

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

Par : Lorraine BASTIEN

Titre : *Utilisation de la théorie de la crédibilité pour la construction d'une table de mortalité prospective sur un portefeuille rentes restreint*

Confidentialité :  NON  (Durée :  1 an  2 ans)

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

*Membres présents du jury de Signature  
l'Institut des Actuaires*

*Entreprise :*

*Nom : Société Générale Assurance*

*Signature :*

*Directeur de mémoire en entre-  
prise :*

*Nom : Reda Jarir*

*Signature :*

*Membres présents du jury de  
l'ISFA*


*Invité :*

*Nom :*

*Signature :*

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

Signature du responsable entreprise



Signature du candidat





# Résumé

**Mots-clés :** risque de longévité, rentes, *Best estimate*, risque d'échantillonnage, théorie de la crédibilité, *Standardized Mortality Ratio*

La réglementation Solvabilité II impose aux assureurs d'effectuer une évaluation *best estimate* du flux des prestations futures actualisées. La connaissance du comportement des assurés est donc primordiale. Dans le cas de l'étude des rentes, la mortalité du portefeuille doit donc être modélisée dans ce sens.

Actuellement, les tables de mortalité prospectives utilisées par SOGECAP sont les tables réglementaires TGH/TGF005. L'objectif de ce travail est d'intégrer l'expérience interne à l'évaluation du risque de longévité encourue par l'entreprise.

Cependant, le portefeuille de l'étude étant restreint, il est nécessaire de trouver des méthodes permettant de limiter le risque d'échantillonnage.

Dans un premier temps, il était essentiel de compléter l'information disponible en étudiant d'autres portefeuilles de l'entreprise. Pour les utiliser au mieux, un partitionnement suivant les niveaux de richesse a été réalisé.

Ensuite, deux méthodes issues de la théorie de la crédibilité ont été retenues : la théorie de la crédibilité partielle et une variante de la crédibilité de Bühlmann, adaptée à la mortalité par Klugman. Ces deux méthodes ont ensuite été intégrées à la construction de tables de mortalité prospective à l'aide de la méthode *Standardized Mortality Ratio* appliquée avec comme référence les tables réglementaires.



# Abstract

**Keywords :** longevity risk, life annuity, Best estimate, sampling risk, credibility theory, Standardized Mortality Ratio

The Solvency II regulations require insurers to carry out a *best estimate* assessment of the flow of discounted future benefits. Knowing insured's behavior is therefore essential. In the case of the life annuity, a mortality rate study should take into account this behavioural component.

Currently, the prospective mortality tables used by SOGECAP are the regulatory tables TGH / TGF005. The objective of this work is to integrate internal experience into the assessment of the longevity risk run by the company.

At first, it was essential to complete the information available by reviewing other portfolios of the company. To make the best use of this information, a partitioning according to the wealth levels was carried out.

Then, two methods from the theory of credibility were selected : the theory of partial credibility and a variant of the credibility of Bühlmann, adapted to mortality by Klugman. These two methods were then integrated into the construction of prospective mortality tables using the *Standardized Mortality Ratio* method applied with the regulatory tables as a reference.



# Remerciements

Tout d'abord, je tiens à remercier Monsieur Reda Jarir pour avoir encadré cette alternance, mais aussi pour sa bienveillance, ses conseils avisés et son partage des connaissances.

Mes remerciements vont aussi à Madame Elodie Durand, manager de l'équipe Pilotage technique et réassurance, pour m'avoir permis de faire cette alternance au sein de son équipe, et d'avoir contribué à son bon déroulement.

Je remercie également Bastien Laizet pour son écoute et ses remarques pertinentes, ainsi que l'ensemble du service pour leur accueil et leur chaleur. Chacun d'entre eux m'a aidée à réaliser cette alternance dans d'excellentes conditions.

Enfin, je remercie Monsieur Christian Robert, professeur à l'ISFA, pour ses relectures rigoureuses.





# Table des matières

<b>1</b>	<b>Introduction</b>	<b>11</b>
1.1	Le risque de longévité . . . . .	11
1.2	Les contrats de retraite supplémentaire . . . . .	11
1.3	Les tables de mortalité . . . . .	12
1.4	Problématique . . . . .	13
<b>2</b>	<b>Le portefeuille de l'étude</b>	<b>15</b>
2.1	Les données à disposition . . . . .	15
2.2	Composition du portefeuille . . . . .	16
2.2.1	Les produits et contrats de l'étude . . . . .	16
2.2.2	Les assurés . . . . .	18
2.2.3	Sinistres observés . . . . .	20
2.3	Critères pour la construction d'une table de mortalité . . . . .	20
2.3.1	Critères théoriques . . . . .	20
2.3.2	Validation des critères précédents . . . . .	23
2.4	Pistes envisagées . . . . .	23
<b>3</b>	<b>Le portefeuille épargne</b>	<b>25</b>
3.1	Les données à disposition . . . . .	25
3.2	Construction d'une table de mortalité du moment . . . . .	26
3.2.1	Principe de construction . . . . .	26
3.2.2	Estimation des taux bruts par la méthode de Kaplan-Meier . . . . .	27
3.2.3	Lissage par la méthode de Whittaker-Henderson . . . . .	28
3.2.4	Prolongement d'une table d'expérience . . . . .	31
3.2.5	Validation d'une table d'expérience . . . . .	32
3.3	Niveau de gamme . . . . .	32
3.3.1	Principe des méthodes de partitionnement . . . . .	33
	Inerties et nuage de points . . . . .	33
	Méthode des k-moyenne . . . . .	35
	Méthode de classification par ascendant hiérarchique . . . . .	37
	Choix du nombre de classes . . . . .	38
3.3.2	Application du partitionnement . . . . .	39
	Partitionnement pour les hommes . . . . .	39
	Partitionnement pour les femmes . . . . .	44

3.3.3	Comparaison avec le portefeuille rentes . . . . .	48
3.3.4	Conclusion . . . . .	53
<b>4</b>	<b>Utilisation de la théorie de la crédibilité partielle pour la construction de tables prospectives</b>	<b>55</b>
4.1	La théorie de la crédibilité partielle . . . . .	55
4.2	Méthodes de construction des tables prospectives . . . . .	57
4.2.1	Le modèle Standardized Mortality Ratio . . . . .	57
4.2.2	Fermeture des tables . . . . .	57
4.2.3	Utilisation de la crédibilité partielle . . . . .	58
4.3	Application aux données . . . . .	59
<b>5</b>	<b>Utilisation de la théorie de Klugman</b>	<b>67</b>
5.1	Principe de la méthode . . . . .	67
5.1.1	Etude du facteur de crédibilité . . . . .	68
5.1.2	Estimation de $\mu$ et $\sigma^2$ . . . . .	71
5.2	Application à nos données . . . . .	73
5.2.1	Les portefeuilles de l'étude . . . . .	73
5.2.2	Validation des tables obtenues . . . . .	74
5.2.3	Comparaison des tables obtenues . . . . .	77
Première intuition	. . . . .	78
Méthodologie de comparaisons	. . . . .	78
Application	. . . . .	79
<b>6</b>	<b>Conclusion</b>	<b>81</b>
<b>7</b>	<b>Annexes</b>	<b>83</b>
7.1	Annexe 1 - Statistiques descriptives du portefeuille rentes . . . . .	83
7.2	Annexe 2 - Méthode de Wittaker-Henderson en dimension 2 . . . . .	84
7.3	Annexe 3 - Etude graphique pour les hommes Théorie de la crédibilité partielle	85
7.4	Annexe 4 - Etude graphique pour les femmes Théorie de la crédibilité partielle	87
7.5	Annexe 5 - Etude graphique pour les homme Théorie de la crédibilité de Klugman	89

# Chapitre 1

## Introduction

### 1.1 Le risque de longévité

La longévité est définie par *Le Larousse* comme la "longue durée de la vie".

En France, en 2018, un homme de 60 ans vivrait encore, en moyenne, 23,2 ans, et une femme 27,6 ans. En vingt ans, l'espérance de vie à 60 ans a progressé de 2,3 ans pour les femmes et 3,2 ans pour les hommes. Ainsi, en moyenne, les français qui atteignent l'âge de 60 ans vivent de plus en plus vieux. En d'autres termes, la longévité des Français augmente.

Si cet allongement de la durée de vie moyenne est synonyme de pays développé et d'une qualité de vie correcte, il est en contrepartie associé à un risque financier important pour l'assureur. En effet, vivre plus longtemps au-delà d'un certain âge implique de percevoir une retraite ou une rente viagère plus longtemps, d'engendrer des frais d'assistance et de dépendance plus élevés.

Le risque de longévité se définit comme un risque de long terme, qui correspond au risque financier associé au fait que les individus vivent en moyenne significativement plus longtemps que prévu.

Dans un contexte où le progrès médical est permanent, comprendre et appréhender le risque de longévité est devenu essentiel pour les acteurs du milieu financier et assurantiel.

### 1.2 Les contrats de retraite supplémentaire

En assurance, les contrats qui sont directement liés au risque de longévité sont ceux qui entraînent le versement d'une rente viagère, ou dans une certaine mesure une rente temporaire. Ces contrats font partie des contrats de retraite supplémentaire.

Les contrats de retraite supplémentaire se répartissent en deux grandes familles :

- les **retraites d'entreprise** : Elles font partie du pilier 2 des retraites. Ces contrats sont mis en place par l'employeur pour une catégorie objective de salariés. Les sorties en rentes peuvent être aussi bien obligatoires que facultatives.
- les **retraites individuelles** : Elles font partie du pilier 3 des retraites. Ces contrats sont souscrits par les assurés directement auprès d'organismes d'assurances.

La retraite supplémentaire fonctionne par capitalisation. Elle recouvre un ensemble varié de dispositifs d'épargne qui se répartissent en deux principaux régimes, selon le mode de calcul de la rente ou du capital à la suite de la liquidation du contrat :

- Les régimes à **cotisations définies** : L'engagement porte sur le montant des cotisations versées périodiquement par le crédirentier et non sur le montant des prestations futures, qui sera déterminé au moment de la liquidation de la rente.
- Les régimes à **prestations définies** : L'engagement porte sur le montant de la prestation déterminé à la souscription du contrat qui est garanti. Ce régime ne concerne que des retraites supplémentaires d'entreprise. Ces contrats sont très souvent connus sous le nom "article 39" du fait de la fiscalité qui dépend de l'article 39 du code général des impôts (CGI). Les rentes viagères bénéficient d'un abattement de 10% pour l'impôt sur le revenu et les cotisations versées par l'entreprise ne sont pas soumises à l'impôt sur le revenu.

### 1.3 Les tables de mortalité

Pour modéliser le risque de longévité, démographes et actuaires ont construit des tables de mortalité qui fournissent une estimation de la probabilité de décès pour chaque individu d'une population.

Selon la segmentation choisie, différents types de table existent :

- Les *tables de mortalité du moment* : ces tables sont calculées à partir de l'observation d'une population sur une période de trois à cinq ans. La probabilité de décès dépend uniquement de l'âge de l'assuré.
- les *tables de mortalité prospectives* : ces tables ont pour but de modéliser la mortalité d'une population génération par génération. Ces tables nécessitent un volume plus important de données. La probabilité de décès dépend donc de l'âge de l'assuré et de son année de naissance.

Pour la construction des tables de mortalité, deux grands types de méthodes se distinguent :

- Les **modèles endogènes** : Les paramètres de ces modèles dépendent uniquement de la population observée. Ils nécessitent un grand nombre de données. On peut citer pour les tables de mortalité prospectives le modèle de Lee-Carter introduit en 1992 pour étudier la mortalité de la population américaine, ou bien le modèle de Cairns-Blake-Dowd développé en 2006.
- Les **modèles exogènes** : Ces modèles font appel à une table de mortalité de référence externe. On les utilise dans le cas où peu de données sont disponibles. On peut par exemple citer le modèle de Brass.

## 1.4 Problématique

La réglementation Solvabilité II, mise en place en janvier 2016, impose aux assureurs d'effectuer une évaluation *best estimate* du flux des prestations futures actualisées. La connaissance du comportement des assurés est donc primordiale. Dans le cas de l'étude des rentes, la mortalité du portefeuille doit donc être modélisée dans ce sens.

Actuellement, les tables de mortalité prospectives utilisées par SOGECAP sont les tables réglementaires TGH/TGF005 construites en 2005 à partir des données de dix-neuf portefeuilles de rentiers français provenant de différents assureurs et organismes. Ces tables ont été construites à l'aide d'un modèle de Brass segmenté suivant des classes d'âges et positionné par rapport à une table de référence construite à partir des données de l'INSEE pour la population française entre 1962 et 2000.

L'objectif de ce travail est d'intégrer l'expérience interne à l'évaluation du risque de longévité encourue par l'entreprise.



# Chapitre 2

## Le portefeuille de l'étude

Dans cette partie, nous étudions les caractéristiques intrinsèques du portefeuille rentes : le profil des assurés, les sinistres observés, les contrats étudiés ... afin d'effectuer les choix les plus judicieux pour, par la suite, modéliser au mieux la mortalité des assurés. Nous devons notamment établir à quelle maille réaliser cette étude, si l'expérience de notre portefeuille est suffisante ou bien si nous devons utiliser une autre source d'information.

### 2.1 Les données à disposition

Pour réaliser cette étude, nous commençons par analyser les données disponibles dans les bases de l'entreprise. Nous extrayons les informations des contrats depuis 2010 concernant les produits permettant une sortie en rentes. Pour certains contrats, la sortie en rentes est facultative. L'assuré peut choisir entre un versement unique ou au contraire le versement d'une rente temporaire ou viagère.

Aussi, nous disposons d'une base contenant environ 60 000 lignes. Cependant, 76% des polices donnent lieu à une sortie en capital. Ces contrats ne rentrant pas dans l'étude de la mortalité, la base de l'étude compte finalement 14 000 polices observées entre le **01/01/2010** et le **30/06/2018**.

Nous avons aussi regroupé le portefeuille par tête. En effet, certains assurés possèdent plusieurs contrats ce qui pourrait biaiser l'étude de la mortalité. Nous avons environ 13 000 têtes.

#### Informations à disposition

Pour ces contrats, nous disposons des informations suivantes :

- Informations sur les assurées (Nom, Date de naissance, Genre)
- Information sur le bénéficiaire (Nom, Date de naissance, Genre)
- Taux de réversion (si bénéficiaire)
- Capital constitutif et montant des provisions mathématiques au 31/12/2017.

- Date d'entrée, Date de sortie, Date de décès. Dans le cas des rentes temporaires, une sortie peut ne pas être dû à un décès.

## Prise en compte des réversions

Les rentes en cours de réversion sont intégrées à l'étude, car présentant le même risque de longévité. Concrètement, au décès du crédientier, la rente est clôturée avec comme date de sortie la date décès. Une nouvelle rente est ajoutée au nom du bénéficiaire avec comme date d'entrée le lendemain du décès. Le montant de la rente versée est le taux de réversion multiplié par le montant de la rente d'origine. Le montant en capital de la nouvelle rente est égal à celui de la rente d'origine. Ce montant est choisi en prévision d'une segmentation future. Ces contrats représentent 2% de l'exposition totale.

## 2.2 Composition du portefeuille

### 2.2.1 Les produits et contrats de l'étude

#### Exposition

Le portefeuille de l'étude n'est pas fermé. Chaque année, des contrats ont été liquidés et sont entrés en phase de restitution. Inversement, des assurés sont sortis. Cela peut être dû au décès, mais aussi à la fin du versement d'une rente temporaire ou bien pour cause de rachat.

Pour ne pas perdre d'informations, nous choisissons de calculer l'exposition suivant le nombre de jours observés pour chaque individu. Plus précisément, dans le cas d'une censure, la durée d'observation est calculée au prorata des jours observés. Dans le cas d'un décès, la date de sortie est la date de fin d'observation, et non la date de décès. En effet, un individu décédé au cours de la période d'exposition est observé entièrement dans le sens où on connaît son état à chaque instant.

Finalement, nous calculons l'exposition sur chaque année civile :

Années	2010	2011	2012	2013	2014	2015	2016	2017
Exposition	4516	5126,4	5897,7	6781,2	7778	8813,6	9975,5	11227,2

#### Types de contrats

Les contrats rentes sont variés et présentent souvent des caractéristiques qui leur sont propres. Toutefois, une étude à une maille élevée est utile pour comprendre les types de contrats composant le portefeuille, et donc des profils des assurés, pour ainsi mieux cerner les risques qui lui sont intrinsèques.



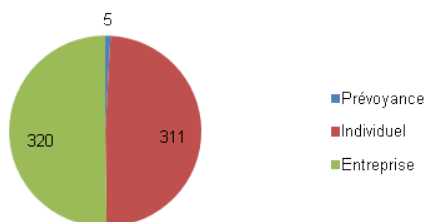


FIGURE 2.1 – Type de contrats au 31/12/2017

Comme le montre le graphique 2.1, nous disposons d’une répartition homogène des contrats individuels et collectifs qui évolue peu durant les années d’observation. Nous choisissons donc de ne pas faire de distinctions entre les types de contrats dans la suite de ce travail.

### Types de fiscalité

De même, la fiscalité donne des informations sur le type de rentes versées. Celles-ci peuvent être à prestations ou à cotisations définis. Les premières sont souvent destinées à des personnes ayant des postes importants et donc un niveau de richesses plus élevé. Or niveau de richesse et longévité sont des variables souvent corrélées.

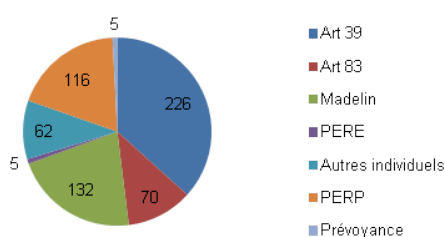


FIGURE 2.2 – Type de fiscalité au 31/12/2017

Ce graphique 2.2 illustre le poids important que représentent les contrats à prestations définies dans notre base. En effet, nous observons 1583 contrats de fiscalité ART39 ce qui correspond à environ 11% des contrats. Pour autant, ces contrats correspondent à 33% des capitaux constitués. En outre, le montant moyen des rentes au 31/12/2017 était à 2900€ pour le portefeuille alors qu’il s’élevait à 7500€ pour les contrats à prestation définie.

Les assurés de notre portefeuille présentent donc des profils socio-économiques disparates ce qui nous incite à approfondir leurs caractéristiques.

## 2.2.2 Les assurés

### Genres

Le niveau de mortalité, au sein de la population française, des hommes et des femmes diffèrent en fonction de l'âge. Ainsi, suivant les parts d'hommes et de femmes observées pour un âge donné, le taux de mortalité peu varier. Il est donc important de savoir comment la répartition hommes/femmes évolue d'une année sur l'autre.

Le graphique suivant présente l'évolution de la répartition des sexes au sein du portefeuille entre 2010 et 2017.

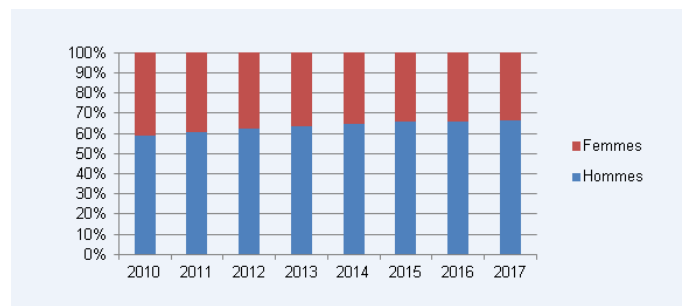


FIGURE 2.3 – Évolution de la répartition hommes/femmes

La part des hommes est de 58% en 2010 alors qu'en 2016, elle s'élève à 67%. En six ans, nous constatons une augmentation de 10% de la part des hommes. La répartition hommes/-femmes évolue significativement durant les années d'observation de notre portefeuille, due à une entrée plus importante d'hommes que de femmes.

### Âges

La mortalité du portefeuille est corrélée à l'âge moyen des assurés. De même, le risque de longévité d'un assuré est d'autant plus important qu'il présente un poids financier significatif. C'est pourquoi nous étudions l'âge moyen et l'âge moyen pondéré par le montant des rentes en distinguant les hommes des femmes.

