



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

le 26 septembre 2013

Par : Mehdi BOUEDDINE

Titre : Assurance Automobile : Analyse de l'impact d'une variation du tarif sur le comportement des assurés lors de l'acte de souscription et de résiliation

Confidentialité : oui, deux ans. *Les signataires s'engagent à respecter la confidentialité indiquée.*

**Membre présent du jury de l'Institut
des Actuaire**

Signature :

Entreprise

Signature :

Membres présents du jury de l'EURIA

Directeur de mémoire en entreprise

Signature :

Invité

Signature :

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

Signature du responsable entreprise

Signature du candidat

Bibliothèque :

Secrétariat :

Table des matières

Résumé	3
Abstract	4
Remerciements	5
Introduction	6
I Contexte	7
1 Marché de l'assurance automobile en France	8
2 Principes de tarification	11
3 Description de l'étude	15
3.1 L'acte de concrétisation d'un devis et l'acte de résiliation d'un contrat	15
3.2 Notions d'élasticité	15
3.3 Périmètre et Démarche	16
II Éléments théoriques	18
4 Modèle linéaire classique	19
4.1 Définition et exemples	19
4.2 Estimation des paramètres du modèle	21
4.3 Limites	23
5 Modèles linéaires généralisés	24
5.1 Définition d'un modèle GLM	24
5.2 Exemple d'un GLM : la Régression logistique	28
5.2.1 Le modèle LOGIT	28
5.2.2 Estimation des coefficients	29
5.2.3 Significativité du modèle et des variables	30
5.2.4 Validation du modèle	32
5.3 Sélection des variables	33
III Présentation et étude préliminaire du portefeuille	35
6 Le produit «Assurance Auto»	36
6.1 Le contrat	36
6.2 Les garanties	37
6.2.1 Description des garanties	37

6.2.2	Des garanties limitées	39
7	Construction des bases de données	40
7.1	Base «Concrétisation» et «Résiliation»	40
7.2	Présentation des variables	43
8	Analyse Descriptive	47
8.1	Profils des individus	47
8.1.1	La concrétisation	47
8.1.2	La résiliation	49
8.2	Corrélations et Mesures d'association	51
IV	Résultats de la modélisation	55
9	L'acte de concrétisation	56
9.1	Introduction de la concurrence	56
9.2	Modélisation du taux de concrétisation d'un devis	64
9.3	Élasticité au prix	68
10	L'acte de résiliation	75
10.1	Modélisation du taux de résiliation en fonction du niveau de prime	75
10.2	Calcul de l'élasticité pour le modèle 1	77
10.3	Modélisation du taux de résiliation en fonction de la variation de prime	78
10.4	Calcul de l'élasticité pour le modèle 2	81
10.5	Autres approches sur l'étude de l'élasticité-prix	85
	Conclusion	93
	Bibliographie	95
	Lexique	97
V	Annexes	99
	Synthèse	124
	Synthesis	130

Résumé

Le marché de l'assurance automobile est un marché pilier de l'assurance «Incendie, Accidents et Risques Divers» (IARD) et il devient de plus en plus concurrentiel, ainsi une tarification précise des primes demandées aux assurés se présente comme un enjeu majeur pour chaque assureur.

La sensibilité au prix, variant fortement d'un client à l'autre, a un impact sur les taux de souscription et de résiliation d'un contrat ainsi que sur la rentabilité d'un portefeuille. Le but de l'étude est de prendre en compte cette sensibilité des clients dans la tarification du produit d'assurance Auto d'Aviva Assurances. Il s'agit aussi de modéliser l'impact du tarif sur le comportement de l'assuré.

Ainsi, le mémoire se présente de la manière suivante :

Une première partie décrit le marché de l'assurance automobile français, des principes de tarification et de segmentation pratiqués dans cette branche, ainsi qu'une description préparative des méthodes nécessaires pour atteindre notre objectif.

