	4	21	12897
35	13	1474	542965	69	259	19858	7342714	103	4	14	7584
36	6	1766	646757	70	319	19200	7124223	104	4	10	5097
37	11	2041	748817	71	304	18540	6878001	105	1	8	3248
38	19	2376	874308	72	310	17806	6612447	106	1	7	2913
39	22	2749	1011365	73	359	16961	6321606	107	1	6	2549
40	30	3175	1169542	74	368	16133	6022722	108	1	5	2184
41	24	3683	1353151	75	339	15281	5701332	109	0	5	1820
42	27	4226	1552154	76	397	14361	5386655	110	0	5	1734
43	36	4831	1776469	77	377	13544	5080888	111	0	4	1456
44	40	5553	2041194	78	416	12745	4803808	112	0	4	1456
45	42	6357	2335625	79	429	11930	4510785	113	0	4	1456
46	31	7173	2629511	80	439	11109	4214793	114	0	4	1456
47	48	8175	3001312	81	431	10315	3922351	115	0	4	1456
48	73	9236	3397902	82	473	9532	3651940	116	0	4	1456
49	73	10445	3839044	83	448	8728	3349210	117	1	3	1457
50	98	11667	4294171	84	515	7840	3049515	118	1	2	1093
51	261	12797	4766058	85	512	6964	2728597	119	0	2	728
52	133	14193	5228979	86	492	6086	2401173	120	0	2	728
53	121	15658	5759311	87	490	5270	2102570				