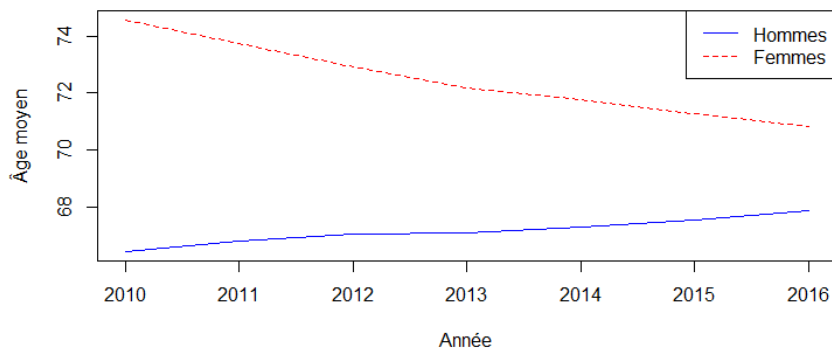


FIGURE 2.4 – Évolution de l'âge moyen des hommes et des femmes au sein du portefeuille rentes

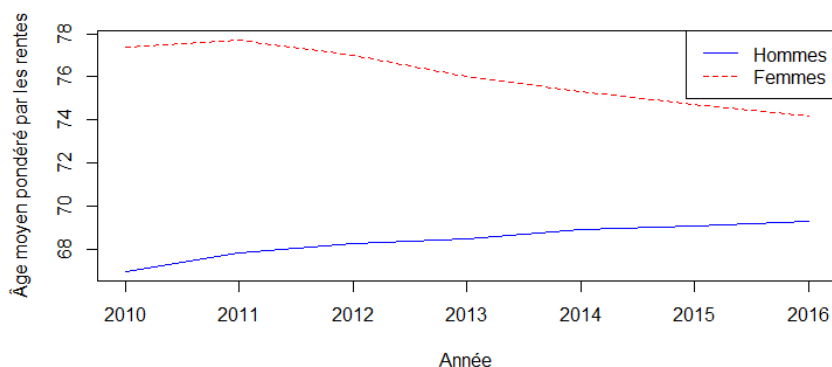


FIGURE 2.5 – Évolution de l'âge moyen des hommes et des femmes pondéré par les montants de rentes au sein du portefeuille rentes

L'âge moyen du portefeuille n'évolue pas de la même façon pour les hommes que pour les femmes. Celui des hommes est stable avec le temps alors que l'âge moyen des femmes chute de 75 à 71 ans. Ces âges élevés sont cohérents avec le produit étudié et logiquement notre exposition concerne des assurés de 60 ans et plus.

L'âge pondéré par les rentes permet d'identifier les âges des assurés ayant le plus grand poids financiers. Chez les hommes, ce sont les assurés d'environ 70 ans qui ont les contrats les plus élevés. Pour les femmes, ce sont les assurées d'environ 80 ans. Cela s'explique par la longévité moyenne plus importante des femmes, et qui sont, par conséquent, plus concernées par les rentes de réversion. Notre portefeuille de rentes semble donc assez hétérogène concernant les capitaux engagés.

### 2.2.3 Sinistres observés

Les sinistres concernent, dans notre cas, les décès d'un assuré ayant une rente en cours de restitution.

L'extraction des données a eu lieu au 30/06/2018, aussi la compagnie peut ne pas avoir encore eu connaissance à cette date de l'ensemble des décès de l'année 2017. Nous choisissons donc d'exclure cette année de notre étude. Notre période d'observation s'étend finalement du 01/01/2010 au 31/12/2016. Nous observons les sinistres suivants :

Années	2010	2011	2012	2013	2014	2015	2016	Total
Hommes	37	65	59	54	73	90	91	469
Femmes	70	66	73	45	81	87	73	495

Une chute du nombre de décès a eu lieu au cours de l'année 2013 aussi bien pour les femmes. Aucune explication particulière n'ayant été trouvée, et après vérification des données avec différentes entités de l'entreprise, nous choisissons de garder cette année dans l'étude.

On s'intéresse maintenant à la répartition des âges des décès. Ayant un nombre restreint de données, nous observons des trous dans les données, c'est-à-dire que pour certains âges nous observons très peu, voir pas du tout de décès. Cela nous contraint d'effectuer nos analyses sur des plages d'âges limitées :

- [60, 99] pour les hommes
- [76, 105] pour les femmes

### Risque d'estimation

Ces informations nous mettent en alerte face au risque d'estimation que présente notre échantillon. Ce risque correspond au fait que les observations ne permettent pas de modéliser correctement le comportement de la population étudiée. Nous allons donc devoir réaliser la modélisation de la mortalité de manière à limiter ce risque.

Nous nous intéressons maintenant à la possibilité de construire une table de mortalité à partir de nos données.

## 2.3 Critères pour la construction d'une table de mortalité

### 2.3.1 Critères théoriques

Nous avons mis en évidence les principales caractéristiques du portefeuille rente. Nous définissons par la suite trois critères pour voir si la construction d'une table de mortalité est envisageable sur le portefeuille.

## Profondeur d'historique

Nous avons vu que la mortalité évolue au fil des ans. La profondeur d'historique se doit d'être choisi de manière judicieuse suivant qu'on souhaite construire une table de mortalité du moment ou prospective.

Pour une table d'expérience, le résultat ne doit pas être biaisé par une éventuelle tendance. L'institut des actuaires recommande dans sa note [IA, 2006] d'avoir un historique de 3 à 5 ans d'ancienneté.

Dans le cadre de la construction d'une table de mortalité prospective, le but est de prendre en compte la tendance de la mortalité en calculant des taux de mortalité par génération. Il faut donc une profondeur d'historique conséquente. Pour des modèles endogènes, 10 à 20 ans d'historique sont recommandés contre 5 à 15 ans pour des modèles exogènes.

## Critère de Cochran

Le critère de Cochran permet de déterminer si on dispose d'un nombre suffisant de données pour construire une table de mortalité.

Pour une table du moment, on doit observer sur au moins 80% des classes d'âge cinq décès et sur chaque classe d'âge au moins un décès. Formellement, avec les notations usuelles :

$$N_x \widehat{q}_x > 5, N_x(1 - \widehat{q}_x) > 5$$
$$d_x > 0$$

avec :

- $N_x$  le nombre d'individus d'âge  $x$  observés
- $\widehat{q}_x$  le taux de mortalité à l'âge  $x$

Pour une table de mortalité prospective, le critère se construit sur les classes d'âge par génération. On a alors :

$$N_{xt} \widehat{q}_{xt} > 5, N_{xt}(1 - \widehat{q}_{xt}) > 5$$
$$d_{xt} > 0$$

Ce critère est issu des travaux de W. G. Cochran dans son article "Some Methods for Strengthening the Common  $\chi^2$  Tests" [Cochran, 1954].

Notons que le choix de la plage de construction se doit d'être judicieux afin de satisfaire les conditions, ce qui en fait un critère assez souple.

## Crédibilité totale

La théorie de la crédibilité totale a pour objectif de définir un critère pour que l'estimateur du taux de mortalité  $\hat{q}$  respecte une précision choisie  $k$  avec une probabilité  $p$ . Si ce critère est vérifié, le risque d'estimation lié à la modélisation réalisée est limité.

Il est usuellement retenu  $k = 0.05$  et  $p = 0.90$ .

Formellement,

$$P\left(\frac{\hat{q} - q}{q} \leq k\right) \geq p$$

$$P(-kq \leq |\hat{q} - q| \leq kq) \geq p$$

Le théorème central limite permet, pour un nombre conséquent d'observations, de faire l'hypothèse que  $\hat{q} - q$  suit une loi normale  $N(q, \sigma_q^2)$ .

Par réduction,

$$P\left(\frac{-kq}{\sigma_q} \leq Z \leq \frac{kq}{\sigma_q}\right)$$

où  $Z \sim N(0, 1)$ .

Puis, par symétrie de la loi normale centrée, nous avons :

$$2P\left(Z \leq \frac{kq}{\hat{q}}\right) \geq p$$

Donc, en notant  $\epsilon_{p/2}$  le quantile d'ordre  $p/2$  :

$$\frac{kq}{\sigma_q} \geq \epsilon_{p/2}$$

ce qui constitue un critère pour savoir si l'estimateur  $\hat{q}$  est crédible. Or  $\sum_{i=1}^N D_i \sim B(N, q)$  donc  $\sigma_q^2 = \frac{q(1-q)}{N}$ . Ainsi :

$$kq * \sqrt{\frac{N}{q(1-q)}} \geq \epsilon_{p/2}$$

Et en supposant  $q$  proche de 0 :

$$N * q \geq \left(\frac{\epsilon_{p/2}}{k}\right)^2$$

Cette formulation a l'avantage de donner une condition sur le nombre de décès à observer. Pour  $k = 0.05$  et  $p = 0.90$ , 1082 sinistres doivent être observés.

### 2.3.2 Validation des critères précédents

Nous regardons maintenant si notre portefeuille vérifie ces critères pour la construction d'une table du moment.

Nous disposons de six années d'historique exploitable, ce qui nous permet d'envisager la construction d'une table du moment ou bien une table prospective par une méthode à référence externe.

Le nombre de sinistres observés est limité, d'autant que nous devons distinguer les hommes des femmes, ce qui restreint la plage d'âge sur laquelle le critère de Cochran est vérifié. Nous obtenons pour une table d'expérience :

- [60, 99] pour les hommes
- [76, 105] pour les femmes

Enfin, le critère de crédibilité totale n'est ni vérifié pour les hommes, ni pour les femmes. Nos observations ne sont pas suffisantes pour avoir confiance dans l'estimation des taux de mortalité.

## 2.4 Pistes envisagées

L'étude du portefeuille rentes nous a permis de comprendre pourquoi il était nécessaire dans la suite de l'étude de considérer les hommes et les femmes séparément. En effet, la répartition hommes/femmes évolue significativement durant la période d'observation. De plus, les caractéristiques de leur profil divergent fortement.

Nous avons finalement :

	Plage des âges exploitable	Nombre de décès	Période d'étude
Hommes	[60,99]	488	[2010,1016]
Femmes	[76,105]	533	[2010,1016]

Enfin, notre portefeuille ne vérifie pas tous les critères pour permettre la construction d'une table de mortalité du moment. En particulier, le critère de crédibilité totale n'est pas satisfait.

Les informations données par le portefeuille rentes ne nous permettent pas d'envisager la construction de tables de mortalité en nous basant uniquement sur son expérience. En effet, construire une table de mortalité revient à estimer une moyenne des décès. Cependant, avec un nombre limité d'observations, l'erreur induite par l'estimateur peut être importante. Idéalement, nous aimerions trouver une autre source d'information qui présente des caractéristiques proches pour compléter nos données afin de limiter cette erreur.

Nous décidons donc nous intéresser aux autres portefeuilles de notre entreprise.

Les produits rentes peuvent être vus comme des produits épargne dans leur phase de constitution. Quels liens existent-ils entre les contrats en sortie en capital et ceux en rentes ?

### **Rentes contre capital**

Intéressons-nous à ces deux possibilités.

D'une part, la sortie en rentes permet de toucher un versement régulier jusqu'au décès. C'est un choix intéressant si on vit au-delà de l'espérance de vie estimée. Son principal désavantage est qu'il ne permet pas la transmission d'un capital, si ce n'est une rente de réversion à un bénéficiaire désigné. D'après les chiffres de l'[INSEE, 2018], il existe un lien entre niveaux de vie et mortalité. Ces contrats devraient donc concerner une population aisée. Toutefois, certains contrats d'épargne-retraite imposent une sortie en rentes.

D'un autre côté, la sortie en capital permet de retirer en une ou plusieurs fois le capital épargné. De plus, les capitaux restants se transmettent au décès du crédirencier.

Finalement, ces deux populations ont pour point commun d'avoir constitué un capital. Il n'existe pas d'étude permettant de lier la mortalité de rentiers à celle d'un portefeuille épargne. Cependant, nous faisons l'hypothèse que le capital constitué nous donne une indication sur le niveau de richesse de l'assuré.

Nous décidons de nous intéresser au portefeuille épargne de l'entreprise afin de voir si son information peut nous aider à étudier les lois de mortalité du portefeuille rentes.



# Chapitre 3

## Le portefeuille épargne

Disposant d'observations limitées sur le portefeuille rentes, nous nous intéressons au portefeuille épargne. L'objectif de ce chapitre est de déterminer s'il est possible d'utiliser l'expérience du portefeuille épargne pour compléter celle du portefeuille de l'étude. Pour cela, nous devons comparer les caractéristiques de chacun des deux portefeuilles, et éventuellement déterminer un portefeuille fictif plus pertinent.

### 3.1 Les données à disposition

La gamme de produits épargne à notre disposition est composée de dizaines de produits, ventilés de manière plus fine dans le système d'information, afin de décrire les spécificités en terme de montant minimum, de taux garantis et de profil de gestion.

Nous procédons à une extraction afin d'observer les sinistres sur la période 01/01/2010 - 31/12/2016 qui correspond à la période sur laquelle nous avons connaissance des décès au moment de l'extraction.

Nous disposons pour chaque contrat des informations suivantes :

- Nom, Prénom, Date de Naissance
- Date d'entrée, Date de sortie
- Montant en PM
- Portefeuille

Des retraitements sont effectués pour regrouper par tête ces informations. Le montant de l'encours considéré est donc la somme des montants de chaque contrat.

Pour l'étude de la mortalité, nous ne retenons que les contrats concernant des personnes physiques. Les personnes morales sont donc exclues.

Enfin, nous scindons directement ce portefeuille en deux sous-ensembles suivant les genres

des assurés. En effet, nous avons vu que l'étude du portefeuille rentes nécessite de distinguer les hommes et les femmes, ce qui doit être aussi le cas pour le portefeuille épargne si nous souhaitons utiliser cette source d'information complémentaire.

Ces deux portefeuilles ont les caractéristiques suivantes.

Années	2010	2011	2012	2013	2014	2015	2016
Exposition hommes	787 526	802 994	812 523	808 451	811 850	824 236	849 116
Décès hommes	12 927	13 211	14 409	14 084	13 609	14 707	14 187
Exposition femmes	915 201	915 201	915 201	908 925,3	910 570	921 617	946 314
Décès femmes	18 358	18 358	18 358	18 201	17 090	17 803	17 793

Pour les besoins de modélisation de la mortalité, une segmentation selon les différents produits n'est pas nécessaire. Cependant, une étude selon les niveaux de richesse peut s'avérer pertinente, d'autant que la répartition de ces niveaux peut varier de celle du portefeuille rentes.

Pour que les informations du portefeuille Épargne soient exploitables, nous regardons maintenant si celui-ci vérifie les critères de construction des tables de mortalité défini au chapitre précédent.

Pour la profondeur d'historique, nous disposons de six années, ce qui est suffisant pour envisager la construction d'une table de mortalité du moment ou bien prospective. Le critère de Cochran est quant à lui vérifié sur la plage  $[40, 100]$  pour chaque genre. Nous disposons également d'un grand nombre de sinistres qui permettent de vérifier le critère de crédibilité totale pour chaque âge.

Le portefeuille épargne dispose de suffisamment d'informations pour construire des tables de mortalité par genre en se basant uniquement son expérience.

Après avoir décrit la méthode de construction des tables d'expérience, nous essayons de déterminer différents niveaux de gamme pour ensuite comparer les lois de décès du portefeuille épargne aux lois décès observés du portefeuille rentes.

## 3.2 Construction d'une table de mortalité du moment

### 3.2.1 Principe de construction

Dans cette section, nous expliquons la démarche utilisée pour construire les tables de mortalité du moment.

La construction d'une table de mortalité d'expérience se déroule en trois principales étapes que nous détailleront par la suite :

**Étape 1 :** Les taux bruts de mortalité sont estimés. Différents estimateurs ont été développés dans la littérature actuarielle. Pour notre étude, nous choisissons d'utiliser celui de Kaplan-Meier.

**Étape 2 :** Pour éliminer les disparités liées à la volatilité des données, un lissage est effectué. Nous choisissons d'appliquer un lissage de Whittaker-Henderson.

**Étape 3 :** La table est fermée en complétant les données aux âges extrêmes avec une source externe d'information.

### 3.2.2 Estimation des taux bruts par la méthode de Kaplan-Meier

Nous sommes en présence de données censurées dans notre échantillon. Pour estimer les taux bruts, nous choisissons d'appliquer la méthode de Kaplan Meier qui permet de prendre en compte les censures. Cette méthode non-paramétrique a été introduite en 1958 par [Kaplan and Meier, 1958]. Nous aurions également pu choisir d'utiliser l'estimateur paramétrique d'[Hoem, 1971] et une comparaison entre ces deux outils aurait pu être intéressante.

Soient  $S$  la probabilité de survie à l'instant  $t$  et  $T$  l'instant de décès. Par définition, pour tout  $t > s$  :

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

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

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

C'est sur cette propriété de la fonction de survie que repose la méthode de Kaplan-Meier.

Pour chaque âge, l'exposition se fait sur  $[t_x, t_{x+1}[$  et la probabilité de survivre sur cet intervalle s'écrit :

$$p_x = P(t > t_{x+1} | T > t_x)$$

Par la propriété précédente :

$$p_x = \frac{P(T > t_{x+1})}{P(T > t_x)}$$

On subdivise l'intervalle  $[t_x, t_{x+1}[$  en  $t_1 = t_x, \dots, t_i, \dots, t_n = t_{x+1}$  où chaque instant  $t_i$  correspond à un évènement (censure droite, entrée ou décès) et on décompose :

$$p_x = \frac{P(T > t_n)}{P(T > t_{n-1})} * \dots * \frac{P(T > t_i)}{P(T > t_{i-1})} * \dots * \frac{P(T > t_2)}{P(T > t_1)} = \prod_{i=1}^{n-1} \frac{S(t_{i+1})}{S(t_i)}$$

On peut alors associer la probabilité de décès à l'âge  $x$  :

$$q_x = 1 - p_x$$

Un estimateur naturel de  $q_x$  est  $\hat{q}_x = \frac{d_x}{n_x}$  avec :

- $d_x$  le nombre de décès à l'âge  $x$
- $n_x$  l'exposition à l'âge  $x$

De même, pour chaque intervalle  $[t_i, t_{i-1}[$ , on a la probabilité de décès :  $q_i = 1 - p_i = \frac{d_i}{n_i}$  On déduit par itération :

$$\frac{S(t_n)}{S(t_1)} = \prod_{i=1}^{n-1} (1 - q_i) = \prod_{i=1}^{n-1} \left(1 - \frac{d_i}{n_i}\right)$$

Ainsi l'estimation de la probabilité de décès à l'âge  $x$  est :

$$\hat{q}_x = 1 - \frac{\widehat{S}(t_{x+1})}{\widehat{S}_{t_x}}$$

$$\hat{q}_x = 1 - \left(\prod_i \left(1 - \frac{d_i}{n_i}\right)\right)$$

avec

$$n_i = n_{i-1} - d_{i-1} - c_{i-1} + e_{i-1}$$

où :

- $d_{i-1}$  le nombre de décès entre  $[t_{i-1}, t_i]$
- $c_{i-1}$  le nombre de données censurées à droite entre  $[t_{i-1}, t_i]$
- $e_{i-1}$  le nombre d'entrée entre  $[t_{i-1}, t_i]$

### 3.2.3 Lissage par la méthode de Whittaker-Henderson

Nous venons d'estimer les taux bruts de mortalité. Cependant, on remarque en traçant la représentation graphique un bruit parasite. Cette volatilité est liée aux fluctuations de l'échantillon. Nous allons donc chercher à éliminer ce bruit en appliquant une méthode de lissage et ainsi obtenir une courbe de mortalité qui croît avec l'âge.

Ces méthodes ont été largement détaillées dans la littérature actuarielle. Nous choisissons d'appliquer la méthode de lissage non-paramétrique de Whittaker-Henderson. Dans le cadre de la construction d'une table d'expérience, nous appliquons ce modèle à une dimension. Cette méthode peut ensuite se généraliser dans le cas à deux dimensions, ce qui peut se révéler utile pour la construction d'une table prospective.

Nous présentons dans la suite de ce paragraphe la cas à une dimension. Le cas de la dimension 2 est expliqué en annexe.

La méthode de Whittaker-Henderson repose sur deux critères :

- un critère de fidélité
- un critère de régularité

Le **critère de fidélité** mesure la distance, pondérée par un poids  $w_x$ , entre les taux lissés  $\tilde{q}_x$  et les taux bruts  $\hat{q}_x$  :

$$F = \sum_{x_{min}+1}^{x_{max}} w_x (\tilde{q}_x - \hat{q}_x)^2 \quad (3.1)$$

La pondération peut être définie de différentes manières.

Si on remarque que des données sont absentes, on peut poser  $w_x = 0$  si  $q_x = 0$  et  $w_x = 1$  sinon.

Dans notre cas, nous avons choisi un intervalle où le nombre de données est a priori suffisant, d'après les critères de Cochran et de crédibilité totale. Nous choisissons de définir la pondération de la manière suivante :

$$w_x = \frac{N_x}{\bar{N}}$$

où :

- $N_x$  est l'effectif observé à l'âge  $x$ .
- $\bar{N} = \frac{\sum_{x_{min}+1}^{x_{max}} N_x}{x_{max} - x_{min} + 1}$  l'effectif moyen sur la plage d'âge étudiée.

Cette pondération présente deux avantages :

- Elle permet de limiter le poids donné aux points où on observe peu d'individus.
- Elle permet d'obtenir le même nombre total de décès et le même âge moyen de décès pour les taux lissés que les taux bruts.