Dans une deuxième partie, nous présentons quelques éléments théoriques dont nous avons eu besoin lors de nos modélisations en détaillant l'explication des modèles linéaires généralisés, avec une attention particulière sur la régression logistique.

Une troisième partie est dédiée à la description des bases de données : la première consacrée à l'acte de concrétisation d'un devis et la deuxième à l'acte de résiliation d'un contrat de la part de l'assuré. De plus, des statistiques descriptives permettent de déterminer les profils d'individus présents dans nos bases et les critères ayant un lien significatif sur le taux de concrétisation et le taux de résiliation.

Une dernière partie est consacrée à l'application des éléments théoriques sur les données et à la présentation des résultats des modèles mis en place. Nous cherchons à estimer, dans un premier temps, la probabilité de concrétiser un devis, puis la probabilité qu'un assuré résilie son contrat en fonction de critères concernant l'assuré, son véhicule et particulièrement la prime payée. Nous finissons cette partie par le calcul d'un indicateur mesurant la sensibilité d'un assuré au prix.

Mots clés : *Assurance automobile, tarification, segmentation, sensibilité au prix, prime commerciale, taux de concrétisation, taux de résiliation, modèles linéaires généralisés.*

Abstract

Car insurance is a pillar of the French property and casualty insurance market. As becoming increasingly competitive, accurate pricing of premiums charged to policyholders is a major issue for each insurer.

Price sensitivity is high regarding car insurance. This characteristic, which varies greatly from one customer to another, has an impact on the subscription rate and the termination rate of a contract and the return of a portfolio. The purpose of the study is to take into account the price sensitivity of customers in the pricing of Aviva Insurance insurance product, then to model the impact of the tariff on the behavior of the insured.

Thus, this paper is structured as follows :

The first part describes the French car insurance market, pricing and segmentation principles performed in this branch, and the preparative study to know the structure and the process needed to achieve our purpose.

In a second part, theoretical elements needed for the modeling are presented, with detailed explanation of generalized linear models, especially the logistic regression.

A third part is devoted to the description of our databases : the first dedicated to the act of subscription and the second to the act of termination of a contract on the part of the insured. In addition, through descriptive statistics, we can determine the profiles of individuals in our bases and criteria with a significant link to the subscription rate and the cancellation rate.

The last part applies theoretical elements to the data and describes the modelling and results. We seek to estimate the probability of subscribing, and the probability that an insured cancels the contract according to criteria concerning the insured's vehicle and particularly the premium paid. To end this section, we calculate an indicator that measures the price sensitivity of an insured.

Key words : *Car pricing, pricing, segmentation, price sensitivity, market premium, subscription rate, cancellation rate, generalized linear models.*

Remerciements

Je tiens tout d'abord à remercier mon maître de stage M. Alexandre MURATET, responsable du service actuariat auto et ma tutrice pédagogique Mlle. Sophie BOURDET, pour leur suivi et leurs conseils tout au long de l'élaboration de ce mémoire.

Je souhaite remercier également l'ensemble du plateau Dommages Particuliers et Actuariat, pour leur accueil et leur aide, notamment :

- Mlle. Sophie GARCIN
- M. Maher MAITI
- M. Sébastien PENVERNE

Je n'oublie pas de remercier aussi mes proches pour leur soutien lors de la rédaction du mémoire.

Enfin, je remercie mes professeurs, les membres du collège de direction et Monsieur Hervé LE BORGNE, ancien professeur et fondateur de l'EURIA, pour leur contribution à ma formation d'actuariat à Brest.

En vous souhaitant bonne lecture.

Introduction

Le marché de l'assurance automobile est le secteur rapportant le plus grand chiffre d'affaires en assurances dommages. Près de 19 milliards d'euros de cotisation en 2011¹, ont été perçus par les assureurs. L'obligation d'assurance est tout d'abord à signaler, de même que l'importance du parc automobile français, car ceci porte alors à un niveau conséquent le nombre d'assurés potentiels. Malgré l'importance de ce marché, il est très compétitif si bien qu'il n'est pas aisé pour un assureur de se développer et d'attirer la clientèle tout en conservant ses assurés d'origine.