Tableau 12 : Critères de Cochran et critère d'exposition pour chaque âge

**Annexe B : Résultats de la validation de la méthode du Logit**

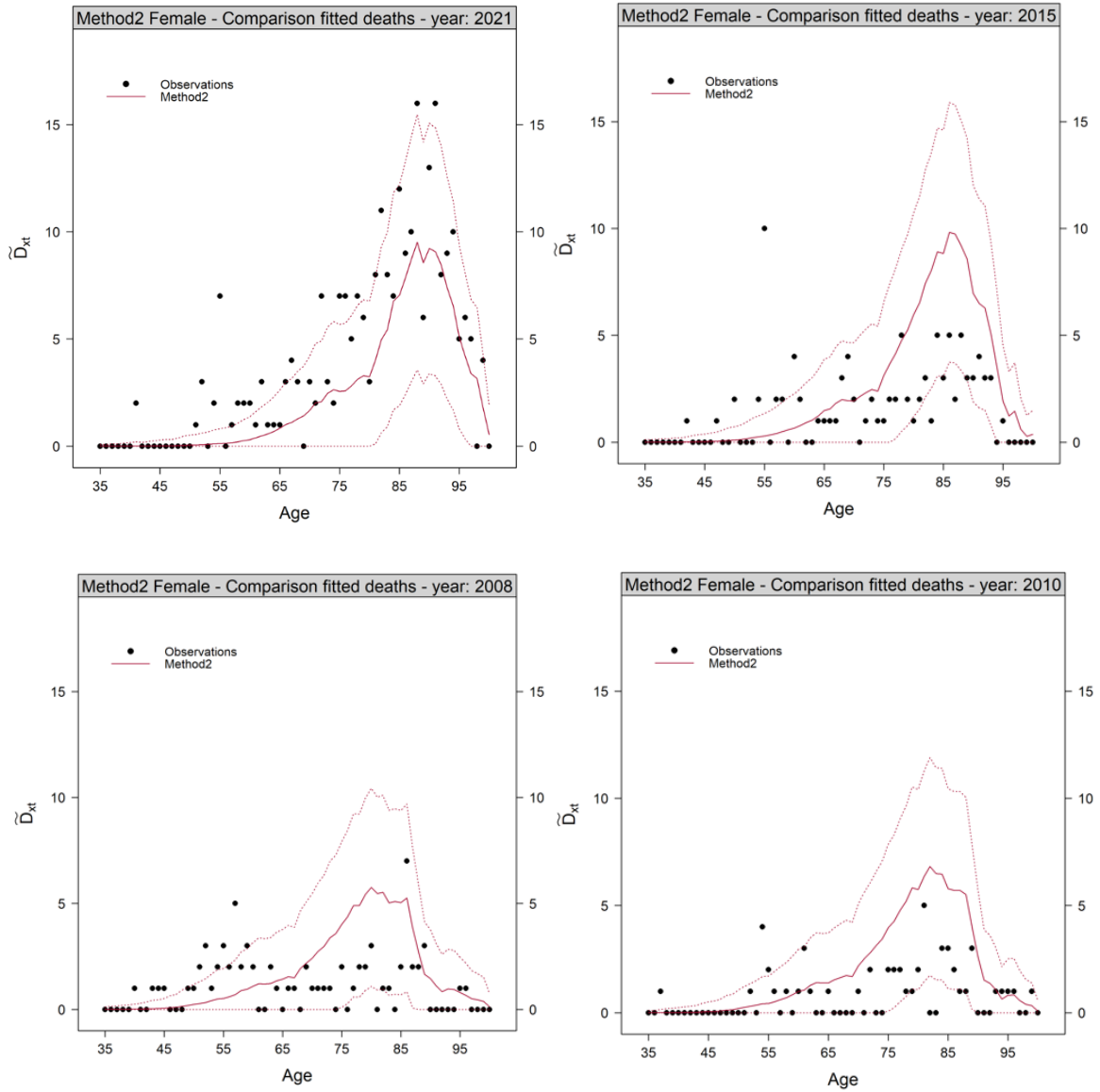
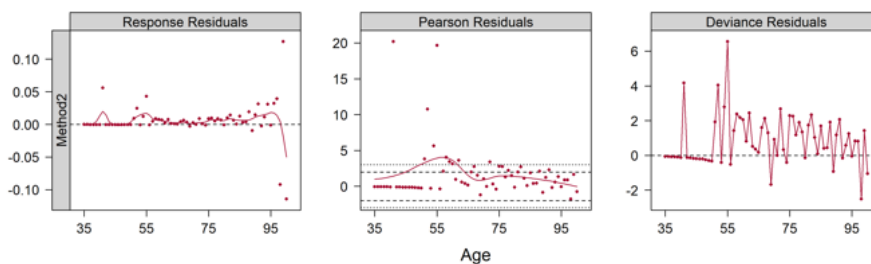
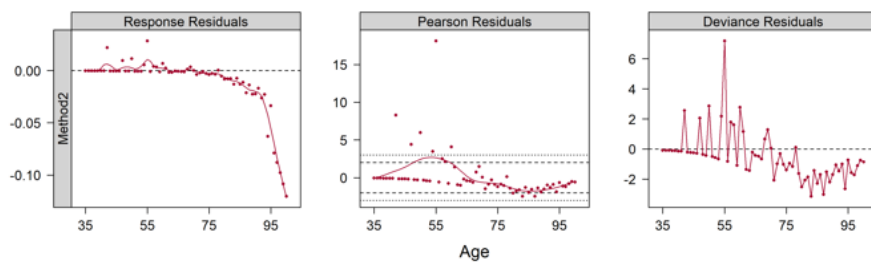


Figure 52 : Graphiques des comparaisons des décès prédits et observés sur les années 2008, 2010, 2015 et 2021

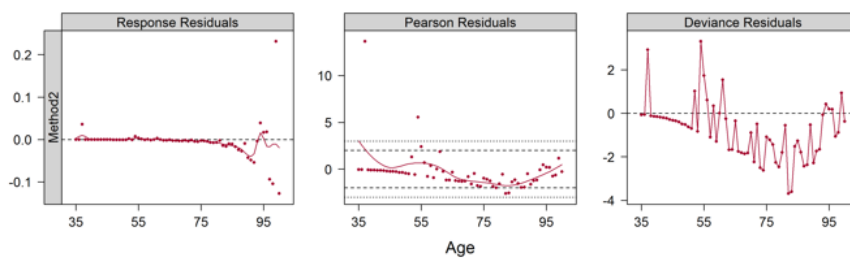
Method2 Female - Plot of the residuals - year: 2021