Le **critère de régularité** mesure une distance entre les taux lissés :

$$R = \sum_{x_{min}+1}^{x_{max}} (\Delta^z \tilde{q}_x)^2 \quad (3.2)$$

où  $z$  est un paramètre du modèle et son choix est à l'appréciation de l'actuaire. On a donc :

- Si  $z = 1$ ,  $(\Delta \tilde{q}_x)^2 = (\widetilde{q_{x+1}} - \tilde{q}_x)^2$ . On retrouve la distance euclidienne.
- Si  $z = 2$ ,  $(\Delta^2 \tilde{q}_x)^2 = (\widetilde{q_{x+2}} - 2\widetilde{q_{x+1}} + \tilde{q}_x)^2$ .

— Si  $z$ ,  $(\Delta^z \tilde{q}_x)^2 = \nabla(\nabla^{z-1} \tilde{q}_x)^2$ .

Généralement,  $z$  est compris entre 2 et 5. Cela permet de tenir compte de suffisamment d'informations consécutives sans pour autant perdre en trop, puisque la plage d'âge doit être restreinte à  $[x_{inf}, x_{sup}]$ .

Dans le cadre de ce travail, nous choisissons  $z = 2$ .

Le critère de Whittaker-Henderson est la combinaison linéaire de ces deux critères de fidélité et régularité. On ajoute le paramètre  $h$  pour donner plus d'importance à l'un ou l'autre d'un des deux critères. On obtient :

$$M = F + hR = \sum_{x_{min}+1}^{x_{max}} w_x (\tilde{q}_x - \hat{q}_x)^2 + h \sum_{x_{min}+1}^{x_{max}} (\Delta^z \tilde{q}_x)^2$$

Souhaitant donner autant de poids aux deux critères, nous posons  $h = 1$ .

Les taux lissés  $\tilde{q}_x$  sont ensuite choisis de manière à minimiser l'expression  $M$ . Pour déterminer les  $\tilde{q}_x$  solutions, nous résolvons le système suivant

$$\forall x, \frac{\delta M}{\delta \tilde{q}_x} = 0$$

en passant en écriture matricielle. On considère :

$$\begin{aligned} \hat{q} &= {}^t(\widehat{q}_{x_{inf}}, \dots, \widehat{q}_{x_{sup}}) \\ \tilde{q} &= {}^t(\widetilde{q}_{x_{inf}}, \dots, \widetilde{q}_{x_{sup}}) \\ W &= \text{diag}(w_{x_{inf}}, \dots, w_{x_{sup}}) \end{aligned}$$

En notation matricielle, le critère de fidélité se réécrit

$$F = {}^t(\tilde{q} - \hat{q})W(\tilde{q} - \hat{q})$$

et le critère de régularité

$$R = {}^t(\Delta^z \tilde{q})(\Delta^z \tilde{q})$$

Remarquons que le vecteur  $\Delta^z \tilde{q}$  peut se réécrire  $\Delta^z \tilde{q} = K_z \tilde{q}$  avec  $K_z$  la matrice des coefficients binomiaux avec alternance de signe.  $K_z$  est de taille  $(x_{sup} - x_{inf} + 1, x_{sup} - x_{inf} + 1 - z)$ . Pour cette étude, nous avons choisi  $z = 2$ . On a alors :

$$K_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}$$

On peut alors écrire :

$$M = F + hR$$

$$M = {}^t(\tilde{q} - \hat{q})W(\tilde{q} - \hat{q}) + h{}^t\tilde{q}{}^tK_zK_z\tilde{q}$$

Puis, en dérivant vectoriellement, le système à résoudre devient :

$$\frac{\delta M}{\delta \tilde{q}} = 2W\tilde{q} - 2W\hat{q} + 2h{}^tK_zK_z\tilde{q} = 0$$

Ce qui se réécrit :

$$(W + {}^tK_zK_z)\tilde{q} = W\hat{q}$$

$$c\tilde{q} = W\hat{q}$$

Dans le cas où  $c$  est inversible, on obtient :

$$\tilde{q} = c^{-1}W\hat{q}$$

Cette inversion peut se révéler fastidieuse, c'est pourquoi la décomposition de Cholesky est utilisée en pratique.

On obtient ainsi les taux lissés recherchés. Notons qu'il est possible de jouer sur les paramètres  $((w_x), z, h)$  afin d'obtenir la courbe la plus cohérente.

Maintenant que nous avons la loi de décès sur l'intervalle  $[40, 100]$ . Il reste à prolonger cette courbe en haut pour étendre la plage jusqu'à 120 ans.

### 3.2.4 Prolongement d'une table d'expérience

Pour cela, nous effectuons une régression linéaire par rapport à une table de référence en passant au *logit* :

$$\text{logit}(\tilde{q}_x) = a \text{logit}(q_x^{ref}) + b$$

La détermination des coefficients  $a$  et  $b$  peut s'effectuer par le critère des moindres carrés. La fonction perte à minimiser est alors :

$$e = \sum_{x_{min}+1}^{x_{max}} (\text{logit}(\hat{q}_x) - a \text{logit}(q_x^{ref}) + b)^2$$

L'intérêt d'utiliser la fonction *logit* est qu'elle permet de passer de l'intervalle  $[0,1]$  à  $]-\infty, +\infty[$ , et donc de réaliser l'approximation par une droite réelle. De plus, la fonction *logit* est inversible, ce qui permet de déduire facilement les taux  $\tilde{q}_x$  en appliquant  $f(x) = \frac{1}{1+e^{-x}}$ .

### 3.2.5 Validation d'une table d'expérience

Pour valider la courbe obtenue, on vérifie qu'on obtient bien des valeurs proches de 0 pour les deux critères énoncés précédemment.

Graphiquement, on vérifie que les décès observés se trouvent bien dans l'intervalle de confiance à 95% induit par la loi des décès.

Enfin, une étude des résidus peut s'avérer pertinente.

## 3.3 Niveau de gamme

L'espérance de vie d'un individu et son niveau de richesse sont corrélés positivement. Aussi est-il intéressant d'étudier la répartition des richesses du portefeuille épargne afin de déterminer une segmentation adéquate.

L'objectif est de pouvoir par la suite de comparer les proportions de chaque segment entre les portefeuilles épargne et rentes. Si ces proportions diffèrent fortement, il sera pertinent de considérer un portefeuille fictif créé en agrégeant les niveaux de gamme suivant la répartition du portefeuille rentes.

Deux approches se présentent pour réaliser cette segmentation :

**Méthode 1 :** Nous travaillons en vision par tête et déterminons les montants pivots de PM afin de regrouper les assurés par caractéristiques similaires. Pour cela, nous divisons le portefeuille en de nombreux sous-ensembles de petites tailles suivant le montant de l'encours pour ensuite les regrouper à l'aide de méthodes de partitionnement.

**Méthode 2 :** Nous travaillons sur les principaux produits épargnes et nous les regroupons par caractéristiques similaires. En effet, au lieu de partitionner le portefeuille par montant d'en cours, on divise les données suivant les familles de produit pour ensuite les regrouper suivant le montant moyen d'encours de chaque produit.

Cependant, le sujet de notre étude est le portefeuille des rentes. Les familles de produits des deux portefeuilles sont différentes, c'est pourquoi nous privilégions la méthode 1.

Pour réaliser le partitionnement, nous devons déterminer les variables explicatives à utiliser. Celles-ci doivent permettre de faire le lien entre le niveau des encours et le taux de



mortalité.

Nous retenons :

- Âge moyen de l'exposition : Plus l'âge moyen est élevé, plus la probabilité de décès est élevée et l'épargne constituée est élevée.
- Âge moyen des décès
- Ratio observés/attendus par rapport aux tables réglementaires du moment : Ce coefficient permet de mesurer la distance entre les groupes et la table de mortalité réglementaire. Nous choisissons d'utiliser les tables réglementaires du moment th002/tf002.

Ces variables ont des ordres de grandeur différents. Nous choisissons donc de les normaliser.

Nous nous intéressons maintenant aux méthodes de partitionnement retenues.

### 3.3.1 Principe des méthodes de partitionnement

Les méthodes de partitionnement font partie des méthodes de classification non supervisées d'un point de vue exploratoire. L'objectif de ces méthodes est de regrouper les observations en classes homogènes et distinctes sur des considérations géométriques.

On commence par définir ces deux notions ainsi que la notion de proximité dans un cadre géométrique.

#### Inerties et nuage de points

On se place dans un espace euclidien muni de sa distance usuelle  $d(.,.)$  et on considère un nuage de  $n$  points  $w_1, \dots, w_n$ .

Soit  $g$  le centre de gravité de ce nuage, l'**inertie totale** de cet ensemble de points est définie par

$$I_{total} = \sum_{l=1}^n d^2(w_l, g)$$

Soit une partition en  $K$  classes  $C_1, \dots, C_K$  de centre de gravité  $g_1, \dots, g_K$  et de nombres d'observations  $n_1, \dots, n_K$ . On définit l'**inertie de la classe**  $C_i$  par :

$$I_i = \frac{1}{n_i} \sum_{w_j \in C_i} d^2(g_i, w_j)$$

On définit l'**inertie interclasse** par :

$$I_{inter} = \sum_{i=1}^k n_i d^2(g, g_i)$$

L'inertie intraclasse mesure l'homogénéité de chaque classe. Plus sa valeur est faible, plus les éléments proches ont été rassemblés dans une même classe. On cherche donc à minimiser cette valeur.

On définit l'**inertie intraclasse** par :

$$I_{intra} = \sum_{i=1}^k I_i$$

L'inertie interclasse mesure l'hétérogénéité entre les classes. Une bonne partition doit donc avoir une inertie interclasse élevée.

Les méthodes de partitionnement reposent sur la **décomposition de Huygens** :

$$I_{total} = I_{inter} + I_{intra}$$

On se propose maintenant de démontrer ce résultat. Pour cela, on part de l'inertie totale

$$I_{total} = \sum_{l=1}^n d^2(w_l, g)$$

En ré-indexant suivant les classes de la partition :

$$I_{total} = \sum_{k=1}^K \sum_{i \in C_k} d^2(w_i, g)$$

$$I_{total} = \sum_{k=1}^K \sum_{i \in C_k} \|w_i - g_k + g_k - g\|^2$$

D'après le théorème de Huygens :

$$I_{total} = \sum_{k=1}^K \sum_{i \in C_k} \|w_i - g_k\|^2 + \sum_{k=1}^K \sum_{i \in C_k} \|g_k - g\|^2$$

$$I_{total} = \sum_{k=1}^K I_i + \sum_{k=1}^K n_k d^2(g_k, g)$$

Ce qui achève la démonstration.

Comme l'inertie totale est indépendant de la partition choisie, minimiser l'inertie intra-classe est équivalent à maximiser l'inertie interclasse.

Formellement, la partition optimale  $C_k^*$  est définie par :

$$C_k^* = \operatorname{argmin}_{C \in \mathcal{C}_K} \sum_{k=1}^K \sum_{i \in C_k} d^2(w_i, g_k)$$

ou de manière équivalente :

$$C_k^* = \operatorname{argmax}_{C \in \mathcal{C}_K} \sum_{k=1}^K n_k d^2(g_k, g)$$

où  $\mathcal{C}_K$  est l'ensemble des partitions constituées de  $K$  classes.

L'ensemble  $\mathcal{C}_K$  étant finie, la partition optimale existe. La complexité combinatoire rend cependant ce problème difficile à résoudre. Le but des algorithmes de partitionnement est donc de réduire le temps de calcul en ne testant pas toutes les partitions une à une. Nous présentons, par la suite, les deux algorithmes liés aux méthodes des k-moyennes et de classification par ascendant hiérarchique.

### Méthode des k-moyenne

On considère un ensemble de données. Le but de cette méthode est de regrouper les données en  $K$  classes,  $K$  étant fixé a priori.

Quatre étapes peuvent être distinguées :

**Étape 1** : On choisit aléatoirement  $K$  individus qui seront les centres des  $K$  classes.

**Étape 2** : On attribue chaque objet à la classe la plus proche.

**Étape 3** : Connaissant les membres de chaque classe, on recalcule le centre d'inertie de chaque classe.

**Étape 4** : On redistribue les objets dans la classe qui leur est la plus proche en tenant compte des nouveaux centres.

On continue le traitement tant que les classes évoluent.

A chaque itération, l'algorithme fait diminuer l'inertie interclasses, ce qui justifie la convergence de l'algorithme.

En effet, notons  $g_1^j, \dots, g_K^j$  les centres des classes à la  $j$ -ème itération et considérons la quantité :

$$Q_j = \sum_{k=1}^K \sum_{i \in C_k^j} d^2(w_i, g_k^j)$$

On recalcule les nouveaux centres  $g_1^{j+1}, \dots, g_K^{j+1}$  en fonction des points  $C_1^j, \dots, C_K^j$  d'effectif  $n_k^j$ . On a alors

$$g_K^{j+1} = \frac{1}{n_k^j} \sum_{i \in C_k^j} w_i$$

L'inertie intraclasse de la partition  $C_1^j, \dots, C_K^j$  est donc

$$I_{intra}^{j+1} = \sum_{k=1}^K \sum_{i \in C_k^j} \|w_i - g_k^{j+1}\|^2$$

On a donc

$$Q_j = \sum_{k=1}^K \sum_{i \in C_k^j} \|w_i - g_k^j\|^2$$

$$Q_j = \sum_{k=1}^K \sum_{i \in C_k^j} \|w_i - g_k^{j+1} + g_k^{j+1} - g_k^j\|^2$$

$$Q_j = \sum_{k=1}^K \sum_{i \in C_k^j} \|w_i - g_k^{j+1}\|^2 + \|g_k^{j+1} - g_k^j\|^2 + 2\langle w_i - g_k^{j+1}, g_k^{j+1} - g_k^j \rangle$$

Ce qui montre

$$Q_j \geq I_{intra}^{j+1}$$

La redistribution des classes lors de l'étape 2 entraîne

$$Q_{j+1} \leq I_{intra}^{j+1}$$

On a donc bien

$$I_{intra}^j \geq I_{intra}^{j+1}$$

Les avantages de cette méthode sont qu'on peut choisir le nombre de classes et que cet algorithme demande un faible nombre d'itérations. Cet algorithme a une complexité en  $O(n)$ .

L'inconvénient est le choix aléatoire des centres des classes à l'initialisation de l'algorithme. Il faut donc appliquer plusieurs fois l'algorithme afin de vérifier qu'on obtienne le même résultat.

## Méthode de classification par ascendant hiérarchique

Le but de cette méthode est d'obtenir une hiérarchie, c'est-à-dire une collection de groupes d'observations.

L'algorithme se décompose en trois étapes :

### Étape 1 : Phase d'initialisation

On initialise l'algorithme en choisissant les  $n$  individus comme les  $n$  centres des  $n$  classes et on calcule la matrice des distances deux à deux.

### Étape 2 : Phase d'itération

On regroupe les deux éléments les plus proches au sens du critère choisi. On met à jour la matrice des distances avec ces nouveaux éléments.

### Étape 3 : Phase finale

On obtient l'agrégation de tous les individus en une seule classe.

Lors de l'étape 2, nous sommes amenés à déterminer les deux ensembles de points les plus proches. Il nous reste à définir le critère pour évaluer la distance entre deux classes.

Pour cela, analysons le déroulement de l'algorithme.

À l'initialisation, chaque classe est composée d'un unique élément et l'inertie intraclasse est nulle. À la fin de l'algorithme, tous les éléments sont regroupés en une unique classe. L'inertie interclasse est alors nulle, alors que l'inertie intraclasse est maximale. On peut montrer qu'à chaque étape on augmente l'inertie intraclasse et on diminue de l'inertie interclasse puisqu'on ajoute de l'hétérogénéité en fusionnant deux classes.

Cependant, notre objectif est d'obtenir pour chaque nombre de classes  $K$  la partition minimisant l'inertie intraclasse. Il nous faut donc à chaque étape regrouper les classes deux à deux manières à minimiser la perte d'inertie interclasse.

Calculons la perte d'inertie interclasse pour passer de  $k$  à  $k-1$  classes en fusionnant les classes  $C_i$  et  $C_j$ , les  $k-2$  autres classes n'étant pas modifiées.

On a

$$I_{inter}^k = \sum_{m \neq i,j} n_m d^2(g_m, g) + n_j d^2(g_j, g) + n_i d^2(g_i, g)$$

Puis à l'étape suivante

$$I_{inter}^{k-1} = \sum_{m \neq i,j} n_m d^2(g_m, g) + (n_i + n_j) d^2(g_{ij}, g)$$

avec  $g_{ij}$  le barycentre de la nouvelle classe définie par

$$g_{ij} = \frac{n_i g_i + n_j g_j}{n_i + n_j}$$

Ainsi la perte d'inertie entre les deux étapes se quantifie par

$$I_{inter}^k - I_{inter}^{k-1} = n_j d^2(g_j, g) + n_i d^2(g_i, g) - (n_i + n_j) d^2(g_{ij}, g)$$

Ce qui peut se réécrire

$$\begin{aligned} &= n_i \|g_i - g\|^2 + n_j \|g_j - g\|^2 - \frac{1}{n_i + n_j} \|n_i g_i + n_j g_j - (n_i + n_j)g\|^2 \\ &= n_i \|g_i - g\|^2 + n_j \|g_j - g\|^2 - \frac{1}{n_i + n_j} [n_i^2 \|g_i - g\|^2 + n_j^2 \|g_j - g\|^2 + 2n_i n_j \langle g_i - g, g_j - g \rangle] \end{aligned}$$

Puis,

$$= \frac{n_i n_j}{n_i + n_j} [\|g_i - g\|^2 + \|g_j - g\|^2 - \langle g_i - g, g_j - g \rangle]$$

Finalement,

$$I_{inter}^k - I_{inter}^{k-1} = \frac{n_i n_j}{n_i + n_j} \|g_i - g_j\|^2$$

Nous avons ainsi établi le critère de Ward pour mesurer la distance entre deux classes  $C_i$  et  $C_j$  de barycentre respectif  $g_i$  et  $g_j$  :

$$D_{ward}(g_i, g_j) = \frac{n_i n_j}{n_i + n_j} d^2(g_i, g_j)$$

*Remarques :*

- On réalise à chaque étape la fusion optimale, mais cela ne garantit en rien l'optimalité de la partition obtenue pour chaque nombre de classe.
- on obtient une hiérarchie de partition et non une partition.
- Enfin, l'inconvénient de cet algorithme est qu'il nécessite en moyenne un nombre d'itérations plus élevé que celui présenté précédemment, sa complexité étant en  $O(n^3)$ .

## Choix du nombre de classes

On a pour l'instant considéré que le nombre de classes  $K$  était connu a priori, notamment pour appliquer la méthode des k-means.

L'objectif in fine est de mettre en évidence la structure sous-jacente aux données. Il est donc pragmatique de choisir une partition dont on peut interpréter les classes et qui présente un ratio  $\frac{I_{intra}}{I_{totale}}$  le plus élevé possible.

Pour identifier le nombre de classes optimal, on représente ce ratio en fonction du nombre de classes  $K$  et on essaie d'identifier les étapes qui accélèrent la dégradation de l'inertie intraclasse, due à l'hétérogénéité induite par la diminution du nombre de classes. Ce critère est connu sous le nom "critère du coude" du fait de la forme de la courbe.

### 3.3.2 Application du partitionnement

#### Partitionnement pour les hommes

Pour réaliser le partitionnement sur les données hommes, nous commençons par regrouper les têtes en 21 classes d'environ 20 000 assurés par ordre croissant des montants. Nous avons choisi de considérer uniquement les contrats qui représentent plus de 3 000€ d'encours.

L'âge joue un facteur important dans le montant de l'encours. Pour les âges les plus jeunes, l'épargne constituée est faible et donne une information peu pertinente sur le niveau de richesse de l'assuré. C'est pour cela que nous choisissons d'effectuer le partitionnement sur l'intervalle [55, 95] ans.

Pour rappel, nous avons construit trois variables afin de caractériser chaque sous-groupe :

- Âge moyen des assurés
- Âge moyen des décès
- Ratio observés sur attendus par rapport aux tables réglementaires th002. Pour chaque âge, nous avons donc  $Ratio O/A = \frac{\sum D_x}{\sum E_x * q_x^{th002}}$

Notre portefeuille est donc divisé de la manière suivant le tableau 3.1.

#### Recherche du nombre de classe

On cherche le nombre de classes pertinentes en s'intéressant au rapport de l'inertie :  $\frac{I_{intra}}{I_{tot}}$ . Le graphique 3.1 représente ce rapport en fonction du nombre de classes.

A partir de 4 ou 5 classes, l'apport d'une nouvelle classe n'est plus significative par la méthode des k-means.

#### Notation des partitions

Pour plus de clarté, nous noterons par la suite les partitions de la même manière que pour les exemples suivants :

- Si nous avons une partition composée de deux classes regroupant d'une part les groupes 1 à 10 et, d'autre part, les groupes 11 à 21, nous noterons la partition  $\boxed{1-10} \boxed{11-21}$ .
- Si nous avons une partition composée de trois classes regroupant d'une part les groupes 1 à 7, d'autre part, les groupes 8 à 13 et enfin les groupes 14 à 21, nous la noterons  $\boxed{1-7} \boxed{8-13} \boxed{14-21}$ .