Dans ce contexte économique, marqué par une forte pression concurrentielle, la sensibilité au prix des assurés semble être une information décisive pour une compagnie d'assurance afin d'ajuster au mieux ses tarifs. Un tarif trop élevé par rapport à la concurrence sera un obstacle à la souscription ou bien une incitation à la résiliation. Le comportement des consommateurs est d'autant plus difficile à prévoir qu'il varie d'un individu à un autre selon de nombreux critères.

L'évolution de la taille ou de la structure d'un portefeuille et les conséquences, à plus ou moins long terme, engendrées par une modification de ses tarifs, semblent difficiles à maîtriser même si la majorité des risques techniques et financiers sont globalement bien identifiés et contrôlés par les compagnies d'assurance.

L'objectif de ce mémoire est d'étudier la sensibilité au prix des assurés pour le produit « Assurance Automobile » d'Aviva Assurances. Il s'agit de modéliser l'impact du tarif sur le comportement de l'assuré, en ce qui concerne l'acte de souscription et de résiliation, étapes primordiales dans la vie d'un contrat. Pour cela, nous utilisons un indicateur mesurant la variation du taux de souscription ou de résiliation en fonction de la variation du prix. Cet outil de mesure, couramment appelé élasticité, pourra ensuite être utilisé comme aide à la décision ou être directement intégré dans le processus de tarification des primes.

1. D'après les chiffres de la Fédération Française des Sociétés d'Assurances (FFSA).

Première partie

Contexte

Chapitre 1

Marché de l'assurance automobile en France

L'assurance de biens et de responsabilité représente près de 25,2% des cotisations de l'assurance française en 2011 (soit un montant de 47,9 milliards d'euros). Avec une progression de +4% (contre +2,2 % en 2010), la croissance des assurances de biens et de responsabilité atteint un niveau record depuis 2004. En effet, ce fort montant de cotisation est dû principalement à la forte présence de l'assurance automobile, devenu le premier marché de l'assurance de biens et de responsabilité. Comme en témoignent les chiffres de la Fédération Française des Sociétés d'Assurances (FFSA), les cotisations de l'assurance automobile s'élèvent à 19 milliards d'euros en 2011, avec une croissance de 4% par rapport à 2010. Ce marché se place juste devant l'assurance multirisque habitation (MRH) qui totalise 8,4 milliards d'euros de cotisations en 2011 [cf Table 1.1].

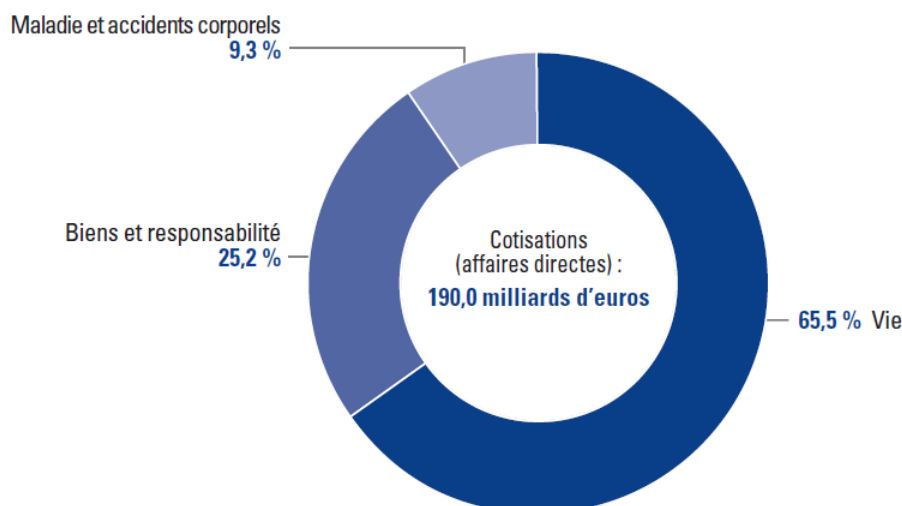


FIGURE 1.1 – Assurance française en 2011 (*Source FFSA*)

