

**Mémoire présenté pour la validation de la Formation
« Certificat d'Expertise Actuarielle »
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires
le**

Par: Driss ASRARY

Titre: Modélisation des rachats et réductions pour un portefeuille Vie entière

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

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

Membres présents du jury de l'Institut des
actuaires :

Membres présents du jury de l'Institut du Risk
Management :

Secrétariat :

Bibliothèque :

Entreprise : Mutuelle Mieux-Etre

Nom :

ASRARY

Signature et Cachet :



Directeur de mémoire en entreprise :

Nom :

HUYGHE

Signature :



Invité :

Nom :

Signature :

**Autorisation de publication et de mise en
ligne sur un site de diffusion de documents
actuariels**

(après expiration de l'obligation de confidentialité)

**MUTUELLE
MIEUX-ETRE**

Signature du responsable entreprise

171 Avenue Ledru Rollin
75544 PARIS Cedex 11



Signature(s) du candidat(s)

Résumé

Mots Clés : vie entière, rachat, réduction, gestion Actif Passif, solvabilité II, comptes sociaux, SCR, provisions, facteurs de risque, régression logistique.

Le produit vie entière est un contrat individuel qui vise à garantir le versement d'un capital au(x) bénéficiaire(s) désigné(s) par l'assuré quelle que soit la date du décès, en échange de quoi le souscripteur prend l'engagement révoquant de verser des primes en fonction de la réalisation d'évènements viagers. Il comporte des options et des garanties pour alléger l'engagement de l'assuré et rendre liquide le produit. Parmi ces options on trouve le rachat et la réduction.

Ces options exposent l'assureur à un réel risque de faillite, si elles ne sont pas prises en compte dans la modélisation du passif. Elles conduisent également à des problématiques liées à la gestion Actif Passif.

La mise en place de la réglementation solvabilité II en 2016, a renforcé la nécessité d'approfondir la connaissance et la compréhension de ces risques, au point qu'un module entier est dédié à quantifier l'impact du rachat via des stress tests dans le SCR de souscription. Ces options sont également intégrées dans l'évaluation au plus juste des provisions, alors que dans les comptes sociaux ces provisions sont calculées avec l'hypothèse que l'assuré continue de régler les cotisations jusqu'à son décès.

La bonne maîtrise de ces risques est donc importante. C'est dans ce contexte que s'inscrit ce mémoire, dans l'objectif de modéliser le comportement de rachat et celui de la réduction.

Cette étude se traduit d'abord par une analyse du portefeuille, pour détecter les facteurs de risques potentiels, ensuite une méthodologie de segmentation est appliquée afin de choisir les variables pertinentes susceptibles d'impacter le comportement de l'assuré.

Le mémoire se termine par l'implémentation du modèle de la régression logistique pour quantifier et mesurer les effets de chacune de ses variables sur le comportement de rachat et la réduction.

Abstract

Keywords : whole life, surrender, reduction, asset and liability management, solvency II, social accounts, SCR, provisions, risk factors, logistic regression.

The whole life product is an individual contract which aims to guarantee the payment of a capital to the beneficiarie (s) designated by the insured regardless of the date of death, in exchange for which the subscriber makes a revocable undertaking to pay premiums according to the realization of life events. It includes options and guarantees to reduce the undertaking of the insured and make the product liquid. Among these options we find surrender and reduction.

These options expose the insurer to a real risk of bankruptcy, if they are not taken into account in the modeling of liabilities, they also lead to issues of asset liability management.

The implementation of the of the Solvency II in 2016 reinforced the need to deepen the knowledge and understanding of these risks, to the point that an entire module is dedicated to the evaluation of the surrender in the underwriting SCR. These options are also integrated into the fair valuation of provisions, while in the social accounts these provisions are calculated with the assumption that the insured continues to pay these premiums until his death.

That is why it is essential to control these risks. It is in this context that this thesis takes place, with the objective of modeling the behavior of surrender and reduction.

This study begins with a portfolio analysis to detect potential risk factors, then a segmentation methodology is applied in order to choose the relevant variables likely to impact the behavior of the insured.

The thesis ends with the implementation of the Logistic regression model to quantify and measure the effects of each of its variables on the surrender and reduction variables.

Synthèse

Mots Clés : coefficient de Pearson, coefficient de détermination R^2 , indice de Cramer V, CART, CHAID, C4.5, échantillon test, validation croisée, STEPDISC, test de Fisher, Wald, rapport de vraisemblance, AIKAIKE, SCHWARTZ, SCR

Le contrat vie entière comporte des options dites cachées. Par options cachées, nous entendons les droits accordés aux assurés par la réglementation ou par des clauses contractuelles destinées à rendre le contrat plus souple et plus attractif. Parmi ces options nous citons le rachat et la réduction. Le rachat permet au souscripteur de résilier son contrat à tout moment, et de disposer d'une somme d'argent avant l'échéance du contrat, quant à la réduction, elle permet à l'assuré de cesser de payer les cotisations tout en restant assuré avec un capital garanti réduit.

La réglementation Solvabilité II fait intervenir les lois de rachat et la réduction calibrées par les assureurs, dans le calcul des provisions en Best Estimate, elles interviennent également dans certains stress tests pour le calcul du SCR de souscription. Elle a renforcé de ce fait la nécessité d'approfondir la connaissance et la compréhension de ces risques. C'est à ce titre que nous abordons dans ce mémoire les problématiques de la modélisation du rachat et la réduction, dans le but de mieux connaître et prendre en compte l'ensemble des facteurs clés qui jouent sur les décisions des assurés.

On distingue deux types de facteurs clés, les facteurs conjoncturels ou économiques liés au marché financier, et les facteurs structurels liés au comportement de l'assuré, à ses caractéristiques personnelles, mais aussi aux caractéristiques du contrat.

Les facteurs économiques sont écartés de l'étude, d'une part car le contrat vie entière est un contrat de prévoyance et non un contrat de placement, il est destiné à couvrir des frais d'obsèques en cas de décès de l'assuré, et d'autre part car les montants des cotisations et capitaux garantis sont faibles comparés à ceux de l'épargne, et sont donc moins sensibles à la variation des taux.

L'objectif principal de ce mémoire est de quantifier les lois de rachat et la réduction, en fonction des facteurs structurels choisis par des méthodes statistiques.

Le portefeuille étudié est fermé à la commercialisation depuis 2013, il est composé de trois produits différents selon leurs modes de commercialisation : courté via un courtier délégué de gestion, direct vendu par la mutuelle mais géré par un délégué de gestion, et en acceptation assuré par le

Conservateur et réassuré par la mutuelle. Le produit courté constitue l'essentiel du portefeuille de la mutuelle Mieux-être.

La période d'observation correspond à l'intervalle de temps durant lequel les taux de rachat et la réduction sont observés. Elle est choisie de telle sorte que les taux sont réguliers ou stables durant cette période. Pour l'analyse du rachat, la période retenue est du 01/01/2007 au 31/12/2018, pour l'analyse de la réduction elle est du 01/01/2007 au 31/12/2013.

L'étude commence par un retraitement des données afin de supprimer les données aberrantes dues en grande partie à des erreurs de saisie, à corriger les erreurs détectées, et à créer de nouvelles variables explicatives.

Afin de s'assurer qu'il n'y a pas de risque de biais dans la sélection des données, une étude d'interaction est appliquée entre les différentes variables, dont l'objectif est de supprimer les variables dépendantes. Des indicateurs statistiques sont utilisés selon le type de la variable :

- Coefficient de Pearson pour la liaison entre les variables quantitatives,
- Coefficient de détermination R^2 , basé sur la statistique de Fisher pour la liaison entre variables qualitatives et quantitatives,
- L'indice de Cramer V pour la liaison entre variables qualitatives.

Cette analyse a permis de supprimer des variables telles que les cotisations, l'âge au rachat et le montant de la provision.

L'analyse descriptive de la base de données donne une bonne vision du portefeuille dans sa globalité. On constate dans la population globale étudiée la présence majoritaire des femmes avec 54%, l'âge moyen à la souscription est de 59 ans, avec une durée de vie moyenne du contrat de 5 ans. L'analyse graphique des taux de rachat et de réduction ressort des effets de linéarité et de saisonnalité avec certaines variables explicatives.

Les méthodes de classification par arbre de décision : CART, CHAID, et C4.5 ont permis de sélectionner parmi les variables existantes, celles les plus discriminantes des taux de rachat et la réduction.

Pour juger de la qualité des algorithmes de classification, des techniques d'échantillonnage bien connues y sont appliquées :

- La technique d'échantillon test : les observations sont divisées en deux échantillons, le premier appelé apprentissage, servira à construire le modèle, et le deuxième appelé test, servira à évaluer et à comparer les modèles,

- La technique de la validation croisée : permet d'utiliser l'ensemble des observations pour la construction et la validation du modèle. Les observations sont découpées en k parties à peu près égales. Tour à tour, chacune des k parties est utilisée comme jeu de test, le reste est utilisé pour la construction du modèle.

La procédure STEPDISC est une autre méthode de sélection des variables explicatives, son principe consiste à effectuer une analyse discriminante pas à pas pour sélectionner un sous-ensemble des variables quantitatives à utiliser pour discriminer une variable binaire. Elle est basée sur la significativité du test de Fisher.

Les méthodes de classification et d'analyse discriminante ont permis de sélectionner les variables telles que l'ancienneté du contrat, le capital garanti et l'âge à la souscription comme étant les variables les plus pertinentes pour expliquer les comportements de rachat et la réduction.

La dernière partie du mémoire traite la modélisation des taux de rachat et la réduction par l'utilisation de la régression logistique. Cette méthode au-delà de sa disponibilité dans le logiciel SAS, permet d'une part de tester la significativité de l'ensemble des variables explicatives sélectionnées en entrée vis-à-vis de la variable à expliquer, et d'autre part de quantifier cette liaison par une estimation ponctuelle de la probabilité de rachat et réduction connaissant la réalisation des variables explicatives. La validité du modèle est mesurée à la fois grâce aux tests statistiques de Wald, du rapport de vraisemblance, et des indicateurs statistiques tels que les critères d'AIKAIKE et de SCHWARTZ.

L'application des taux issus du modèle de la régression logistique permettent une baisse des provisions techniques, conduisant à une légère baisse du SCR global.

Synthesis

Keywords : Pearson coefficient, the R^2 coefficient of determination, Cramer's V, CART, CHAID, C4.5, Test sample, Cross-validation, STEPDISC, Fisher, Wald, Likelihood Ratio statistical, AIKAIKE, SCHWARTZ, BSCR.

The whole Life product includes so-called hidden options. By hidden options, we mean the rights granted to policyholders by regulations or by contractual clauses intended to make the contract more flexible and more attractive. They generally depend on a decision by the insured. Among these options, we cite the surrender and reduction.

The surrender allows the subscriber to terminate his contract at any time, and to have a sum of money before the expiry of the contract, as for the reduction it allows him to stop paying premiums while remaining insured with a capital guaranteed reduced.

Solvency II involve surrender and reduction laws calibrated by insurers in the calculation of best estimate provisions, they are also involved in certain stress tests for the calculation of SCR. It therefore reinforced the need to deepen the knowledge and understanding of these risks. It is for this reason that we address in this brief the problems of modeling surrenders and reductions, with the aim of better understanding and taking into account all the key factors that influence the decisions of policyholders.

There are two types of key factors, cyclical or economic factors related to the financial market, and structural factors related to the behavior of the policyholder, his personal characteristics, but also the characteristics of the contract.

Economic factors are excluded from this study, because on the one hand the whole life contract is a Foresight contract and not an investment contract, intended to cover funeral costs in the event of the death of the insured, and on the other hand, the amounts of premiums and guaranteed capital are low compared to savings, and are therefore less impacted by rate fluctuations.

The main objective of this report is to quantify the laws of surrender and reduction according to the structural factors chosen by statistical methods.

The portfolio studied has been canceled since 2013, and is made up of three products distinguished according to their marketing methods: Broker via a broker-delegated management, Direct sold by the Mutual but managed by a delegated management, and Accepted reinsurance insured by the Conservateur and reinsured by the Mutual. The brokerage product constitutes the bulk of the MME portfolio.

The observation period is the time interval during which the surrender and reduction rates are observed. It is chosen so that the rates are regular or stable during this period. For the surrender analysis, the period used is from 01/01/2007 to 12/31/2018. For the reduction analysis, it is from 01/01/2007 to 12/31/2013.

The study begins by reprocessing the data in order to remove outliers due in large part to entry errors, correct the errors detected, and create new explanatory variables.

In order to ensure there is no risk of bias in the selection of data, an interaction study is applied between the different variables, the objective of which is to remove the dependent variables. Statistical indicators are used depending on the type of the variable :

- Pearson coefficient for the link between quantitative variables
- R^2 coefficient based on Fisher statistic for the link between qualitative and quantitative variables
- Cramer's V for the link between qualitative variables

With this analysis, the variables like premiums, surrender's age and the amount of provisions are removed.

The descriptive analysis of the database gives a good view of the portfolio as a whole. In the overall population studied, the majority presence of women is observed with 54%, the average age of subscription is 59 years, with an average contract life of 5 years. The graphical analysis of surrender and reduction rates emerges from the effects of linearity and seasonality with certain explanatory variables.

The classification methods: CART, CHAID, and C4.5 permit to select among the existing variables, those most discriminating in terms of surrender and reduction rates.

To judge the quality of the classification algorithms, some sampling techniques are applied:

- The test sample: the observations are divided into two samples, the first called training, will be used to build the model, and the second called test, will be used to evaluate and compare the models.
- The cross-validation technique: use all the observations for the construction and validation of the model. The observations are divided into k roughly equal parts. In turn, each of the k parts is used as a test set, the rest is used for building the model.

The STEPDISC procedure is another method of selecting explanatory variables, its principle consists of performing a step-by-step discriminant analysis to select a subset of the quantitative variables to be used to discriminate a binary variable. It is based on the significance of the Fisher test. The classification and discriminant analysis methods made it possible to select the variables as the

age of the contract, the guaranteed capital, the subscription's age as being the most relevant to explain surrender and reduction behaviors.

The last part of the report deals with the modeling of surrender and reduction rates by using logistic regression. This method, beyond its availability in the SAS software, permit on the one hand to test the significance of all the explanatory variables selected as input with respect to the variable to be explained, and to quantify this link by a point estimate of the probability of surrender or reduction knowing the achievement of the explanatory variables. Both Wald and Likelihood Ratio statistical tests, and statistical indicators such as the AIKAIKE and SCHWARTZ criteria measure the validity of the model.

the probability calculated from the logistic regression model, allows a decrease in provisions S2, and a slight decrease in SCR.

Remerciements

Je tiens à remercier Jean-Christophe HUYGHE, directeur technique et financier de la mutuelle Mieux-être pour m'avoir donné la possibilité de faire ce mémoire.

Je remercie également mon collègue Mohamed NEGROUZ pour sa relecture du mémoire et ces remarques.

Un grand merci à Yani CAMBOURIS ancien élève du CEA pour m'avoir encouragé à suivre cette formation et ainsi d'évoluer dans mon domaine.

Mes remerciements vont également à l'ensemble des enseignants de l'Institut du Risk Management, pour le partage de leur expérience. Formation qui m'a permis de progresser dans mon domaine et ainsi atteindre mes objectifs.

Et pour finir, un très grand merci à ma famille de m'avoir toujours soutenu et encouragé au cours de la réalisation de ce mémoire. Merci d'avoir supporté mon absence durant toutes ces années de formation.

Sommaire

Table des matières

Résumé	3
Abstract	4
Synthèse	5
Synthesis	8
Remerciements.....	11
Sommaire.....	12
Introduction.....	13
Préliminaire	15
I. Présentation de la mutuelle Mieux-être	15
II. Quelques chiffres de l'assurance obsèques en France	17
III. Présentation des phénomènes de rachat et de réduction.....	19
IV. Description technique du produit vie entière	22
V. Méthodologie actuelle de calcul	25
Base de données : Traitement-Analyse descriptive	30
I. Délégation de Gestion.....	30
II. Choix de la période d'observation	31
III. Présentation du portefeuille étudié	33
IV. Extraction Et Traitement des Données.....	34
V. Etude de liaison entre les variables explicatives.....	39
VI. Statistique descriptive des données.....	41
Sélection des variables	52
I. Classification avec les arbres de décision	52
II. Méthode STEPDISC	64
Modélisation par régression logistique et études d'impact.....	70
I. Modélisation par régression logistique	70
II. Etudes d'impact et tests de sensibilité.....	84
Conclusion générale	89
Bibliographie.....	91
Annexe 1	92
Annexe 2	93
Annexe 3	94

Introduction

Le portefeuille de La mutuelle mieux-être est composé principalement des garanties de frais de soins. Aujourd'hui l'un des axes de sa stratégie consiste à diversifier ces produits et s'orienter vers la prévoyance, notamment la prévoyance individuelle.

Le marché de l'assurance obsèques connaît une croissance importante, en raison du vieillissement de la population française, et de la hausse du nombre de décès annuels.

La hausse de sinistralité liée à l'explosion du nombre de décès en 2020 à cause de la crise sanitaire, a contribué à la prise de conscience des Français de l'avantage de prévoir en avance, l'organisation et le financement des obsèques notamment chez les personnes touchées par le décès d'un proche.

Au-delà de ces indicateurs, le contrat obsèques offre une légèreté et souplesse à l'assuré par les options qu'il comporte : le rachat et la réduction.

Bien que l'engagement de l'assureur soit irrévocable, le rachat donne droit à l'assuré de résilier son contrat à tout moment, tout en récupérant la provision mathématique constituée à la date de la demande après déduction des éventuelles pénalités.

La réduction donne droit à l'assuré de cesser de payer ces cotisations à tout moment, dans ce cas un nouveau capital garanti sera calculé en fonction du montant de la provision mathématique calculée à la date de la demande.

Ces options peuvent contribuer si elles ne sont pas prises en compte ou mal modélisées par l'assureur, à l'insolvabilité de l'assureur en cas de rachat ou de réduction massif.

La réglementation solvabilité II porte une attention particulière aux options de rachat et réduction, certains modules du SCR de souscription font intervenir des stress tests pour quantifier le rachat.

Ce mémoire a pour objectif de comprendre les mécanismes incitant les assurés à racheter et à réduire leurs contrats, à quantifier la probabilité de la réalisation de ces événements à travers un modèle statistique.

Le mémoire est organisé en quatre parties. Dans la première partie une présentation générale de la mutuelle, avec une présentation rapide du marché de l'obsèques est introduite, ensuite le contexte de l'étude et la description technique du produit sont abordés. Cette partie se termine par une présentation de la méthodologie actuelle pour le calcul des lois de rachat et réduction.

La deuxième partie commence par une présentation de l'environnement des données, puis aux différents traitements effectués pour constituer une base fiable. Une première sélection des variables explicatives est abordée, puis elle est suivie d'une étude de corrélation des différentes variables pour supprimer les variables dépendantes entre elles. Cette partie se termine par l'analyse des statistiques descriptives des variables, avec une description graphique du taux de rachat et réduction en fonction des variables susceptibles de les expliquer.

Dans la troisième partie, nous nous intéressons aux causes de rachat et réduction, nous abordons deux techniques de sélections des variables explicatives : la classification par arbres de décision, et l'analyse discriminante avec la procédure STEPDISC. Cette partie permet de choisir les variables les plus discriminantes pour le déclenchement de rachat et réduction.

L'étude sera terminée par l'introduction du modèle de régression logistique appliqué sur les variables présélectionnées. La base de données est divisée en deux parties, une première partie qu'on appelle base d'apprentissage utilisée pour l'estimation des paramètres du modèle, et une deuxième partie qu'on appelle échantillon test utilisée pour la validation du modèle. La robustesse du modèle est appréciée par le calcul du taux d'erreur de classement sur la base d'apprentissage et sur la base d'échantillon test.

Préliminaire

I. Présentation de la mutuelle Mieux-être

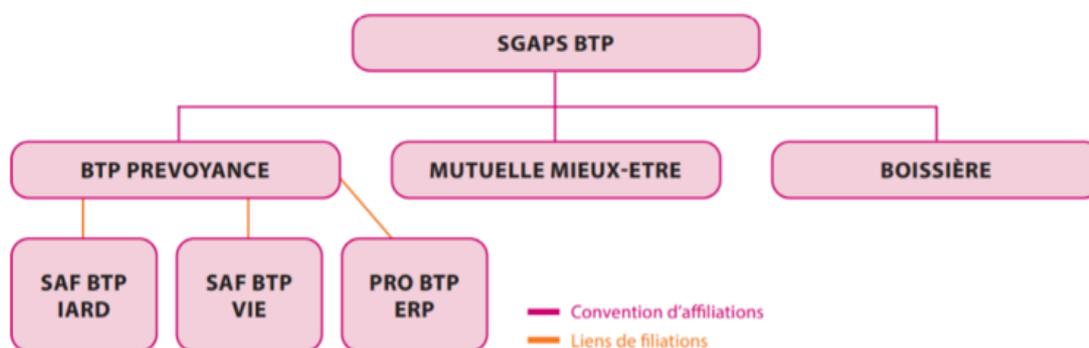
1) Présentation générale

La mutuelle Mieux-être, créée en 1927 sous le nom de SMBTP (Société Mutuelle du Bâtiment et des Travaux Publics), est un organisme à but non lucratif, régie par le livre 2 du Code de la mutualité, et agréée pour proposer des garanties accidents, maladie, vie, décès, natalité et nuptialité (branches 1, 2, 20 et 21).

Pour mieux répondre aux spécificités de chaque secteur la mutuelle a développé deux marques :

- Mutuelle Mieux-être : pour son activité de mutuelle interprofessionnelle, proposant des solutions en santé et en prévoyance collective et individuelle,
- SMBTP : proposant des solutions en santé et prévoyance aux professionnels du BTP.

Depuis fin 2016, la mutuelle Mieux-être est affiliée à la SGAPS BTP (société de groupe assurantiel de protection sociale) qui structure les liens financiers entre ses 3 affiliés (BTP Prévoyance, mutuelle Boissière, et mutuelle Mieux-être). Cette affiliation qui matérialise l'adossement de la mutuelle au groupe PRO BTP, permet à la mutuelle de renforcer son offre de service et sa capacité d'innovation.



▪ Figure n°1 : Structuration de la SGAPS BTP

Quelques chiffres et Indicateurs clés de la mutuelle en 2020 :

- Un Chiffre d'affaires de 132 M€,
- 234 990 bénéficiaires en 2020,
- Fonds propres solvabilité II de 44,6 M€,
- Capital de solvabilité requis (SCR) de 15,7 M€,
- Une implantation nationale avec 12 agences.