Groupes	Montant	Âge moyen as- surés	Âge moyen décès	Ratio O/A
1	[3 000, 4 200]	69,0	78,8	0,78
2	[4 200, 5 700]	69,4	79,2	0,82
3	[5 700, 7 500]	69,8	79,9	0,84
4	[7 500, 9 300]	69,7	80,0	0,81
5	[9 300, 11 200]	69,8	80,1	0,79
6	[11 200, 13 900]	71,3	81,4	0,76
7	[13 900, 16 900]	71,5	81,9	0,78
8	[16 900, 20 300]	72,2	82,0	0,77
9	[20 300,24 600]	72,4	82,3	0,72
10	[24 600, 29 100]	72,8	82,8	0,76
11	[29 100,34 200]	73,3	82,8	0,78
12	[34 200,40 900]	73,6	83,1	0,78
13	[40 900,48 800]	73,8	83,1	0,72
14	[48 800, 58 700]	73,2	82,8	0,70
15	[58 700, 71 500]	73,8	83,1	0,73
16	[71 500, 87 900]	73,8	83,5	0,70
17	[87 900, 108 400]	73,6	83,3	0,65
18	[108 400, 137 800]	74,1	84,7	0,67
19	[137 800, 180 000]	74,3	84,0	0,66
20	[180 000, 257 400]	74,1	83,9	0,58
21	[257 400, $\infty$ ]	73,6	83,8	0,52

TABLE 3.1 – Groupes en prévision du partitionnement pour les hommes



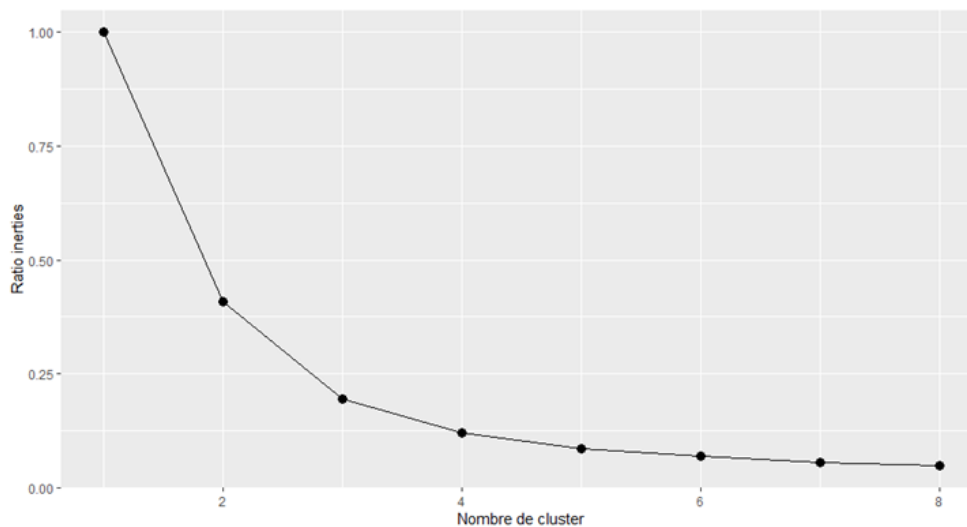


FIGURE 3.1 – Méthode du coude appliquée au portefeuille hommes

*Application de la méthode des k-means*

On commence par appliquer cette méthode pour répartir en quatre groupes nos sous-groupes. On réalise 10 000 simulations afin de mettre en évidence le maximum de partitions possibles. Nous choisirons en suite la partition qui maximise l’inertie interclasse et qui est cohérente avec la répartition de montants. Le tableau 3.2 récapitule les partitions obtenues.

Partition	% $I_{inter}$
1-5 6-12 13-19 20-21	90,6%
1-5 6-9 10-16 17-21	87,5%

TABLE 3.2 – Méthode K-means pour les hommes,  $K = 4$

Nous appliquons maintenant la méthode des k-means pour cinq classes (tableau 3.3).

Partition	% $I_{inter}$
1-2 3-5 6-9 10-19 20-21	88,5%
1-2 3-6 7-15 16-19 20-21	90,3%
1-2 3-5 6-12 13-19 20-21	91,6%
1-5 6-12 13-19 20 21	92,0%
1-5 6-9 10-15 16-19 20-21	93,6%



Gammes	Montants €
Gamme 1	$\leq 12\ 000$
Gamme 2	$[12\ 000, 25\ 000[$
Gamme 3	$[25\ 000, 90\ 000[$
Gamme 4	$[90\ 000, 180\ 000[$
Gamme 5	$\geq 180\ 000$

TABLE 3.5 – Gammes obtenus après le partitionnement

Il reste à vérifier si cette segmentation fait sens du point de vue des lois de décès. Pour chaque gamme, nous construisons, suivant la méthode décrite précédemment, les tables du moment. Pour comparer les courbes obtenues, nous simulons une exposition aléatoire et nous regardons le nombre de décès obtenus pour chaque gamme.

Par exemple, la gamme 1 concerne des contrats de capitaux inférieurs à ceux de la gamme 2. Pour que notre partitionnement soit valide, nous voulons que, pour une exposition identique, le nombre de décès donné par la loi de mortalité de la gamme 1 soit supérieur au nombre de décès obtenus par la loi de mortalité de la gamme 2. Nous souhaitons également que l'écart relatif soit suffisamment significatif. En effet, il serait peu pertinent d'avoir deux gammes avec des niveaux de mortalité similaires.

Nous avons effectué 10 000 simulations à chaque fois.

De même, nous comparons les gammes 2 et 3, puis les gammes 3 et 4, et les gammes 4 et 5. Les résultats sont regroupés dans le tableau 3.6.

Gammes comparées	Ecart relatif	Nombre de décès décroissants
1-2	2,0%	Vrai
2-3	0,8%	Vrai
3-4	2,6%	Vrai
4-5	7,3%	Vrai

TABLE 3.6 – Comparaison des niveaux de gamme

Les gammes 2 et 3 donnent un nombre de décès similaire. Nous choisissons de fusionner ces deux gammes.

Finalement, nous obtenons quatre niveaux de gamme que nous définissons :

Niveaux de gamme homme	Montant encours €
Gamme 1	$\leq 12\ 000$
Gamme 2	$[12\ 000, 90\ 000[$
Gamme 3	$[90\ 000, 180\ 000[$
Gamme 4	$\geq 180\ 000$

TABLE 3.7 – Niveaux de gamme retenus pour les hommes

### Partitionnement pour les femmes

Dans le cas des femmes, nous regroupons les données en 27 classes d'environ 20 000 têtes. Comme pour les hommes, nous effectuons le partitionnement sur les âges de 55 à 95 ans qui possède au moins 3 000€ d'encours.

Notre portefeuille est donc divisé de la manière suivant le tableau 3.8.

#### *Recherche du nombre de classe*

Nous commençons par regarder l'évolution du pourcentage moyen de l'inertie intraclasse afin d'identifier le nombre de classes à chercher. Le graphique 3.3 nous indique de privilégier les partitions à trois ou quatre classes.

#### *Application de la méthode des k-means*

Nous commençons par appliquer la méthode des k-moyennes pour des partition à trois classes. Pour chaque combinaison obtenue, nous calculons le pourcentage de l'inertie inter-classe afin de les comparer comme le montre le tableau 3.9.

Partition	% $I_{inter}$
1-6   7-22   23-27	81,1%
1-10   11-24   25-27	77,7%

TABLE 3.9 – K-Means pour les femmes avec  $K = 3$

Nous passons maintenant aux partitions composées de quatre classes. Cependant, on observe que pour les trois partitions obtenues les sous-groupes ne sont pas tous regrouper par

Groupes	Montant	Âge moyen assurés	Âge moyen décès	coefficient O/A
1	[3 000, 4 000]	71,1	83,1	0,85
2	[4 000, 5 300]	71,5	83,7	0,88
3	[5 300, 6 700]	72,2	84,0	0,83
4	[6 700, 8 100]	72,0	84,1	0,83
5	[8 100, 9 700]	72,2	84,1	0,88
6	[9 700, 11 200]	71,7	84,1	0,80
7	[11 200, 13 200]	73,5	85,0	0,81
8	[13 200, 15 600]	74,0	85,4	0,87
9	[15 600, 18 000]	74,2	85,4	0,85
10	[18 000, 20 700]	74,5	85,5	0,92
11	[20 700, 23 700]	74,8	85,8	0,78
12	[23 700, 27 300]	75,2	85,8	0,83
13	[27 300, 31 200]	75,4	85,6	0,85
14	[31 200, 35 300]	75,6	85,9	0,80
15	[35 300, 40 400]	76,1	86,2	0,86
16	[40 400, 45 800]	76,3	86,3	0,81
17	[45 800, 52 700]	75,7	85,3	0,81
18	[52 700, 60 800]	75,9	85,9	0,80
19	[60 800, 70 400]	76,2	86,1	0,80
20	[70 400, 81 900]	76,1	86,1	0,79
21	[81 900, 95 500]	76,5	86,2	0,77
22	[95 500, 111 600]	76,3	86,3	0,76
23	[111 600, 133 000]	77,3	86,6	0,75
24	[133 000, 161 100]	77,4	86,6	0,75
25	[161 100, 201 600]	77,6	86,6	0,69
26	[201 600, 269 900]	77,8	86,5	0,66
27	[269 900, $\infty$ ]	78,0	86,9	0,67

TABLE 3.8 – Groupes en prévision du partitionnement pour les femmes

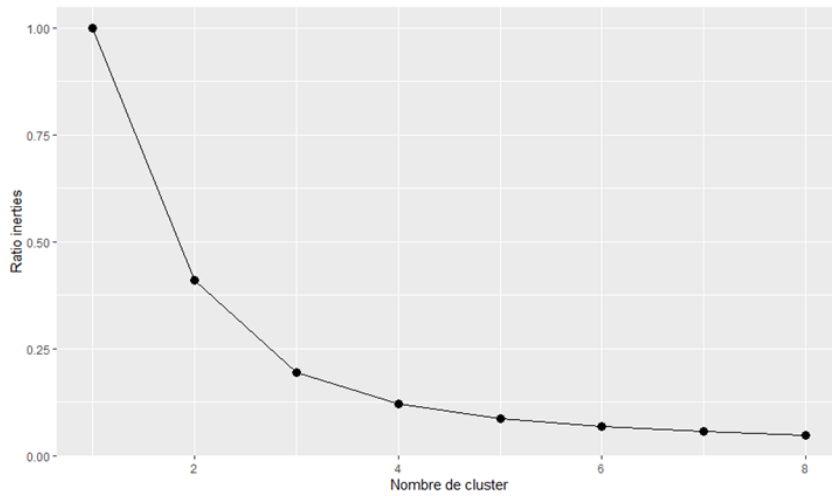


FIGURE 3.3 – Méthode du coude appliquée au portefeuille femmes

montants croissants. Nous rejetons donc ces partitions et calculons les inerties pour les partitions cohérentes les plus proches sur le tableau 3.10.

Partition	% $I_{inter}$
1-2   3-7   8-22   23-27	83,1%
1-6   7-13   14-22   23-27	87,3%
1-6   7-16   17-24   25-27	86,8%

TABLE 3.10 – K-means pour les femmes avec  $K = 4$

### *Application de la méthode CAH*

Nous passons maintenant à la seconde méthode. Pour nos données, nous obtenons le dendrogramme 3.4 et on identifie les partitions du tableau 3.11.

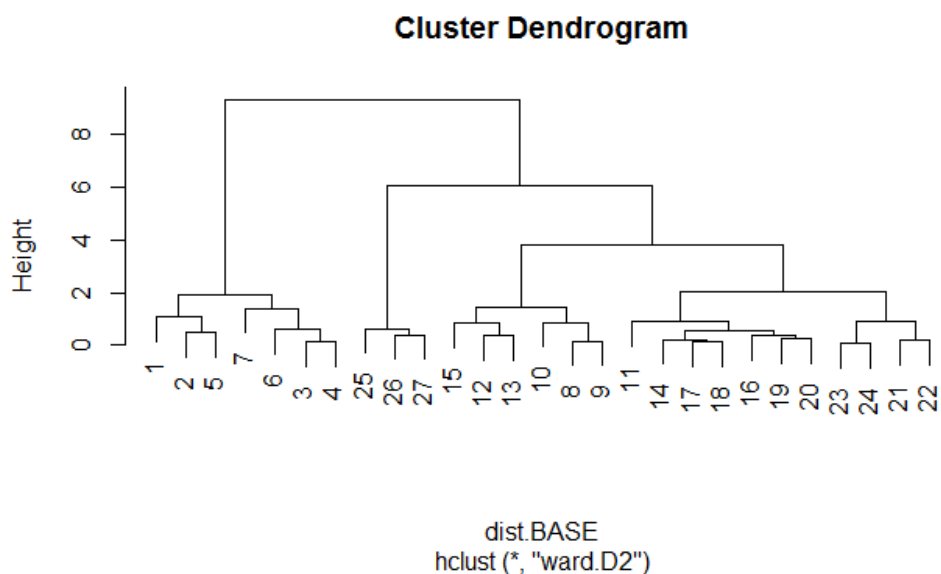


FIGURE 3.4 – Dendrogrammes pour les femmes

Nombre de classe	Partition	% $I_{inter}$
3	<span style="border: 1px solid black; padding: 2px;">1-7</span> <span style="border: 1px solid black; padding: 2px;">8-24</span> <span style="border: 1px solid black; padding: 2px;">25-27</span>	78,9%
4	<span style="border: 1px solid black; padding: 2px;">1-7</span> <span style="border: 1px solid black; padding: 2px;">8-15</span> <span style="border: 1px solid black; padding: 2px;">16-24</span> <span style="border: 1px solid black; padding: 2px;">25-27</span>	86,9%

TABLE 3.11 – Méthode CAH pour les femmes

Finalement, nous retenons la partition 1-6 7-13 14-22 23-27 qui donne les quatre niveaux de gamme suivant.

Gammes	Montants €
Gamme 1	$\leq 12\ 000$
Gamme 2	$[12\ 000, 35\ 000[$
Gamme 3	$[35\ 000, 110\ 000[$
Gamme 4	$\geq 110\ 000$

TABLE 3.12 – Gammes obtenus après le partitionnement

Il reste à vérifier que cette partition est cohérente d'un point de vue des lois de décès. Comme pour les hommes, nous simulons des portefeuilles pour comparer les nombres de décès obtenus par chacune des quatre courbes. Nous réalisons à chaque fois 10 000 simulations.

Gammes comparées	Ecart relatif	Nombre de décès décroissants
1-2	2,7%	FAUX
2-3	1,1%	VRAI
1-3	1,6%	FAUX
3-4	4,7%	VRAI

TABLE 3.13 – Comparaison des gammes 1 à 4

Les gammes 2 et 3 présentent pour chaque simulation un nombre de décès supérieur à celui de la gamme 1, alors que celle-ci concerne des assurés supposés d'un niveau de richesse moins élevé. De plus, nous remarquons que l'écart relatif moyen entre ces trois gammes est assez faible. Pour que les gammes construites soient cohérentes, nous choisissons de fusionner les gammes 1, 2 et 3.

Nous effectuons les mêmes vérifications avec les deux nouvelles gammes (tableau 3.14).

Gammes comparées	Ecart relatif	Nombre de décès décroissants
123-4	5,3%	VRAI

TABLE 3.14 – Comparaison des gammes après fusion des gammes 1, 2 et 3

Finalement, nous obtenons deux niveaux de gamme.

Niveau de gamme femmes	Montant encours €
Gamme 1	$\leq 110000$
Gamme 2	$\geq 110000$

### 3.3.3 Comparaison avec le portefeuille rentes

Nous avons pu segmenter les portefeuilles épargne hommes et femmes en déterminant des niveaux de gammes selon le montant de l'encours (quatre pour les hommes et deux pour les femmes). Il nous faut maintenant étudier si cette segmentation est utile pour l'étude du portefeuille rentes.

Pour appliquer la segmentation déterminée au portefeuille rentes, nous travaillons sur les montants des capitaux constitutifs. Ceux-ci sont donc figés au moment de la liquidation de la rente, contrairement au montant des encours des contrats épargnes qui sont susceptibles de s'accroître avec le temps. Pour limiter cette erreur, nous avons réalisé le partitionnement



sur les âges  $[55, 95]$  ans pour les contrats d'au moins 3 000€ d'en-cours.

Nous commençons par regarder la répartition de ces gammes dans chaque portefeuille.

Niveaux de gamme homme	Proportion rentes	Proportion épargne
Gamme 1	4%	55%
Gamme 2	72%	31%
Gamme 3	14%	7%
Gamme 4	10%	6%

TABLE 3.15 – Proportion des niveaux de gamme dans le portefeuille rentes pour les hommes

Niveaux de gamme femmes	Proportion rentes	Proportion épargne
Gamme 1	88%	86 %
Gamme 2	12%	14 %

TABLE 3.16 – Proportion des niveaux de gamme dans le portefeuille rentes pour les femmes

Les proportions diffèrent entre les portefeuilles rentes et épargne. Pour les hommes, les trois gammes les plus élevées sont proportionnellement plus représentées dans le portefeuille rentes que dans celui de l'épargne. Cela s'explique d'une part par le montant pivot assez faible entre les gammes 1 et 2 (12 000€), mais aussi par le fait que les sorties en rentes concernent principalement des contrats avec un capital constitutif assez conséquent. Pour les contrats avec un capital faible, une sortie en capital est privilégiée au versement d'une rente peu importante.

Pour les femmes, en revanche, la gamme concernant les montants les plus faibles est plus représentée dans le portefeuille rentes que dans le portefeuille épargne. Cependant, le montant pivot entre les deux gammes étant élevés (110 000€), cela n'a rien d'aberrant.

Ces différences de répartition nous incitent à considérer la loi de décès d'un portefeuille fictif obtenu en agrégeant les lois décès de chaque gamme épargne selon les proportions observées dans le portefeuille rentes.

En effet, en faisant l'hypothèse que, pour chaque gamme, les portefeuilles épargnes et rentes ont une loi de décès similaire, ce nouveau portefeuille présente une estimation de la mortalité plus proche de celle du portefeuille rentes que celle obtenue avec le portefeuille épargne d'origine.

Le faible nombre d'observations sur le portefeuille rentes ne nous permet pas de valider ou d'infirmer cette hypothèse. Toutefois, nous décidons de comparer les décès observés du portefeuille rentes aux décès attendus obtenus, d'une part, en appliquant la loi de décès du portefeuille épargne, et, d'autre part, la loi de décès du portefeuille fictif à l'exposition du portefeuille rentes. Nous calculons également les ratios observés sur attendus.

Le tableau 3.17 présente les résultats pour les hommes.

<b>Année</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>Total</b>
Décès observés	37	65	59	54	73	90	91	469
Décès attendus fictif	53,5	61,1	68,4	75,4	87	99,6	114,1	559,1
Ratio O/A	69%	106%	86%	72%	84%	90%	80%	84%
Décès attendus épargne	56	64	72	80	93	106	122	593
Ratio O/A	66%	102%	82%	68%	78%	85%	75%	79%

TABLE 3.17 – Comparaison des décès observés et attendus par année pour les hommes

En premier, ce tableau nous montre la volatilité qu'il existe entre les années, induite par le faible nombre de décès. Nous remarquons également que, hormis pour l'année 2011, les décès issus du portefeuille fictif sont plus proches de ceux observés que ceux donnés par le portefeuille épargne. Par ailleurs, en dehors de l'année 2011, les ratios observés sur attendus sont inférieurs à 1. Les lois des portefeuilles épargne et fictif semblent donc surestimées les décès du portefeuille rentes.

De la même façon, nous étudions les décès observés et attendus pour les femmes dans le tableau 3.18.

<b>Année</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>Total</b>
Décès observés	70	66	73	45	81	87	73	495
Décès attendus fictif	65	66	68	69	74	76	77	495
Ratio	108%	100%	107%	65%	109%	114%	95%	100%
Décès attendus épargne	63	63	66	67	71	74	75	486
Ratio	111%	105%	111%	67%	114%	118%	97%	103%

TABLE 3.18 – Comparaison des décès observés et attendus par année pour les femmes

Pour les femmes, nous remarquons que les lois de décès du portefeuille fictif et du portefeuille épargne sous-estiment le nombre de décès du portefeuille rentes pour la plupart des

années. Globalement, cependant, le ratio est très proche de 1, ce qui nous conforte dans l'idée que la loi de décès du portefeuille fictif est similaire à celle du portefeuille rentes.

Nous nous intéressons maintenant à la mortalité de manière globale. Le graphe 3.5 compare les décès observés aux décès attendus issus du portefeuille fictif pour les hommes sans distinguer par année. Nous traçons également les intervalles de confiance à 95% pour les décès attendus.