Affaires directes (en Mds d'euros)	2010	2011	Variation
Automobile	18,3	19,0	4%
Multirisques habitation	7,9	8,4	6%
Dommages aux biens des professionnels	5,9	6,0	2%
Responsabilité civile générale	3,4	3,5	3%
Construction	2,3	2,4	4%
Catastrophes naturelles	1,4	1,4	-
Transports	1,0	1,0	-
Dommages aux biens agricoles	1,0	1,1	10%
Crédit-caution, PJ, PP, Assistance	4,8	5,1	6%
Total	46,0	47,9	4%

TABLE 1.1 – Montants et variation des cotisations des différents secteurs de l'assurance dommages (*Source FFSA*)

L'usage fréquent (ou quotidien) des véhicules automobiles par la population implique la présence d'un risque « automobile » contre lequel on doit se couvrir. De plus, l'assurance automobile est rendue obligatoire en France par la loi du 27 février 1958, aujourd'hui codifiée au titre 1, livre 2 partie réglementaire du code des assurances. L'obligation d'assurance ne concerne que la garantie « Responsabilité civile », c'est-à-dire les dommages causés aux tiers et aux passagers. C'est pourquoi le marché de l'assurance automobile est devenu vital pour la pérennité d'une compagnie d'assurance (IARD).

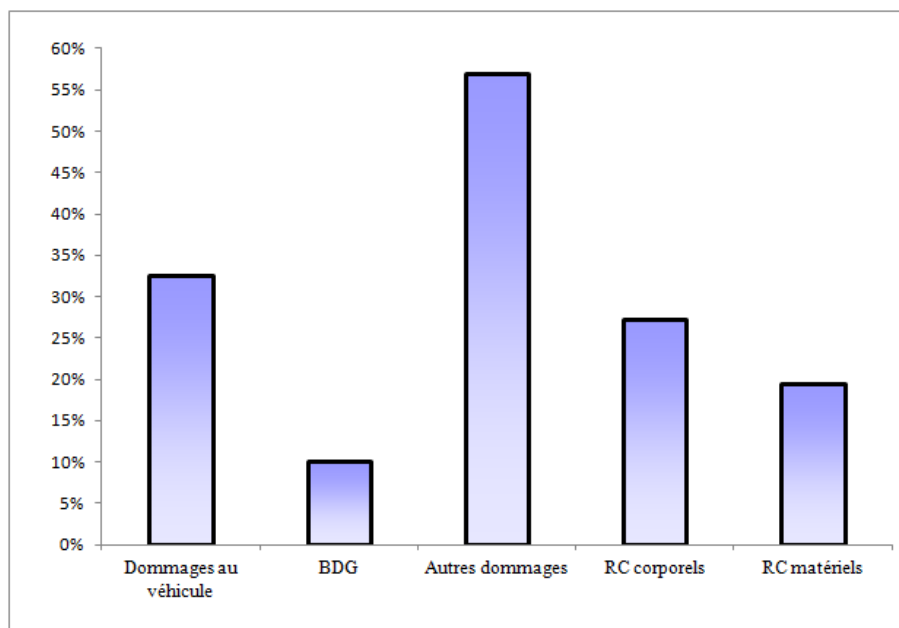


FIGURE 1.2 – Répartition de la charge totale selon le type de sinistre Auto en 2011 (*Source FFSA - GEMA*)

Si les montants de cotisations en assurance automobile peuvent paraître élevés, les coûts des sinistres le sont aussi. En 2011, les assureurs ont enregistré près de 9 millions de sinistres et un montant total de 14,5 milliards d'euros. D'après la FFSA, le coût moyen d'un sinistre automobile est de 1.728 euros et monte jusqu'à 15.000 euros en cas de sinistre grave avec dommages corporels. En effet, nous constatons sur la figure 1.2 que les sinistres de type «RC corporels» (Responsabilité civile corporels) au nombre de 270.000

représentent 27% de la charge totale. Ce type de sinistres est le moins fréquent mais coûte cher aux assureurs. Les bris de glaces, au nombre de 3 millions et 10% de la charge totale, sont les sinistres les plus fréquents mais les moins coûteux.