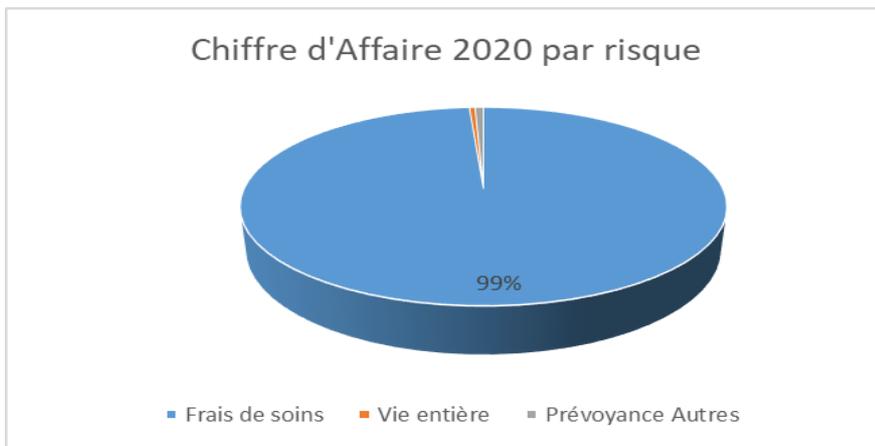
2) Présentation des risques assurés

Les risques assurés par la mutuelle sont décrits ci-dessous :

- Le risque frais de santé : cette garantie permet le remboursement complémentaire du frais de santé pris en charge par la sécurité sociale. Ces remboursements complémentaires concernent les frais chirurgicaux, les frais d'hospitalisation et les frais médicaux pour l'adhérent et sa famille,
- Le risque frais d'obsèques : permet le versement d'une allocation en cas de décès de l'assuré. Il s'agit de contrat de type vie entière à primes périodiques ou uniques. Le contrat est fermé à la commercialisation depuis 2012,
- Le risque collectivités territoriales : permet en cas d'interruption d'activité des agents des collectivités territoriales, en raison d'une incapacité physique par suite de maladie ou d'accident, le versement des prestations spécifiques en fonction du type de maladie concernée : maladie ordinaire, maladie de longue durée, Le portefeuille est en run off depuis le 31/12/2012,
- Le risque décès : L'assurance en cas de décès prévoit le versement des prestations en cas de décès de l'adhérent :
 - ✓ Un capital calculé en fonction de sa rémunération annuelle brute et de sa situation de famille,
 - ✓ Des rentes d'éducation aux enfants à charge,
 - ✓ Une rente de conjoint : une rente temporaire versée au conjoint survivant n'ayant pas droit à la pension de réversion du régime de retraite complémentaire auquel était affilié l'adhérent, et une rente viagère versée tant que le conjoint est en vie, sauf en cas de remariage,
- Le risque arrêt de travail : cette assurance permet en cas d'interruption d'activité, en raison d'une incapacité physique par suite de maladie ou d'accident, le versement de prestations :
 - ✓ Incapacité : des indemnités journalières complémentaires à celles de la sécurité sociale, qui interviennent au terme d'un délai de carence,
 - ✓ Invalidité : d'une rente d'invalidité qui s'ajoute à celle servie par la sécurité sociale.

3) Répartition du chiffre d'affaires par sous risque

L'activité principale de la mutuelle est la complémentaire santé, elle représente 99% du chiffre d'affaires total encaissé en 2020 .la prévoyance représente 1%, et est composée principalement de la garantie vie entière et capital décès.



▪ Figure n°2 : Répartition du chiffre d'affaires en 2020

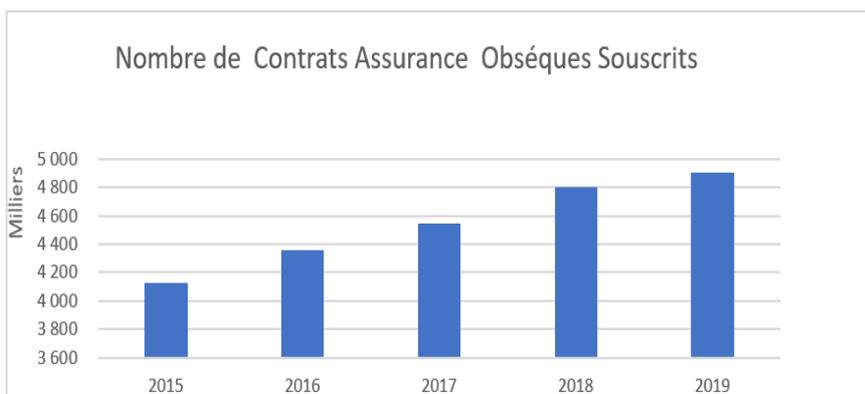
Dans le cadre de ses objectifs stratégiques, la mutuelle Mieux-être souhaite diversifier ses produits, et s'orienter davantage dans la prévoyance individuelle comme le produit vie entière.

C'est dans ce cadre que s'inscrit ce mémoire, dans le but de mieux connaître, et prendre en compte tous les facteurs de risque liés à ce produit.

De plus, Le nouveau régime solvabilité II a imposé aux assureurs l'obligation de tenir compte des options proposées par le produit, pour le calcul des provisions en vision économique, et de modéliser le risque de rachat dans le module de la souscription.

II. Quelques chiffres de l'assurance obsèques en France

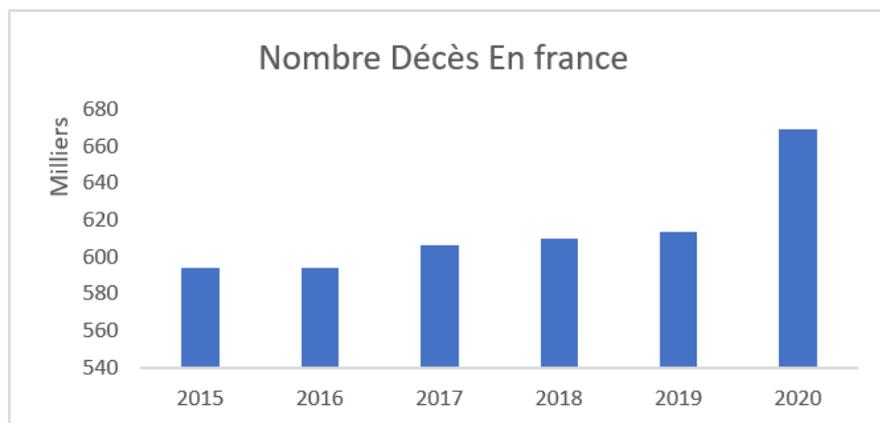
Le Marché de l'assurance obsèques est un secteur en pleine expansion, avec une augmentation du chiffre d'affaires en 2020 de 2,3% pour s'établir à près de 1,6 milliard d'euros¹. Le nombre de contrats souscrits est en croissance comme le montre le graphique suivant :



¹ Source : Voir l'annexe 1

▪ Figure n°3 : Evolution du nombre de contrats souscrits

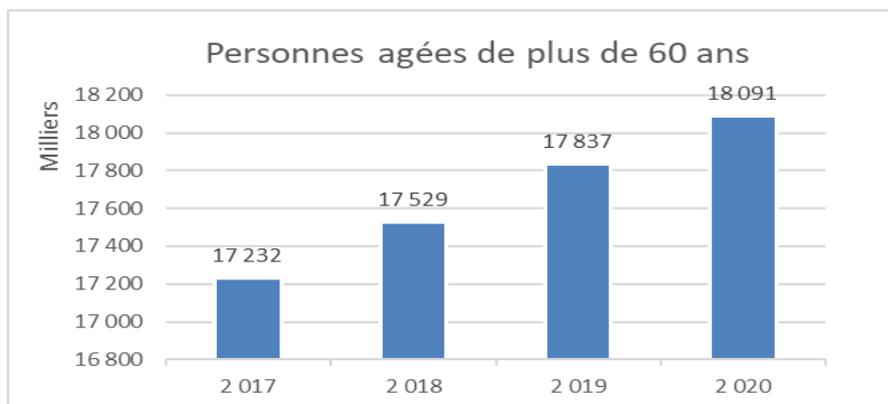
Dans le même temps, les prestations ont augmenté de +6,7% par rapport à 2019 pour s'établir à 685 Millions d'euros en 2020². Cette augmentation de sinistralité est liée au contexte marqué par la pandémie, qui a entraîné une explosion du nombre de décès en 2020 de 9,09%, contre 0,59% en 2019. Le graphique suivant illustre l'augmentation du décès sur ces dernières années.



▪ Figure n°4 : Evolution du nombre de décès entre 2015 et 2020

L'évolution du nombre de décès passe de 3,29% entre 2015 et 2019, à 9,08% entre 2019 et 2020. L'assurance obsèques est considérée comme un secteur très porteur, qui dispose d'un fort potentiel de croissance, et ce en raison des facteurs suivants :

- Comme conséquence du vieillissement de la population française, le nombre de seniors de plus de 60 ans, considéré comme la clientèle cible des assureurs ne cesse d'augmenter (1,42% en 2020). Le graphique suivant trace l'évolution des seniors entre 2017 et 2020,



² Source : Voir l'annexe 1

▪ Figure n°5 : Evolution des séniors de plus de 60 ans

- La crise sanitaire, au-delà de ses effets immédiats traduits par l'explosion du nombre de décès en 2020 aura favorisé la prise de conscience des personnes touchées par le décès d'un proche, de l'avantage de financer les obsèques à l'avance,
- La progression du coût des obsèques, incite logiquement à la souscription de telles assurances.
- Le prix des assurances obsèques reste relativement abordable au grand public.

Pour répondre aux attentes des séniors en quête de solutions simples et fluides, Les assureurs n'ont pas ménagé d'efforts pour renouveler leurs offres, renforcer les prestations d'accompagnement (tiers payant), déployer des garanties obsèques dans les packs santé et prévoyance, multiplier les offres promotionnelles, développer des outils en ligne (simulateur, souscription.), et avoir recours à la digitalisation des services et prestations.

III. Présentation des phénomènes de rachat et de réduction

1) Description du produit vie entière

Le produit vie entière est un contrat individuel qui vise à garantir le versement d'un capital au(x) bénéficiaire(s) désigné(s) par l'assuré quelle que soit la date du décès, en échange de quoi le souscripteur prend l'engagement révocable de verser des primes en fonction de la réalisation d'évènements viagers. Des options et garanties sont proposées pour rendre l'engagement de l'assuré allégé et le produit plus liquide. Parmi ces options on trouve le rachat et la réduction :

- Option de rachat : il s'agit du droit de l'assuré de résilier son contrat totalement ou partiellement, le montant restitué correspond au montant de la provision mathématique. Eventuellement des pénalités pour couvrir les frais encourus par l'assureur sont déduites.
- Option de réduction : elle permet à l'assuré de cesser le paiement de ces cotisations périodiques à tout moment, pendant la durée du contrat tout en restant assuré jusqu'à son décès, pour une valeur de réduction proportionnelle aux versements effectués depuis le début du contrat.

Ces options accordées aux assurés présentent un risque majeur pour l'assureur si elles sont négligées ou mal modélisées.

2) Les risques inhérents aux phénomènes de rachat et de réduction

Parmi les risques liés aux phénomènes de rachat et de réduction on peut citer :

- Risque de liquidité : dans le cadre de la gestion Actif Passif, l'assureur doit disposer des liquidités pour payer ses prestations futures. Dans le cas d'un rachat massif, l'assureur court le risque de ne pas disposer des sommes nécessaires pour honorer ces engagements. Pour faire face à cette situation, l'assureur peut investir dans des produits dont la maturité coïncide avec les sortis en rachat,
- Risque de Taux :
 - ✓ En cas d'une hausse des taux : la concurrence offre des taux garantis meilleurs, Engendrant un mouvement de rachat important, obligeant l'assureur à vendre ces actifs même à perte, ou à proposer un taux supérieur que la concurrence,
 - ✓ En cas d'une baisse des taux : l'assureur doit tenir ses engagements et verser aux assurés un taux garanti supérieur au rendement de ses actifs, le risque étant que les rachats sont beaucoup moins nombreux que prévu et que l'assureur manque de liquidité.

La garantie vie entière est une opération de prévoyance, orientée principalement à financer les frais d'obsèques en cas de décès, les sommes investies, ainsi que les capitaux garantis sont beaucoup plus faibles comparés à l'assurance classique d'épargne. et de ce fait, le risque de taux n'affecte pas tellement ce risque.

3) Facteurs de risque

Globalement, les facteurs de risque pouvant expliquer le comportement de l'assuré peuvent se résumer en deux grandes catégories, facteurs structurels qui dépendent de la conviction de l'assuré, et facteurs conjoncturels liés à la conjoncture économique.

Parmi les facteurs structurels on peut citer ceux liés au contrat, et ceux liés à l'assuré :

- Caractéristiques liées au contrat : type de contrat, participation aux bénéfices, ancienneté du contrat, capital garanti, primes versées, commissionnement, le réseau de distribution...
- Caractéristiques liées à l'assuré : l'âge de l'assuré, sexe, saisonnalité, catégorie socio professionnelle, statut marital, salaire, adresse...

Parmi les facteurs conjoncturels on cite :

- Tout ce qui est lié au contexte économique et financier : l'évolution du taux d'intérêt, taux de chômage, l'inflation, la croissance,

- Changement de la législation,
- L'écart du spread de taux avec la concurrence,
- La notation et l'image de l'assureur.

S'agissant d'une opération de prévoyance, le but du souscripteur étant de couvrir les frais d'obsèques en cas de décès de l'assuré, ce n'est nullement un produit d'épargne ou d'investissement, où l'assuré cherche à fructifier son épargne. De plus les montants en jeu, sont faibles comparés aux produits d'épargne, ce qui nous amène à supposer que les acteurs économiques et financiers n'ont pas d'influence ou d'impact sur les décisions de l'assuré en cas de rachat ou réduction.

Dans la suite du mémoire, seuls les facteurs structurels seront pris en compte dans la modélisation du comportement de l'assuré.

4) Focus sur Solvabilité II

Les provisions mathématiques en norme S1 sont calculées sans tenir comptes des options implicites prévues dans le contrat (rachat, réduction, participation aux bénéficiaires...), elles sont calculées sous l'hypothèse que l'assuré ne rachète pas son contrat, et qu'il continue de payer ses cotisations jusqu'à l'échéance du contrat ou à son décès.

En norme S2, l'évaluation des engagements de l'assureur se fait au prix du marché c'est à dire en Best Estimate, elle devra refléter tous les coûts potentiels liés au contrat. Il s'agit de prendre en compte les garanties financières et les différentes options qui permettent aux assurés de modifier les conditions de leurs contrats.

L'exigence de capital pour le risque de rachat « cessation », telle que définie dans la directive solvabilité II est égale à la plus élevée des exigences de capital suivantes :

- L'exigence de capital pour risque de hausse permanente des taux de cessation,
- L'exigence de capital pour risque de baisse permanente des taux de cessation,
- L'exigence de capital pour risque de cessation de masse.

Pour la modélisation des rachats structurels, les assureurs doivent utiliser des données d'expérience si celles-ci sont conformes aux observations passées, ou à défaut des données de marché.

Le choix des paramètres pris en compte dans la modélisation, est laissé libre à l'assureur, qui devra néanmoins justifier ses choix.

IV. Description technique du produit vie entière

1) Tarification

Nous supposons dans cette partie que les cotisations périodiques sont viagères.

Calcul de la prime unique pure

La prime unique pure CUP_x est déterminée sur la base (A 212-10 ; A 212-12) :

- Des tables de mortalité TF 00-02 et TH 00-02
- D'un taux d'intérêt technique égal au moins élevé des deux taux suivants : 3,5% du capital ou 60% du dernier taux moyen des emprunts d'Etat français calculé sur une base semestrielle.

CUP_x représente l'engagement de l'assureur d'âge x , c'est à dire le versement de l'allocation obsèques en cas de décès du bénéficiaire. Il dépend de l'âge du bénéficiaire à la souscription et du montant du capital garanti C .

Pour un bénéficiaire d'âge x , l'engagement de l'assureur est donc : $CUP_x = C \times A_x$;

$$\text{Où } A_x = \sum_{k=0}^{\omega-x} v^{k+1/2} \times {}_k p_x \times q_{x+k} \times (1+i)^k \quad ;$$

Correspond à la prime demandée pour une couverture viagère d'un capital décès de 1 € avec :

x : l'âge de l'assuré à la souscription du contrat calculé par différence de millésimes,

C : le Capital garanti,

i : le Taux technique,

$$v = \frac{1}{1+i} \text{ le taux d'escompte,}$$

${}_k p_x$: la probabilité pour un individu d'âge x de survivre k années,

$${}_k q_x = 1 - {}_k p_x$$

ω : l'âge limite lu dans la table de mortalité.

Calcul de la prime périodique pure

La prime périodique pure CPP_x est déterminée sur la base (A 212-10 ; A 212-12) :

- Des tables de mortalité TF 00-02 et TH 00-02,

- D'un taux d'intérêt technique égal au moins élevé des deux taux suivants : 3,5% du capital ou 60% du dernier taux moyen des emprunts d'Etat français calculé sur une base semestrielle.
- D'une périodicité m : Cela consiste à niveler la Cotisation unique pure :

$$CPP_x = \frac{C \times A_x}{\ddot{a}_x^{(m)}} ; \quad \text{Où} \quad \ddot{a}_x^{(m)} = \sum_{k=0}^{n-x} v^{k+1/2} \times {}_k p_x \times (1+i)^k - \frac{m-1}{2m} ;$$

m correspond à la périodicité des cotisations, et prend les valeurs :

m = 1 si annuel, 4 si trimestriel, et 12 si mensuel

Calcul de la prime d'inventaire

La prime d'inventaire est égale à la prime pure augmentée des chargements de gestion. Les chargements de gestion sont regroupés en deux catégories :

- Les chargements de gestion du contrat. On les note g_1 , ils sont destinés à couvrir les frais liés à la gestion du contrat, tels que la relation client et les relevés annuels de situation,
- Les chargements de gestion liés aux primes. On les note g_2 , c'est un pourcentage du montant du capital C. Ces frais sont prélevés pendant toute la durée de paiements des primes. Ils sont destinés à couvrir les frais liés à la gestion des paiements tels que les encaissements des primes et leurs comptabilisations.

Si l'on note respectivement CUI_x et CPI_x la cotisation unique d'inventaire et la cotisation périodique d'inventaire, les formules sont données par :

$$CUI_x = (A_x + g_1 \times \ddot{a}_x) \times C ; \quad CPI_x = \frac{A_x + g_1 \times \ddot{a}_x + g_2 \times \ddot{a}_x^{(m)}}{\ddot{a}_x^{(m)}} \times C ;$$

Calcul de la prime commerciale

La prime commerciale est égale à la prime d'inventaire augmentée des chargements d'acquisition. Les chargements d'acquisition sont destinés à couvrir les commissions et les frais de commercialisation.

Si l'on note respectivement CUC_x et CPC_x la cotisation unique commerciale et la cotisation périodique commerciale, avec α chargement d'acquisition, les formules sont données par :

$$CUC_x = \frac{A_x + g_1 \times \ddot{a}_x}{(1-\alpha)} \times C ; \quad CPC_x = \frac{A_x + g_1 \times \ddot{a}_x + g_2 \times \ddot{a}_x^{(m)}}{\ddot{a}_x^{(m)} \times (1-\alpha)} \times C ;$$

2) Provisionnement

La réglementation impose le calcul d'une provision mathématique, définie à l'article R212-26-1 du Code de la Mutualité comme la : « différence entre les valeurs actuelles des engagements pris par la mutuelle ou l'union, d'une part, et par les membres participants ou souscripteurs de contrats collectifs, d'autres part ».

Les normes actuarielles utilisées pour ce calcul sont les mêmes que celles utilisées en tarification.

Provision mathématique pour les cotisations uniques

$$PM_x^k = CUP_{x+k} ; \text{ avec } k \text{ la durée écoulée en années}$$

Provision mathématique pour les cotisations périodiques

$$PM_x^k = CUP_{x+k} - CPP_x \times \ddot{a}_{x+k}^{(m)} ; \text{ avec } k \text{ la durée écoulée en années}$$

Provision mathématique interpolée

Le membre participant pouvant souscrire son contrat en cours d'année, et non forcément au premier janvier, les provisions mathématiques sont interpolées en fonction de la date de souscription du membre participant et de la date de calcul de la provision mathématique (ici 31/12/Y).

Soit $m \in [1, 12]$ le mois de souscription du contrat et Y son année. La provision mathématique au 31/12/Z (avec $Z \geq Y$) pour un individu d'âge à la souscription est ainsi calculée :

$$PM = \frac{(13-m)}{12} \times PM_x^{k+1} + \frac{m-1}{12} \times PM_x^k ;$$

3) Valeur de rachat

Mode de calcul (R 212-27) : Pour tout bulletin d'adhésion ou contrat collectif relatif à une opération d'assurance sur la vie comportant une valeur de rachat et pour les opérations de capitalisation, la valeur de rachat est égale à la provision mathématique du bulletin d'adhésion ou du contrat collectif diminuée, éventuellement, d'une indemnité qui ne peut dépasser 5 % de cette provision mathématique. Cette indemnité doit être nulle à l'issue d'une période de dix ans à compter de la date d'effet du bulletin d'adhésion ou du contrat. La valeur de rachat est donc :

Si le rachat intervient dans les 10 premières années de l'adhésion :

$$RACH_x^k = PM_x^k - 0,05 PM_x^k ;$$

Si le rachat intervient après les 10 premières années de l'adhésion :

$$RACH_x^k = PM_x^k$$

4) Valeur de réduction

Mode de calcul : Les modalités proprement dites de calcul de la valeur de réduction sont déterminées par le règlement général mentionné dans le contrat d'assurance vie. Elles consistent à considérer la provision mathématique du contrat comme la prime unique d'une nouvelle assurance, qui est nécessairement inférieure à celle initialement prévue. Il ne s'agit toutefois pas de la conclusion d'un contrat nouveau. Dès lors, le nouveau capital assuré sera déterminé à partir non pas de l'âge de l'assuré au jour de la réduction, mais à partir du tarif d'inventaire en vigueur au jour de la souscription initiale.

La réduction peut ainsi être définie comme une révision du contrat d'origine qui, sous réserve de la diminution des prestations assurées, continue avec ses autres effets.

La valeur de réduction est donc :

$$RED_x^k = \frac{PM_x^k}{CUP_x}$$

V. Méthodologie actuelle de calcul

L'objet de cette partie est de détailler le calcul actuel des lois de rachat et réduction empiriques sur les produits vie entière au sein de la direction technique de la mutuelle Mieux-être.

Ces lois sont utilisées pour la projection des flux de passif pour le calcul des Best Estimate. Les données utilisées pour l'estimations des lois, correspondent au nombre de rachats et réductions survenus entre le 01/01/2010 et le 31/12/2016, soit sur une période d'observation de 7 années.

La méthodologie est la même pour l'estimation des lois de rachat et réduction, elle est basée sur l'estimation des taux de chute en fonction de l'ancienneté du contrat, sans tenir compte des autres caractéristiques du contrat ou de l'assuré (âge, capital, réseau, saisonnalité, ...).

1) Notations

Nous avons les notations suivantes :

T : est la période d'observation allant du 01/01/2010 au 31/12/2016,

T_i : la période d'observation annuelle avec $i=1, 2, \dots, 7$,

k : ancienneté du contrat en année,

$Exp(i, k)$: nombre d'individus présents en début de périodes susceptibles d'être rachetés avec une ancienneté k ,

$Nrach(i, k)$: nombre de rachats survenus en période i avec une ancienneté k ,

$Nred(i, k)$: nombre de réductions survenues en période i avec une ancienneté k ,

$\alpha(i, k)$: la pondération de l'effectif de l'année T_i d'ancienneté k sur l'effectif total de l'ensemble des années.

2) Formules de Calcul

Dans un premier temps, nous calculons un taux de rachat pour la période d'observation annuelle T_i , et avec l'ancienneté k , ensuite nous calculons un taux de rachat pour la période T et l'ancienneté k par pondération des taux de rachat annuels.

Taux de rachat Annuel

La probabilité à chercher est la suivante : probabilité [contrat d'ancienneté $k-1$ en début de période T chute totalement en k], correspondant au rapport du nombre de rachats survenus pendant la période T_i avec ancienneté k sur le nombre de personnes exposées en début de période T_i avec ancienneté $k-1$.

Autrement dit,

$$Trach(i, k) = \frac{Nrach(i, k)}{Exp(i, k)} \quad ;$$

Taux de rachat sur la période T

Le taux de rachat global sur la période T d'un contrat avec ancienneté K est donné par la formule suivante :

$$Trach(T, k) = \sum_{i=1}^T \alpha(i, k) \times Trach(i, k) \quad ;$$

avec la pondération :

$$\alpha(i, k) = \frac{Exp(i, k)}{\sum_{i=1}^T Exp(i, k)} \quad ;$$

On vérifie bien que pour une ancienneté donnée k : $\sum_{i=1}^T \alpha(i, k) = 1$;

Taux de réduction Annuel

La formule de calcul du taux de réduction pour une ancienneté k est le même que celui du rachat, en remplaçant $Nrach(i, k)$ par $Nred(i, k)$ La formule devient :

Taux de réduction sur la période T

$$Tred(i, k) = \frac{Nred(i, k)}{Exp(i, k)}$$

Le Taux de réduction total sur la période T pour une ancienneté k est donné par la formule suivante :

$$Tred(T, k) = \sum_{i=1}^T \alpha(i, k) \times Tred(i, k);$$

3) Résultats

Les taux de rachat et réduction sont communiqués par ancienneté. L'ancienneté maximale des lois est 30 ans. Le nombre des contrats rachetés et réduits au-delà de 18 ans étant trop faible, nous avons retenu le dernier taux calculé non nul pour les anciennetés supérieures.