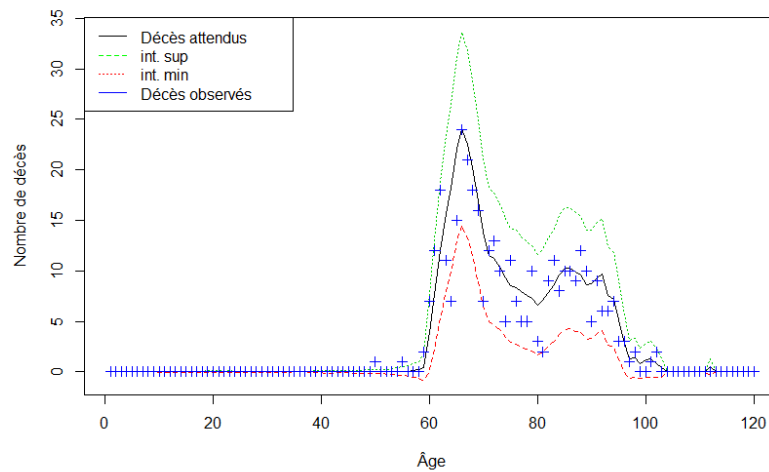


FIGURE 3.5 – Décès observés pour les rentes par rapport aux décès attendus selon la loi épargne

Les décès observés sont majoritairement dans l'intervalle de confiance. Nous poursuivons notre analyse en étudiant les résidus de la réponse (graphe 3.6). Ceux-ci ne présentent pas de tendance.

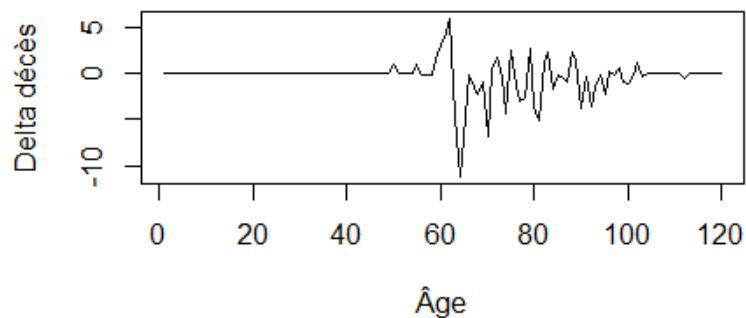


FIGURE 3.6 – Résidus de la tendance observés pour les rentes par rapport aux décès attendus selon la loi du portefeuille fictif

De même que pour les hommes, nous effectuons une analyse globale pour les femmes en représentant graphiquement les décès observés et ceux attendus issus du portefeuille fictif. Nous ajoutons les intervalles de confiance à 95%.

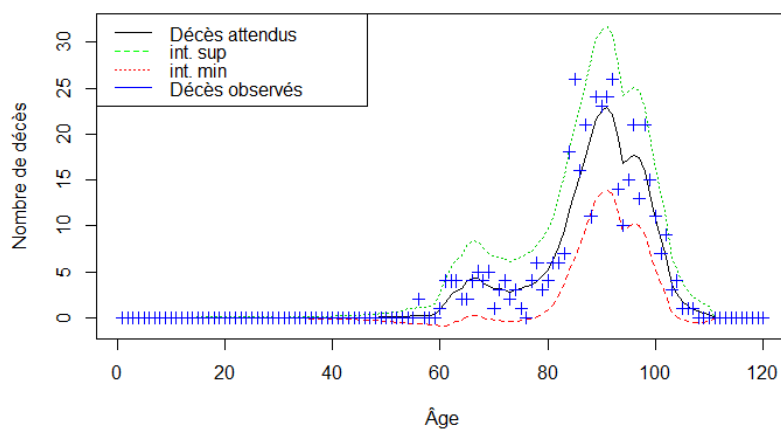


FIGURE 3.7 – Décès observés pour les rentes par rapport aux décès attendus selon la loi du portefeuille fictif pour les femmes

Les décès observés se trouvent majoritairement dans l'intervalle de confiance à 95%. Une étude des résidus de la réponse ne laisse apparaître aucune tendance.

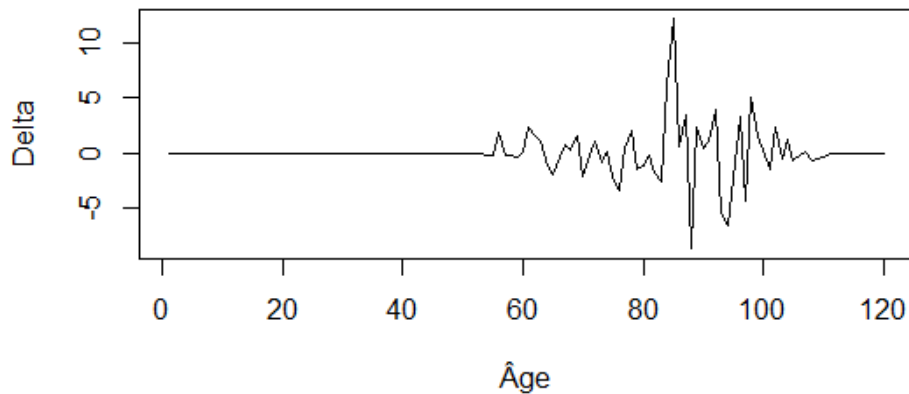


FIGURE 3.8 – Résidus de la tendance observés pour les rentes par rapport aux décès attendus selon la loi épargne pour les femmes

### 3.3.4 Conclusion

L'étude du portefeuille épargne nous a permis de réaliser une segmentation selon le montant de l'encours. De cette segmentation, nous avons pu construire un portefeuille fictif possédant comme loi de décès la loi agrégée des niveaux de gamme selon les proportions observés dans le portefeuille rentes. En étudiant pour chaque année les décès observés et attendus mais aussi de façon plus globale, nous avons vu que les lois de décès de ces deux portefeuilles étaient similaires pour les femmes comme pour les hommes. Cependant, ces résultats doivent toutefois être relativisés au vu du nombre de décès.

Ainsi, nous possédons un portefeuille fictif dont la mortalité est "proche" du portefeuille de l'étude. Nous devons maintenant étudier comment utiliser cette source d'information afin de limiter l'erreur induite par le risque d'estimation pour la construction d'une table de mortalité prospective sur le portefeuille rentes.



# Chapitre 4

## Utilisation de la théorie de la crédibilité partielle pour la construction de tables prospectives

Pour limiter les fluctuations liées au risque d'estimation lors de la construction d'une table de mortalité, nous nous intéressons à la théorie de la crédibilité partielle. Nous commençons par poser le cadre théorique pour ensuite l'intégrer à la construction d'une table de mortalité prospective.

### 4.1 La théorie de la crédibilité partielle

La théorie de la crédibilité partielle a été par [Whitney, 1918] pour tenir compte de l'expérience individuelle dès lors que la crédibilité totale n'était pas garantie.

Le principe de la théorie de la crédibilité partielle est de déterminer un barycentre entre le taux de mortalité issu de l'expérience et un taux de référence qu'on utilisera en l'absence de données. Formellement, cela revient à écrire :

$$\tilde{q}_x = \lambda \hat{q}_x + (1 - \lambda) \hat{q}_x^{ref}$$

L'objectif est d'estimer le facteur de crédibilité  $\lambda$  qui représente la confiance que nous accordons à nos données. Ce paramètre est donc dans l'intervalle  $[0, 1]$ . Si  $\lambda = 1$ , nous retrouvons le cas de la crédibilité complète.

En reprenant les notations de la partie 2.3.1, nous souhaitons que  $\tilde{q}_x$  vérifie :

$$P\left(\frac{|\tilde{q}_x - q_x|}{q_x} \leq k\right) \geq p$$

Nous pouvons également écrire le taux de mortalité réelle comme  $q_x = \lambda q'_x + (1 - \lambda) q_x^{ref}$ , ce

qui entraîne :

$$P \left( \frac{|\lambda \widehat{q}_x + (1 - \lambda) \widehat{q}_x^{ref} - (\lambda q'_x + (1 - \lambda) q_x^{ref})|}{q_x} \leq k \right) \geq p$$

Nous faisons alors l'hypothèse que  $q_x^{ref} = \widehat{q}_x^{ref}$ . Nous ne considérons pas l'erreur induite par la table de référence, mais uniquement celle issue de notre estimateur  $\widehat{q}_x$ . On obtient :

$$P \left( \frac{\lambda |\widehat{q}_x - q'_x|}{q_x} \leq k \right) \geq p$$

Le théorème central limite permet, pour un nombre conséquent d'observations, de faire l'hypothèse que  $\widehat{q}_x - q'_x$  suit une loi normale  $N(q'_x, \sigma_{q'_x}^2)$ .

Par réduction,

$$P \left( \frac{-kq_x}{\lambda \sigma_{q'_x}} \leq Z \leq \frac{kq_x}{\lambda \sigma_{q'_x}} \right)$$

où  $Z \sim N(0, 1)$ .

Souhaitant donner le plus grand poids possible à notre propre expérience, on cherche le plus grand  $\lambda$  vérifiant

$$\frac{kq_x}{\lambda \sigma_{q'_x}} \geq \epsilon_{p/2}$$

avec  $\epsilon_{p/2}$  le quantile d'ordre  $p/2$  de la loi normale.

Ce qui donne

$$\lambda = \frac{kq_x}{\epsilon_{p/2} \sigma_{q'_x}}$$

Comme  $\lambda \in [0, 1]$  en tant que paramètre d'un barycentre, on impose

$$\lambda = \min \left( \frac{kq_x}{\epsilon_{p/2} \sigma_{q'_x}}, 1 \right)$$

De la même manière que pour la crédibilité totale, nous pouvons travailler en nombre de sinistres. En posant  $N_0$  le nombre de décès minimal à observer dans le cadre de la crédibilité totale, et  $E$  l'exposition, on a avec  $\sigma_{q'_x}^2 = \frac{q'_x(1-q'_x)}{E}$

$$\lambda = \sqrt{\frac{1}{N_0}} \sqrt{\frac{q_x}{q'_x(1-q'_x)}} \sqrt{q_x E}$$

Et donc, en notant  $N$  le nombre de décès observés, et en supposant que les taux de mortalité sont proches de 0, on obtient :

$$\lambda = \sqrt{\frac{N}{N_0}}$$

Dans notre cas, la table de référence choisie serait celle caractérisant la loi de décès du portefeuille fictif construit au chapitre précédent à l'aide des données du portefeuille épargne.



## 4.2 Méthodes de construction des tables prospectives

Nous détaillons le modèle retenu pour la construction d'une table de mortalité prospective et expliquons comment l'adapter à l'utilisation de la théorie de la crédibilité partielle.

### 4.2.1 Le modèle Standardized Mortality Ratio

La première méthode utilisée consiste à positionner la mortalité du portefeuille par rapport à la table de référence selon un unique coefficient appelé *Standardized Mortality Ratio*.

Formellement, on définit notations suivantes, pour  $x \in [x_{min}, x_{max}]$  et  $t \in [t_{min}, t_{max}]$  :

- $L_{x,t}$  le nombre d'individus d'âge  $x$  vivant au 1er janvier de l'année  $t$ .
- $D_{x,t}$  le nombre de décès enregistré à l'âge  $x$  au cours de l'année  $t$ .
- $E_{x,t}$  l'exposition au risque à l'âge  $x$  au cours de l'année  $t$ .
- $q_{x,t}$  est la probabilité qu'un individu d'âge  $x$  décède au cours de l'année  $t$ .

Le modèle SMR repose sur la relation

$$\widetilde{q}_{x,t} = SMR * q_{x,t}^{ref}$$

Où :

$$SMR = \frac{\sum D_{x,t}}{\sum E_{x,t} * q_{x,t}^{ref}}$$

Le choix de la plage  $[x_{min}, x_{max}]$  est primordial, car le coefficient  $SMR$  est uniquement valable sur cette plage. Il est donc nécessaire de définir une procédure de fermeture de la table pour les âges élevés. Nous retenons la procédure décrite par [Denuit and Goderniaux, 2005].

Remarquons qu'on peut choisir de calculer un unique coefficient SMR ou bien de calculer un coefficient SMR propre à chaque classe d'âge.

### 4.2.2 Fermeture des tables

La méthode de [Denuit and Goderniaux, 2005] repose sur les quotients de mortalité et introduit une contrainte pour la fermeture des tables de mortalité. Elle consiste à ajuster par la méthode des moindres-carrés le modèle sur les âges élevés.

$$\log \widehat{q}_x = a + bx + cx^2 + \epsilon_x$$

avec  $\epsilon_x \sim N(0, \sigma^2)$  On impose les contraintes suivantes :

$$q_{130} = 1 \text{ et } q'_{130} = 0$$

Ces deux conditions conduisent à

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

Ainsi, ces contraintes imposent à la courbe des taux de mortalité une forme concave aux grands âges. La tangente horizontale au point  $x = 130$  empêche une décroissance éventuelle des  $q_x$  aux âges élevés.

Il reste à déterminer l'âge à partir duquel les  $\widehat{q}_x$  bruts sont remplacés par les valeurs ajustées par le modèle. Idéalement, il faudrait choisir l'âge sur la plage de l'étude qui maximise le coefficient de détermination  $R^2$ . Cependant, en pratique, il ne saurait pas optimal d'avoir des âges de raccords différents suivant les années d'observation. C'est pourquoi nous choisissons de fixer de manière arbitraire l'âge de raccord en fonction de la plage d'âges étudiée.

Si nécessaire, un lissage géométrique est appliqué autour de l'âge raccord.

### 4.2.3 Utilisation de la crédibilité partielle

Nous allons maintenant voir comment prendre en compte la théorie de la crédibilité partielle.

Intuitivement, nous serions tenté d'appliquer la théorie de la crédibilité partielle, âge par âge. Cependant, vu le nombre de données à disposition, cela se révèle peu pertinent. Nous allons donc nous intéresser à la mortalité globale du portefeuille.

Considérons  $D_i$  la variable aléatoire du nombre de décès à l'âge  $i$ . Celle-ci suit une loi de Bernoulli de paramètre  $q_i$ . Les variables  $D_i$  sont donc indépendantes mais pas identiquement distribuées. On note alors  $\widehat{q} = \frac{1}{N} \sum_{i=1}^N D_i$  l'estimateur de la mortalité moyenne par âge et  $q$  la mortalité moyenne réelle.

Nous cherchons toujours à minimiser l'erreur

$$P\left(\frac{|\widehat{q} - q|}{q} \leq k\right) \geq p$$

Bien que les  $D_i$  ne soient pas identiquement distribués et que nous sortons du champ d'application du théorème centrale limite, nous réalisons l'approximation gaussienne classique.

La suite du raisonnement est alors analogue à celui présenté à la partie 4.1.

Nous avons donc

$$\widetilde{q} = \lambda \widehat{q} + (1 - \lambda)q$$

Ce qui en divisant par la mortalité obtenue en utilisant la table de référence permet de trouver une relation pour le ratio SMR

$$\widetilde{SMR} = \lambda \widehat{SMR} + (1 - \lambda)SMR$$

## 4.3 Application aux données

Nous allons maintenant appliquer la méthode *Standardized Mortality Ratio* à nos données afin de construire une table de mortalité prospective.

D'un point de vue pratique, nous avons utilisé le package R "ELT" développé par J. TOMAS et F. PLANCHET pour construire les tables de mortalité prospective, et avons adapté les fonctions à la prise en compte de la crédibilité partielle.

### Choix des tables de référence

Comme vu au chapitre 3, le portefeuille rentes et le portefeuille fictif<sup>1</sup> présentent une mortalité proche sur les années 2010 à 2016. Nous allons donc utiliser les décès issus du portefeuille fictif comme référence pour la crédibilité partielle.

En l'absence de tables propres à notre expérience, nous utilisons les tables réglementaires TGH/TGF05. Ce sont donc ces tables que nous utiliserons comme référence pour l'application du taux SMR.

### Calcul du coefficient de crédibilité

Nous choisissons d'appliquer la théorie de la crédibilité partielle pour que l'estimateur  $\hat{q}$  de la mortalité globale respecte une précision de  $k = 5\%$  avec une probabilité de  $p = 90\%$ . Ainsi, pour être dans le cadre de la théorie de la crédibilité totale, il nous faudrait observer au moins  $N_0 = 1082$  décès.

Rappelons que le facteur de crédibilité partielle  $\lambda$  est donné par la formule

$$\lambda = \min\left(1, \sqrt{\frac{N}{N_0}}\right)$$

avec  $N$  le nombre de décès observés.

Pour calculer le facteur de crédibilité, nous devons déterminer au préalable la plage d'âge à utiliser. Dans le cadre des tables du moment, cette plage était établie à l'aide du critère de Cochran. Cependant, pour une table prospective, ce critère n'est plus vérifié pour le portefeuille rentes. Nous étudions donc la sensibilité en calculant ce facteur pour différentes plages.

Pour les hommes, nous choisissons d'étudier les plages :

- [40, 100] ans qui correspond à la plage de l'étude du portefeuille épargne.
- [60, 90] ans qui est une plage logique pour un portefeuille composé de rentes.

---

1. Nous avons considéré le portefeuille fictif qui a pour loi de décès l'agrégation des lois de chaque niveau de gamme selon les proportions présentes dans le portefeuille de l'étude.

- [65, 85] ans qui est la plage d'âge où ce concentre nos observations.
- [60, 99] ans la plage établie à l'aide du critère de Cochran pour la construction d'une table du moment.

Ainsi, nous obtenons les facteurs de crédibilité suivant pour les hommes :

Plage d'âges	40-100	60-90	65-85	60-99
Décès observés	364	323	222	360
Facteur de crédibilité	58%	55%	45%	58%

TABLE 4.1 – Facteur de crédibilité partielle pour les hommes

De la même façon, pour les femmes, nous choisissons d'étudier les plages :

- [40, 100] ans qui correspond à la plage de l'étude du portefeuille épargne.
- [60, 90] ans qui est une plage logique pour un portefeuille composé de rentes.
- [80, 100] ans qui est la plage d'âge où ce concentre nos observations.
- [76, 105] ans la plage établi à l'aide du critère de Cochran pour la construction d'une table du moment.

Ainsi nous obtenons :

Plage d'âges	40-100	60-90	80-100	76-105
Décès observés	340	189	278	297
Facteur de crédibilité	56%	42%	51%	52%

TABLE 4.2 – Facteur de crédibilité partielle pour les femmes

Le facteur de crédibilité dépendant du nombre de décès, plus une plage d'âge est étendu, plus ce facteur est élevé.

### Détermination des coefficients SMR

Maintenant que nous avons nos facteurs de crédibilité partielle potentiels, nous pouvons déterminer le coefficient SMR donné par la formule

$$\widehat{SMR} = \lambda \widehat{SMR}_{rentes} + (1 - \lambda) \widehat{SMR}_{fictif}$$

On remarque que ces coefficients sont relativement stables d'une plage d'âges à l'autre aussi bien pour les hommes que pour les femmes.

Nous passons maintenant au calcul du SMR pour les rentes. Nous gardons les mêmes plages d'âge d'étude. Puis, nous déduisons le SMR global.

Pour les hommes, nous avons :

Plage d'âges	40-100	60-90	65-85	60-99
$\widehat{SMR}_{rentes}$	1,11	1,14	1,11	1,10
$\widehat{SMR}_{fictif}$	1,23	1,24	1,24	1,22
Facteur de crédibilité	58%	55%	45%	58%
$\widehat{SMR}$	1,16	1,19	1,18	1,15

TABLE 4.3 – Calcul du SMR épargne pour les hommes

Pour les femmes :

Plage d'âges	40-100	60-90	80-100	76-105
$\widehat{SMR}_{rentes}$	1,11	1,08	1,04	1,03
$\widehat{SMR}_{fictif}$	1,16	1,17	1,14	1,15
Facteur de crédibilité	56%	42%	51%	52%
$\widehat{SMR}$	1,08	1,13	1,09	1,09

TABLE 4.4 – Calcul du SMR épargne pour les femmes

Nous devons maintenant réaliser un choix quant aux plages d'âges sur lesquelles vont être construites les tables de mortalité prospective.

Nous aimerions au maximum utiliser l'expérience du portefeuille rente tout en gardant une plage cohérente avec nos données de l'étude. C'est pour quoi nous choisissons l'intervalle [65, 99] ans pour les hommes.

Pour les femmes, en revanche, l'intervalle présentant le facteur de crédibilité le plus élevé est [40, 100] ans. Cependant, nous avons vu que nous observons principalement les décès au-delà de 75 ans. Nous choisissons donc de retenir l'intervalle [76, 105] ans.

On applique ensuite ces coefficients SMR aux tables réglementaires TGH05/TGF05.

## Validation des tables construites

Se pose maintenant la question de la validation du modèle obtenue. De part sa méthode de construction, il est normal que ce modèle ne soit pas parfaitement ajusté à nos données. Néanmoins, nous devons nous assurer que la loi obtenue n'est pas complètement absurde.

Pour cela, nous effectuons une analyse graphique.

Pour les hommes, nous choisissons de travailler sur les âges [65, 85] ans et pour les femmes [76, 96] ans. Ce choix est effectué en vu de la fermeture de la table.

Nous commençons par l'analyse résidus. Pour ne pas alourdir ce rapport, nous ne présentons ici que les résidus obtenus pour l'année 2016.