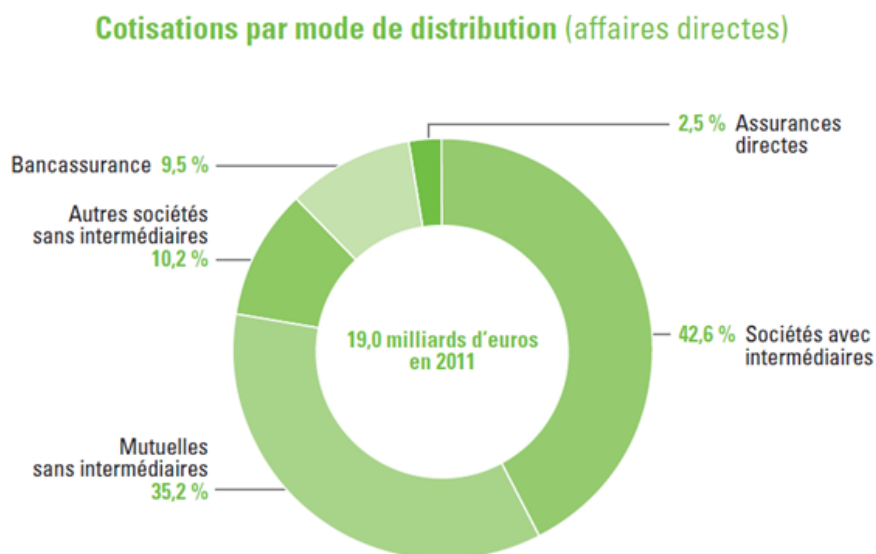


FIGURE 1.3 – Cotisations perçues par les principaux acteurs du marché automobile (*Source FFSA*)

L'importance de cette branche entraîne la présence de différents acteurs qui se livrent une guerre de prix pour gagner de nombreuses parts de marché. En effet, ce marché est convoité par : les sociétés d'assurance, les mutuelles et les bancassureurs, dont certaines traitent avec leurs clients sans intermédiaires, tandis que d'autres s'appuient sur les agents généraux ou des courtiers. Les sociétés avec intermédiaires détiennent à 42,6% les cotisations perçues en 2011, devant les mutuelles (à 35,2%). Ainsi, le marché automobile est un marché très concurrentiel au sein duquel les assureurs doivent mettre à jour et innover leur tarification s'ils veulent préserver leurs marges et ne pas perdre leurs clients. Cette concurrence s'alourdit avec la présence des comparateurs d'assurances, qui évaluent ces différents acteurs en fonction de plusieurs critères (les garanties, le tarif, la qualité des prestations fournies, etc), ce qui influe fortement sur le choix du client.

En 2012, le chiffre d'affaires de l'assurance automobile a augmenté d'environ 3%. Cette augmentation s'explique par une croissance organique liée à l'évolution du parc automobile, et également par l'amélioration du taux d'équipement des assurés en matière de garanties.

Chapitre 2

Principes de tarification

Très schématiquement, on peut voir l'assurance comme la gestion de deux flux financiers :

- les primes qui vont des assurés à l'assureur ;
- le coût des sinistres qui vont en sens inverse.

La tarification consiste à estimer le premier flux pour qu'il compense le second ; c'est à dire qu'il s'agit là d'une activité cruciale de l'assureur. Les assurés payent une prime commerciale définie de la manière suivante :

$$\text{Prime commerciale} = \text{Prime Pure} + \text{Chargements} - \text{Produits financiers}$$

La prime pure représente le coût du risque, c'est le montant de prime payé par l'assuré qui est nécessaire pour payer son sinistre. Les chargements se composent de plusieurs frais : les frais de gestion, de commercialisation, de personnel, etc. Les produits financiers viennent en diminution de la prime commerciale grâce aux placements effectués sur les marchés. L'assureur est en mesure de déterminer le montant de chargements et des produits financiers.