<i>Ancienneté</i>	<i>Taux Rachat</i>	<i>Taux réduction</i>
1	6,56%	4,39%
2	10,07%	4,66%
3	12,28%	2,86%
4	9,32%	2,03%
5	8,41%	1,83%
6	8,46%	1,94%
7	5,77%	1,62%
8	5,12%	1,04%
9	3,95%	0,94%
10	3,10%	1,06%
11	8,47%	0,72%
12	14,11%	0,94%
13	14,55%	0,83%
14	0,92%	0,35%
15	1,11%	0,79%
16	0,56%	0,38%
17	1,59%	1,06%
18	2,42%	0,61%
19	2,42%	0,61%
20	2,42%	0,61%
21	2,42%	0,61%
22	2,42%	0,61%
23	2,42%	0,61%
24	2,42%	0,61%
25	2,42%	0,61%
26	2,42%	0,61%
27	2,42%	0,61%
28	2,42%	0,61%
29	2,42%	0,61%
30	2,42%	0,61%

- [Tableau n°1 : Taux de rachat et réduction par ancienneté](#)

Le taux de rachat est volatile et varie en fonction de l'ancienneté k . Pendant les premières années du contrat $[1,4]$ le taux de rachat augmente, puis il baisse pour $k \in [5,10]$. Pour $k \in [11,13]$ le taux augmente très fortement, pour baisser et se stabiliser à partir de 18 ans d'ancienneté.

Le taux de réduction augmente pour les deux premières années de contrat, puis il baisse pour se stabiliser à partir de 18 ans d'ancienneté.

4) Backtesting

A partir des taux estimés précédemment, nous calculons le nombre de rachats et réductions, puis nous les comparant avec le nombre réel constaté lors des exercices 2016 et 2017.

<i>Exercices 2016-2017</i>	<i>Rachat</i>	<i>Réduction</i>
<i>Ecart En %</i>	<i>8,19%</i>	<i>24,12%</i>

Des écarts significatifs sont constatés entre le nombre réel et l'estimé.

L'estimation des lois de rachat et de réduction basée uniquement sur l'ancienneté du contrat ne s'avère donc pas comme un choix suffisant pour expliquer le comportement des assurés.

L'objectif principal de ce mémoire est d'étudier le comportement de rachat et de réduction, de sélectionner les variables susceptibles de les expliquer, et de proposer un modèle plus pertinent qui quantifie les probabilités de survenance de ces comportements en fonction de ces variables.

Conclusion

La mutuelle Mieux-être couvre principalement les garanties « frais de soins ». Aujourd'hui l'un des axes de sa stratégie consiste à se développer en prévoyance, notamment la prévoyance individuelle.

Le vieillissement de la population française, l'accroissement du nombre de décès annuel, avec l'effet de la crise sanitaire sur la prise de conscience des gens sur l'importance de financer l'obsèques à l'avance, sont autant de facteurs qui ont rendu le marché de l'assurance obsèques dynamique et disposant d'un fort potentiel.

La souscription d'un contrat vie entière par un assuré, engage l'assureur de façon irrévocable à régler les prestations en cas du décès, ou en cas de rachat. La cessation de paiement des cotisations engendre une baisse du chiffre d'affaires pour l'assureur. Les phénomènes de rachat et de réduction s'ils sont ignorés peuvent présenter un risque majeur pour la solvabilité de l'assureur, ils

sont liés au comportement de l'assuré, à ces caractéristiques personnelles, mais aussi aux caractéristiques du contrat.

Les facteurs économiques liés au marché financier, sont écartés de l'étude car les montants des capitaux ainsi que les cotisations sont faibles comparés à l'assurance de l'épargne, et de ce fait, ils sont moins sensibles aux fluctuations des taux.

La réglementation Solvabilité II, impose aux assureurs d'intégrer les options implicites prévues dans les contrats pour le calcul des engagements, et le capital réglementaire requis SCR.

Actuellement, au sein de la direction d'actuariat, les lois de rachat et la réduction sont déterminées empiriquement en fonction de l'ancienneté du contrat, et sont utilisées pour la projection du passif pour le calcul des Best Estimate. Les taux sont mis à jour d'une année sur l'autre. Le Backtesting appliqué sur les exercices 2016 et 2017 nous montre des écarts entre les taux estimés et les taux réels, et laissent penser que d'autres variables outre que l'ancienneté du contrat peuvent influencer le comportement de l'assuré.

Base de données : Traitement-Analyse descriptive

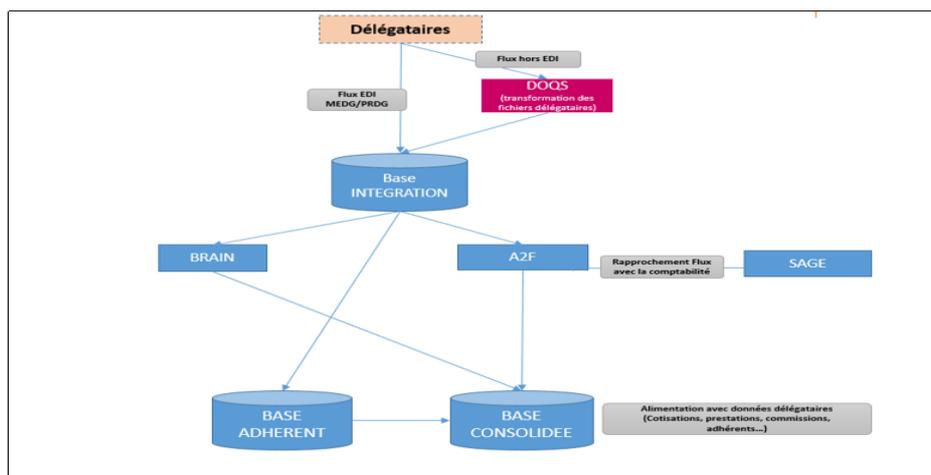
Dans cette partie nous allons décrire les points suivants :

- Problématique de la délégation de gestion,
- Choix de la période d'observation de l'étude,
- Présentation du portefeuille sujet de l'étude,
- Extraction et traitement des données,
- Analyse de corrélations entre les variables sélectionnées,
- Liste des variables sélectionnées pour la modélisation.

I. Délégation de Gestion

La mutuelle est en gestion totalement déléguée, l'ensemble de ces contrats sont gérés par des délégataires de gestion, qui transmettent l'ensemble des données, soit en flux automatiques injectés directement dans l'infocentre de gestion de la mutuelle, soit en flux manuels contrôlés et validés avant d'être intégrés dans le système de gestion interne de la mutuelle.

Le schéma fonctionnel d'intégration des données délégataires jusqu'à l'alimentation de la base « Adhérents » est représenté ci-après :



▪ Figure n°6 : Schéma d'intégration des données délégataires

Cette particularité expose la mutuelle à un risque sur la qualité de la donnée : Insuffisance des données externes/absence.

Les Données externes sont soumises à des conditions strictes imposées par la réglementation :

- Elles doivent être plus appropriées que les données internes et cohérentes avec ces dernières,

- Les mécanismes de retraitement de la donnée par le délégataire doivent être compris par l'assureur,
- Capacité de l'assureur à détecter des évolutions des méthodes et hypothèses de production de ces données par le délégataire de gestion,
- L'assureur doit avoir la possibilité de mener des audits de la qualité des données auprès du délégataire de gestion.

Consciente de l'importance de la problématique de la qualité des données, la mutuelle Mieux-être a entamé un projet de la qualité des données décliné en cinq chantiers :

- Chantier 1 : traitement de la donnée chez le délégataire,
- Chantier 2 : échanges de données entre l'environnement délégataire et l'environnement de la mutuelle,
- Chantier 3 : intégration des données dans le système d'information de la mutuelle,
- Chantier 4 : traitement et exploitation des données dans le système d'information de la Mutuelle,
- Chantier 5 : livrables Solvabilité 2 – traçabilité de la donnée.

Pour chacun de ces chantiers, est établie une cartographie des flux, un dictionnaire des données, et un fichier de contrôles. Ces travaux ont pour objectif de fiabiliser et assurer la traçabilité, l'exactitude, et l'exhaustivité de la donnée depuis sa source jusqu'à son utilisation finale (Bilan S1, Bilan prudentiel et SCR, production QRT).

II. Choix de la période d'observation

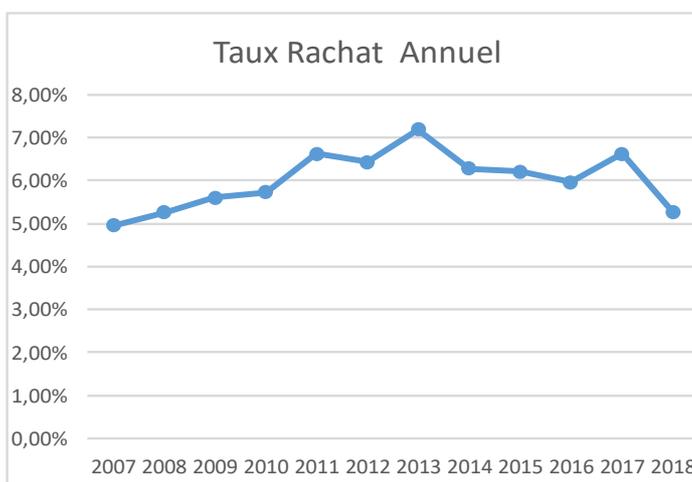
Il s'agit de l'intervalle du temps retenu pour l'étude. A la suite des recommandations de l'institut des actuaires « Lignes directrices Mortalité », il faut prendre en compte certains nombres de précautions dans le choix de la période d'observation :

- Prise en compte des sinistres tardives : pour tenir compte des déclarations tardives des sinistres, et ne pas sous-estimer le risque étudié, il faut veiller à ce qu'un certain décalage ait été conservé entre la date de fin d'observation et la date d'extraction des données. Une analyse sur l'écart entre la date de survenance et la date de déclaration de sinistre est recommandé afin de valider le décalage utilisé,
- La durée de l'observation doit être un multiple de douze mois,
- La longueur de la période d'observation : il est préférable que la période d'observation comporte plusieurs années, cela permet d'augmenter le volume d'observations, ce qui est souvent utile pour améliorer la qualité des estimations. Il faut toutefois éviter de tomber dans l'excès et accepter une période d'étude trop longue, sans quelques précautions préalables.

En effet, pour une période choisie au-delà de cinq ans, il conviendra de vérifier qu'il n'y avait pas d'évolution sensible du risque étudié sur la période retenue.

Les taux de rachat sur la période du 01/01/2007 au 31/12/2018 sont donnés dans le tableau suivant :

<i>Année rachat</i>	<i>Taux rachat Annuel</i>
2007	4,96%
2008	5,26%
2009	5,61%
2010	5,73%
2011	6,63%
2012	6,44%
2013	7,19%
2014	6,29%
2015	6,22%
2016	5,96%
2017	6,63%
2018	5,28%

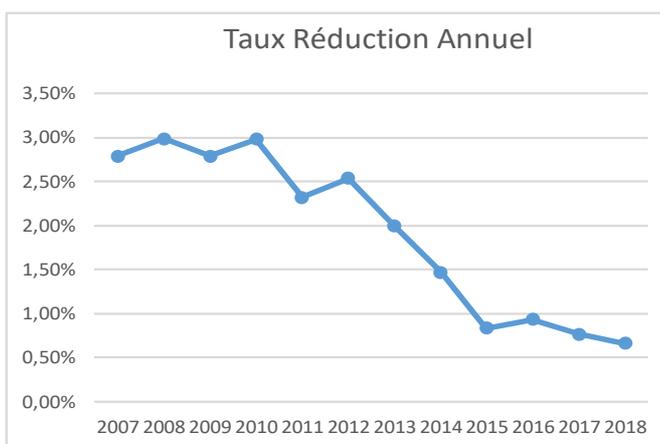


▪ Figure n°7 : Taux de rachat annuel sur la période de 2007 à 2018

Les taux de rachat annuels entre 2007 et 2018 sont stables et varient entre 4,96% pour atteindre un taux maximal de 7,17%.

De la même façon nous présentons les taux de réduction sur la même période de 2007 à 2018 :

<i>Année réduction</i>	<i>Taux réduction Annuel</i>
2007	2,79%
2008	2,99%
2009	2,79%
2010	2,98%
2011	2,32%
2012	2,53%
2013	1,99%
2014	1,47%
2015	0,83%
2016	0,93%
2017	0,77%
2018	0,66%



▪ Figure n°8 : Taux de réduction annuel sur la période de 2007 à 2018

Les taux de réduction sont quasi stables entre 2007 et 2013. Ils baissent considérablement à partir de 2014.

Pour prendre en considération les recommandations de l'institut des actuaires, nous avons choisi pour la modélisation du rachat la période du 01/01/2007 au 31/12/2018, et pour la modélisation de la réduction la période du 01/01/2007 au 31/12/2013, périodes dont les taux de rachat et réduction sont supposés stables.

III. Présentation du portefeuille étudié

Dans cette section nous allons présenter les données utilisées ainsi que les produits sélectionnés pour effectuer les modélisations. Les montants de rachat et réduction sont ceux survenus durant la période d'observation du 01/01/2007 au 31/12/2018 pour les rachats, et la période du 01/01/2007 au 31/12/2013 pour les réductions.

Le portefeuille étudié est constitué de la liste des adhérents ayant souscrit un contrat vie entière, de la liste des rachats, réductions et décès. Ces adhésions peuvent être classées selon leur mode de commercialisation (Direct, Courté ou en Acceptation) :

- Direct : la Mutuelle a commercialisé ses premiers contrats vie entière via son délégataire historique entre 1989 et 2012, il s'agit des contrats dont le capital garanti se situe entre 2000 € et 20 000 €, les cotisations sont soit à prime unique ou périodique, l'âge à la souscription est entre 19 ans et 85 ans.
Dans la suite de l'étude, ce produit sera nommé « produit 1 », seuls les frais de gestion sont supportés par l'assureur.
Ce produit a été fermé à la souscription en 2012.

- Courté : à partir de 2005, la mutuelle a commercialisé via un délégataire courtier deux gammes de produit vie entière, dont le capital se situe entre 1500 € et 10 000 €, le premier est commercialisé depuis 2005 jusqu'à 2009, et le deuxième est commercialisé depuis 2009 jusqu'à 2012. Les cotisations sont périodiques, et l'âge à la souscription est entre 18 ans et 85 ans.
Dans la suite de l'étude, ce produit sera nommé « produit 2 ». Sur ce produit l'assureur supporte les frais de gestion, les frais d'apport dont le taux moyen s'élève à 35%.
Ce produit a été fermé à la souscription en 2012.

- Acceptation : la mutuelle est réassureur sur ce portefeuille, le capital garanti est le même pour l'ensemble des adhérents 3400 €, les cotisations sont soit à prime unique ou périodique. Ce produit est commercialisé entre 2001 et 2004.
Dans la suite de l'étude, ce produit sera nommé « produit 3 ».
Ce produit a été fermé à la souscription en 2004.

Ci-dessous le tableau récapitulatif des contrats, capitaux sous risque, nombre de rachats, nombre de réductions par produit :

<i>Produits</i>	<i>Nombre Contrats</i>	<i>Capitaux sous risque</i>	<i>Nombre Rachats</i>	<i>Nombre Réductions</i>
<i>Produit1</i>	1 190	5 164 241	209	75
<i>Produit2</i>	14 163	47 048 000	6 635	2 465
<i>Produit3</i>	382	1 899 211	65	10
<i>Total</i>	15 735	54 111 452	6 909	2 550

▪ [Tableau n°2 : Détails par produit](#)

Le « produit 2 » centralise plus de 80% de l'ensemble des contrats vie entière, 96% de rachats, et 95% de réductions. Il s'agit du portefeuille le plus important, auquel il faut prêter une attention particulière lors de la phase d'extraction et traitement de données.

IV. Extraction Et Traitement des Données

La phase extraction de données et leurs traitements représente une part considérable du temps consacré à l'étude, et ce pour s'assurer d'obtenir des résultats fiables et cohérents.

1) Extraction des Données

Cette phase consiste à extraire les données brutes, à extraire les variables relatives aux assurés, aux sinistres, et aux types de produits. Le cadrage des flux techniques et comptables s'avère difficile à valider, en raison de la non-disponibilité de ce niveau de détail dans les flux comptables.

Les données à extraire sont :

- La liste des adhérents,
- La liste des rachats survenus,
- La liste des réductions survenues,
- La liste des autres mouvements (décès, sans effet, annulation...).

Ces données peuvent provenir de plusieurs sources :

- La base « Adhérents » : cette base comporte la liste des adhérents vue à une date donnée avec l'ensemble des informations sur l'assuré et le contrat. Cette base est alimentée à partir des fichiers transmis par les délégataires, soit en flux automatiques, soit en flux Excel. Cette base est alimentée dans SAS depuis le 31/12/2019. Les informations sur l'état du Contrat, et la date de résiliation n'étant pas disponibles antérieurement à cette date, sauf pour les contrats pour lesquels les délégataires envoient régulièrement l'historique des données depuis l'origine. Les données concernant le « produit 2 » sont exportées depuis cette base.

✚ 90% des données nécessaires à l'étude sont renseignées, et sont disponibles

- Données en Acceptations : elles représentent moins de 3% des données Vie entière. Ces données ne figurent pas dans la base « Adhérents ». Cette liste est transmise par l'assureur avec tous les mouvements survenus depuis l'origine. Les données concernant le « produit 3 » sont renseignées à partir de cette liste.
- Fichiers Excel issus des données d'inventaires : ils représentent moins de 7% des observations, il s'agit du portefeuille le plus ancien, avec des dates d'effet allant de 1989 à 2012. C'est la partie la plus compliquée, car ce portefeuille a été géré dans le passé par un autre délégataire avec qui la mutuelle a arrêté de travailler depuis 2017. Ce même délégataire avait changé son système informatique en 2009, ce qui avait induit une perte importante d'informations notamment sur les dates de rachat et réduction. La seule alternative possible pour récupérer l'historique, consistait à exploiter manuellement les différents fichiers d'inventaires depuis 2004. C'est sur cette partie que le traitement des données a été lourd et conséquent. Les données concernant le « produit 2 » sont récupérées à partir des fichiers d'inventaires.

✚ Bien que ces données ne représentent que 7% de l'ensemble des observations, leur intégration dans l'étude est très utile, il s'agit d'un portefeuille très ancien, le seul qui permettrait d'avoir des informations importantes sur les lois de rachat et réduction pour des anciennetés au-delà de 15 ans.

2) Liste des données disponibles dans la base

Nous dressons la liste des variables extraites des différentes bases :

Informations relatives à l'assuré :

- Identifiant,
- Date de naissance,
- Sexe,
- Code postal.

Informations relatives au Contrat :

- Numéro de contrat,
- Nom de produit,
- Nom du délégataire de gestion,
- Date d'adhésion,
- Etat contrat : en cours, rachat, décès, réduction, annulation, sans effet ...
- Date Mouvement,

- Capital garanti,
- Périodicité des cotisations.

3) Traitement et fiabilisation des données

La phase de traitement est essentielle pour la suite de l'étude, les bases de données présentent souvent des anomalies qu'il convient de détecter et de les corriger. Ces anomalies se manifestent par des données erronées, incomplètes, ou obsolètes.

Nous allons détailler dans la suite, pour chaque variable extraite les traitements effectués et les corrections apportées. A défaut, les lignes non corrigées seront supprimées de la base pour ne pas biaiser le résultat.

a) Traitement des variables

- « Sexe » : elle n'est pas renseignée pour les « produits 2 » et « produit 3 ». Elle est complétée à partir du numéro de sécurité sociale, ou le prénom de l'assuré,
- « Date de naissance » : Cette variable est souvent absente pour le « produit 2 », soit 5% d'observations. Seule l'année de naissance est renseignée. Cette variable sera donc remplacée par l'année de naissance. Elle est contrôlée avec les conditions d'âge à la souscription,
- « Date d'adhésion » : cette donnée est absente pour le « produit 2 », soit 3% d'observations. Elle sera remplacée par l'année d'effet,
- « Date mouvement » : correspond à la date de la survenance d'un évènement (décès -rachat-réduction-...), cette variable est contrôlée avec la variable « Etat contrat ». Elle est vide lorsque le motif est « actif ». la modalité « actif » représente un assuré n'ayant subi aucun évènement durant la période d'observation,
- « Etat contrat » : correspond à un évènement survenu pendant la période d'observation, il peut s'agir d'un décès, rachat, réduction, défaut de paiement, contentieux... La modalité « en cours » fait référence à un contrat n'ayant pas subi d'évènements. Cette variable est contrôlée avec la variable « date mouvement »,
- « Périodicité des cotisations » : cette variable prend deux modalités, soit périodique « P », ou Unique « U »,
- « Capital garanti » : Cette variable est contrôlée avec la liste des capitaux proposés pour chaque produit,
- « Code postal » : cette donnée n'est pas disponible pour les « produits 2 » et « produit 3 ». soit 10% des observations. Nous ne la retenons pas pour la suite de l'étude.

b) Création de nouvelles variables

Dans cette partie nous allons créer des nouvelles variables dites explicatives à partir des variables déjà extraites :

- « Age_ souscription » : Année d'effet-Année de naissance,
- « Age_rachat » : Année demande rachat-Année de naissance,
- « Age_réduction » : Année demande réduction-Année de naissance,
- « Ancienneté » : Date de point-Année d'effet,
La date de point est la date de fin de la période d'observation pour un contrat Actif soit le 31/12/2018, ou la date de la survenance de l'évènement (rachat, réduction) pour un contrat non Actif,
- « Adhésion_Couple » : prend deux modalités {1,2} ,1 si une souscription unique d'un membre de famille ,2 si plusieurs souscriptions de la même famille. Cette variable est déterminée, soit à partir du numéro de contrat, soit à partir du nom de famille pour une date d'effet et nom de produit identique,
- « Réseau » : prend trois modalités, courté, direct, et acceptations,
- « Saisonnalité » : correspond au mois de la survenance de l'évènement. 12 modalités proposées de 1 à 12,
- « Montant_Cotisations » : correspond à la cotisation annuelle théorique y compris chargements,
- « Montant_PM » : correspond au montant de la provision mathématique calculée à la date de point. Il s'agit de la provision mathématique pure calculée comme la différence entre l'engagement de l'assureur et l'engagement de l'assuré. L'engagement de l'assuré est calculé à partir de la cotisation théorique.

c) Classification des variables qualitatives

Les différentes valeurs des variables qualitatives/Binaires ont été classifiées dans les tableaux suivants :

Classification de la variable « Adhésion_Couple »

<i>Modalité</i>	<i>Nombre Contrats Souscrits par Couple</i>
<i>Couple_1</i>	<i>Non</i>
<i>Couple_2</i>	<i>Oui</i>

Classification de la variable « Périodicité des cotisations »

<i>Modalité</i>	<i>Périodicité des cotisations</i>
<i>Periodicite_1</i>	<i>P (périodique)</i>
<i>Periodicite_2</i>	<i>U (unique)</i>

Classification de la variable « Réseau »

<i>Modalité</i>	<i>Réseau de distribution</i>
<i>Reseau_1</i>	<i>Courté</i>
<i>Reseau_2</i>	<i>Direct</i>
<i>Reseau_3</i>	<i>Acceptations</i>

Classification de la variable « Sexe »

<i>Modalité</i>	<i>Sexe de l'assuré</i>
<i>Sexe_1</i>	<i>M (masculin)</i>
<i>Sexe_2</i>	<i>F (féminin)</i>

d) Liste finale des variables sélectionnées pour l'étude

Le tableau suivant dresse la liste des variables sélectionnées, leurs notations, et leurs Types :

<i>nom de la variable</i>	<i>Notation</i>	<i>Type</i>
<i>Identifiant individu</i>	<i>Identifiant</i>	<i>Qualitative</i>
<i>Année de naissance</i>	<i>Année_Naiss</i>	<i>Discrete</i>
<i>Sexe</i>	<i>Sexe</i>	<i>Binaire</i>
<i>Année d'adhésion</i>	<i>Année_Adh</i>	<i>Discrete</i>
<i>Date mouvement</i>	<i>Année_Mvt</i>	<i>Discrete</i>
<i>Etat contrat</i>	<i>Etat_Contrat</i>	<i>qualitative</i>
<i>Périodicité des cotisations</i>	<i>Period</i>	<i>Binaire</i>
<i>Capital garanti</i>	<i>Capital</i>	<i>Discrete</i>
<i>Age_souscription</i>	<i>Age_Init</i>	<i>Discrete</i>
<i>Age_rachat</i>	<i>Age_Rachat</i>	<i>Discrete</i>
<i>Age_réduction</i>	<i>Age_réduction</i>	<i>Discrete</i>
<i>Ancienneté</i>	<i>Anc</i>	<i>Discrete</i>
<i>Adhésion-Couple</i>	<i>Adhesion_C</i>	<i>binaire</i>
<i>Réseau</i>	<i>Reseau</i>	<i>Qualitative</i>
<i>Saisonnalité</i>	<i>Mois_Rad</i>	<i>Discrete</i>
<i>Montant_Cotisations</i>	<i>Cotis</i>	<i>Continue</i>
<i>Montant_PM</i>	<i>PM</i>	<i>Continue</i>

▪ Tableau n°3 : Liste des variables présélectionnées

Dans la partie qui suit, nous allons étudier la liaison entre les différentes variables sélectionnées, et supprimer celles qui sont corrélées entre elles.