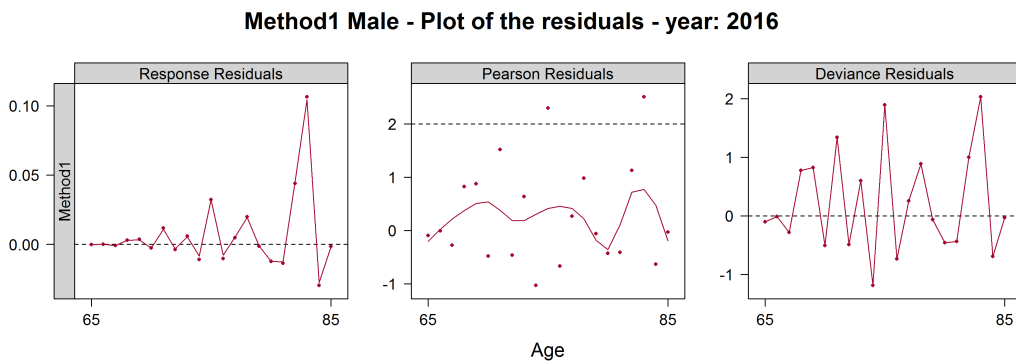


FIGURE 4.1 – Résidus de l'année 2016 pour les hommes

Pour les hommes, les résidus de la réponse ne présentent pas de tendances tout comme les résidus de Pearson. Cependant, pour deux âges, les résidus de Pearson sortent à deux reprises de l'intervalle  $[-2, 2]$ . La variabilité des données n'est donc pas parfaitement modéliser. Enfin, les résidus de la déviance oscillent autour de 0. Il n'y a donc pas de sur-lissage localement.

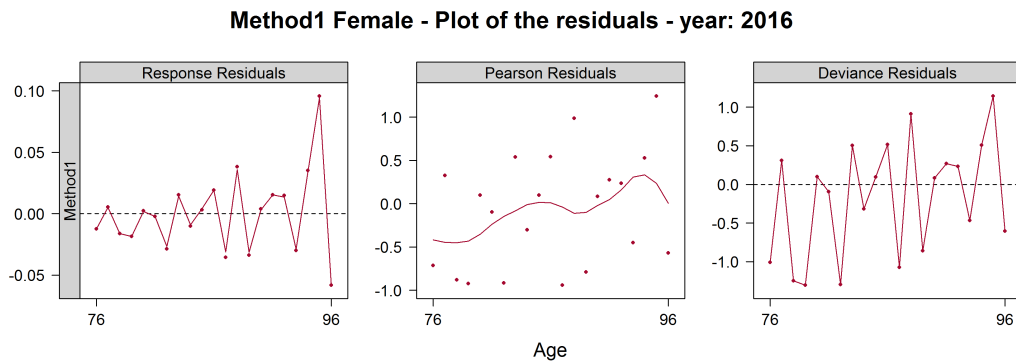


FIGURE 4.2 – Résidus de l'année 2016 pour les femmes

Pour les femmes, les résidus de la réponse et de Pearson ne présentent pas non plus de tendances. tout Les résidus de Pearson sont bien compris l'intervalle  $[-2, 2]$ . Enfin, les résidus de la déviance n'indiquent pas de sur-lissage localement.

Nous passons maintenant à l'étude des intervalles de confiance pour les décès. Pour que notre modélisation donne une représentation fidèle de la réalité, il faut que les décès observés se situent entre les deux traits représentant les intervalles de confiance. Le graphique 5.2 représente les intervalles de confiance pour les hommes et les femmes.

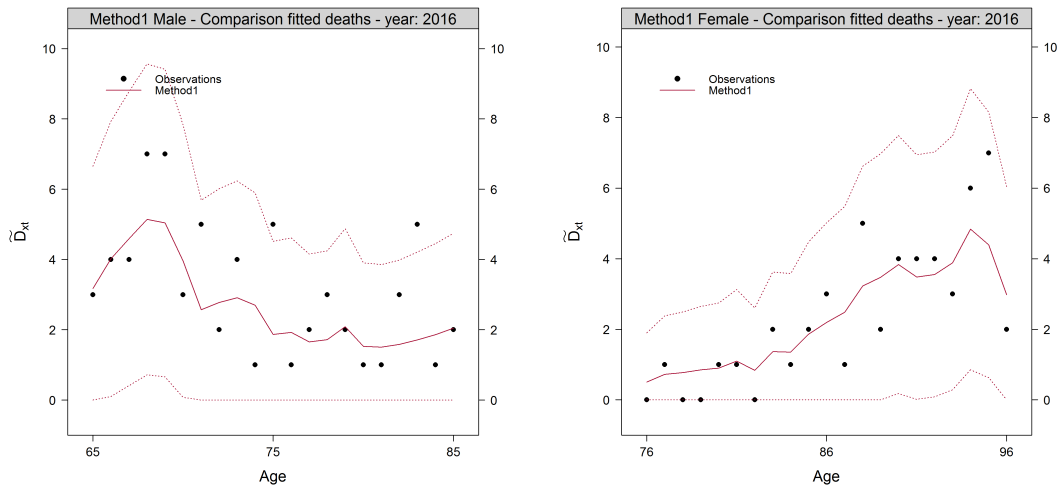


FIGURE 4.3 – Intervalles de confiance à 95% de l'année 2016 pour les hommes et les femmes

Malgré leur méthode de construction, les modèles obtenus captent bien la sinistralité de nos données d'origine.

Nous comptons pour chaque année le nombre de points en dehors de l'intervalle de confiance.

Année	2010	2011	2012	2013	2014	2015	2016
Hommes	0	1	0	0	0	2	2
Femmes	0	1	0	1	1	0	0

TABLE 4.5 – Nombre de points en dehors des intervalles de confiance à 95%

Enfin, nous regardons l'ajustement obtenu pour l'année 2016.

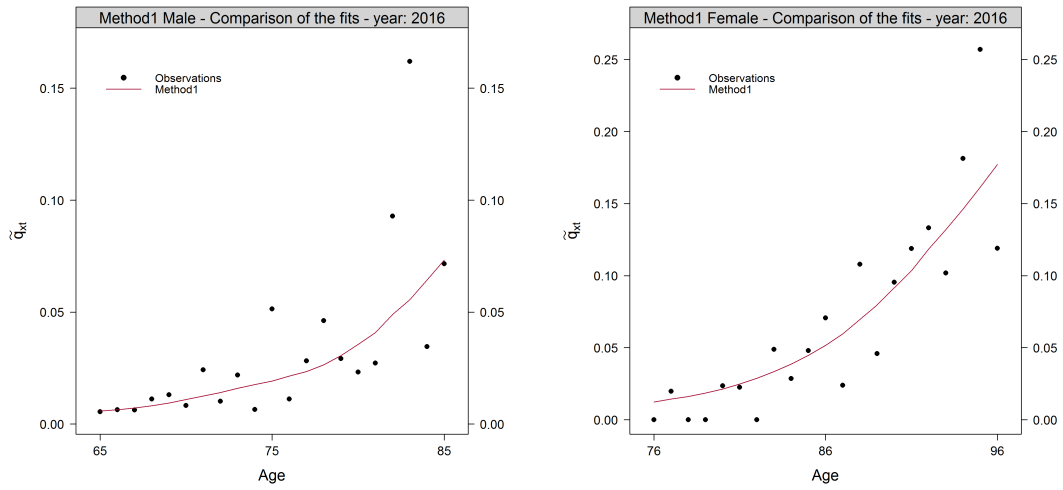


FIGURE 4.4 – Ajustement pour l’année 2016 pour les hommes et les femmes

Pour les hommes, nous remarquons que les observations sont très écartés de l’ajustement à partir de 75 ans.

### Fermeture de la table

D’après la répartition de nos observations, l’âge pivot pour la fermeture est fixé à 85 ans pour les hommes et à 95 ans pour les femmes.

Nous appliquons donc la méthode de [Denuit and Goderniaux, 2005]. Pour l’année 2016, la fermeture de la table

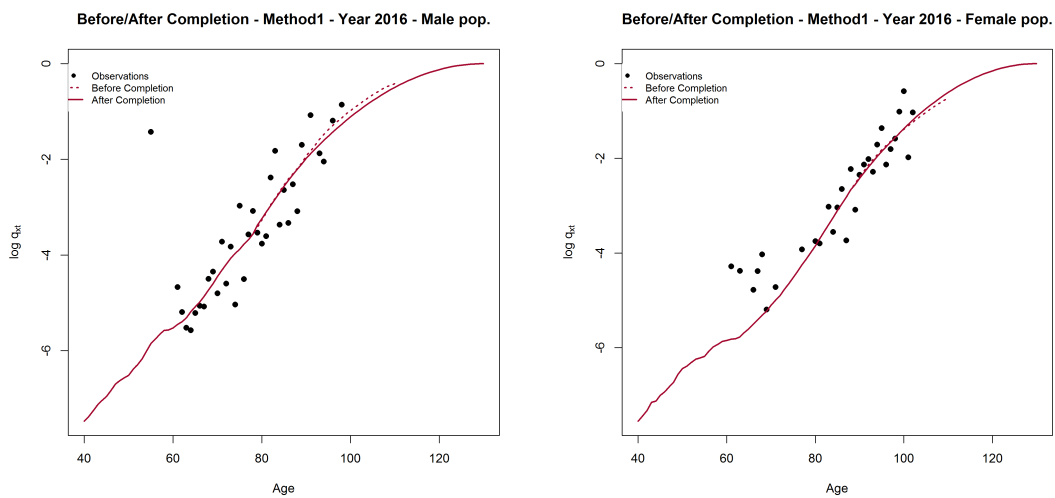


FIGURE 4.5 – Fermeture de la table pour l’année 2016 pour les femmes



Nous obtenons ainsi les tables de mortalité suivantes :

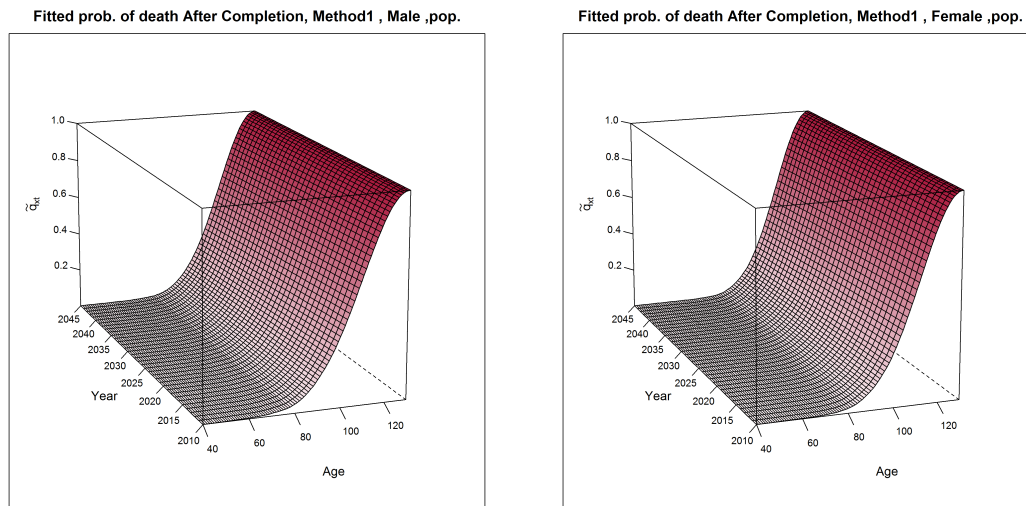


FIGURE 4.6 – Table prospective obtenu pour les hommes et les femmes



# Chapitre 5

## Utilisation de la théorie de Klugman

Dans le chapitre précédent, nous avons construit une table de mortalité en appliquant la théorie de la crédibilité partielle. Cela nous a conduits à faire des hypothèses pour construire un portefeuille de référence. Cependant, dans cette méthode, nous n'avons pas tenu compte de l'hétérogénéité des risques entre les portefeuilles. Nous nous intéressons donc maintenant à une autre méthode de crédibilité développée par [Klugman et al., 2009], inspiré de la théorie de la crédibilité de Bühlmann.

### 5.1 Principe de la méthode

Nous commençons par décrire la méthodologie de [Klugman et al., 2009] que nous adaptons au cas d'une table de mortalité prospective.

Cette méthode est basée sur le modèle linéaire bayésien qui relie les deux premiers moments de la distribution des observations passées d'un ensemble de portefeuille. Cela permet ensuite de prédire le ratio observé sur attendus du portefeuille de l'étude.

Cette méthode de crédibilité repose donc sur l'estimation des différents paramètres du modèle, au contraire de la théorie de la crédibilité partielle où certains paramètres étaient fixés a priori.

Pour appliquer cette méthode, nous avons donc besoin de connaître les informations suivantes :

- $r$  le nombre de portefeuilles à notre disposition. Chaque portefeuille  $h$  est composé de  $n_{ht}$  observations au cours de l'année  $t$ .
- $n_T$  le nombre d'années d'observations.

Pour chaque assuré  $i$ , nous disposons, au cours de l'année  $t$ , des informations suivantes :

- $f_{hit}$  la fraction de l'année  $t$  durant laquelle cet assuré est observé

- $b_{hit}$  le montant associé à l'assuré. Dans notre cas, il s'agit du capital constitutif de la rente ou du montant de l'encours pour un contrat épargne. Nous le supposons constant au fil des ans.
- $D_{hit}$  la variable aléatoire représentant le décès de l'assuré  $i$  du portefeuille  $h$  au cours de l'année  $t$ .
- $q_{hit}$  le taux de mortalité réel pour l'année  $t$  de l'assuré  $i$  du portefeuille  $h$  et  $q_{hit}^{ref}$  le taux issu de la table de référence.

L'introduction de la variable  $t$  nous permet de travailler en montant ou bien en fréquence en posant  $b_{hit} = 1$ .

Pour chaque portefeuille, nous calculons les quantités suivantes :

- $A_h = \sum_{t=1}^{n_T} \sum_{i=1}^{n_{ht}} b_{hit} D_{hit}$  Cette variable aléatoire représente le poids financier associé aux décès du portefeuille  $h$ .
- $E_h = \sum_{t=1}^{n_T} \sum_{i=1}^{n_{ht}} b_{hit} f_{hit} q_{hit}^{ref}$  Cette variable aléatoire représente le poids financier associé aux décès du portefeuille  $h$ .
- $\widehat{m}_h = \frac{A_h}{E_h}$  L'estimateur du ratio O/A.

Nous faisons également les deux hypothèses suivantes :

- Pour chaque portefeuille  $h$ , il existe ratio O/A  $m_h$ . Conditionnellement à ce paramètre,  $D_{hit}$  suit une loi de Bernoulli de paramètre  $m_h f_{hit} q_{hit}^{ref}$ . Nous avons donc  $E(D_{hit}/m_h) = m_h f_{hit} q_{hit}^{ref}$ .
- Les  $m_h$  sont des variables iid dont nous supposons connaître  $E(m_h) = \mu$  et la variance  $Var(m_h) = \sigma^2$

Nous cherchons le meilleur estimateur de  $m_h$  qui s'écrit sous la forme linéaire

$$\widetilde{m}_h = \lambda_h \widehat{m}_h + (1 - \lambda_h) \mu$$

Dans un premier temps, nous supposons connaître les estimateurs de  $\mu$  et  $\sigma^2$  et cherchons le facteur de crédibilité  $\lambda_h$ . Dans un second temps, nous étudierons comment estimer  $\mu$  et  $\sigma^2$ .

### 5.1.1 Etude du facteur de crédibilité

Nous voulons déterminer le meilleur estimateur de  $m_h$  qui s'écrit sous la forme

$$\widetilde{m}_h = \lambda_h \widehat{m}_h + (1 - \lambda_h) \mu$$

Nous cherchons donc la valeur de  $\lambda_h$  qui minimise la distance quadratique de l'estimateur final à  $m_h$ . Ainsi, nous pouvons écrire

$$L = E[(m_h - \lambda_h \widehat{m}_h - (1 - \lambda_h) \mu)^2]$$

En développant,

$$L = E[m_h^2] - \lambda_h E[2m_h \widehat{m}_h] - (1 - \lambda_h) \mu E[2m_h] + \lambda_h^2 E[m_h^2] + 2\lambda_h(1 - \lambda_h) E[\widehat{m}_h] + (1 - \lambda_h)^2 \mu^2$$

Pour trouver le facteur de crédibilité cherché, nous nous intéressons donc à la dérivée de L par rapport à  $\lambda_h$  et résolvons :

$$\frac{\delta L}{\delta \lambda_h} = 0$$

Ce qui donne

$$E[m_h \widehat{m}_h] - \lambda_h E[\widehat{m}_h^2] - (1 - \lambda_h) \mu^2 = 0$$

Nous nous intéressons maintenant à chacun des termes de l'équation. Dans les étapes suivantes, nous allons utiliser les propriétés de l'espérance conditionnelle.

*Premier terme*

En premier, nous avons

$$E[m_h \widehat{m}_h] = E[m_h E[\widehat{m}_h | m_h]]$$

Or nous avons vu que  $\widehat{m}_h = \frac{A_h}{E_h}$ . Donc, en remplaçant,

$$E[m_h \widehat{m}_h] = \frac{1}{E_h} E[m_h E[A_h | m_h]]$$

Comme  $A_h = \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi} D_{hit}$

$$E[m_h \widehat{m}_h] = \frac{1}{E_h} E[m_h E[\sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi} D_{hit} | m_h]] = \frac{1}{E_h} E[m_h \sum_{i=1}^{n_{ht}} b_{hi} E[\sum_{t=1}^{n_t} D_{hit} | m_h]]$$

D'après les hypothèses effectuées précédemment,  $D_{hit} | m_h$  suit une loi de Bernoulli de paramètre  $m_h f_{hit} q_{it}^r e f$ .

$$E[m_h \widehat{m}_h] = \frac{1}{E_h} E[m_h \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi} f_{hit} q_{hit}]$$

$$E[m_h \widehat{m}_h] = \frac{1}{E_h} E[m_h^2 E_h] = \mu^2 + \sigma^2$$

*Second terme*

Nous nous intéressons maintenant au second terme et procédons de la même façon.

$$E[\widehat{m}_h^2] = E[E[\widehat{m}_h^2 | m_h]]$$

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} E[E[A_h^2 | m_h]]$$

Puis, en développant le carré :

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} E\left[\sum_{t_1=1}^{n_{t_1}} \sum_{i=1}^{n_{ht_1}} \sum_{t_2=1}^{n_{t_2}} \sum_{j=1}^{n_{ht_2}} b_{hi} b_{hj} D_{hit_1} D_{hjt_2} | m_h\right]$$

Ensuite, nous réordonnons les termes :

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} E\left[\sum_{(i,t_1) \neq (j,t_2)} b_{hi} b_{hj} D_{hit_1} D_{hjt_2} | m_h\right] + \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} E[b_{hi}^2 D_{hit}^2 | m_h]$$

En réutilisant que  $D_{hit} | m_h$  suit une loi de Bernoulli  $m_h f_{hit} q_{it}^{ref}$ .

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} E\left[\sum_{(i,t_1) \neq (j,t_2)} b_{hi} b_{hj} m_h^2 f_{hit_1} q_{hit_1} f_{hjt_2} q_{hjt_2} + \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi}^2 f_{hit} m_h q_{it}^{ref} (1 - f_{hit} m_h q_{it}^{ref}) - (f_{hit} m_h q_{it}^{ref})^2\right]$$

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} E\left[\sum_{(i,t_1) \neq (j,t_2)} b_{hi} b_{hj} m_h^2 f_{hit_1} q_{hit_1} f_{hjt_2} q_{hjt_2} + \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi}^2 f_{hit} m_h q_{it}^{ref}\right]$$

Or, par les propriétés de la variance, nous avons  $E[m_h^2] = Var[m_h] + E[m_h] = \sigma^2 + \mu^2$ .