Method2 Female - Plot of the residuals - year: 2015



Method2 Female - Plot of the residuals - year: 2010



Method2 Female - Plot of the residuals - year: 2008

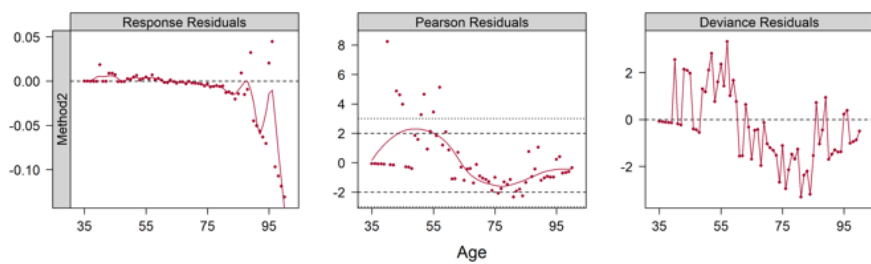


Figure 53 : Graphiques des résidus pour la méthode du Logit

## **Table des figures :**

Figure 1 - Méthode appliquée pour la construction de la table .....	3
Figure 2 : Estimation des taux bruts par différentes méthodes .....	4
Figure 3 : Taux ajustés par deux méthodes de positionnement .....	5
Figure 4 : Taux ajustés et complétés sur la période [2007 ;2060] .....	6
Figure 5 : Histogramme des ax moyens femmes de la TGF et de la table d'expérience .....	7
Figure 6 : Method applied for the construction of the table .....	9
Figure 7 : Estimation of the raw rates .....	10
Figure 8 : Adjusted rates using two positioning methods .....	11
Figure 9: Adjusted and completed rates over the period [2007,2060] .....	12
Figure 10 : Average annuities for women using TGF 05 and the experience table .....	13
Figure 1: La prévoyance en France.....	18
Figure 2: Ventilation des recette 2019 .....	20
Figure 3 : Pourcentage de cotisations pour les contrats de prévoyance .....	21
Figure 4: Différents types de régimes .....	21
Figure 5: Schéma de fonctionnement d'un contrat collectif.....	22
Figure 6: Les différentes garanties.....	23
Figure 7 : Pourcentage de dossiers par périmètre .....	31
Figure 8 : Nombre de dossiers par type de rente .....	32
Figure 9 : Nombre de dossiers par sexe pour les rentes temporaires .....	32
Figure 10 : Nombre de dossiers par sexe pour les rentes viagères .....	33
Figure 11 : Ages moyens de l'assuré et du bénéficiaire lors du décès de l'assuré.....	33
Figure 12 : Répartition des bénéficiaires selon l'âge au sinistre pour les rentes viagères .....	34
Figure 13 : Répartition des bénéficiaires selon l'âge au sinistre selon le type de rente .....	34
Figure 14 : Nombre de dossiers par survénance pour les rentes temporaires .....	35
Figure 15 : Nombre de dossiers par survénance pour les rentes viagères .....	35
Figure 16 : Nombre de dossiers par âge à la survénance pour les rentes temporaires .....	36
Figure 17 : Nombre de dossiers par âge à la survénance les rentes viagères .....	36
Figure 18 : Nombre de dossiers par âge à l'inventaire pour les rentes temporaires .....	37
Figure 19 : Nombre de dossiers par âge à l'inventaire pour les rentes viagères .....	37
Figure 20 : Nombre de dossiers par âge max de garantie pour les rentes temporaires .....	38
Figure 21 : Nombre de dossiers par âge au décès du bénéficiaire pour les rentes viagères .....	38
Figure 22 : Durée d'exposition en fonction âge pour les rentes viagères .....	39
Figure 23 : Etapes de construction d'une table.....	48
Figure 24 : Nombre de dossiers par date d'entrée .....	52
Figure 25 : Nombre de dossiers entrants par date d'entrée après correction .....	52
Figure 26 : Exposition par âge des bénéficiaires selon le sexe.....	53
Figure 27 : Nombre de décès par genre .....	54
Figure 28 : Nombre de décès des bénéficiaires par année de décès.....	55
Figure 29 : Schéma des censures et troncatures .....	57
Figure 30 : Taux bruts par âge estimés par Kaplan-Meier .....	65
Figure 31 : Taux bruts par âge estimés par la méthode d'Hoem du moment .....	68
Figure 32 : Taux bruts de Hoem par âge par génération .....	69
Figure 33 : Comparaison des taux bruts de mortalité .....	70
Figure 34 : Taux bruts de mortalité de la table de référence .....	71
Figure 35 : Taux ajustés et extrapolés par la méthode du SMR .....	73
Figure 36 : Taux ajustés et extrapolés par la méthode de la régression logistique .....	74
Figure 37 : Graphiques des résidus pour l'année 2021 .....	76
Figure 38 : Graphique des décès théoriques et observés pour l'année 2021 pour les deux méthodes...77	