V. Etude de liaison entre les variables explicatives

Une première étape pour la sélection des variables discriminantes pour les phénomènes rachat et réduction, consiste à éliminer les variables potentiellement corrélées entre elles. La liaison se mesure par un coefficient statistique selon le type de la variable à analyser.

1) Variable quantitative contre variable quantitative

Le coefficient de Pearson permet d'évaluer la relation linéaire entre deux variables quantitatives sa formule est donnée par : $\rho = \frac{cov(x, y)}{\sigma_x \sigma_y}$; le degré de la liaison dépend de la valeur du ρ :

- Si $|\rho|=1$ la corrélation est dite parfaite,
- Si $0.8 \leq |\rho| < 1$ la corrélation est forte,
- Si $0.5 \leq |\rho| < 0.8$ la corrélation est moyenne,
- Si $0 \leq |\rho| < 0.5$ la corrélation est faible,
- Si $\rho=0$ la corrélation est nulle,

Le tableau suivant présente le coefficient de Pearson entre les différentes variables quantitatives :

<i>force de liaison</i>	<i>Anc</i>	<i>Age_Init</i>	<i>Age_Rachat</i>	<i>PM</i>	<i>Cotis</i>
<i>Anc</i>	100,00%	0,33%	24,91%	62,90%	55,83%
<i>Age_Init</i>	0,33%	100,00%	96,88%	22,52%	39,75%
<i>Age_Rachat</i>	24,91%	96,88%	100,00%	37,59%	52,41%
<i>PM</i>	62,90%	22,52%	37,59%	100,00%	88,67%
<i>Cotis</i>	55,83%	39,75%	52,41%	88,67%	100,00%

▪ Tableau n°4 : Corrélation des variables quantitatives

Le montant de la provision est moyennement corrélé avec l'ancienneté du contrat, et est fortement corrélée avec les cotisations. L'âge à la souscription et l'âge au rachat sont fortement corrélés.

2) Variable quantitative contre variable qualitative

L'outil ANOVA permet d'étudier la liaison entre une variable qualitative à k modalités et une variable quantitative. Il est basé sur l'analyse des variances intergroupe et intragroupe, et sur la statistique de Fisher F.

La force de la liaison peut se mesurer avec le coefficient de détermination R^2 , sa formule est donnée par :

$$R^2 = \frac{\sum_{i=1}^k n_i (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^k \sum_j n_{ij} (y_{ij} - \bar{y})^2} ;$$

Plus le coefficient R^2 est proche de 1, plus la liaison est forte.

Le tableau suivant présente le coefficient R^2 entre les différentes variables quantitatives et qualitatives :

<i>force de liaison</i>	<i>Anc</i>	<i>Age_Init</i>	<i>Age_Rachat</i>	<i>PM</i>	<i>Cotis</i>
<i>Adhesion_C</i>	0,598%	0,107%	13,987%	0,082%	0,025%
<i>Capital</i>	0,020%	3,524%	3,390%	12,425%	7,218%
<i>Mois_Rad</i>	1,120%	0,294%	0,263%	1,061%	0,569%
<i>Period</i>	5,238%	20,863%	15,902%	10,323%	4,156%
<i>Reseau</i>	17,703%	0,519%	1,641%	3,372%	1,539%
<i>Sexe</i>	0,021%	0,004%	0,001%	0,002%	0,000%

▪ Tableau n°5 : Corrélation des variables quantitatives et qualitatives

Le taux maximal du coefficient de corrélation R^2 est de 21%, et concerne la périodicité des cotisations et l'âge à la souscription. On déduit que ces variables analysées ne sont pas corrélées.

3) Variable qualitative contre variable qualitative

L'outil utilisé est le test de Khi deux noté χ^2 , avec un seuil d'erreur α , l'intensité de la liaison est mesurée par l'indice de Cramer V, sa formule est donnée par :

$$V = \sqrt{\frac{\chi^2}{n \times [\min(I, J) - 1]}} ;$$

La valeur de V varie entre 0 et 1, le degré de la force du lien entre les variables qualitatives se mesure selon le tableau suivant :

<i>Valeur de V</i>	<i>Intensité du lien</i>
<i>[0,000;0,045[</i>	<i>Très faible</i>
<i>[0,045;0,090[</i>	<i>Faible</i>
<i>[0,090;0,180[</i>	<i>Moyenne</i>
<i>[0,180;0,360[</i>	<i>Forte</i>
<i>[0,360;1,000[</i>	<i>Très forte</i>

Plus V est proche de zéro, plus il y a indépendance entre les deux variables étudiées. Il vaut 1 en cas de liaison complète.

Le tableau suivant présente l'indice de Cramer V entre les différentes variables qualitatives :

<i>force de liaison</i>	<i>Sexe</i>	<i>Adhesion_C</i>	<i>Period</i>	<i>Mois_Rad</i>	<i>Reseau</i>
<i>Sexe</i>	100,00%	7,13%	0,42%	3,50%	0,60%
<i>Adhesion_C</i>	7,13%	100,00%	0,36%	4,10%	0,95%
<i>Period</i>	0,42%	0,36%	100,00%	5,45%	23,05%
<i>Mois_Rad</i>	3,50%	4,10%	5,45%	100,00%	18,77%
<i>Reseau</i>	0,60%	0,95%	23,05%	18,77%	100,00%

▪ Tableau n°6 : Corrélation des variables qualitatives

Le coefficient maximal de Cramer est de 23%, et concerne le réseau de distribution et la périodicité des cotisations.

En conclusion, les variables corrélées entre elles sont :

- Le montant de la provision et l'ancienneté,
- Le montant de la provision et les cotisations,
- L'âge à la souscription et l'âge au rachat.

Pour la suite de l'analyse, nous retirons les cotisations, l'âge au rachat et le montant de la provision. Nous allons maintenant procéder à une analyse statistique profonde de notre portefeuille, et essayer de ressortir graphiquement le lien entre les variables sélectionnées avec le comportement du rachat et la réduction.

VI. Statistique descriptive des données

Nous commençons par présenter quelques statistiques sur les variables qui seront utilisées dans la modélisation. Nous commençons par les variables quantitatives, puis nous terminerons par l'analyse des variables qualitatives.

1) Statistiques élémentaires des Variables quantitatives

Le tableau suivant présente les statistiques descriptives élémentaires de l'ancienneté du contrat, le montant de la provision, l'âge à la souscription, et le capital garanti.

	<i>Moyenne</i>	<i>Médiane</i>	<i>Ecart type</i>	<i>Minimum</i>	<i>Maximum</i>
<i>Anc</i>	6	4	5	1	29
<i>PM</i>	563	276	808	7	11 989
<i>age_Init</i>	59	61	14	18	86
<i>Capital</i>	3 447	3 000	1 898	1 500	17 747

▪ Tableau n°7 : Statistiques élémentaires des variables quantitatives

Le 01/01/2007 la Mutuelle compte 15 735 contrats vie entière. Le produit est ouvert à la souscription pour les âges de 18 ans à 86 ans, La moyenne d'âge à la souscription est de 59 ans, avec un écart type de 14 ans. Les souscriptions en âges extrêmes de 18 ans et à partir de 80 ans sont peu fréquentes.

L'ancienneté moyenne du contrat est de 6 ans, avec une ancienneté inférieure à 4 ans pour plus de 50% de la population étudiée.

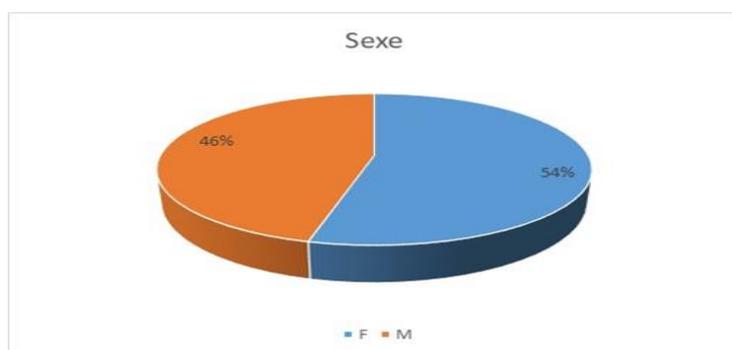
Le montant de la provision est faible par adhérent, avec un minimum de 7 € pour les contrats résiliés le mois même d'adhésion, et un maximum de 12 000 € pour les contrats à prime unique pour lesquels l'engagement de l'assuré est nul.

Le Capital moyen garanti s'élève à 3 447 €, avec un minimum de 1 500 € et un maximum de 18 000 €.

2) Répartition des variables qualitatives

Nous nous intéressons à la répartition du portefeuille en fonction des différentes modalités des variables qualitatives. Cela concerne le sexe de l'adhérent, le type d'adhésion, le réseau de distribution, et la périodicité des cotisations.

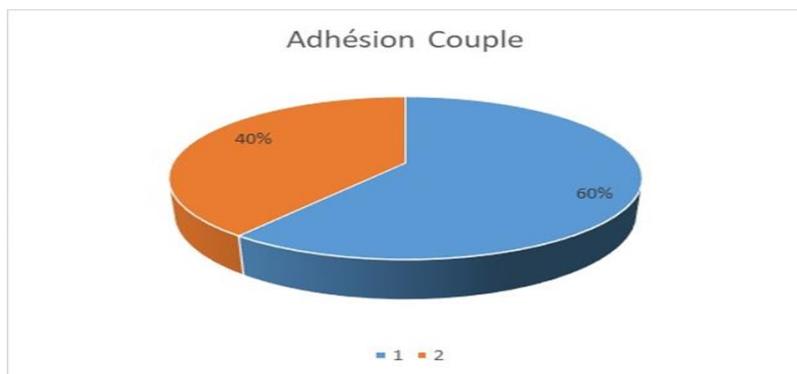
Répartition par Sexe



▪ Figure n°9 : Répartition par Sexe

Nous constatons une répartition du portefeuille légèrement majoritaire chez les femmes.

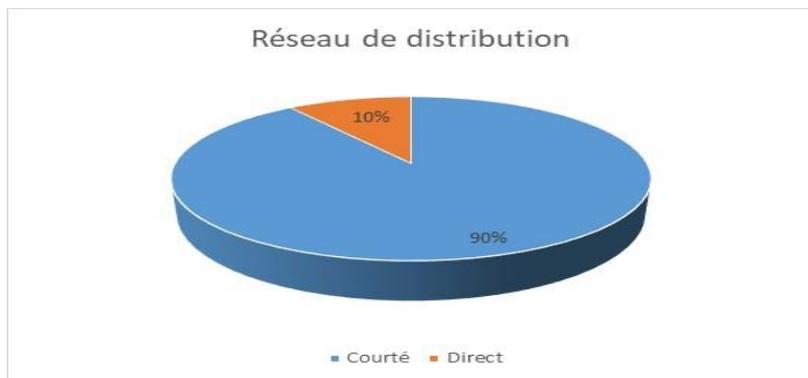
Répartition par type d'adhésion



▪ [Figure n°10 : Répartition par type d'adhésion](#)

Dans le portefeuille, on dénombre 60% des contrats souscrits par des personnes seules, et 40% sont souscrits par un couple.

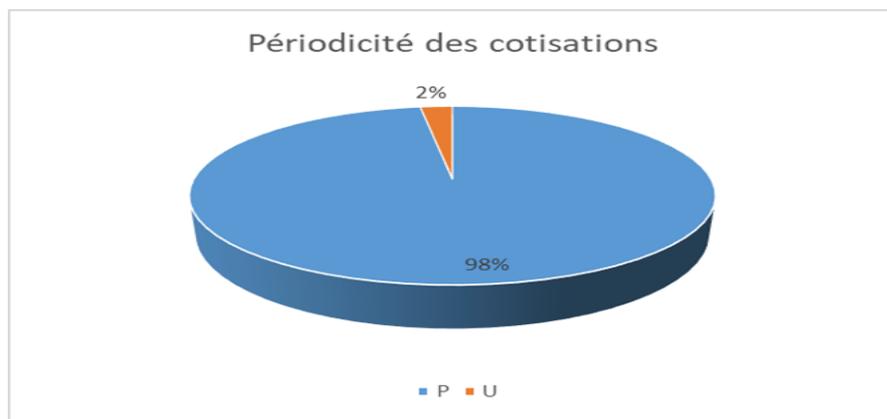
Répartition par réseau de distribution



▪ [Figure n°11 : Répartition par Réseau de distribution](#)

90% des contrats sont distribués par voie de courtage, et 10% sont distribués en direct (dont 3% en Acceptations). Cette variable n'est donc pas discriminante pour le comportement de rachat ou réduction.

Répartition par périodicité des cotisations



▪ [Figure n°12 : Répartition par périodicité des cotisations](#)

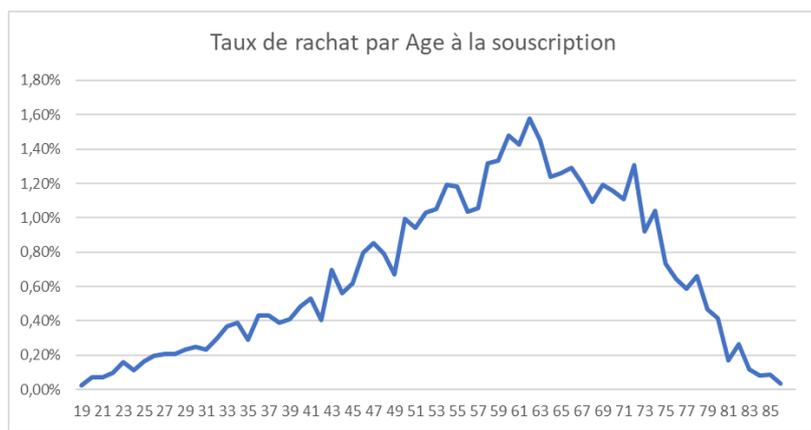
Le versement périodique est majoritaire pour 98% des contrats. Cette variable ne peut donc pas être une variable discriminante pour le comportement de rachat ou la réduction.

Par la suite, nous allons décrire graphiquement les taux de rachat et la réduction en fonction des différentes variables disponibles.

3) Description du taux de rachat

Nous allons décrire graphiquement la variation du taux de rachat en fonction des variables : l'âge à la souscription, l'ancienneté du contrat, le sexe de l'adhérent, le type d'adhésion, le réseau de distribution, la périodicité des cotisations, et le mois de rachat,

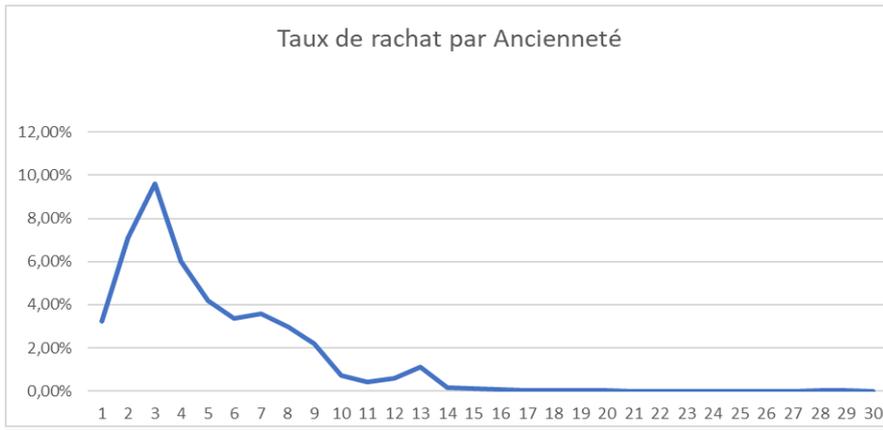
Répartition du taux de rachat en fonction de l'âge à la souscription



▪ [Figure n°13 : Variation du taux de rachat en fonction de l'âge à la souscription](#)

Le taux de rachat augmente avec l'âge à la souscription, jusqu'à l'âge de 61 ans où il atteint son pic, puis il se stabilise entre l'âge de 63 ans jusqu'à l'âge de 72 ans, après on constate une baisse linéaire jusqu'à l'âge de 85 ans.

Répartition du taux de rachat en fonction de l'ancienneté

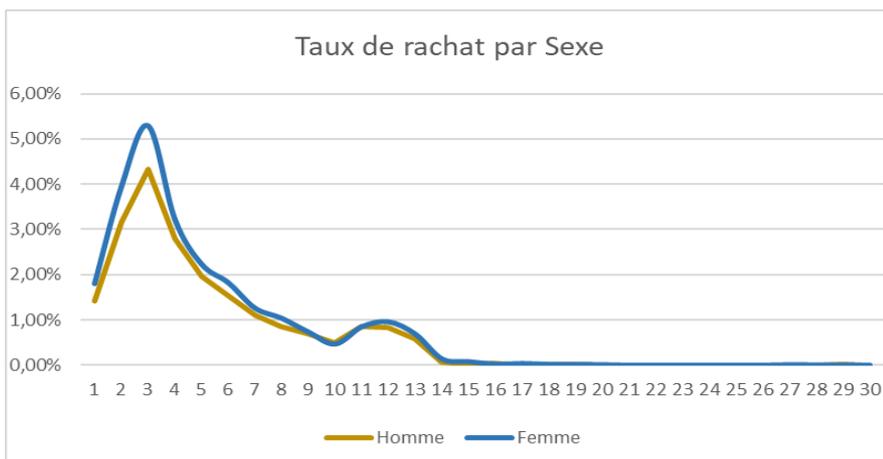


▪ Figure n°14 : Variation du taux de rachat en fonction de l'ancienneté

Le taux de rachat augmente linéairement jusqu'à l'ancienneté 3 ans où il atteint son pic, ensuite nous observons une diminution nette. On constate aussi que le pic fiscal (après 8 ans d'ancienneté) n'est pas visible, fait confirmant que la fiscalité n'est pas une variable discriminante pour le comportement de rachat.

La hausse du taux de rachat peut jusqu'à 3 ans d'ancienneté être expliquée par le fait que parmi les assurés, certains ne faisaient pas la différence entre une garantie vie entière dont la finalité est de couvrir les frais d'obsèques après le décès, et l'épargne qui est une opération d'investissement. Ces assurés décident de mettre fin à leurs contrats à partir du moment où ils comprennent le mécanisme de l'assurance vie entière.

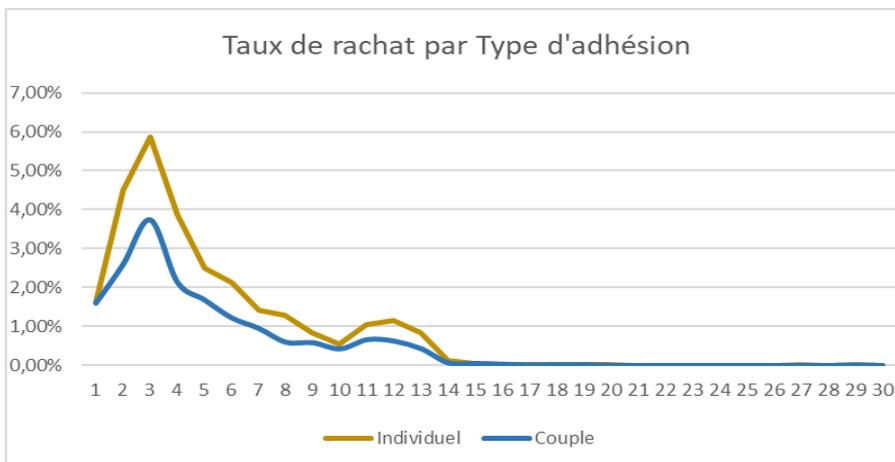
Répartition du taux de rachat en fonction du sexe



▪ Figure n°15 : Variation du taux de rachat en fonction du sexe

Le taux de rachat chez les femmes est supérieur au taux de rachat chez les hommes.

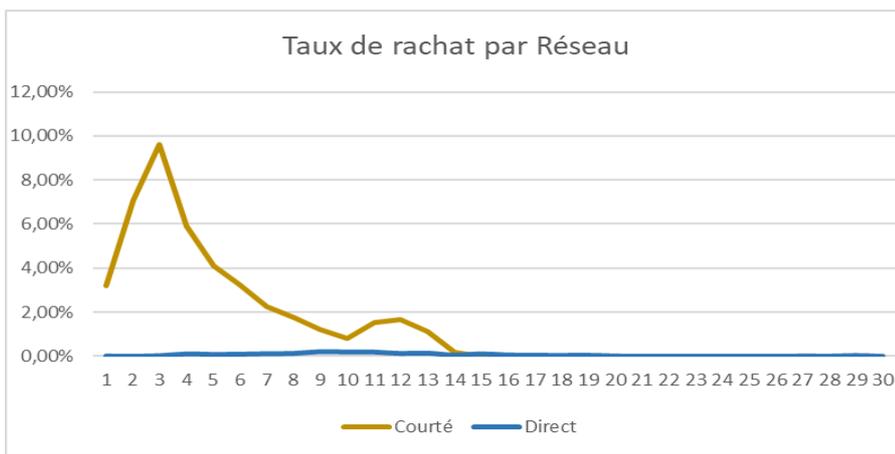
Répartition du taux de rachat en fonction du type d'adhésion



▪ Figure n°16 : Variation du taux de rachat en fonction du type d'adhésion

Le taux de rachat chez les assurés ayant souscrits individuellement un contrat est plus important que chez les assurés ayant souscrits en couple.

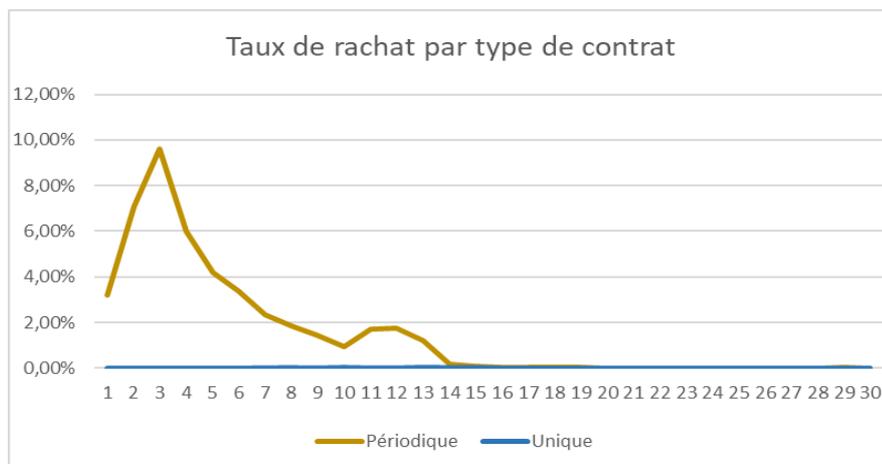
Répartition du taux de rachat en fonction du réseau de distribution



▪ Figure n°17 : Variation du taux de rachat en fonction du réseau de distribution

Les contrats distribués par voie de courtage représentent 90% du portefeuille, et 96% des contrats rachetés.