Ainsi

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} \sum_{(i,t_1) \neq (j,t_2)} b_{hi} b_{hj} f_{hit_1} q_{hit_1} f_{hjt_2} q_{hjt_2} (\sigma^2 + \mu^2) + \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi}^2 f_{hit} q_{it}^{ref} \mu$$

A nouveau, en réindexant, on peut écrire

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} \sum_{(i,t_1,j,t_2)} b_{hi} b_{hj} f_{hit_1} q_{hit_1} f_{hjt_2} q_{hjt_2} (\sigma^2 + \mu^2) + \sum_{t,i} b_{hi}^2 f_{hit} q_{it}^{ref} \mu - b_{hi}^2 f_{hit}^2 q_{it}^{ref^2} (\sigma^2 + \mu^2)$$

Ce qui donne :

$$E[\widehat{m}_h^2] = \frac{1}{E_h^2} E_h^2 (\sigma^2 + \mu^2) + \frac{\mu}{E_h^2} \sum_{t,i} b_{hi}^2 f_{hit} q_{it}^{ref} - \frac{\sigma^2 + \mu^2}{E_h^2} \sum_{t,i} b_{hi}^2 f_{hit}^2 q_{it}^{ref^2}$$

Pour simplifier ces expressions, nous notons

$$B_h = \sum_{t,i} b_{hi}^2 f_{hit} q_{it}^{ref} \text{ et } C_h = \sum_{t,i} b_{hi}^2 f_{hit}^2 q_{it}^{ref^2}$$

Nous avons donc

$$E[\widehat{m}_h^2] = \sigma^2 + \mu^2 + \frac{\mu}{E_h^2} B_h - \frac{\sigma^2 + \mu^2}{E_h^2} C_h$$

Pour rappel , nous avons

$$E[m_h \widehat{m}_h] - \lambda_h E[\widehat{m}_h^2] - (1 - \lambda_h) \mu^2 = 0$$

Donc en remplaçant avec les valeurs des espérances trouvées précédemment

$$\sigma^2 + \mu^2 - \lambda_h (\sigma^2 + \mu^2 + \frac{\mu}{E_h^2} B_h - \frac{\sigma^2 + \mu^2}{E_h^2} C_h) - (1 - \lambda_h) \mu^2 = 0$$

$$\sigma^2 - \lambda_h (\sigma^2 + \frac{\mu}{E_h^2} B_h - \frac{\sigma^2 + \mu^2}{E_h^2} C_h) = 0$$

La résolution de cette équation nous donne

$$\lambda_h = \frac{\sigma^2}{\sigma^2 + \frac{\mu}{E_h^2} B_h - \frac{\sigma^2 + \mu^2}{E_h^2} C_h}$$

En dépendant de  $\sigma$ , cette expression prend en compte la volatilité qu'il existe entre les portefeuilles. Ce qui n'est pas le cas dans le cadre de la théorie de la crédibilité partielle, le facteur de crédibilité ne dépendant pas uniquement du nombre de sinistres observés.

### 5.1.2 Estimation de $\mu$ et $\sigma^2$

Il nous reste maintenant à estimer les paramètres  $\mu$  et  $\sigma^2$  du problème.

*Estimation de  $\mu$*

Pour  $\mu$ , nous disposons de l'estimateur naturel.

$$\widehat{\mu} = \frac{\sum_{h=1}^r \sum_{t=1}^{n_t} \sum_{i=1}^{n_{ht}} b_{hi} D_{hit}}{\sum_{h=1}^r E_h} = \frac{A}{T}$$

On peut facilement montrer que cet estimateur est sans biais.

*Estimation de  $\sigma^2$*

On s'intéresse maintenant à la variance  $\sigma^2$  dont l'estimation n'est pas intuitive.

Nous considérons

$$S = \sum_{h=1}^r E_h (\widehat{m}_h - \widehat{\mu})^2$$

Puis développons cette quantité

$$S = \sum_{h=1}^r E_h(\widehat{m}_h - \widehat{\mu})^2 = \sum_{h=1}^r E_h \widehat{m}_h^2 - 2\widehat{m}uA + T\widehat{\mu}^2$$

$$S = \sum_{h=1}^r E_h(\widehat{m}_h - \widehat{\mu})^2 = \sum_{h=1}^r E_h \widehat{m}_h^2 - T\widehat{\mu}^2$$

Nous nous intéressons maintenant à l'espérance de cette expression, car elle dépend de  $\sigma$ . Nous examinons chacun des termes de l'expression.

*Premier terme*

Nous avons déjà calculer  $E[\widehat{m}_h^2]$  précédemment. En remplaçant,

$$E\left(\sum_{h=1}^r E_h(\widehat{m}_h^2)\right) = \sum_{h=1}^r E_h\left[(\sigma^2 + \mu^2) + \frac{\mu}{E_h^2}B_h - \frac{\sigma^2 + \mu^2}{E_h^2}C_h\right]$$

$$E\left(\sum_{h=1}^r E_h(\widehat{m}_h^2)\right) = T(\sigma^2 + \mu^2) + \mu \sum_{h=1}^r \frac{B_h}{E_h} - (\sigma^2 + \mu^2) \sum_{h=1}^r \frac{C_h}{E_h}$$

*Second terme*

Pour le second terme, nous obtenons

$$E(T\widehat{\mu}^2) = E\left[T\frac{A^2}{T^2}\right]$$

$$E(T\widehat{\mu}^2) = \frac{1}{T}E\left[\left(\sum_{h=1}^r A_h\right)^2\right]$$

$$E(T\widehat{\mu}^2) = \frac{1}{T}E\left[\left(\sum_{h=1}^r E_h \widehat{m}_h\right)^2\right]$$

$$E(T\widehat{\mu}^2) = \frac{1}{T}E\left[\left(\sum_{h=1}^r \sum_{g=1}^r E_h E_g \widehat{m}_h \widehat{m}_g\right)\right]$$

En regroupant les termes où  $g = h$  :

$$E(T\widehat{\mu}^2) = \frac{1}{T}E\left[\left(\sum_{h \neq g} E_h E_g \widehat{m}_h \widehat{m}_g + \sum_{h=1}^r E_h^2 \widehat{m}_h^2\right)\right]$$



En utilisant les calculs précédents et le fait que  $E[m_h] = \mu$ ,

$$E(T\hat{\mu}^2) = \frac{1}{T} \sum_{h \neq g} (E_h E_g \mu^2 + \sum_{h=1}^r E_h^2 (\mu^2 + \sigma^2) + \mu B_h - (\mu^2 + \sigma^2) C_h]$$

$$E(T\hat{\mu}^2) = \frac{1}{T} [T^2 \mu^2 + \sum_{h=1}^r (E_h^2 \sigma^2 + \mu B_h - (\mu^2 + \sigma^2) C_h)]$$

Ainsi, en réinjectant les espérances calculées,

$$\begin{aligned} & T(\mu^2 + \sigma^2) + \mu \sum_{h=1}^r \frac{B_h}{E_h} - (\mu^2 + \sigma^2) \sum_{h=1}^r \frac{C_h}{E_h} - T\mu^2 - \frac{1}{T} \sum_{h=1}^r (\sigma^2 E_h^2 + \mu B_h - (\mu^2 + \sigma^2) C_h) \\ &= \sigma^2 (T - \frac{1}{T} \sum_{h=1}^r E_h^2 - \sum_{h=1}^r \frac{C_h}{E_h} + \frac{1}{T} \sum_{h=1}^r C_h) + \mu (\sum_{h=1}^r \frac{B_h}{E_h} - \frac{1}{T} \sum_{h=1}^r B_h) - \mu^2 (\sum_{h=1}^r \frac{C_h}{E_h} - \frac{1}{T} \sum_{h=1}^r C_h) \end{aligned}$$

Ce qui donne

$$E(\sum_{h=1}^r E_h (\hat{m}_h - \hat{\mu})^2) = \sigma^2 (T - \frac{1}{T} \sum_{h=1}^r E_h^2 - \sum_{h=1}^r \frac{C_h}{E_h} + \frac{1}{T} \sum_{h=1}^r C_h) + \mu (\sum_{h=1}^r \frac{B_h}{E_h} - \frac{1}{T} \sum_{h=1}^r B_h) - \mu^2 (\sum_{h=1}^r \frac{C_h}{E_h} - \frac{1}{T} \sum_{h=1}^r C_h)$$

On peut alors isoler la variance et on a l'estimateur

$$\sigma^2 = \frac{E(\sum_{h=1}^r E_h (\hat{m}_h - \hat{\mu})^2) - \mu (\sum_{h=1}^r \frac{B_h}{E_h} - \frac{1}{T} \sum_{h=1}^r B_h) - \mu^2 (\sum_{h=1}^r \frac{C_h}{E_h} - \frac{1}{T} \sum_{h=1}^r C_h)}{(T - \frac{1}{T} \sum_{h=1}^r E_h^2 - \sum_{h=1}^r \frac{C_h}{E_h} + \frac{1}{T} \sum_{h=1}^r C_h)}$$

Nous ne pouvons calculer  $\mu$ , nous faisons donc l'approximation  $\hat{\mu} = \mu$ . Nous obtenons finalement

$$\sigma^2 = \frac{\sum_{h=1}^r E_h (\hat{m}_h - \hat{\mu})^2 - \hat{\mu} (\sum_{h=1}^r \frac{B_h}{E_h} - \frac{1}{T} \sum_{h=1}^r B_h) - \hat{\mu}^2 (\sum_{h=1}^r \frac{C_h}{E_h} - \frac{1}{T} \sum_{h=1}^r C_h)}{(T - \frac{1}{T} \sum_{h=1}^r E_h^2 - \sum_{h=1}^r \frac{C_h}{E_h} + \frac{1}{T} \sum_{h=1}^r C_h)}$$

## 5.2 Application à nos données

### 5.2.1 Les portefeuilles de l'étude

Pour appliquer la méthode précédemment présentée, nous choisissons en premier les portefeuilles sur lesquelles seront réalisés l'étude.

Pour les hommes, nous avons construit au chapitre trois un ensemble de quatre portefeuilles à partir des produits d'épargne. En ajoutant le portefeuille de l'étude des rentes, nous disposons donc de cinq portefeuilles, ce qui est suffisant pour appliquer la théorie de la crédibilité de Klugman.

Pour les femmes, en revanche, nous ne disposons que de deux portefeuilles issus des portefeuilles rentes. Appliquer la théorie de Klugman n'est donc pas très judicieux.

Nous devons également déterminer si nous appliquons la méthode en raisonnant sur les montants ou les fréquences d'apparition. De part la façon dont nous avons construit les portefeuilles épargne, en segmentant selon les niveaux d'encours, nous choisissons l'approche fréquentielle.

Nous passons maintenant à l'estimation des différents paramètres du problème et obtenons les taux d'abattements suivants.

Pour les hommes, nous avons :

Plage d'âges	40-100	60-90	65-85	60-99
$\hat{\mu}$	1,28	1,28	1,29	1,25
$\hat{\sigma}^2$	0.05	0.05	0.05	0.05
$\widehat{SMR}_{rentes}$	1,11	1,14	1,11	1,10
Facteur de crédibilité	91%	93%	91%	93%
$\widehat{SMR}$	1,09	1,15	1,12	1,11

TABLE 5.1 – Calcul du SMR épargne pour les hommes

Comme pour la théorie de la crédibilité partielle, nous choisissons de travailler sur les âges [60 – 99] ans.

### 5.2.2 Validation des tables obtenues

De même que pour les tables construites avec la théorie de la crédibilité partielle, les tables issues de la théorie de la crédibilité de Klusmann ne sont pas ajustées aux données. De la même façon que dans le chapitre précédent, nous nous assurons qu'elles sont cohérentes avec l'étude menée.

Nous suivons la même démarche en réalisant une analyse graphique sur les âges [65, 85] ans pour les hommes. Ce choix est effectué en vu de la fermeture de la table.

Nous commençons par l'analyse résidus. Pour ne pas alourdir ce rapport, nous ne présentons ici que les résidus obtenus pour l'année 2016.

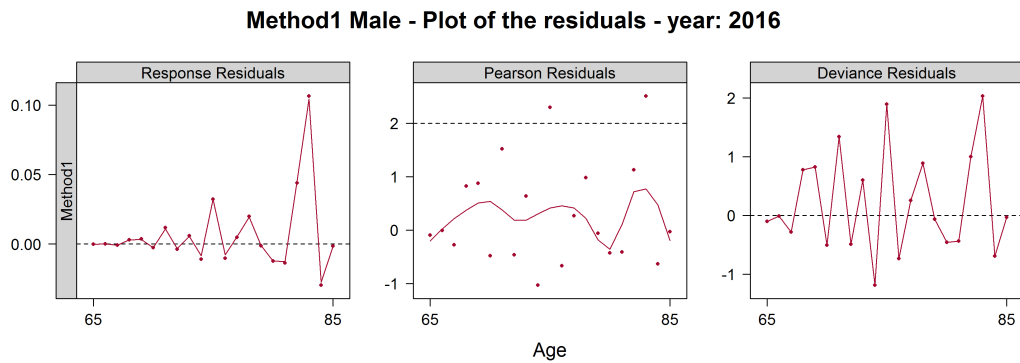


FIGURE 5.1 – Résidus de l’année 2016 pour les hommes

Les résidus de la réponse ne présentent pas de tendances tout comme les résidus de Pearson. Cependant, pour deux âges, les résidus de Pearson sortent à deux reprises de l’intervalle  $[-2,2]$ . La variabilité des données n’est donc pas parfaitement modélisée. Enfin, les résidus de la déviance oscillent autour de 0. Il n’y a donc pas de sur-lissage localement.

Nous passons maintenant à l’étude des intervalles de confiance pour les décès. Pour que notre modélisation donne une représentation fidèle de la réalité, il faut que les décès observés se situent entre les deux traits représentant les intervalles de confiance. Le graphique 5.2 représente les intervalles de confiance pour les hommes.

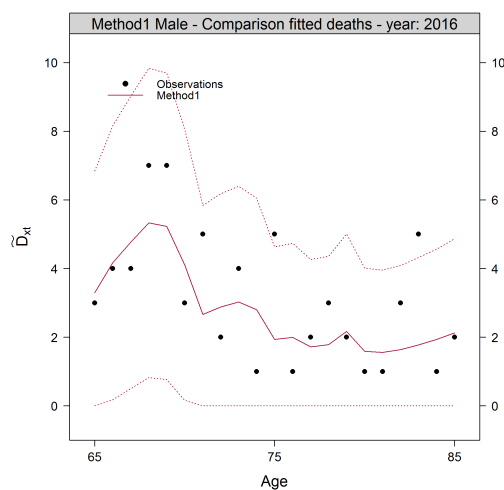


FIGURE 5.2 – Intervalles de confiance à 95% de l’année 2016 pour les hommes et les femmes

Malgré leur méthode de construction, les modèles obtenus captent bien la sinistralité de nos données d’origine.

Nous comptons pour chaque année le nombre de points en dehors de l’intervalle de confiance.

Année	2010	2011	2012	2013	2014	2015	2016
Hommes	0	1	0	0	0	0	3

TABLE 5.2 – Nombre de points en dehors des intervalles de confiance à 95%

Enfin, nous regardons l'ajustement obtenu pour l'année 2016.

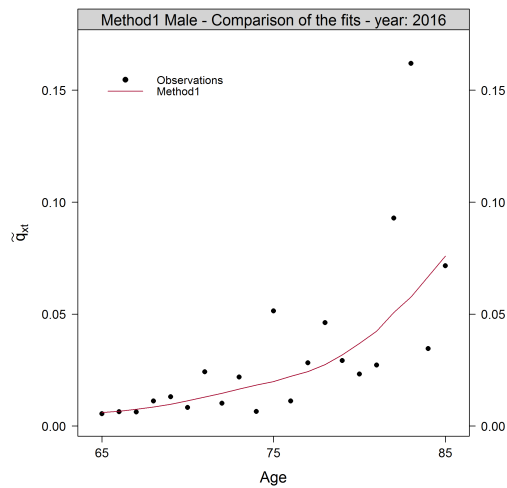


FIGURE 5.3 – Ajustement pour l'année 2016 pour les hommes et les femmes

Pour les hommes, nous remarquons que les observations sont très écartées de l'ajustement à partir de 75 ans.

### Fermeture de la table

D'après la répartition de nos observations, l'âge pivot pour la fermeture est fixé à 85 ans pour les hommes.

Nous suivons à nouveau la méthode de [Denuit and Goderniaux, 2005]. Pour l'année 2016, la fermeture de la table donne :

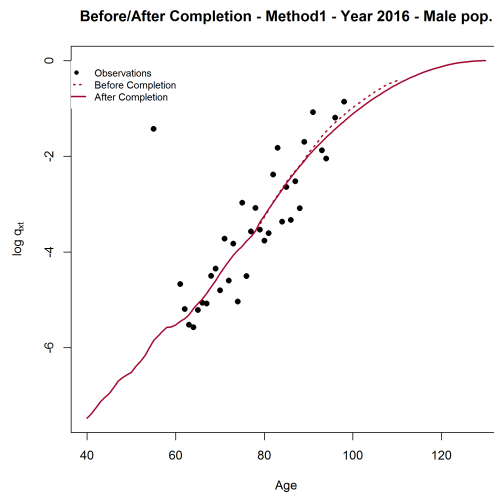


FIGURE 5.4 – Fermeture de la table pour l’année 2016 pour les femmes

Finalement, la théorie de la crédibilité de Klugman nous permet d’obtenir la table de mortalité suivante pour les hommes :

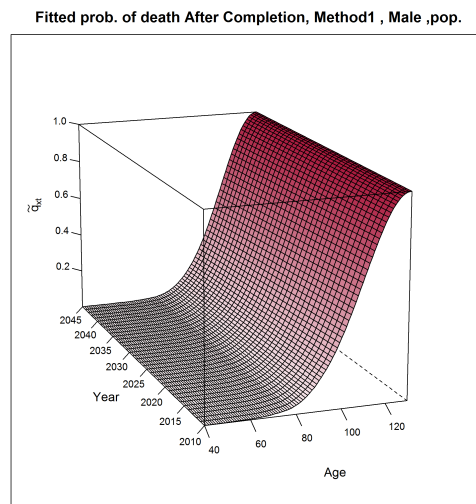


FIGURE 5.5 – Table prospective obtenu pour les hommes à l’aide de la théorie de la crédibilité de Klugman

### 5.2.3 Comparaison des tables obtenues

Les théories de la crédibilité partielle et de Klugman nous ont permis de construire deux tables prospectives pour les hommes. Nous devons maintenant déterminer laquelle retenir *in fine*.

## Première intuition

Pour rappel, le coefficient SMR retenu pour leur construction des tables est de 1,15 dans le cas de la théorie de la crédibilité partielle et de 1,11 dans le cadre de la crédibilité de Klugman.

L'apport de la théorie de la crédibilité de Klugman est la prise en compte de l'hétérogénéité entre les portefeuilles utilisés pour l'étude. Intuitivement, cela nous semble important et nous privilégions donc la table obtenue par la théorie de la crédibilité de Klugman.

Cependant, gardons à l'esprit qu'il est toujours mieux de favoriser le modèle le plus simple, c'est pourquoi si l'hétérogénéité des portefeuilles est faible, il paraît plus judicieux de choisir la théorie de la crédibilité partielle.

Idéalement, il nous faut un critère permettant de déterminer en fonction de l'hétérogénéité des portefeuilles de l'étude quelle méthode choisir.

Dans son mémoire, [Laizet, 2018] a proposé un critère permettant de choisir entre les théories de la crédibilité partielle et de Klugman dans le cadre de l'étude de portefeuilles emprunteurs. Cette étude a été réalisée pour des taux moyens d'abattement allant de 0,2 à 0,8.

Ce critère établit un seuil d'hétérogénéité en fonction du niveau moyen du taux d'abattement au-delà duquel la théorie de Klugman donne une meilleure estimation de la mortalité que la théorie de la crédibilité partielle.

Nous nous proposons de reprendre la méthodologie utilisée et de l'appliquer au cas de notre étude.

## Méthodologie de comparaisons

Nous commençons par construire N groupes de 4 portefeuilles dont nous simulons la mortalité.

**Étape 1 :** Nous générons l'exposition initiale de chaque portefeuille simulé. Nous choisissons un nombre d'individus pour chaque portefeuille et nous tirons leur âge suivant une loi normale centrée de caractéristiques :

Portefeuille	1	2	3	4
Espérance	70	70	70	70
Variance	7	7	7	7
Exposition	20000	100000	100000	100000

TABLE 5.3 – Caractéristiques des portefeuilles simulés

Nous ne retenons que les individus entre 60 et 100 ans. Les choix de l'âge moyen, de la variance et de la limite d'âge sont parfaitement arbitraires et sont établis afin de correspondre approximativement aux portefeuilles de l'étude.

**Étape 2 :** Nous déterminons à l'aide de la loi uniforme les taux d'abattement à appliquer sur la table réglementaire pour obtenir les véritables taux de mortalité de chacun des portefeuilles.

**Étape 3 :** Nous simulons une observation possible de la sinistralité de chaque portefeuille. Le décès de chaque individu suit une loi binomiale. La probabilité de décès n'est autre que la probabilité de décès suivant la table réglementaire abattue du coefficient généré à l'étape 2.

Pour chaque ensemble de portefeuilles simulés, nous appliquons les théories de la crédibilité partielle et de Klugman. Nous comparons ensuite les résultats obtenus par rapport au taux de mortalité déterminé à l'étape 2.

### Application

Nous appliquons la méthode décrite pour des niveaux de mortalité de 1,25 et 1,30. Ceux-ci correspondent au niveau moyen du ratio O/A des portefeuilles suivant la plage d'âge choisie.

Dans le tableau suivant, nous présentons en fonction du niveau de mortalité et de la variance des portefeuilles le pourcentage de fois où la théorie de la crédibilité partielle donne une meilleure estimation de la mortalité que la théorie de la crédibilité de Klugman.

$\sigma^2$	$\mu = 1,25$	$\mu = 1,30$
0,045	29,8%	29,3%
0,0475	31,3%	30,5%
0,05	33,5%	32,3%
0,0525	35,2%	33,3%
0,055	36,2%	34,9%

TABLE 5.4 – Caractéristiques des portefeuilles simulés

Nos portefeuilles présentent une variance de  $\sigma^2 = 0,05$ . En moyenne, dans notre cas, le théorie de Klugman conduit à une meilleure estimation de la mortalité.