Figure 39 : Rapport des taux ajustés par la méthode du Logit sur la méthode du SMR .....	82
Figure 40 : Taux bruts de mortalité ajustés et complétés .....	83
Figure 43 : Histogramme des ax moyens de la TGF et de la table d'expérience.....	88
Figure 44 : Graphiques des comparaisons des décès prédits et observés sur les années 2008, 2010, 2015 et 2021 .....	92
Figure 45 : Graphiques des résidus pour la méthode du Logit .....	93

### Liste des tableaux :

Tableau 1 : Tableau des rapports entre les ax de la table d'expérience et la TGF 05 .....	7
Tableau 2 : Table of ratios between the ax of the experience table and the TGF 05 .....	13
Tableau 1 : Ages moyen par genre .....	33
Tableau 2 : Résultats du test du Log-Rank.....	41
Tableau 3 : Ratios moyens annuels sur une moyenne des années d'observations des différents modèles .....	70
Tableau 4 : Résultats obtenus par tranche d'âges pour la méthode du SMR .....	72
Tableau 5 : Résultats des coefficients obtenus pour la méthode du Logit .....	74
Tableau 6 : Résultats des tests statistiques pour comparer la méthode du SMR et la méthode du Logit .....	79
Tableau 7 : Résultats obtenus pour le test des runs .....	81
Tableau 8 : Résultats obtenus pour le test des signes.....	81
Tableau 9 : Taux technique par année de survenance.....	<b>Error! Bookmark not defined.</b>
Tableau 10 : Rapports entre les ax de la table d'expérience et la TGF 05 .....	88
Tableau 11 : Critères de Cochran et critère d'exposition pour chaque âge.....	91

## **Bibliographie :**

- [1] Frédéric Planchet - Modèles de durée, applications actuarielles, code R, <https://www.ressources-actuarielles.net/gtmortalite>.
- [2] Frédéric Planchet (2017) - Cours Isfa modèles de durée : "tables de mortalité".
- [3] Frédéric Planchet et Julien Tomas - Méthodes de positionnement : Aspects méthodologiques. Institut des Actuaire, Note de travail II1291-12 v1.5.
- [4] Julien Tomas & Frédéric Planchet (2013) – Critère de validation : aspects méthodologiques. Institut des Actuaire, Note de travail II1291-14.
- [5] Julien Tomas et Frédéric Planchet (2014) – Construction d'une table de mortalité par positionnement : Mode d'emploi. Institut des Actuaire, Note de travail II1291-15v1.5.
- [6] Aki Quashie & Michel Denuit (2005) – Modèles d'extrapolation de la mortalité aux grands âges.
- [7] Frédéric Planchet (2005) - Tables de mortalité d'expérience pour les portefeuilles de rentiers, notice de présentation.
- [8] Despierres, E (2010) : Construction et validation de tables d'expérience pour le provisionnement d'un portefeuille de rentes éducation, Mémoire d'Actuariat, ISUP.
- [9] Benjamin BALTESAR (2013) - Construction d'une table de mortalité sur un portefeuille de temporaires décès, Mémoire d'Actuariat, ISFA.
- [10] Qi ZUO - Construction d'une table de mortalité pour l'assurance Décès Vie Entière, Mémoire d'Actuariat, ISUP.
- [11] Julien Tomas et Frédéric Planchet - Construire une table de mortalité prospective : le package ELT.
- [12] El hadji Babacar NDIAYE (2016) - Construction d'une table de mortalité et lissage par positionnement, Mémoire d'Actuariat, ISUP.
- [13] Rania BOUYKHF – Construction et validation de tables de mortalité prospectives en cas de petit échantillons, Mémoire d'Actuariat, ISUP.
- [14] Xiaolin ZHANG (2011) - Construction des tables de mortalité d'expérience en cas de décès, Mémoire d'Actuariat, ISUP.
- [15] El Hadji Abdoulaye FALL – Construction de table de mortalité prospective d'expérience pour les contrats de rentes viagères en périmètre retraite, Mémoire d'Actuariat, ISFA.