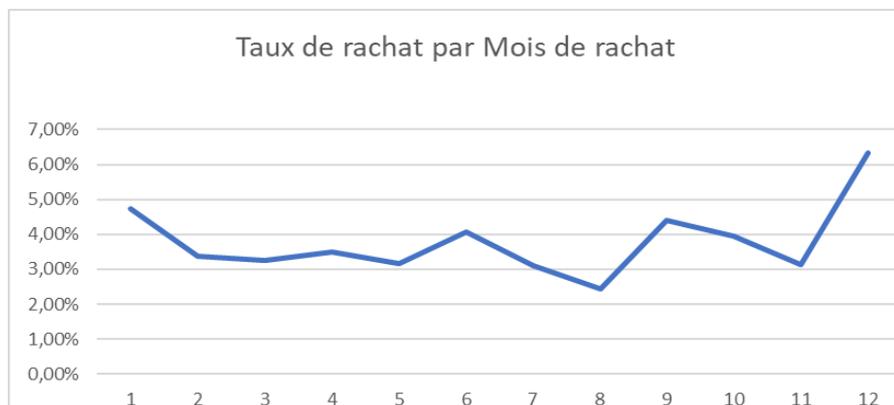
Répartition du taux de rachat en fonction de la périodicité des cotisations



▪ Figure n°18 : Variation du taux de rachat en fonction de la périodicité des cotisations

Les contrats périodiques représentent 98% du portefeuille, il est donc cohérent que le taux de rachat des contrats périodiques soit plus important comparé aux contrats à prime unique.

Répartition du taux de rachat en fonction du mois de rachat



▪ Figure n°19 : Variation du taux de rachat en fonction du mois de rachat

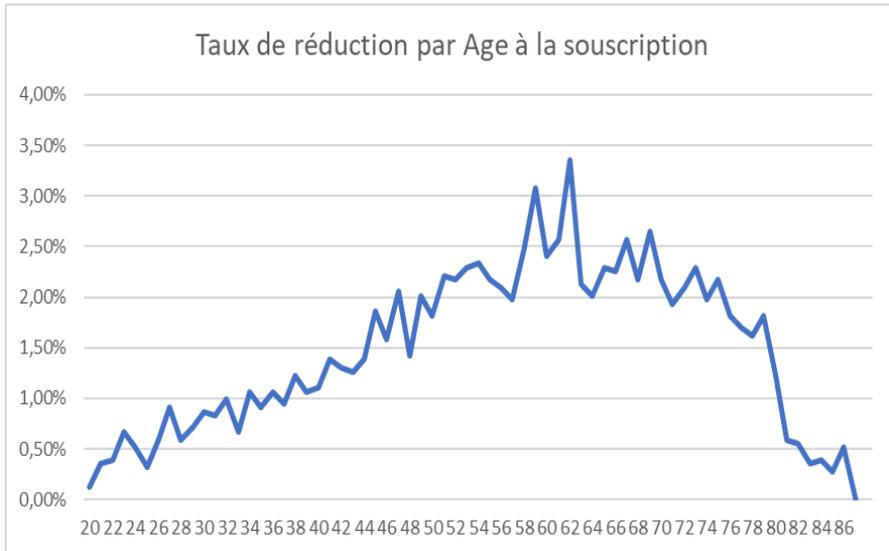
On constate une augmentation du taux de rachat durant les mois de janvier, juin, et septembre, avec un pic en mois de décembre.

Ces mois correspondent à des périodes de début ou de fin de vacances, on est bien en présence d'un effet de saisonnalité.

4) Description du taux de réduction

Nous allons décrire graphiquement la variation du taux de réduction en fonction de l'âge à la souscription, l'ancienneté du contrat, le sexe de l'adhérent, le type d'adhésion, le mois de la réduction.

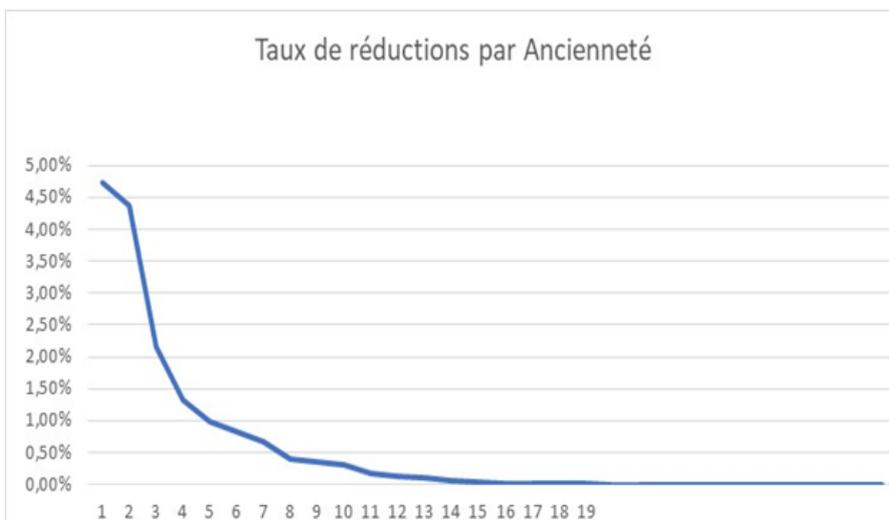
Répartition du taux de réduction en fonction de l'âge à la souscription



▪ Figure n°20 : Variation du taux de réduction en fonction de l'âge à la souscription

Le taux de réduction augmente avec l'âge, il atteint son maximum à l'âge de 62 ans, puis se stabilise entre les âges de 66 ans à 75 ans, ensuite il baisse linéairement à partir de 75 ans.

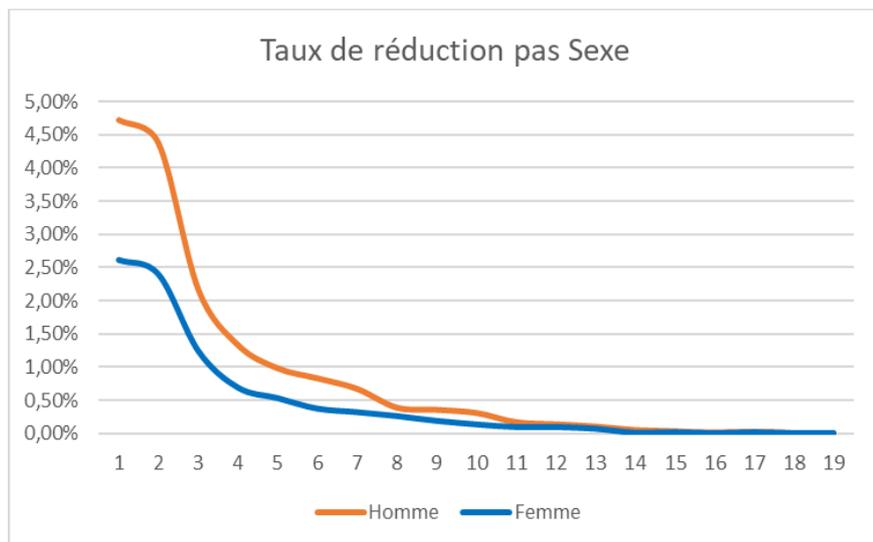
Répartition du taux de réduction en fonction de l'ancienneté



▪ Figure n°21 : Variation du taux de réduction en fonction de l'ancienneté

On observe un taux maximum à l'ancienneté 1 ans, puis une baisse importante après la première année, les taux sont quasi nuls entre l'ancienneté 12 ans à 19 ans.

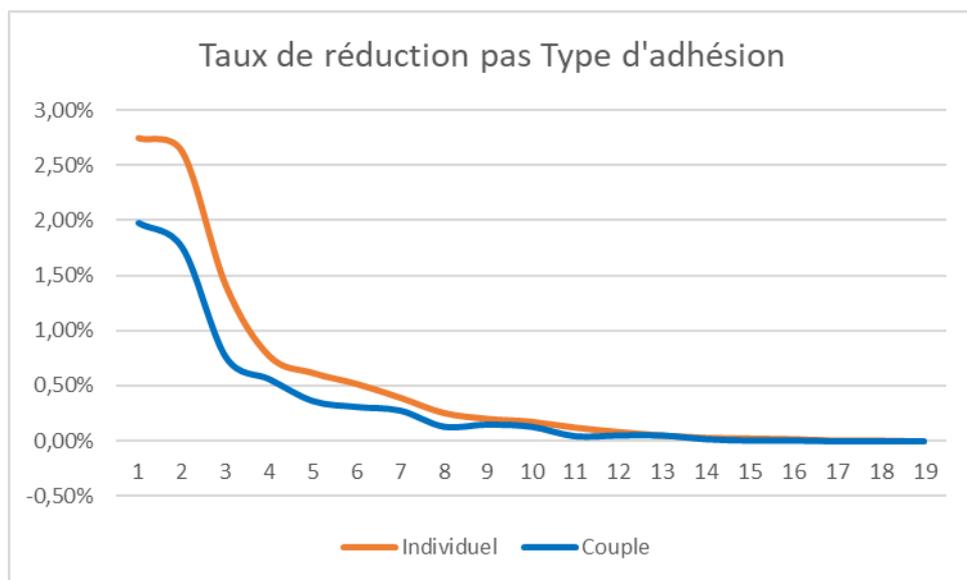
Répartition du taux de réduction en fonction du sexe



▪ Figure n°22 : Variation du taux de réduction en fonction du sexe

Le taux de réduction est supérieur chez les hommes que chez les femmes.

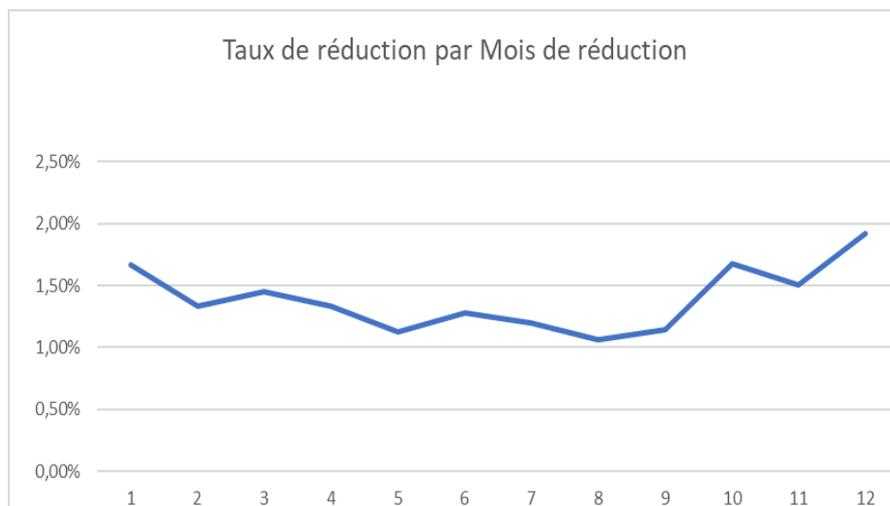
Répartition du taux de réduction en fonction du type d'adhésion



▪ Figure n°23 : Variation du taux de réduction en fonction du type d'adhésion

Le taux de réduction chez les assurés ayant souscrits individuellement un contrat, est plus important que chez les assurés ayant souscrits en couple.

Répartition du taux de réduction en fonction du mois de réduction



▪ Figure n°24 : Variation du taux de réduction en fonction du mois de réduction

L'effet de saisonnalité est visible pour les mois de janvier et décembre.

Conclusion

La phase de traitement préliminaire a permis de définir la période d'observation de l'étude durant laquelle les taux de rachat et réduction sont quasiment stables, et de créer une base de données fiable, exploitable, sans anomalies majeures.

L'étude d'interaction entre les différentes variables, a permis de supprimer les variables fortement corrélées entre elles.

L'analyse des statistiques descriptives donne une bonne vision du portefeuille dans sa globalité.

L'analyse du rachat en fonction des différentes variables explicatives permet de constater que :

- Le taux de rachat augmente linéairement avec l'âge à la souscription jusqu'à 61 ans, puis il baisse au-delà,
- Le taux de rachat chez les femmes est plus important que chez les hommes,
- Le taux de rachat est important durant les trois premières années d'adhésion, puis il baisse fortement au-delà,
- Le taux de rachat est plus important chez les personnes ayant souscrits individuellement un contrat d'assurance, que chez les personnes ayant souscrits en couple,
- L'effet de saisonnalité est visible avant et après les périodes des fêtes et des vacances.

L'analyse de la réduction en fonction des différentes variables explicatives nous montre que :

- Le taux de réduction augmente progressivement avec l'âge à la souscription jusqu'à 62 ans, puis il diminue au-delà,
- Le taux de réduction chez les hommes est supérieur à celui des femmes,
- Le taux de réduction est très élevé la première année d'adhésion, puis il baisse fortement après,
- Le taux de réduction est important chez les personnes ayant souscrits individuellement un contrat d'assurance, que chez les adhésions en couple,
- Effet de saisonnalité pendant et après la période des fêtes de fin d'année,
- Les modalités 'Courté' de la variable réseau de distribution, et 'Périodique' de la variable périodicité des cotisations, représentent chacune plus de 90% des données, par conséquent, l'analyse du rachat et la réduction en fonction de ces variables ne sera pas objective. Elles seront donc retirées de la modélisation.

Nous pouvons à présent procéder à la sélection des variables explicatives du rachat et réduction par l'utilisation des techniques de classification et d'analyse discriminante.

Sélection des variables

Cette partie a pour objectif de sélectionner les variables les plus explicatives du comportement de rachat et réduction parmi toutes les variables retenues lors de la section précédente.

Nous utilisons deux techniques de segmentation : la technique de classification par les arbres de décision, et l'analyse discriminante par la procédure STEPDISC.

Le choix est porté sur ces deux méthodes, en raison de leur simplicité à mettre en œuvre, leur accessibilité dans le logiciel SAS, et ce sont des méthodes qui ne nécessitent pas d'hypothèses sur les données.

I. Classification avec les arbres de décision

1) Introduction

L'arbre de décision est un outil de classification et de prédiction, sa popularité repose en grande partie sur sa simplicité. Un arbre de décision est composé d'un nœud racine par lequel entrent les données, de nœuds feuilles qui correspondent à un classement de questions et de réponses qui conditionnent la question suivante.

C'est un processus interactif d'induction de règles qui permet d'aboutir à une affectation bien justifiée. Le branchement des nœuds passe par le calcul de différents critères selon l'algorithme choisi.

- Le nœud racine contient toutes les données de l'apprentissage,
- A chaque étape un nœud est divisé pour construire deux nouveaux nœuds les plus homogènes possible au sens de la variable à expliquer,
- L'arbre maximal est obtenu lorsqu'aucun nœud ne peut plus être divisé. Un nœud terminal est appelé feuille,
- Chaque feuille est alors affectée à l'une des modalités de la variable à expliquer.

Il existe plusieurs méthodes de discrimination par arbre. En général ces algorithmes suivent le même schéma et nécessitent :

- La définition d'un critère permettant de choisir la variable de segmentation, celle qui permet la meilleure division parmi toutes celles admissibles,
- Une règle permettant de décider si un nœud est terminal, il devient alors une feuille,
- Choix de la taille optimale de l'arbre : Les arbres de décision sont très raffinés et ont tendance à produire un classifieur trop complexe voire instable car fortement dépendant des échantillons d'apprentissage qui a permis leurs constructions. Bien que les feuilles soient pures, elles sont composées de très peu d'individus pour être fiable lors de la

prédiction. C'est le phénomène de surajustement.

Pour éviter ce phénomène, des règles d'arrêts sont fixées selon l'algorithme choisi pour stopper ou réduire l'arbre de manière à obtenir un classifieur optimal, nous parlons ici des méthodes de pré-élagage ou post-élagage de l'arbre.

D'autres critères plus empirique relatifs au nombre de feuilles, à la profondeur de l'arbre, au nombre d'effectifs dans un nœud peuvent être mis en place. L'objectif est d'éviter l'apparition de sommets d'effectifs trop faibles pour espérer obtenir une prédiction fiable.

Dans ce qui suit, nous présentons les trois algorithmes de classification les plus connus CHAID, CART, et C4.5, ensuite nous passons à la mise en œuvre pratique de ces méthodes sur le portefeuille de l'étude, puis à la validation de la qualité de la segmentation.

2) Description des algorithmes des arbres de décision

Dans cette partie nous présentons trois méthodes connues, à savoir CART, CHAID, et C4.5. Ces méthodes ont l'avantage d'être disponibles dans SAS avec la procédure PROC Hpsplit, et comme résultat le tableau de toutes les variables prédictives triées par ordre décroissant de l'indicateur de qualité de la segmentation.

a) Algorithme CHAID

CHAID (CHi-squared Automatic Interaction Detector) a été publiée en 1980 par Gordon V. Kass

➤ Choix de la variable de segmentation

- Cas d'une variable prédictive Catégorielle

On note Y la variable à expliquer à K modalité, et X une variable catégorielle prédictive avec L modalités, le tableau de croisement des variables est le suivant :

Y / X	x_1	x_l	x_L	Σ
y_1		\vdots		
y_k	\dots	n_{kl}	\dots	$n_{.l}$
y_K		\vdots		
Σ		$n_{.k}$		n

Pour évaluer la pertinence d'une variable dans la segmentation, CHAID utilise le Khi-2 d'écart à l'indépendance dont la formule est la suivante :

$$\chi^2 = \sum_{k=1}^K \sum_{l=1}^L \frac{\left(n_{kl} - \frac{n_k \times n_l}{n} \right)^2}{\frac{n_k \times n_l}{n}} ;$$

Le procédé est le suivant : la mesure de Khi-2 est calculée pour toutes les variables prédictives, la variable retenue est celle qui maximise cette mesure. Ce processus est réitéré sur chaque segment que l'on veut segmenter.

- Cas d'une variable prédictive Continue

La segmentation se fait en 2 étapes :

- Définir un seuil de coupure ou discrétisation qui permet de produire le meilleur partitionnement pour chacune des variables. Pour une variable continue X, il s'agit dans un premier temps de trier les données selon les valeurs de X, puis tester chaque borne de coupure possible entre deux valeurs de la variable en calculant la mesure Khi-2 avec la variable à expliquer (rachat ou réduction),
- Sélectionner globalement la meilleure segmentation en comparant la pertinence de toutes les variables : les variables discrètes et les variables continues qui ont été découpées en 2 intervalles.

➤ Définir la taille optimale de l'arbre

Le pré-élagage consiste à fixer une règle d'arrêt qui permet de stopper la construction de l'arbre lors de la phase de construction. En ce sens, la méthode CHAID accepte la segmentation si le Khi-2 calculé sur un sommet est significativement supérieur à un seuil que l'on se fixe. La formalisation passe par un test d'hypothèse statistique : l'hypothèse nulle est l'indépendance de la variable de segmentation avec la variable à prédire. Si le Khi-2 calculé est supérieur au seuil théorique correspondant au risque de première espèce que l'on s'est fixé arbitrairement noté α , on accepte la segmentation.

b) Algorithme CART

CART (Classification and Regression Tree) a été introduit par Breiman, Friedman, Olsen, et Stone en 1984.

➤ Choix de la variable de segmentation

Pour évaluer la pertinence d'une variable dans la segmentation, CART utilise une fonction d'impureté appelé indice de GINI dont la formule pour un nœud donné est la suivante :

$$I(Y) = - \sum_{k=1}^K \frac{n_{k.}}{n} \times \left(1 - \frac{n_{k.}}{n}\right) ;$$

avec n le nombre d'observations du nœud, $n_{k.}$ est le nombre d'observations du nœud ayant la modalité k . Le procédé est le suivant :

- L'indice de GINI est calculé pour toutes les règles de divisions possibles,
- Calcul du cout du nœud avec la formule suivante :

$$\text{Coût du nœud} = \frac{n_G}{n} \times G_{droite} + \frac{n_d}{n} \times G_{gauche} ;$$

Avec n le nombre d'observations du nœud, n_G le nombre d'observations du nœud fils gauche, n_d est le nombre d'observation du fils droit.

La règle de division retenue est celle qui minimise le coût du nœud. Ce processus est réitéré sur chaque sommet que l'on veut segmenter. L'algorithme s'arrête lorsque le nœud est pur.

D'autres règles d'arrêt peuvent être utilisées :

- La profondeur de l'arbre atteint la valeur maximale définie,
- La taille du nœud est inférieure à la taille minimale définie,
- La taille d'un des nœuds fils qui résulterait de la séparation optimale est inférieure à la taille minimale définie.

➤ Définir la taille optimale de l'arbre

Le principe du Post-élagage consiste à :

- Construire un arbre maximal qui soit le plus pur possible, et dans lequel nous acceptons toutes les segmentations même si elles ne sont pas pertinentes,
- Construire une suite emboîtée de sous arbres à partir de l'arbre maximal par élagage successif
- Choisir parmi cette suite, un sous arbre optimal au sens du critère coût -complexité qui mesure un compromis entre la taille de l'arbre et son coût de mauvais classement.

c) Algorithme C4.5

Cet algorithme a été proposé par Quinlan en 1993

➤ Choix de la variable de segmentation

La fonction d'impureté utilisée pour mesurer la dispersion des classes est l'entropie de Shannon dont la formule est la suivante :

$$E(Y) = - \sum_{k=1}^K \frac{n_{k.}}{n} \times \log_2 \left(\frac{n_{k.}}{n} \right)$$

Le procédé est le même que l'indice de Gini.

➤ Définir la taille optimale de l'arbre

C4.5 élague l'arbre construit afin de supprimer les règles inutiles, c'est le principe du Post-élagage qui est retenu pour la construction de l'arbre optimal comme pour CART, à la différence que l'estimation du taux d'erreur du classement est calculée sur le même échantillon d'apprentissage mais pénalisée par la taille de l'effectif du sommet à traiter, alors que CART utilise une évaluation du taux d'erreur avec un échantillon test.

3) Evaluation de la qualité du modèle

Il s'agit de confronter la prédiction du modèle avec les valeurs observées sur un échantillon de la population. L'indice de mesure utilisé est le taux d'erreur de classement. Pour que cet indice soit objectif, il est impératif de le mesurer sur un échantillon qui n'a pas servi à la construction du modèle.

a) Echantillon test

Les observations sont divisées en deux échantillons, le premier appelé apprentissage, servira à construire le modèle, et le deuxième appelé test, qui servira à évaluer et à comparer les modèles.

b) Validation Croisée

Une autre technique de rééchantillonnage est souvent appliquée pour la validation du modèle, il s'agit de la validation croisée qui permet d'utiliser l'ensemble des observations pour la construction et la validation du modèle. On découpe les observations en k parties à peu près égales. Tour à tour, chacune des k parties est utilisée comme jeu de test. Le reste (autrement dit l'union des k-1 autres parties) est utilisé pour la construction du modèle. Le taux d'erreur de classement correspond à la moyenne des erreurs sur les k parties.

c) Critère AUC-Courbe ROC

La courbe ROC est un outil très riche pour l'évaluation et la comparaison des performances des classifieurs. Elle met en relation le taux de vrais positifs avec le taux de faux positifs dans un graphique nuage de points.

Le critère AUC (Area under curve) correspond à la surface située sous la courbe ROC, elle exprime la probabilité de placer un individu positif devant un négatif. Ainsi pour $AUC=1$ la discrimination est parfaite, et les positifs sont surs d'être placés devant les négatifs.