Finalement, nous choisissons de retenir la table issue de la théorie de la crédibilité de Klugman pour modéliser la mortalité des hommes.





# Chapitre 6

## Conclusion

Ce travail nous a permis de construire des tables de mortalité prospectives en tenant compte de l'expérience du portefeuille rentes et en limitant le risque d'échantillon.

Après avoir étudié les caractéristiques du portefeuille rentes, sujet de notre étude, nous nous sommes tournés vers le portefeuille épargne. Nous avons réalisé une segmentation suivant les niveaux de richesse de chaque assuré pour créer un portefeuille fictif dont la loi de décès est proche de celle du portefeuille rentes. Puis, nous avons intégré les théories de la crédibilité partielle et de Klugsmann au modèle SMR pour construire des tables de mortalité prospectives.

Dans le cas des femmes, seule la théorie de la crédibilité partielle nous a permis d'aboutir à une table de mortalité prospective. Pour les hommes, après comparaison des tables issues des deux méthodes, nous avons choisi de retenir la table issue de la crédibilité de Klugman.

Cependant, nous devons garder à l'esprit que, de part la méthode utilisée, les modèles obtenus ne sont pas *best estimate*, mais uniquement plus proche de la réalité que les tables réglementaires.



# Chapitre 7

## Annexes

### 7.1 Annexe 1 - Statistiques descriptives du portefeuille rentes

Années	2010	2011	2012	2013	2014	2015	2016	2017
Âges Moyen hommes	66,4	66,8	67,0	67,1	67,3	67,5	67,9	68,2
Âges Moyen femmes	74,6	73,7	73,0	72,2	71,8	71,3	70,9	70,6
Âges rentes hommes	67,0	67,8	68,2	68,5	68,9	69,0	69,3	69,6
Âges rentes femmes	77,4	77,7	77,0	76,0	75,3	74,7	74,2	74,0

## 7.2 Annexe 2 - Méthode de Wittaker-Henderson en dimension 2

Nous avons vu au chapitre 3 la méthode de Wittaker-Henderson en dimensions. Nous l'étendons à la dimension 2 pour pouvoir l'utiliser pour, par exemple, des tables prospectives.

Pour cela, nous généralisons le critère de fidélité par :

$$F = \sum_t \sum_{x_{min}+1}^{x_{max}} w_{xt} (\widetilde{q}_{xt} - \widehat{q}_{xt})^2$$

Le critère de régularité se généralise en deux sommes pour distinguer la régularité verticales (suivant les  $x$ ) et la régularité horizontale (suivant les  $t$ ).

$$R_v = \sum_t \sum_{x_{min}+1}^{x_{max}-z} (\Delta^z \widetilde{q}_{xt})^2$$

$$R_h = \sum_{x_{min}+1}^{x_{max}} \sum_t^{n-y} (\Delta^y \widetilde{q}_{xt})^2$$

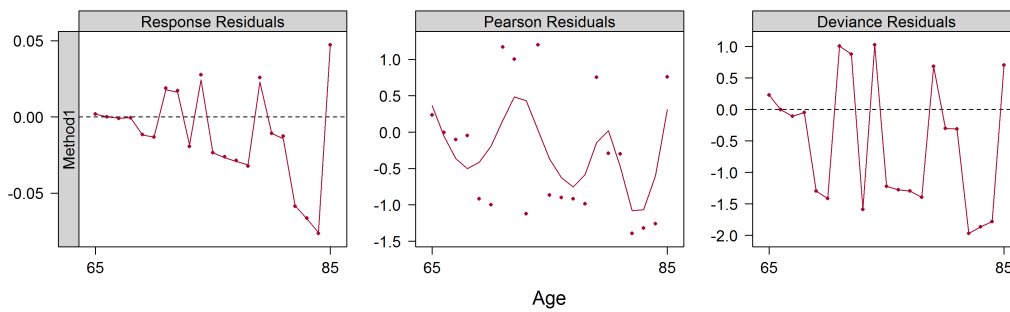
La fonction à minimiser devient alors :

$$M = F + h_v R_v + h_h R_h$$

Par un changement d'indexation judicieux, il est possible de se ramener au cas unidimensionnel.

## 7.3 Annexe 3 - Etude graphique pour les hommes Théorie de la crédibilité partielle

Method1 Male - Plot of the residuals - year: 2010



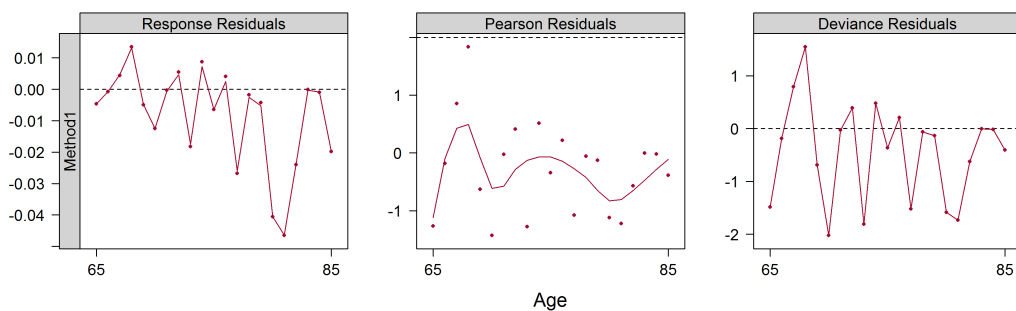
Method1 Male - Plot of the residuals - year: 2011



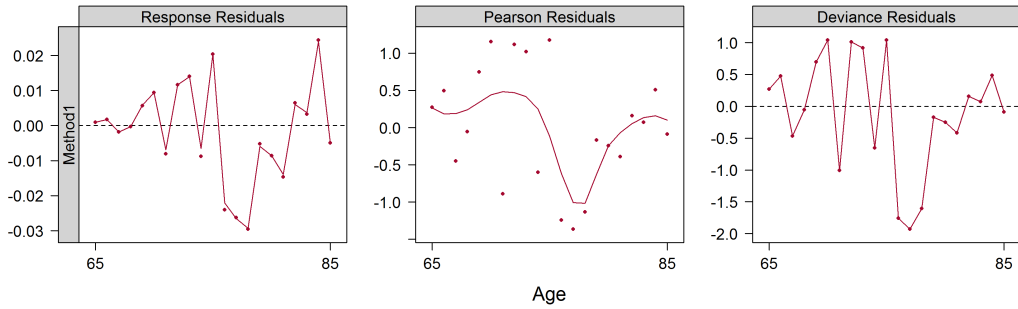
Method1 Male - Plot of the residuals - year: 2012



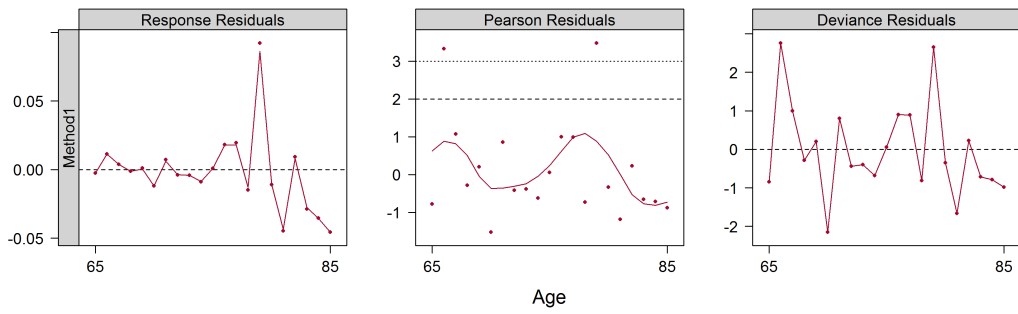
Method1 Male - Plot of the residuals - year: 2013



**Method1 Male - Plot of the residuals - year: 2014**



**Method1 Male - Plot of the residuals - year: 2015**



**Method1 Male - Plot of the residuals - year: 2016**



## 7.4 Annexe 4 - Etude graphique pour les femmes Théorie de la crédibilité partielle

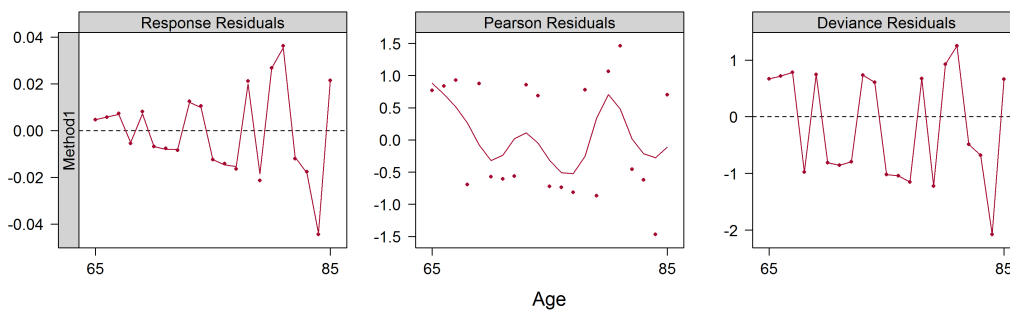
Method1 Female - Plot of the residuals - year: 2010



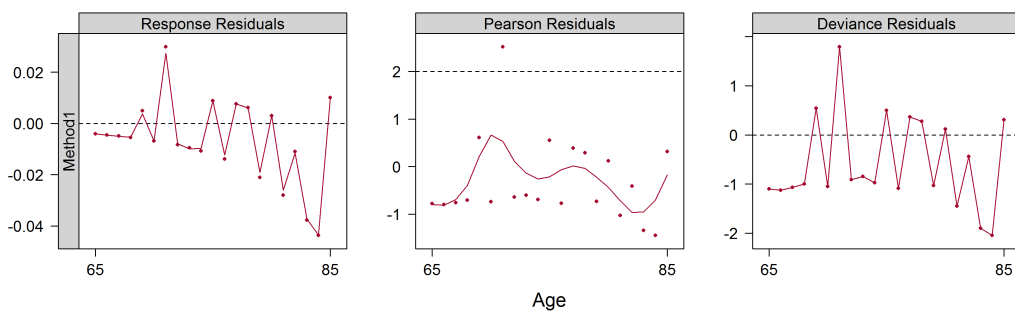
Method1 Female - Plot of the residuals - year: 2011



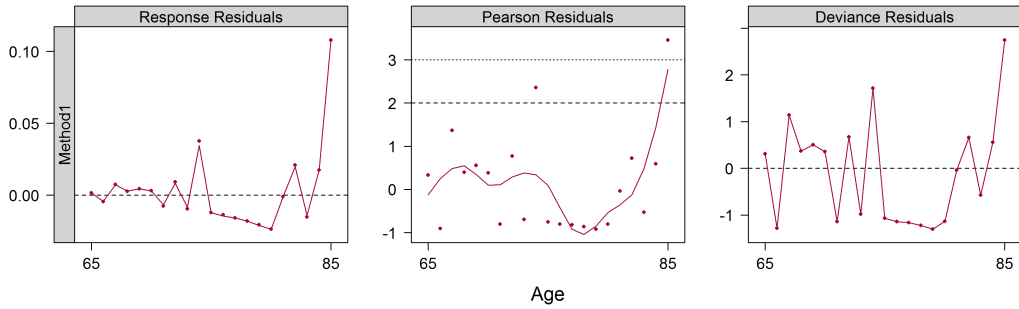
Method1 Female - Plot of the residuals - year: 2012



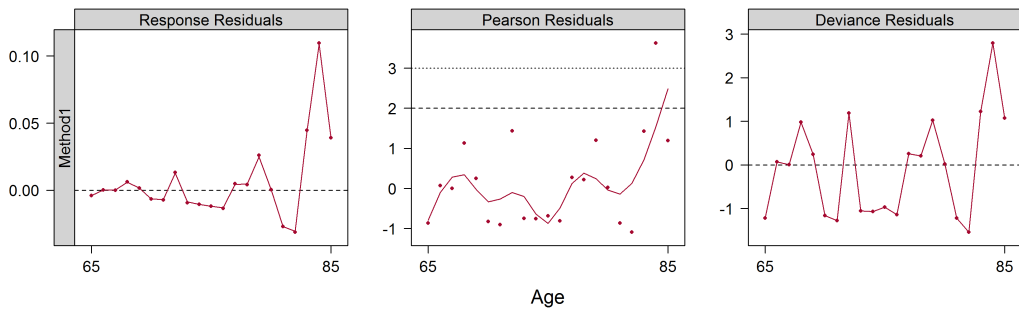
Method1 Female - Plot of the residuals - year: 2013



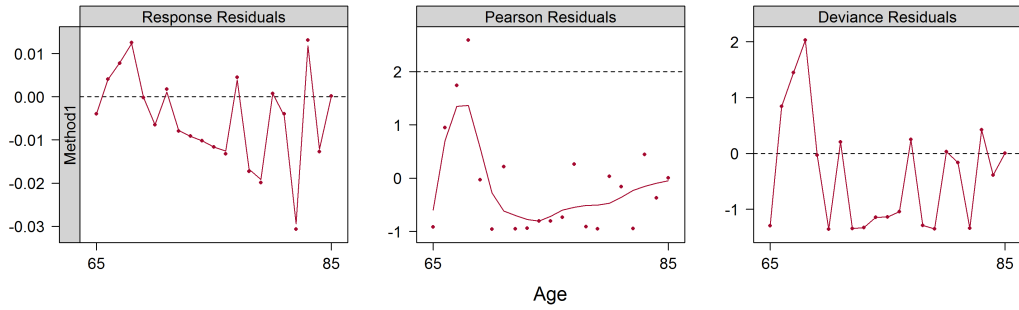
**Method1 Female - Plot of the residuals - year: 2014**



**Method1 Female - Plot of the residuals - year: 2015**



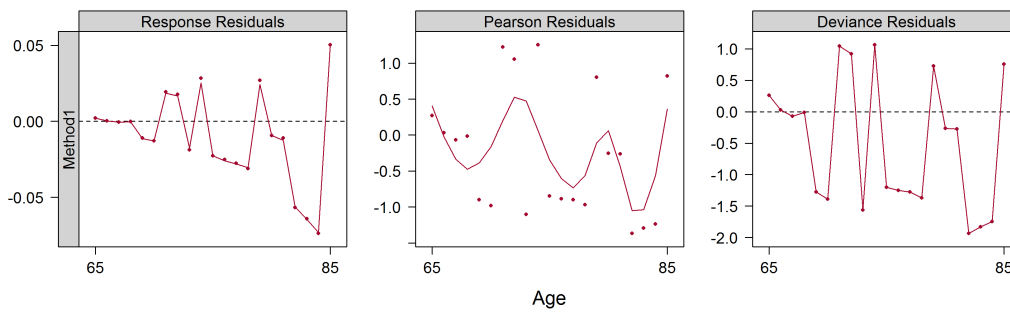
**Method1 Female - Plot of the residuals - year: 2016**



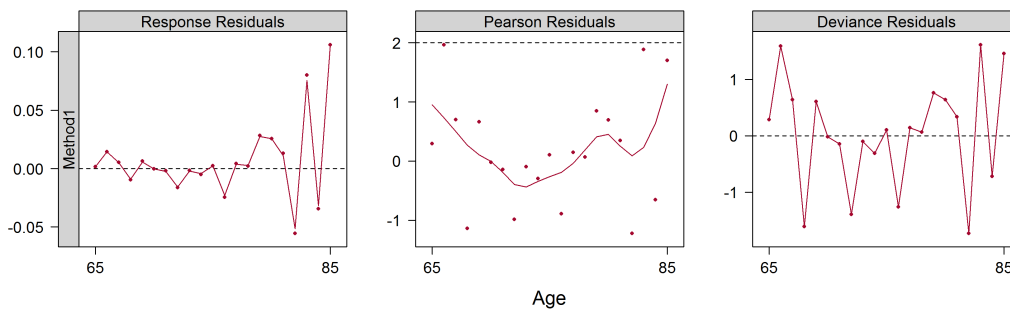


## 7.5 Annexe 5 - Etude graphique pour les homme Théorie de la crédibilité de Klugman

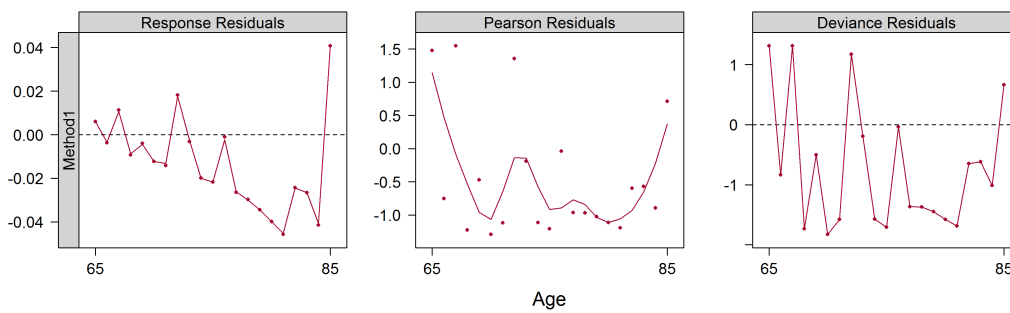
Method1 Male - Plot of the residuals - year: 2010



Method1 Male - Plot of the residuals - year: 2011



Method1 Male - Plot of the residuals - year: 2012



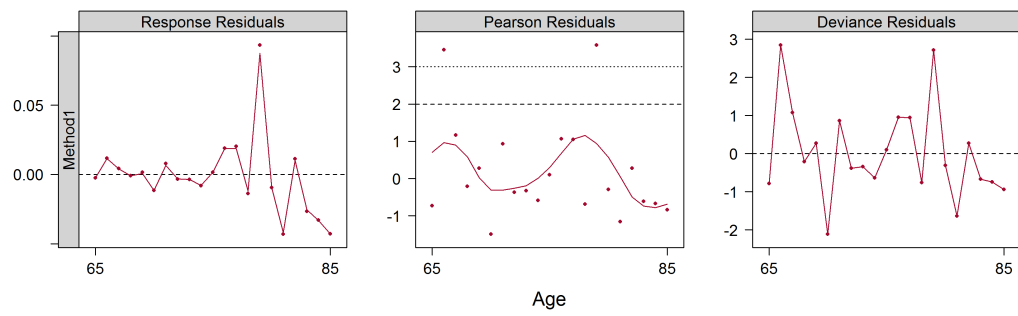
Method1 Male - Plot of the residuals - year: 2013



**Method1 Male - Plot of the residuals - year: 2014**



**Method1 Male - Plot of the residuals - year: 2015**



**Method1 Male - Plot of the residuals - year: 2016**



# Bibliographie

- [Bühlman, 1967] Bühlman, H. (1967). Experience rating and credibility. *ASTIN Bulletin*.
- [Cochran, 1954] Cochran, W. G. (1954). Some methods for strengthening the common  $\chi^2$  tests. *Biometrics*.
- [Denuit and Goderniaux, 2005] Denuit, M. and Goderniaux, A.-C. (2005). Closing and projecting lifetables using log-linear models. *Bulletin of the Swiss association of Actuaries*, pages 29–49.
- [Henderson, 1924] Henderson, R. (1924). A new method of graduation. *Transactions of the Actuarial Society of America*.
- [Hoem, 1971] Hoem, J. M. (1971). Point estimation of forces of transition in demographic models. *Journal of the Royal Statistical Society*.
- [IA, 2006] IA (2006). *Lignes directrices mortalité de la Commission d’Agrément*.
- [INSEE, 2018] INSEE (2018). L’espérance de vie par niveau de vie : chez les hommes, 13 ans d’écart entre les plus aisés et les plus modestes. *Isee Première*.
- [Kaplan and Meier, 1958] Kaplan, E. L. and Meier, P. (1958). Non parametric estimation from incomplete observations. *Journal of the American Statistical Association*.
- [Klugman et al., 2009] Klugman, S., Rhodes, T., Purushotham, M., and Gill, S. (2009). *Credibility Theory Practices*.
- [Laizet, 2018] Laizet, B. (2018). Apports de la théorie de la crédibilité à l’estimation de la mortalité. *Mémoire d’actuariat*.
- [MacQueen, 1967] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*.
- [Mowbray, 1914] Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*.
- [Planchet, 2006] Planchet, F. (2006). *Tables de mortalité d’expérience pour les portefeuilles de rentiers*.
- [Tomas and Planchet, 2014a] Tomas, J. and Planchet, F. (2014a). *Critère de validation : Aspects méthodologiques (Note III291-14v1.4)*.
- [Tomas and Planchet, 2014b] Tomas, J. and Planchet, F. (2014b). *Méthode de positionnement : Aspects méthodologiques (Note III291-12v1.7)*.

[Whitney, 1918] Whitney, A. (1918). The theory of experience rating? *Proceedings of the Casualty Actuarial Society*.

[Whittaker, 1923] Whittaker, E. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*.