Le tableau ci-dessous présente la qualité de la discrimination en fonction de la valeur du critère AUC :

Valeur de l'AUC	Commentaire
$AUC = 0.5$	Pas de discrimination.
$0.7 \leq AUC < 0.8$	Discrimination acceptable
$0.8 \leq AUC < 0.9$	Discrimination excellente
$AUC \geq 0.9$	Discrimination exceptionnelle

- [Tableau n°8 : Qualité de discrimination AUC](#)

4) Application numérique sur le portefeuille de rachat

Nous allons procéder à l'application des algorithmes de classification décrits plus haut sur le portefeuille de rachat. Les données de sortie correspondent à un tableau des variables les plus pertinentes triées par ordre décroissant de l'indicateur de la qualité de la segmentation.

Pour juger de la qualité du modèle, le taux d'erreur de classement est calculé sur un échantillon test (20% du portefeuille), puis calculé avec la technique de validation croisée.

Nous présentons également le critère AUC pour conforter les résultats déjà obtenus.

a) Méthode de CHAID

Variable Importance			
Variable	Training		Count
	Relative	Importance	
ANC	1.0000	57.4591	11
CAPITAL	0.3148	18.0898	41
AGEINIT	0.1727	9.9236	38
ADH	0.0331	1.9000	3
SEXE	0.0307	1.7647	2

- [Tableau n°9 : Classification des variables avec la méthode CHAID](#)

L'ancienneté est de loin la variable la plus discriminante, le capital garanti et l'âge à la souscription le sont également mais dans une moindre mesure.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Apprentissage	0	4098	160	0.0376
	1	1163	4407	0.2088
Validation	0	1050	52	0.0472
	1	304	1021	0.2294

- Tableau n°10 : Matrice de confusion modèle apprentissage et modèle test du rachat avec méthode CHAID

Il en ressort que le taux de mauvaise classification sur l'échantillon d'apprentissage est de 13%, celui sur l'échantillon test est de 14% soit un écart de 1,03%.

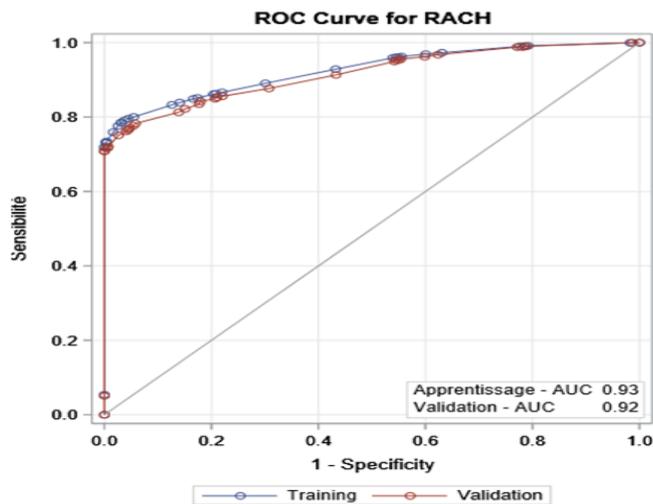
- ✚ La performance du classifieur sur l'échantillon test avoisine 98,97% de classification correcte.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Basé sur le modèle	0	5132	228	0.0425
	1	1348	5547	0.1955
Validation croisée	0	5032	328	0.0612
	1	1474	5421	0.2138

- Tableau n°11 : Matrice de confusion modèle apprentissage et modèle validation Croisée du rachat avec la méthode CHAID

Le taux d'erreur du modèle est de 12,8%, celui de la validation croisée s'élève à 14,7% soit un écart de 1,84%.

- ✚ La performance du classifieur avec la validation croisée avoisine 98,16% de classification correcte.



▪ Figure n° 25 : la courbe ROC du rachat avec la méthode CHAID

La valeur du critère AUC est de 0,93 .la discrimination est donc jugée excellente.

b) Méthode de CART

Variable Importance			
Variable	Training		Count
	Relative	Importance	
ANC	1.0000	57.4522	9
CAPITAL	0.3135	18.0137	40
AGEINIT	0.1758	10.1015	41
ADH	0.0331	1.9000	3
SEXE	0.0307	1.7647	2

▪ Tableau n°12 : Classification des variables avec la méthode CART

Comme pour la méthode CHAID, l'ancienneté, le capital garanti et l'âge à la souscription sont les plus discriminantes.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Apprentissage	0	4135	183	0.0424
	1	1136	4393	0.2055
Validation	0	995	47	0.0451
	1	274	1092	0.2006

▪ Tableau n°13 : Matrice de confusion du rachat modèle apprentissage et modèle test avec méthode CART

Le taux de mauvaise classification sur l'échantillon d'apprentissage est de 13,45%, celui sur l'échantillon test est de 13% soit un écart de 0,36%.

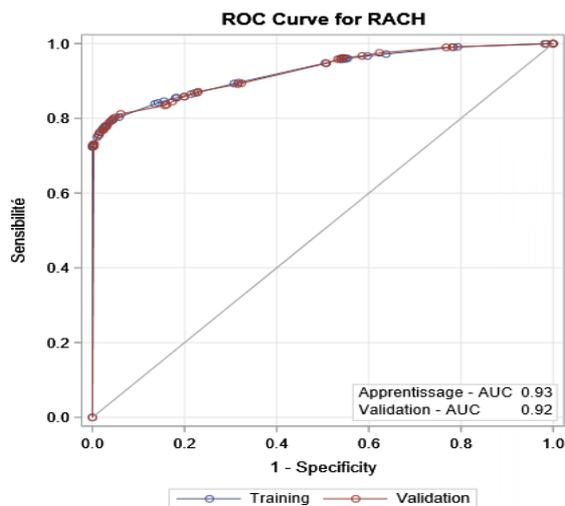
- ✚ La performance du classifieur sur l'échantillon test avoisine 99,64% de classification correcte.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Basé sur le modèle	0	5132	228	0.0425
	1	1347	5548	0.1954
Validation croisée	0	5026	334	0.0623
	1	1474	5421	0.2138

- [Tableau n°14 : Matrice de confusion modèle apprentissage et validation croisée du rachat avec méthode CART](#)

Le taux d'erreur du modèle est de 12,85%, celui de la validation croisée s'élève à 14,75% soit un écart de 1,90%.

- ✚ La performance du classifieur avec la validation croisée avoisine 98,10% de classification correcte.



- [Figure n° 26 : la courbe ROC du rachat avec la méthode CART](#)

La valeur du critère AUC est de 0,93 .la discrimination est donc jugée excellente.

c) *Méthode de C4.5*

Variable Importance			
Variable	Training		Count
	Relative	Importance	
ANC	1.0000	57.4309	13
CAPITAL	0.3205	18.4050	33
AGEINIT	0.1605	9.2155	43
ADH	0.0324	1.8616	3
SEXE	0.0304	1.7466	3

- [Tableau n°15 : classification des variables avec la méthode C4.5](#)

D'après le tableau, l'ancienneté du contrat est la plus discriminante, suivie par le capital garanti, et l'âge à la souscription.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Apprentissage	0	4039	200	0.0472
	1	1182	4341	0.2140
Validation	0	1061	60	0.0535
	1	291	1081	0.2121

- [Tableau n°16 : Matrice de confusion du rachat modèle apprentissage et modèle test avec méthode C4.5](#)

Le taux de mauvaise classification sur l'échantillon d'apprentissage est de 14,09%, celui sur l'échantillon test est de 14,32% soit un écart de 0,2%.

- ✚ La performance du classifieur sur l'échantillon test avoisine 99,7% de classification correcte.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Basé sur le modèle	0	5145	215	0.0401
	1	1390	5505	0.2016
Validation croisée	0	5013	347	0.0647
	1	1474	5421	0.2138

- [Tableau n°17 : Matrice de confusion du rachat Modèle Apprentissage et Validation Croisée avec méthode C4.5](#)

Le taux d'erreur du modèle est de 13%, celui de la validation croisée s'élève à 14,08% soit un écart de 1,7%.

- La performance du classifieur avec la validation croisée avoisine 98,23% de classification correcte.

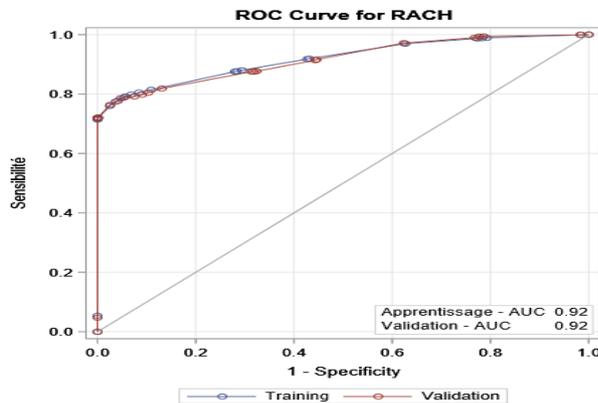


Figure n° 27 : la courbe ROC du rachat avec la méthode C4.5

La valeur du critère AUC est de 0,92 .la discrimination est donc jugée excellente.

Pour résumer, l'application des trois algorithmes de classification par arbre de décision aboutit à des résultats similaires, avec une qualité de modèle très satisfaisante.

Les variables à retenir sont l'ancienneté, le capital garanti, et l'âge à la souscription.

Les variables Sexe et type d'adhésion n'ont aucun effet sur le comportement de rachat.

5) Application numérique sur le portefeuille de réduction

Nous procédons à l'application des trois méthodes d'arbres de décision sur le portefeuille de réduction. La qualité des évaluations est jugée au travers d'un échantillon test correspondant à 20% du portefeuille initial, et aussi de la technique de la validation croisée, et du critère de AUC.

Comme pour le rachat, l'application des 3 méthodes ressortent des résultats similaires, nous nous contenterons de présenter uniquement les résultats de la méthode CHAID.

Variable Importance			
Variable	Training		Count
	Relative	Importance	
ANC	1.0000	25.0635	12
AGEINIT	0.4375	10.9642	54
CAPITAL	0.3426	8.5878	18
ADH	0.1425	3.5726	6
SEXE	0.1122	2.8112	5

Tableau n°18 : Classification des variables avec la méthode CHAID

L'ancienneté, le capital garanti et l'âge à la souscription sont considérées comme des variables discriminantes, le sexe et le type d'adhésion n'ont pas d'effet sur le comportement de la réduction.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Apprentissage	0	2231	15	0.0067
	1	148	1888	0.0727
Validation	0	578	5	0.0086
	1	43	452	0.0869

- [Tableau n°19 : Matrice de confusion modèle apprentissage et modèle test de la réduction avec méthode CHAID](#)

Le taux de mauvaise classification sur l'échantillon d'apprentissage est de 3,8%, celui sur l'échantillon test est de 4,4% soit un écart de 0,65%.

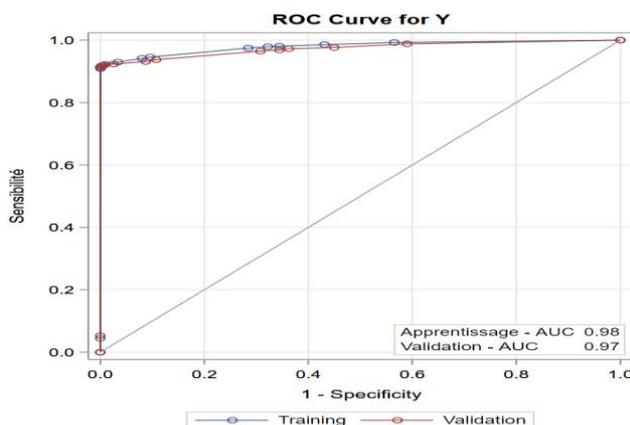
- ✚ La performance du classifieur sur l'échantillon test avoisine 99,35% de classification correcte.

Matrices de confusion				
	Actual	Predicted		Error Rate
		0	1	
Basé sur le modèle	0	2822	7	0.0025
	1	172	2359	0.0680
Validation croisée	0	2797	32	0.0113
	1	216	2315	0.0853

- [Tableau n°20 : Matrice de confusion modèle apprentissage et modèle validation croisée de la réduction avec la méthode CHAID](#)

Le taux d'erreur du modèle est de 3,3%, celui de la validation croisée s'élève à 4,6% soit un écart de 1,3%.

- ✚ La performance du classifieur avec la validation croisée avoisine 98,69% de classification correcte.



▪ Figure n° 28 : la courbe ROC de la réduction avec la méthode CHAD

La valeur du critère AUC est de 0,97 .la discrimination est donc jugée excellente.

Avec les 3 méthodes de classification, les variables : l'ancienneté du contrat, le capital garanti et l'âge à la souscription, sont les variables les plus explicatives du comportement de réduction.

Les modèles d'arbres de décision, malgré leur très grande simplicité, sont particulièrement instables, et très sensibles aux fluctuations des données. Nous avons essayé de pallier cette problématique en introduisant trois algorithmes de classification, le résultat est sans appel, nous avons obtenu la même liste des variables discriminantes. Par la suite, et afin de juger de la qualité de ces modèles, nous avons appliqué trois critères différents, qui ont tous confirmé la qualité de la segmentation.

Dans le chapitre suivant, nous présentons la méthode de l'analyse discriminante (STEPDISC), elle permet d'identifier les variables qui sont pertinentes parmi celles qui sont sélectionnées par la méthode d'arbres de décision.

II. Méthode STEPDISC

1) Description de la méthode

La procédure STEPDISC effectue une analyse discriminante pas à pas pour sélectionner un sous-ensemble de variables quantitatives à utiliser pour discriminer une variable binaire. Elle est basée sur la significativité du test de Fisher appliqué à la statistique du Lambda de Wilks qui n'est d'autre que le rapport de la variance intra classe sur la variance totale.

La procédure regroupe trois méthodes : FORWARD, BACKWARD, et STEPWISE :

- La méthode FORWARD commence sans aucune variable dans le modèle. A chaque étape la procédure PROCDISC entre une variable qui contribue le plus au pouvoir discriminant du modèle au sens de la mesure de Lambda de Wilks. Le processus de sélection s'arrête lorsqu'aucune variable non sélectionnée ne répond au critère d'entrée,
- La méthode BACKWARD commence avec toutes les variables du modèle. A chaque étape, la variable qui contribue le moins au pouvoir discriminant du modèle au sens de la statistique Lambda de Wilks est supprimée. Lorsque toutes les variables restantes remplissent le critère pour rester dans le modèle, le processus BACKWARD s'arrête.
- La méthode STEPWISE commence comme FORWARD sans variables dans le modèle, puis teste à chaque étape si elle ne peut pas éliminer une variable entrée auparavant. Le processus s'arrête lorsque toutes les variables du modèle répondent au critère du maintien au sens de la statistique du Lambda de Wilks, et qu'aucune d'autre variable ne répond au critère d'entrée.

2) Tableau de résultats pour le rachat

a) La méthode BACKWARD

La procédure STEPDISC
Elimination descendante : Etape 1

Statistiques pour la suppression, DDL = 1, 12249			
Variable	R carré partiel	Valeur F	Pr > F
ADH	0.0002	2.41	0.1202
SEXE	0.0002	2.79	0.0951
AGEINIT	0.0003	3.09	0.0787
ANC	0.4302	9247.22	<.0001
CAPITAL	0.0037	45.71	<.0001

Aucune variable ne peut être supprimée.

- Tableau n°21 : Liste des variables discriminantes du rachat par la méthode BACKWARD

Les variables avec des valeurs de la statistique de Lambda F élevées sont retenues. Autrement dit, l'ancienneté du contrat et le capital garanti.

b) La méthode FORWARD

La procédure STEPDISC

Synthèse de la sélection ascendante									
Etape	Nombre dans	Saisi	R carré partiel	Valeur F	Pr > F	Lambda de Wilk	Pr < Lambda	Corrélation canonique moyenne au carré	Pr > ASCC
1	1	ANC	0.4384	9566.13	<.0001	0.56157133	<.0001	0.43842867	<.0001
2	2	CAPITAL	0.0034	42.31	<.0001	0.55963857	<.0001	0.44036143	<.0001
3	3	AGEINIT	0.0003	3.35	0.0672	0.55948556	<.0001	0.44051444	<.0001
4	4	SEXE	0.0002	2.44	0.1182	0.55937408	<.0001	0.44062592	<.0001
5	5	ADH	0.0002	2.41	0.1202	0.55926384	<.0001	0.44073616	<.0001

- Tableau n°22 : Liste des variables discriminantes du rachat par la méthode FORWARD

D'après le tableau, L'ancienneté du contrat et le capital garanti sont les plus discriminantes.

c) La méthode STEPWISE

La procédure STEPDISC

Synthèse de la sélection Stepwise										
Etape	Nombre dans	Saisi	Supprimé	R carré partiel	Valeur F	Pr > F	Lambda de Wilk	Pr < Lambda	Corrélation canonique moyenne au carré	Pr > ASCC
1	1	ANC		0.4384	9566.13	<.0001	0.56157133	<.0001	0.43842867	<.0001
2	2	CAPITAL		0.0034	42.31	<.0001	0.55963857	<.0001	0.44036143	<.0001
3	3	AGEINIT		0.0003	3.35	0.0672	0.55948556	<.0001	0.44051444	<.0001
4	4	SEXE		0.0002	2.44	0.1182	0.55937408	<.0001	0.44062592	<.0001
5	5	ADH		0.0002	2.41	0.1202	0.55926384	<.0001	0.44073616	<.0001

- Tableau n°23 : Liste des variables discriminantes du rachat par la méthode STEPWISE

Comme pour les méthodes BACKWARD et FORWARD, les variables ancienneté et capital garanti sont les plus discriminantes avec la procédure STEPWISE.

En résumé, avec les 3 méthodes de l'analyse discriminante, les variables les plus descriptives du rachat sont : l'ancienneté et le capital garanti.

3) Tableau de résultats pour la réduction

a) La méthode BACKWARD

La procédure STEPDISC
Elimination descendante : Etape 2

Statistiques pour la suppression, DDL = 1, 5355			
Variable	R carré partiel	Valeur F	Pr > F
ADH	0.0005	2.71	0.0999
AGEINIT	0.0284	156.43	<.0001
ANC	0.0580	329.76	<.0001
CAPITAL	0.0021	11.05	0.0009

Aucune variable ne peut être supprimée.

- Tableau n°24 : Liste des variables discriminantes de la réduction par la méthode BACKWARD

L'ancienneté et l'âge à la souscription sont considérées comme variables discriminantes avec la méthode BACKWARD.

b) La méthode FORWARD

La procédure STEPDISC

Synthèse de la sélection ascendante									
Etape	Nombre dans	Saisi	R carré partiel	Valeur F	Pr > F	Lambda de Wilk	Pr < Lambda	Corrélation canonique moyenne au carré	Pr > ASCC
1	1	ANC	0.0837	489.18	<.0001	0.91633888	<.0001	0.08366112	<.0001
2	2	AGEINIT	0.0261	143.60	<.0001	0.89241710	<.0001	0.10758290	<.0001
3	3	CAPITAL	0.0021	11.09	0.0009	0.89057233	<.0001	0.10942767	<.0001
4	4	ADH	0.0005	2.71	0.0999	0.89012219	<.0001	0.10987781	<.0001

- Tableau n°25 : Liste des variables discriminantes de la réduction par la méthode FORWARD

D'après le tableau, on retient également l'ancienneté et l'âge à la souscription avec la méthode FORWARD.

c) La méthode STEPWISE

La procédure STEPDISC										
Synthèse de la sélection Stepwise										
Etape	Nombre dans	Saisi	Supprimé	R carré partiel	Valeur F	Pr > F	Lambda de Wilk	Pr < Lambda	Corrélation canonique moyenne au carré	Pr > ASCC
1	1	ANC		0.0837	489.18	<.0001	0.91633888	<.0001	0.08366112	<.0001
2	2	AGEINIT		0.0261	143.60	<.0001	0.89241710	<.0001	0.10758290	<.0001
3	3	CAPITAL		0.0021	11.09	0.0009	0.89057233	<.0001	0.10942767	<.0001
4	4	ADH		0.0005	2.71	0.0999	0.89012219	<.0001	0.10987781	<.0001

- Tableau n°26 : Liste des variables discriminantes de la réduction par la méthode STEPWISE

Avec la méthode STEPWISE, ce sont l'ancienneté du contrat et l'âge à la souscription qui sont retenues, avec des valeurs de F élevées.

En résumé, avec la technique de discrimination STEPDISC, l'ancienneté et l'âge à la souscription sont retenues pour l'analyse de la réduction.

Conclusion

Pour expliquer le comportement du rachat :

- Les arbres de décision avec les 3 méthodes utilisées, ont abouti à la même liste des variables explicatives : l'ancienneté du contrat, le capital garanti, et l'âge à la souscription,
- L'ancienneté du contrat est la variable la plus importante pour expliquer le comportement du rachat.
- Les variables retenues par l'analyse discriminante STEPDISC avec les 3 méthodes appliquées sur les données sont : l'ancienneté du contrat et le capital garanti.
- Les tests statistiques utilisés confirment que les variables : le sexe, le type d'adhésion n'ont pas de lien avec la variable rachat.

En résumé, les variables retenues pour la modélisation du rachat sont l'ancienneté du contrat, le capital garanti et l'âge à la souscription.

Pour expliquer le comportement de la réduction :

- Les variables retenues avec la technique des arbres de décision sont : l'ancienneté du contrat, le capital garanti, et l'âge à la souscription.
- L'ancienneté est la variable la plus pertinente parmi toutes les autres variables.
- Les variables retenues par la procédure STEPDISC sont : l'ancienneté, l'âge à la souscription.

En conclusion, les variables retenues pour la modélisation de la réduction sont : l'ancienneté du contrat, le capital garanti et l'âge à la souscription.

Modélisation par régression logistique et études d'impact

Après avoir sélectionné les variables explicatives du comportement de rachat et la réduction, nous allons procéder à la quantification de cette liaison à travers l'estimation des probabilités de survenance par contrat. Pour ce faire, nous allons utiliser la régression logistique binaire. 80% des données ont été sélectionnées pour l'estimation des paramètres, et 20% pour la validation du modèle.

Comme dans toute démarche de modélisation, les étapes à suivre sont :

- Choisir la forme de la fonction de lien,
- Estimer les paramètres du modèle à partir d'un échantillon du portefeuille à modéliser,
- Evaluer la précision des estimations,
- Mesurer le pouvoir explicatif du modèle,
- Identifier les variables explicatives les plus significatives,
- Déterminer la formule de la probabilité de rachat et de la réduction.

I. Modélisation par régression logistique

1) Notations

Soit Y une variable aléatoire binaire qui prend ces valeurs dans $\{0,1\}$, soit $x = (x_1, x_2, \dots, x_j)$ le vecteur de j variables aléatoires quantitatives ou qualitatives susceptibles d'expliquer le comportement de l'assuré, il peut s'agir du sexe de l'assuré, l'âge à la souscription, ...

Dans le cadre de notre étude Y correspond à la variable rachat ou réduction, la valeur 0 signifie que le contrat n'est pas racheté ou n'est pas réduit.

On note $\pi(x) = P(Y = 1 | X = x)$ la probabilité de rachat pour un individu ayant les caractéristiques

$$x = (x_1, x_2, \dots, x_j) .$$

On définit $\beta = (\beta_0, \beta_1, \dots, \beta_j)$ comme un vecteur de dimension $j+1$.

a) Principe

Nous cherchons à expliquer la probabilité $\pi(x)$, ou plutôt une transformation de celle-ci, par l'observation conjointe des variables explicatives. L'idée est de chercher un modèle linéaire de la forme :

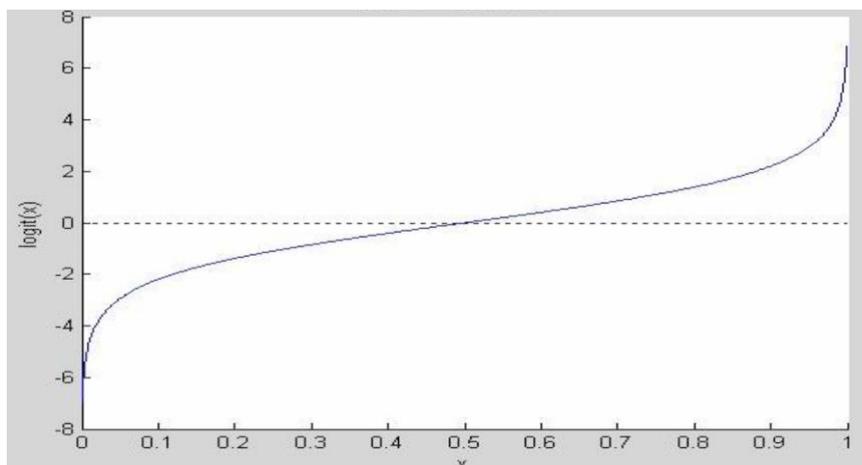
$$f(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j ;$$

b) Fonction de lien

La régression logistique utilise la fonction de lien appelée Logit définie par :

$$\text{logit}(x) = \ln\left(\frac{x}{1-x}\right) \text{ où } x \in]0,1[;$$

Sa représentation graphique est la suivante :



▪ Figure n° 29 : Représentation graphique de la fonction Logit

C'est une fonction monotone, bijective, sa fonction réciproque est donnée par :

$$g(x) = \frac{1}{1 + e^{-x}} \quad \text{Où } x \in]-\infty, +\infty[;$$

c) Formulation mathématique du modèle

La formule mathématique de la régression logistique est donnée par :

$$\ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j$$

Avec β_0, \dots, β_j sont les coefficients du modèle à estimer, la probabilité $\pi(x)$ devient :

$$\pi(x) = P(Y=1|X=x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j}} ;$$

2) Estimation par la méthode du maximum de vraisemblance

a) La fonction de la vraisemblance

Nous notons $\pi(x_i) = P(Y_i|X = x_i)$ comme la probabilité que l'individu i ayant les caractéristiques $x_i = (x_{i1}, x_{i2}, \dots, x_{ij})$ rachète ou réduit son contrat.

La variable $(Y_i|X = x_i)$ prend ces valeurs dans $\{0,1\}$, elle suit une loi binomiale $(1, \pi(x_i))$, sa loi est

$$\text{donnée par } P(Y_i|X = x_i) = (\pi(x_i))^{(Y_i)} (1 - \pi(x_i))^{(1-Y_i)} ;$$

Cette modélisation est cohérente par rapport à ce qui était décrit précédemment, en effet :

Si $Y_i=1$ alors $P(Y_i = 1|X = x_i) = \pi(x_i)$;

Si $Y_i=0$ alors $P(Y_i = 0|X = x_i) = 1 - \pi(x_i)$;

On définit La vraisemblance de l'échantillon ($Y_i|X = x_i$) que l'on note L, par la formule suivante :

$$L(x_1, \dots, x_n) = \prod (\pi(x_i))^{(Y_i)} (1 - \pi(x_i))^{(1-Y_i)} ;$$

Pour simplifier le calcul, nous passons à la Log-vraisemblance noté LogL, dont la relation est donnée par :

$$\text{Log}L(x_1, \dots, x_n) = \text{Log}(\prod (\pi(x_i))^{(Y_i)} (1 - \pi(x_i))^{(1-Y_i)}) ;$$

$$\text{Log}L(x_1, \dots, x_n) = \sum_{i=1}^n \text{Log} (\pi(x_i))^{(Y_i)} (1 - \pi(x_i))^{(1-Y_i)} ;$$

$$\text{Log}L(x_1, \dots, x_n) = \sum_{i=1}^n Y_i \text{Log} (\pi(x_i)) + (1 - Y_i) \text{log}(1 - \pi(x_i)) ;$$

En remplaçant $\pi(x_i)$ par sa formule dans l'équation précédente, la fonction LogL devient :

$$\text{Log}L(x_1, \dots, x_n) = \sum_{i=1}^n (Y_i(\beta_0 + \beta_1 x_{i1} \dots \beta_j x_{ij}) - \log (1 + e^{\beta_0 + \beta_1 x_{i1} \dots + \beta_j x_{ij}})) ;$$

Cette équation définit la loi de la Log-vraisemblance en fonction des paramètres à estimer β_0, \dots, β_j

b) L'Estimateur du Maximum de vraisemblance

La fonction Log étant une fonction croissante, le vecteur qui maximise la vraisemblance L est le même que celui qui maximise la fonction Log-vraisemblance. Rechercher les paramètres $\hat{\beta}$, revient aussi à résoudre l'équation de la vraisemblance suivante :

$$\frac{\partial \text{log}L}{\partial \beta_k} = 0 \quad \forall k \in (0, 1, \dots, j) \quad (1) ;$$

Le vecteur $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_j)$ solution de l'équation (1), s'appelle l'estimateur du maximum de vraisemblance.

L'existence et l'unicité de l'estimateur $\hat{\beta}$ sont assurés par la concavité stricte de la fonction LogL. De plus, l'estimateur $\hat{\beta}$ possède également les propriétés suivantes :

- Il est asymptotiquement sans biais,
- Il est asymptotiquement de variance minimale,
- Il est asymptotiquement normal.

Cependant, il n'existe pas une solution analytique à l'équation (1).

Plusieurs techniques d'optimisation peuvent être utilisées et permettent d'approcher la valeur de $\hat{\beta}$ avec précision à l'aide d'algorithmes itératifs efficaces. Il y a notamment :

- L'algorithme Newton-Raphson,
- L'algorithme de Gauss-Newton,
- Le score de Fisher.

Nous détaillons la méthode de Newton-Raphson car son algorithme est disponible dans SAS.

c) Algorithme de Newton-Raphson

Il est l'une des méthodes numériques les plus utilisées pour optimiser la Log-vraisemblance, il démarre avec une initialisation quelconque du vecteur de paramètre $\hat{\beta}$ pour passer de l'étape (i) à l'étape (i+1), et se rapproche de la solution finale :

$$\beta^{i+1} = \beta^i - \left(\frac{\partial \ln L}{\partial \beta \partial \beta^t} \right)^{-1} \times \frac{\partial \ln L}{\partial \beta} ;$$

Plusieurs règles d'arrêt sont possibles pour stopper le processus :

- Le processus s'arrête lorsque l'écart d'une étape à l'autre de la log-vraisemblance est plus Petit qu'un seuil ε fixé à l'avance,
- Le processus s'arrête lorsque le nombre d'itérations dépasse un nombre fixé à l'avance pour limiter le temps de calcul,
- Le processus s'arrête lorsque l'écart d'une étape à l'autre du vecteur solution $\hat{\beta}$ est faible. Un seuil est fixé à l'avance, auquel on compare la somme des écarts aux carrés à la somme des écarts absolus entre les composantes du vecteur solution.

3) Validité globale du modèle

Nous allons lister brièvement une liste de critères et de tests utilisés dans SAS pour la validation globale du modèle de régression logistique.

a) Critère d'Aikaike

Le critère d'information d'Aikaike noté AIC (Akaike information criterion), est un indicateur de la qualité du modèle statistique. sa formule s'écrit :

$$AIC = 2k - 2Ln(L) ;$$

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

Si l'on considère un ensemble de modèles candidats, le modèle choisi est celui qui aura la plus faible valeur d'AIC. Ce critère repose donc sur un compromis entre la qualité de l'ajustement et la complexité

du modèle, en pénalisant les modèles ayant un grand nombre de paramètres.

b) Critère de SCHWARTZ

Le critère d'information bayésien noté *BIC* (bayésien information criterion), est un critère d'information dérivé du critère d'information d'AKAIKE.

À la différence du critère d'information d'AKAIKE, la pénalité dépend de la taille de l'échantillon et pas seulement du nombre de paramètres. sa formule s'écrit :

$$BIC = -2\ln(L) + k \times \ln(N) ;$$

Avec L la vraisemblance du modèle estimée, N le nombre d'observations dans l'échantillon, et k le nombre de paramètres du modèle.

Comme pour l' *AIC*, le modèle choisi est celui qui aura la plus faible valeur de *BIC* .

c) Test du rapport de vraisemblance

Le test du rapport de vraisemblance consiste à comparer deux modèles emboîtés M_r et M_s . M_r comporte r variables, donc $r+1$ variables à estimer, M_s comporte s variables à $s+1$ variables à estimer . la statistique du test est donnée par : $LR = -2 \times \ln \frac{L(M_s)}{L(M_r)}$;

Avec $L(M_s)$ représente la vraisemblance du modèle avec s paramètres.

Sous l'hypothèse nulle H_0 , les coefficients des variables supplémentaires que l'on retrouve dans avec $L(M_r)$ sont tous nuls.

LR suit une loi de Khi deux à $(r-s)$ degré de liberté. Par conséquent, ce test est rejeté au niveau α lorsque la statistique du test LR est supérieure au quantile d'ordre $1-\alpha$ de la loi $\chi^2(r-s)$.

Dans la pratique $L(M_r)$ représente le modèle complet avec toutes les variables sélectionnées.

d) Test de Significativité des paramètres -Wald

Matrice Hessienne :

Il s'agit de la matrice aux dérivées partielles seconde de la Log-Vraisemblance notée

$H(\beta) = \frac{\partial^2 \ln L}{\partial \beta \partial \beta^t}$, elle est de dimension $(J \times 1 + J \times 1)$, et son expression matricielle s'écrit :

$H = X^t V X$, avec V est la matrice diagonale composée des valeurs de $\pi(x_i) \times (1 - \pi(x_i))$ pour chaque individu i .

Les probabilité $\pi(x_i)$ étant obtenues après estimation des paramètres. X est la matrice des valeurs des différentes variables sélectionnées x_j , avec la première colonne la constante.

Une propriété intéressante est que la matrice de Variance-Covariance correspondant à l'inverse de la matrice Hessienne, soit $\hat{\Sigma} = H^{-1}$. Grâce à cette propriété, nous obtenons facilement les variances des coefficients sur la diagonale principale.

Test de Wald :

Il est fondé sur la propriété de la normalité asymptotique de l'estimateur du maximum de vraisemblance $\hat{\beta}$. On peut tester la significativité, soit pour une variable, soit pour certaines variables, soit pour l'ensemble des variables du modèle.

- Test de Nullité d'un coefficient β_k

La statistique de Wald est donnée par :
$$W_k = \frac{\hat{\beta}_k}{\hat{\sigma}_{\beta_k}^2};$$

Où $\hat{\sigma}_{\beta_k}^2$ est la variance du coefficient β_k lue sur la diagonale principale de la matrice variance covariance. Sous l'hypothèse nulle $H_0 : \beta_k = 0$, la statistique de Wald suit une loi de Khi deux de degré de liberté 1 noté $\chi^2(1)$.

- Test de Nullité des Coefficients $\beta_1 = \beta_2 = \dots = \beta_q$ avec $q < j$

La statistique de Wald est définie par :
$$W_{(q)} = \hat{\beta}_{(q)}^t \times \hat{\Sigma}_{(q)}^{-1} \times \hat{\beta}_{(q)}$$

Où $\hat{\beta}_{(q)}$ est le sous vecteur des valeurs observées des coefficients que l'on souhaite tester. $\hat{\Sigma}_{(q)}^{-1}$ est la sous matrice de Variance-Covariance associée à ces coefficients. Sous l'hypothèse nulle $H_0 : \beta_1 = \beta_2 = \dots = \beta_q = 0$ la statistique de Wald suit la loi de Khi-2 de degré de liberté q noté $\chi^2(q)$.

- Test de Nullité de tous les Coefficients $\beta_1 = \beta_2 = \dots = \beta_j = 0$

La statistique de Wald $W(J)$ est définie par :
$$W(J) = \hat{\beta}^t \times \hat{\Sigma}^{-1} \times \hat{\beta};$$

Sous l'hypothèse nulle $H_0 : \beta_1 = \beta_2 = \dots = \beta_j = 0$ la statistique de Wald suit la loi de Khi-2 de degré de liberté j noté $\chi^2(j)$.

Construction d'intervalle de Confiance pour un coefficient β_k

β_k suit asymptotiquement une loi normale, les bornes de l'intervalle de confiance au niveau de confiance $1-\alpha$ sont obtenues de la manière suivante : $\pm U_{1-\frac{\alpha}{2}} \times \hat{\sigma}_{\beta_k}$ avec $U_{1-\frac{\alpha}{2}}$ est le fractile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite.

e) Rapport des chances ou odds

L'odds ratio (OR), également appelé rapport des cotes, est une statistique exprimant le degré de dépendance entre des variables aléatoires, et permet de mesurer l'effet d'un facteur. Pour une variable réponse binaire à deux modalités 0 et 1, et une variable explicative à deux modalités 0 et 1, l'odds est donné par la formule suivante :

$$\text{Odds} = \frac{\frac{P(Y = 1 / X = 1)}{P(Y = 0 / X = 1)}}{\frac{P(Y = 1 / X = 0)}{P(Y = 0 / X = 0)}} ;$$

- Si odds = 1, la variable X n'a aucun effet sur la variable Y,
- Si odds > 1, cela signifie que le groupe à modalité X=1 à plus de chance à racheter son contrat que le groupe à modalité X=0,
- Si odds < 1, cela signifie que le groupe à modalité X=1 à moins de chance à racheter son contrat que le groupe à modalité X=0.

Après avoir décrit le modèle logistique, nous allons l'appliquer sur le portefeuille de rachat et de la réduction, avec comme variables explicatives celles retenues dans la section précédente grâce aux techniques de classification et de discrimination : l'ancienneté, le montant de la provision, l'âge à la souscription, le capital garanti. Les tableaux de résultats sont issus de SAS³.

4) Application numérique de la régression logistique sur la variable rachat

Dans ce qui suit, nous allons présenter les résultats obtenus dans SAS avec la Procédure PROC LOGISTIC et ses options.

Statistique d'ajustement du modèle		
Critère	Constante uniquement	Constante et Covariables
AIC	13460.602	5024.506
SC	13467.792	5053.268
-2 Log L	13458.602	5016.506

▪ Tableau n°27 : Statistiques d'ajustement du modèle

Le tableau présente les critères d'AIKAIKE et SCHWARTZ, calculés pour le modèle avec la seule constante comme régresseur et pour le modèle spécifié dans la procédure. Les valeurs de ces critères étant inférieures pour notre modèle que dans le cas de la constante seule, le modèle n'est à priori pas trop mauvais.

³ Voir Annexe 2

Test de l'hypothèse nulle globale : BE TA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	8442.0959	3	<.0001
Score	5933.1122	3	<.0001
Wald	2242.1756	3	<.0001

- [Tableau n°28 : Test de l'hypothèse nulle globale du modèle rachat](#)

Le tableau présente le résultat des différents tests (Rapport de vraisemblance, Wald, Score) pour la validité globale du modèle.

Ici le modèle est globalement valide, puisque les 3 tests sont significatifs. En effet, la p-value de la dernière colonne étant inférieure à 5% on rejette l'hypothèse nulle de la nullité des coefficients. Il existe au moins un des coefficients de régression qui est non nul.

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	8.3800	0.2625	1019.5020	<.0001
ANC	1	-0.6570	0.0170	2235.1108	<.0001
AGEINIT	1	-0.00648	0.00254	6.4918	0.0108
CAPITAL	1	-0.00020	0.000019	103.1858	<.0001

- [Tableau n°29 : Valeurs estimées des paramètres du modèle](#)

Le tableau présente l'estimation des paramètres de la régression, avec la méthode du maximum de vraisemblance. On y trouve pour chaque coefficient β_j , sa valeur estimée $\hat{\beta}_j$, son écart type, la valeur de la statistique de Wald, et la p-value de la loi de Khi deux $\chi^2(1)$.

Les variables sont significatives au sens du test de Wald, les p-values sont inférieures au seuil de 5%. On observe également que l'ancienneté est la variable qui a la statistique de Wald la plus élevée, ce qui signifie qu'elle est la variable la plus significative, les autres variables comme le capital garanti et l'âge à la souscription le sont également mais dans une moindre mesure.

Les écarts type mesurent l'incertitude à laquelle soumis chaque estimateur, de plus, ils permettent la construction des intervalles de confiance au seuil de 5%.

Estimation du rapport de cotes			
Effet	Estimation du point	Intervalle de confiance de Wald à 95%	
ANC	0.447	0.432	0.462
AGEINT	0.994	0.989	0.999
CAPITAL	1.000	1.000	1.000

- [Tableau n°30 : Estimation du rapport de cotes](#)

Le rapport des chances de la variable capital garanti est égal à 1, cela signifie que cette variable à priori, n'a pas trop d'effet sur le comportement du rachat.

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	95.7	D de Somers	0.914
Pourcentage discordant	4.3	Gamma	0.914
Pourcentage lié	0.0	Tau-a	0.451
Paires	23703780	c	0.957

- [Tableau n°31 : Statistiques de performance du modèle rachat](#)

Des statistiques complémentaires fournies par SAS pour juger si le modèle est bien adapté aux données, on y trouve la valeur AUC, l'aire sous la courbe ROC, sa valeur est égale à 0,957, confirmant ainsi que le modèle est excellent en termes de discrimination. La statistique Pourcentage concordant indique que les prévisions coïncident à 95,7% avec la réalité observée, donc le modèle est bien adapté aux données.

5) Validation du modèle de la régression du rachat

a) Taux d'erreur de classement

Pour mesurer la performance du modèle, nous utilisons la matrice de confusion issue du modèle de régression, elle permet de comparer les valeurs prédites avec les valeurs réelles de la variable à expliquer ici le rachat, l'indicateur de mesure est le rapport des observations mal classifiées sur l'ensemble de la population (ici c'est le modèle d'apprentissage).

Dans le tableau suivant, on retrouve la matrice de confusion issue du modèle de régression logistique :

Réal	Predicted		Total
	0	1	
Y=0	3993	342	4335
Y=1	494	4974	5468
Total	4487	5316	9804

Le taux d'erreur de classement s'élève à 8,52%, autrement dit, sur 100 observations 8,52 sont mal classifiées par le modèle.

b) Backtesting sur un échantillon-test

Le modèle a été produit à partir d'un échantillon appelé apprentissage correspondant à 80% de l'ensemble des observations. Nous allons tester la performance du modèle sur les 20% des données restantes, appelé échantillon-test, nous calculons ensuite le taux de mauvaise classification :

Réal	Predicted		Total
	0	1	
Y=0	903	108	1011
Y=1	111	1330	1441
Total	1014	1438	2452

Le taux d'erreur de classement s'élève à 8,93%, soit un écart absolu avec le taux d'erreur du modèle de 0,4%, le modèle est donc bien adapté aux données de l'échantillon test, et par conséquent, il est jugé robuste.

6) Application numérique sur le portefeuille réduction

Statistique d'ajustement du modèle		
Critère	Constante uniquement	Constante et Covariables
AIC	5934.492	3232.576
SC	5940.855	3258.031
-2 Log L	5932.492	3224.576

- [Tableau n°32 : Statistiques d'ajustement du modèle](#)

Les critères d'AIKAIKE et de SCHWARTZ sont calculés. Leurs valeurs étant inférieures pour notre modèle que dans le cas de la constante seule. Le modèle est supposé globalement valide.

Test de l'hypothèse nulle globale : BE TA=0			
Test	khi-2	DDL	Pr > khi-2
Rapport de vrais	2707.9160	3	<.0001
Score	1466.7158	3	<.0001
Wald	861.4374	3	<.0001

- [Tableau n°33 : Test de l'hypothèse nulle globale du modèle](#)

On observe que la p-value de la dernière colonne est inférieure à 5%, nous avons aussi les statistiques des tests supérieures à la valeur critique 3,84 correspond au seuil de confiance à 95%. Le modèle est donc valide dans sa globalité au sens des tests du Rapport de vraisemblance, Wald et la Score. Il existe au moins un des coefficients de régression qui n'est pas nul.

Analyse des valeurs estimées du maximum de vraisemblance					
Paramètre	DDL	Estimation	Erreur type	Khi-2 de Wald	Pr > khi-2
Intercept	1	-12.0449	0.5271	522.2076	<.0001
Anc	1	0.6653	0.0289	531.2380	<.0001
Capital	1	0.00131	0.000053	614.4454	<.0001
Age_init	1	0.0876	0.00454	371.8390	<.0001

- [Tableau n°34 : Valeurs estimées des paramètres du modèle](#)

On remarque que tous les paramètres sont significatifs au seuil de 5%, avec une p-value Inférieure à 5%, de plus la statistique de Wald du Capital garanti est la plus élevée, cela signifie qu'elle est la variable la plus significative pour expliquer la réduction. Les écarts type permettent de mesurer les incertitudes liées à l'estimation, et de créer des intervalles de confiance pour les paramètres estimés.

Estimation du rapport de cotes			
Effet	Estimation du point	Intervalle de confiance de Wald à 95%	
Anc	1.024	0.977	1.074
Age_init	1.018	1.012	1.024
Capital	0.996	0.996	0.997

- [Tableau n°35 : Estimation du rapport de cotes](#)

Les rapports de chance de chaque variable explicative avec la variable réponse sont présentés dans le tableau en dessous. On constate que lorsqu'on gagne un an d'ancienneté, le rapport de chance de réduire son contrat sur la modalité de ne pas réduire est multiplié par 1,024.

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	93.2	D de Somers	0.864
Pourcentage discordant	6.8	Gamma	0.864
Pourcentage lié	0.0	Tau-a	0.431
Paires	4584600	c	0.932

- [Tableau n°36 : Statistiques de performance du modèle](#)

Pour juger si le modèle est bien adapté aux données, nous avons les valeurs des statistiques de l'AUC et le pourcentage concordant. Nous observons que les prévisions coïncident à 93,2% avec les données.

7) Validation du modèle de la régression réduction

a) Taux d'erreur de classement

Le taux d'erreur de classement est déterminé à partir de la matrice de confusion suivante :

Réel	Predicted		Total
	0	1	
Y=0	2705	101	2806
Y=1	180	1313	1493
Total	2885	1414	4299

Le taux d'erreur de classement s'élève à 6,536%, autrement dit, sur 100 observations 6,536 sont mal classifiées par le modèle.

b) Backtesting sur un échantillon-test

Comme pour le rachat, les paramètres du modèle sont estimés à partir de l'échantillon d'apprentissage (80% des données), l'échantillon test permet de juger de la robustesse du modèle. La performance est déterminée à partir du taux d'erreur de classement sur l'échantillon test :

Réel	Predicted		Total
	0	1	
Y=0	21	59	80
Y=1	76	905	981
Total	97	964	1061

Le taux d'erreur est de 7,163%, soit un écart absolu de 0,627% par rapport au taux d'erreur du modèle. Le modèle est donc bien adapté aux données de l'échantillon test.

La régression logistique a permis de confirmer la significativité de l'ensemble des variables explicatives sélectionnées en entrée vis-à-vis de la variable à expliquer (rachat ou réduction).

Certaines variables sont plus pertinentes que d'autres, l'ancienneté est de loin la variable la plus importante, en ayant la valeur la plus élevée de la statistique de Wald. Le modèle permet de calculer une estimation ponctuelle de la probabilité de rachat et réduction connaissant la réalisation des variables explicatives.

La validité du modèle est mesurée à la fois grâce aux tests statistiques de (Wald, rapport de vraisemblance), et des indicateurs statistiques tels que les critères d'AIKAIKE, et celui de SCHWARTZ. La robustesse du modèle est validée grâce à l'application du taux d'erreur de classement sur l'échantillon Test.

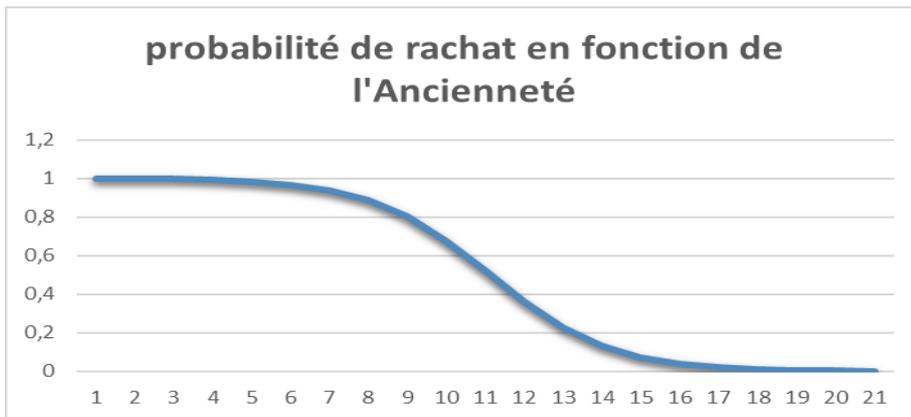
Ainsi, la probabilité $\pi(x_i)$ que le contrat lié à l'individu i ayant les caractéristiques x_i soit racheté est donnée par la formule suivante :

$$\pi(x_i) = \frac{e^{(8,38-0,657 \times Anc - 0,00648 \times Age_{init} - 0,00020 \times Capital)}}{1 + e^{(8,38-0,657 \times Anc - 0,00648 \times Age_{init} - 0,00020 \times Capital)}} ;$$

Ici les caractéristiques de l'individu i sont :

- Anc : l'ancienneté du contrat,
- Age_{init} : l'âge à la souscription du contrat,
- $Capital$: le capital garanti.

Il ressort du modèle que l'ancienneté a un impact important sur la probabilité de rachat chez l'assuré. De plus, le signe négatif des coefficients indique que la probabilité de rachat augmente lorsque l'ancienneté, l'âge à la souscription et le capital garanti diminuent.



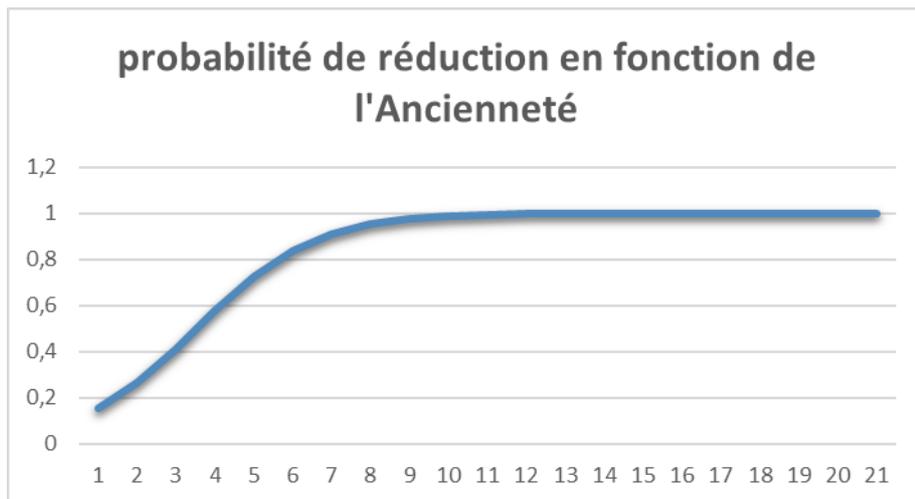
▪ Figure n° 30 : Probabilité de rachat en fonction de l'ancienneté

Durant les premières années du contrat, la probabilité de rachat est élevée, cette tendance s'est confirmée lors de l'analyse descriptive des variables explicatives (figure n°14). Beaucoup d'assurés confondent la vie entière avec de l'épargne, leurs compréhensions du mécanisme de ce produit, les incite à résilier rapidement leurs contrats, ce qui explique le nombre élevé des rachats lors des premières années de souscription. Plus le contrat vieillit, plus il est dans l'intérêt de l'assuré de ne pas racheter son contrat, car les montants des cotisations versées dépassent le montant du rachat en cas de résiliation du contrat.

De même, la probabilité $\pi'(x_i)$ que le contrat lié à l'individu i ayant les caractéristiques x_i soit réduit s'écrit :

$$\pi'(x_i) = \frac{e^{(-12,0449+0,6653 \times Anc+0,0876 \times Age_{init}+0,00131 \times Capital)}}{1 + e^{(-12,0449+0,6653 \times Anc+0,0876 \times Age_{init}+0,00131 \times Capital)}} ;$$

D'après la formule, on Conclue que l'ancienneté a un impact important sur la probabilité de réduction, le signe positif des coefficients estimés indique qu'ils évoluent dans le même sens.



- [Figure n° 31 : Probabilité de réduction en fonction de l'ancienneté](#)

Durant les premières années du contrat La probabilité de réduction augmente linéairement, la même tendance est constatée lors de l'analyse descriptive des variables (Figure n°21). plus le contrat vieillit, plus l'assuré a tout intérêt à réduire son contrat pour constituer un capital en cas de décès, au lieu de demander le rachat de son contrat et perdre une partie des cotisations versées.

II. Etudes d'impact et tests de sensibilité

Dans cette partie, nous allons comparer les lois de rachat issues du modèle de régression logistique avec la méthodologie actuelle utilisée par la mutuelle, basée sur l'historique et l'ancienneté du contrat comme l'unique variable explicative.

Nous allons ensuite procéder à des tests de sensibilité du modèle de l'étude, pour s'assurer que les résultats sont cohérents.

Nous choisissons comme indicateurs, le montant des provisions techniques, le SCR global, et le ratio de solvabilité.

L'analyse est faite sur les données de l'inventaire arrêtées le 31/12/2020.

1) Etudes d'impact du modèle

Le calcul des probabilités de rachat avec le modèle de la régression logistique, nécessite au préalable le calcul des provisions techniques au 31/12/2020.

Les provisions techniques sont calculées comme étant la somme de la meilleure estimation des flux futurs « Best Estimate » et d'une marge pour risque.

a) Best Estimate

Le montant des « Best Estimate » est défini comme la moyenne pondérée par leur probabilité de survenance des flux de trésorerie futurs, en tenant compte de toutes les entrées et sorties de trésorerie nécessaires pour régler les engagements jusqu'au terme de tous les contrats en portefeuille. Il doit également intégrer toutes les options et clauses prévues dans le contrat.

Le Best Estimate est l'actualisation des flux suivants en tenant compte de l'évolution du portefeuille par application des lois de mortalité (table TH 00-02) et lois de rachat :

- Projection annuelle des cotisations théoriques,
- Projection annuelle des prestations capitaux décès et rachats.

L'actualisation s'effectue avec la courbe des taux sans risque au 31/12/2020 fournie par l'EIOPA.

b) Marge pour risque

La marge pour risque représente le coût du capital que devrait lever le cessionnaire pour couvrir son exigence de capital jusqu'à son extinction.

Le calcul de la marge pour risque est réalisé de manière globale pour toutes garanties confondues. Ce calcul est effectué à partir de la méthode simplifiée dite « Duration * BE » proposée par EIOPA dans les spécifications techniques.

$$\text{Marge pour risque} = 6\% \times \text{Duration} \times \frac{SCR_{MR}}{(1 + \text{Taux sans risque}_{1 \text{ an}})}$$

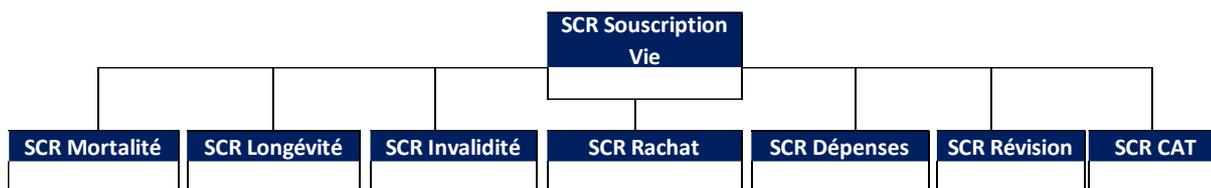
Avec

- La duration correspond à la duration moyenne des engagements nets au 31 /12/2020,
- Le SCR pour la marge pour risque SCR_{MR} correspond à l'agrégation des modules suivants :
 - SCR souscription (y compris catastrophe),
 - SCR de défaut,
 - SCR Opérationnel.

c) Capital de solvabilité requis Vie (SCR Vie)

Le SCR Vie est l'agrégation des sous modules suivants :

- Risque de mortalité,
- Risque de longévité,
- Risque de dépense,
- Risque de révision,
- Risque de rachat (Cessation),
- Risque de catastrophe.



L'agrégation des sous modules se fait avec la matrice de corrélation :

Matrice des corrélations	SCR Mortalité	SCR Longévité	SCR Invalidité	SCR Rachat	SCR Dépenses	SCR Révision	SCR CAT
SCR Mortalité	1,00	-0,25	0,25	0,00	0,25	0,00	0,25
SCR Longévité	-0,25	1,00	0,00	0,25	0,25	0,25	0,00
SCR Invalidité	0,25	0,00	1,00	0,00	0,50	0,00	0,25
SCR Rachat	0,00	0,25	0,00	1,00	0,50	0,00	0,25
SCR Dépenses	0,25	0,25	0,50	0,50	1,00	0,50	0,25
SCR Révision	0,00	0,25	0,00	0,00	0,50	1,00	0,00
SCR CAT	0,25	0,00	0,25	0,25	0,25	0,00	1,00

Les risques auxquels s'appliquent le SCR de souscription vie sont :

- Capitaux Décès ou Frais d'obsèques annuels,
- Vie entière,
- Maintien de la garantie décès en cas d'arrêt de travail,
- Rente de conjoint, et rente éducation.

Les chocs appliqués à la garantie vie entière sont :

- Augmentation de 15% du taux annuel de mortalité,
- Augmentation de 10% du taux de dépenses, et 1% par an au titre de l'inflation,
- Augmentation additive de 0,15% du taux de mortalité.

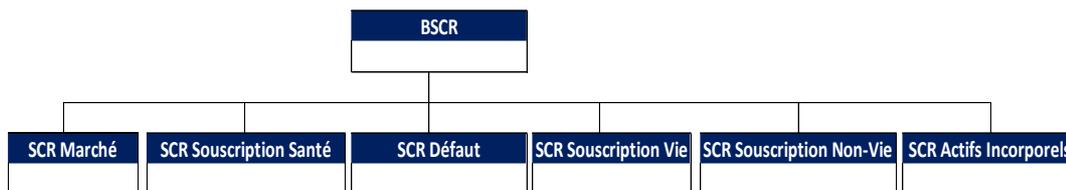
Pour le risque rachat « Cessation » les chocs suivant y sont appliqués :

- Un choc à la hausse du taux de rachat (+50% dans la limite de 100%)
- Un choc à la baisse du taux de rachat (-50% dans la limite de 20% en valeur absolue)
- Un choc dans le cas d'un rachat massif avec un taux de rachat égal à 30%.

Le choc de rachat correspond au maximum des trois chocs présentés ci-dessus.

d) Capital de solvabilité requis de base (BSCR)

Le BSCR est l'agrégation des SCR de modules (souscription vie/non vie/ santé, marché, contrepartie) à laquelle s'ajoute le SCR du module des actifs incorporels. Chaque module est lui-même l'agrégation de sous-modules.



Les agrégations des modules ou sous modules se font à l'aide des matrices de corrélation fixées par la directive (formule en racine carré).

Matrice des corrélations	SCR Marché	SCR Défaut	SCR Souscription Vie	SCR Souscription Santé	SCR Souscription Non-Vie
SCR Marché	1,00	0,25	0,25	0,25	0,25
SCR Défaut	0,25	1,00	0,25	0,25	0,50
SCR Souscription Vie	0,25	0,25	1,00	0,25	0,00
SCR Souscription Santé	0,25	0,25	0,25	1,00	0,00
SCR Souscription Non-Vie	0,25	0,50	0,00	0,00	1,00

e) *Résultat de Calcul*

Pour mesurer l'impact généré par les lois de rachat avec la méthode de la régression logistique, nous allons calculer des indicateurs comme les provisions techniques, le SCR global, et le ratio de solvabilité.

La marge pour risque représente 3,41% des Best Estimate pour le scénario sans rachat, ce taux est appliqué pour les autres scénarios. Les provisions et SCR sont détaillés en Annexe 2.

Méthode de Calcul	Provisions S2 (Best Estimate+ Marge pour risque)	SCR Global	Ratio de Solvabilité
Sans Rachat	62 453 730	15 718 637	284%
Avec Rachat Empirique	62 107 423	15 650 353	286%
Avec Rachat Modèle régression	61 911 514	15 592 266	288%

▪ Tableau n°37 : Comparaison des méthodes de calcul

- L'utilisation de la méthode empirique entraîne une baisse des provisions de 0,6% par rapport à la méthode sans rachat, elle est dûe à la baisse des prestations servies en cas sortie en rachat.

Nous constatons également une baisse du SCR global de 0,4%, et ainsi une amélioration du ratio de solvabilité de 2 points.

- La méthode de la régression logistique entraîne quant à elle une baisse des provisions de 0,9% par rapport à la méthode sans rachat, avec une baisse du SCR de 0,8%, et une amélioration du ratio de solvabilité de 4 points.

L'utilisation des taux de rachat issus du modèle de la régression logistique entraîne une baisse des provisions techniques et du SCR global de la mutuelle.

2) Test de sensibilité

Afin de mesurer l'impact d'une déviation des lois de rachat, nous allons faire varier les probabilités de rachat issues du modèle de la régression logistique, pour mesurer son impact sur les provisions S2, SCR global et le ratio de solvabilité.

Pour analyser la sensibilité des lois de rachat, nous appliquons les deux scénarios suivants :

- Une hausse de +5% des probabilités de rachat,
- Une hausse de +10% des probabilités de rachat.

Puis nous calculons pour chaque scénario les indicateurs cibles.

Le résultat est donné dans le tableau suivant :

<i>Modèle de régression logistique</i>	<i>Provisions S2 (Best Estimate+ Marge pour risque)</i>	<i>SCR Global</i>	<i>Ratio de Solvabilité</i>
<i>Central</i>	<i>61 911 514</i>	<i>15 592 266</i>	<i>288%</i>
<i>Sensibilité à 5%</i>	<i>61 836 707</i>	<i>15 550 316</i>	<i>289%</i>
<i>Sensibilité à 10%</i>	<i>61 763 776</i>	<i>15 507 875</i>	<i>290%</i>

- [Tableau n°38 : Test de sensibilité du modèle de la régression logistique](#)

Il en ressort les remarques suivantes :

- L'augmentation du taux de rachat de 5% conduit à une baisse des provisions de 0,12% et du SCR global de 0,27% par rapport au scénario central. Le ratio de solvabilité passe de 288% à 289%.
- Le Choc à 10% induit une baisse plus importante des provisions de 0,24%, et du SCR global de 0,54%, permettant ainsi à l'amélioration du ratio de solvabilité de 2 points.
-

En conclusion, la déviation à la hausse des lois de rachat induit une baisse des provisions et du SCR global de la mutuelle.

Conclusion générale

Dans un premier temps, le but de ce mémoire a été d'expliquer le comportement de rachat et la réduction, puis par la suite de calculer la probabilité de la survenance à travers l'utilisation d'un modèle mathématique. Ainsi l'explication du comportement de l'assuré passe d'abord par l'identification des facteurs susceptibles d'influencer le rachat et la réduction.

L'analyse de la liaison entre les différentes variables de la base de données, pour supprimer les variables corrélées entre elles a permis d'identifier et de retirer l'âge au rachat, le montant des cotisations et le montant de la provision.

Les statistiques descriptives ont permis de supprimer les variables telles que le réseau de distribution et la périodicité des cotisations.

Les méthodes de classification par arbre de décision (CART, CHAID, C4.5) ont permis d'identifier la même liste des variables explicatives, à savoir : l'ancienneté, le capital garanti et l'âge à la souscription.

Il découle de cette analyse que l'ancienneté est la variable la plus discriminante pour le comportement de rachat et la réduction.

La qualité du modèle de classification a été validée grâce aux techniques d'échantillon test, la validation croisée, et le critère de AUC de la courbe ROC.

L'application de l'analyse discriminante STEPDISC vient conforter les résultats de la classification par arbre de décision.

La régression logistique nous a permis de prendre en compte les facteurs présélectionnés, de quantifier la liaison entre ces facteurs et le comportement de rachat et la réduction.

Le modèle est jugé robuste grâce au calcul du taux d'erreur de classement sur l'échantillon test.

Le modèle ainsi construit, permet à la mutuelle de matérialiser le rachat et la réduction d'un individu au vu de ces caractéristiques propres, par un taux qui indique si le contrat sera racheté ou réduit dans l'année. Ainsi, si le taux est supérieur à 0,5 nous déduisons que le contrat sera racheté ou réduit, sinon le contrat restera actif.

L'outil permet aussi à la mutuelle de répondre aux exigences de l'ACPR pour le calcul du SCR de souscription, et la prise en compte des options de rachat et de réduction pour le calcul des provisions techniques en norme S2.

Enfin, le modèle de régression logistique permet de baisser le niveau des Best Estimate et le SCR global de la mutuelle.

Limites et Axes d'amélioration du mémoire

Le frein majeur qui constitue la limite principale des résultats du modèle, réside dans les données utilisées pour la construction du modèle :

- Taille du portefeuille : Les modèles d'arbre de décision, sont particulièrement instables, et très sensibles à des fluctuations des données, surtout pour des portefeuilles de taille faible.
- Instabilité des taux de rachat et la réduction pour les anciennetés au-delà de 14 ans, en raison de l'effectif très faible pour cette catégorie.
- Problématique de la qualité des données : Certains fichiers ont fait l'objet de traitements manuels et d'hypothèses prises pour les rendre exploitables, notamment pour les souscriptions antérieures à 2004.

Dans le modèle de la régression logistique, nous avons fait l'hypothèse que les variables quantitatives retenues, exercent une influence linéaire sur le comportement de rachat et réduction. or, on peut se poser la question si certaines variables sont bien linéaires ou existe-t-il un effet de seuil ? et dans ce cas-là, il convient de les discrétiser et transformer en variables qualitatives.

Dans cette étude comme dans la plupart des articles, les facteurs conjoncturels n'ont pas d'effet sur le comportement des assurés d'un contrat Vie entière en raison des montants faibles des capitaux garantis, qui sont peu sensibles aux fluctuations des marchés financiers, Néanmoins, on pourrait se demander si un contexte pandémique comme le Covid 19, où les taux de chômage et de l'inflation ne cessent d'augmenter, aurait un impact sur le comportement de rachat ou de réduction de nos assurés.

Bibliographie

✓ MEMOIRES D'ACTUARIATS

- Naoufal RAKAH (2012) « Modélisation des rachats dans les contrats d'épargne »
- Xavier Milhaud (2011) « Segmentation et modélisation des comportements de rachat en assurance vie »
- SOKHNA MAME BOUSSO MBAKE SAKHO « Analyse et modélisation des comportements de rachat d'un portefeuille d'épargne »
- Nicolas Ladias (2013) « Analyse des causes de rachats sur des contrats d'assurance_vie »
- William Nastro « Mise en œuvre et pilotage d'un produit Frais d'obsèques »
- Ophélie Berné (2013) « Estimation des lois de rachat et de mortalité en assurance vie »
- Adrien Suru (2011) « Le rachat : Modélisations et préconisations »

✓ Ouvrages

- Edith Bocquaire « Les grands principes de l'actuariat »
- Pierre Petauton en collaboration avec Michel Fromenteau « Théorie et pratique de l'assurance vie »

✓ Articles

- (2006) « Lignes directrices mortalités »
- Xavier Milhaud- Marie-Pierre Gonon –Stéphane Loisel « Les comportements de rachat en assurance vie en régime de croisière et en période de crise »

Annexe 1

- <https://www.ffa-assurance.fr/etudes-et-chiffres-cles/les-contrats-assurance-obseques-adhesion-individuelle-en-2017>
- [https://www.ffa-assurance.fr/etudes-et-chiffres-cles?f%5B0%5D=field categorie chiffre cle%253Aparents all%3A28](https://www.ffa-assurance.fr/etudes-et-chiffres-cles?f%5B0%5D=field%20categorie%20chiffre%20cle%253Aparents%20all%3A28)
- https://www.xerfi.com/presentationetude/Comment-capter-la-croissance-du-marche-de-l-assurance-obseques_21ABF56
- [http://eric.univ-lyon2.fr/~ricco/cours/slides/Arbres de decision Introduction.pdf](http://eric.univ-lyon2.fr/~ricco/cours/slides/Arbres%20de%20decision%20Introduction.pdf)
- http://eric.univ-lyon2.fr/~ricco/cours/cours_regression_logistique.html

Annexe 2

Détail Provisions Techniques	Scénario sans rachat	Scénario avec Modèle empirique	Scénario avec Modèle de régression	Scénario avec Sensibilité à 5%	Scénario avec Sensibilité à 10%
Best Estimate	60 319 905	59 985 429	59 796 214	59 723 963	59 653 524
Marge pour risque	2 133 826	2 121 993	2 115 300	2 112 744	2 110 252
Provisions techniques	62 453 730	62 107 423	61 911 514	61 836 707	61 763 776

Détail SCR	Scénario sans rachat	Scénario avec Modèle empirique	Scénario avec Modèle de régression	Scénario avec Sensibilité à 5%	Scénario avec Sensibilité à 10%
SCR	15 718 637	15 650 353	15 592 266	15 550 316	15 507 875
BSCR	12 708 021	12 695 338	12 673 195	12 649 532	12 625 276
SCR Marché	2 725 526	2 725 526	2 725 526	2 725 526	2 725 526
SCR Santé	7 222 732	7 222 732	7 222 732	7 222 732	7 222 732
SCR Vie	535 681	501 583	482 294	474 928	467 747
SCR défaut	7 073 496	7 073 496	7 073 496	7 073 496	7 073 496
SCR Opérationnel	3 812 406	3 808 601	3 801 959	3 794 860	3 787 583
Impôts différés	- 801 790	- 853 586	- 882 887	- 894 076	- 904 984

Fonds propres S2	44 672 503	44 805 693	44 910 341	44 909 810	44 937 860
------------------	------------	------------	------------	------------	------------

Annexe 3

Programmes SAS utilisés :

```
graphics on;
proc hpsplit data=work.nvfichier;
  class RACH ADH SEXE PERIO RESEAU;
  model RACH(event='1') = RACH ADH SEXE PERIO RESEAU AGEINIT ANC COTIS PM
CAPITAL NAISS ;
  prune costcomplexity;
  partition fraction(validate=0.20);
RUN;
% CART%
proc hpsplit data=work.nvfichier cvmodelfit seed=123;
  class RACH ADH SEXE PERIO RESEAU;
  model RACH (event='1') =
    RACH ADH SEXE PERIO RESEAU AGEINIT ANC COTIS PM CAPITAL NAISS ;
  grow GINI;
  prune costcomplexity;
run;
% C4.5%
proc hpsplit data=work.nvfichier cvmodelfit seed=123;
  class RACH ADH SEXE PERIO RESEAU;
  model RACH (event='1') =
    RACH ADH SEXE PERIO RESEAU AGEINIT ANC COTIS PM CAPITAL NAISS ;
  grow entropy;
  prune costcomplexity;
run;
% C4.5%
proc hpsplit data=work.nvfichier cvmodelfit seed=123;
  class RACH ADH SEXE PERIO RESEAU;
  model RACH (event='1') =
    RACH ADH SEXE PERIO RESEAU AGEINIT ANC COTIS PM CAPITAL NAISS ;
  grow CHAID;
  prune costcomplexity;
run;

%let _stdvar = ;
proc stepdisc data=work.nvfichier bsscp tsscp
METHOD=BACKWARD;
  class RACH;
  var AGEINIT ANC COTIS PM CAPITAL NAISS ;
run;
%let _stdvar = ;
proc stepdisc data=work.nvfichier bsscp tsscp
METHOD=FORWARD;
  class RACH;
  var AGEINIT ANC COTIS PM CAPITAL NAISS ;
run;
proc discrim data=work.nvfichier;
  class RACH;
```

```

var &_stdvar;
run;
title 'Résultats de la Selection STEPWISE';

%let _stdvar = ;
proc stepdisc method=stepwise data=work.donnees_rachat1 bsscp tsscp;
class y;
var pm Capital Année_Naiss Anc Age_init ;
run;

proc logistic data=work.rachat_valid desc;
model Y=Anc Age_init PM Capital/link=logit;

output out=sortie ;
run;
quit;
proc logistic data=work.réduc2 desc;
model Y=Anc Age_init PM Capital/link=logit;

output out=out2 predicted=pcalcule lower=inf95 upper=sup95 ;
run;
proc print data=out2;
run;
proc logistic data=work.rachatvf desc;
model Y=Anc Age_init PM Capital/link=logit;

output out=out predprobs=I ;
run;
proc print data=out;
run;

```