La prime pure peut se calculer selon deux paramètres : la Fréquence et le Coût moyen.

$$\text{Prime Pure} = \text{Fréquence} \times \text{Coût moyen}$$

La fréquence (annualisée) est le nombre de sinistres divisé par l'exposition (correspondant au nombre d'années) pour une police d'assurance, ou un groupe de polices d'assurance. La plupart des contrats étant annuels, on ramènera toujours le nombre de sinistres à une exposition annuelle lors du calcul de la prime, et on notera N la variable aléatoire associée. Durant la période d'exposition, on notera Y_i les coûts des sinistres, c'est à dire les indemnités versées par l'assureur à l'assuré (ou une tierce personne). La charge totale par police est alors $S = 0$ s'il n'y a pas eu de sinistres, ou sinon :

$$S = Y_1 + \dots + Y_N = \sum_{i=1}^N Y_i$$

Classiquement, les sinistres sont de coût positif, $Y_i > 0$, et N est alors le nombre de sinistres en excluant les sinistres classés sans suite (i.e. de coût nul).

Donc, la prime pure est $E(S) = E(N).E(Y_i)$ dès lors que les coûts individuels sont indépendants et identiquement distribués, et supposés indépendants du nombre de sinistres.

Actuellement, la tarification est souvent liée à la notion de «Segmentation» qui consiste à réaliser une partition de la population à assurer en sous-populations - appelées classes tarifaires ou segments - auxquelles sont attachées des primes différentes. Cette notion est contraire à la «Mutualisation» qui représente la répartition du risque à égalité parmi les membres d'un groupe ; on considère tous les sinistres du portefeuille et on calcule la prime pure globale. Ainsi, on ne tient pas compte du niveau de risque différent entre individus, la prime pure est le reflet du risque global. La mutualisation du risque est dangereuse pour un assureur car il y a un effet d'antisélection : certains individus ne payent pas la prime réelle de leur risque. Prenons un exemple simple (particulièrement en assurance automobile) permettant d'illustrer la nécessité de segmenter pour un assureur :

Un assureur A choisit de mutualiser son risque et un assureur B choisit de segmenter son risque. L'assureur A fait donc un profit avec les clients à faible probabilité de sinistre et une perte avec les assurés à forte probabilité de sinistre. L'assureur B décide de tarifer plus cher les conducteurs à risque, déterminés à l'aide de critères telles que l'âge ou les antécédents de sinistres, et de faire payer moins cher une prime aux assurés à risque faible. Au fur et à mesure de leur activité, comme l'assureur A propose un tarif unique, il verra venir chez lui beaucoup de conducteurs avec une forte probabilité de sinistre et ainsi la rentabilité de son portefeuille sera mauvaise et il fera surement beaucoup de pertes. Cependant, l'assureur B quant à lui, verra rentrer dans son portefeuille la plupart des «bons» conducteurs car son tarif est plus attractif que celui de l'assureur A. Ainsi, l'assureur B fera beaucoup de profits et son portefeuille aura une bonne rentabilité. L'assureur B aura su **maîtriser le risque** de son portefeuille.

Ainsi, les assureurs calculent la prime en fonction de variables, appelées **critères tarifaires**, pour chaque police d'assurance. Dans ce cas, les fréquences et les charges sont hétérogènes, l'hétérogénéité étant caractérisée par une information ω , la prime pure devrait être :

$$E(S|\omega) = E(N|\omega).E(Y_i|\omega)$$

Le facteur d'hétérogénéité ω étant inconnu, on utilise les critères tarifaires à notre disposition pour obtenir un proxy de ces espérances conditionnelles. On cherche alors $X = (X_1, ..., X_k)$ un ensemble de variables explicatives tels que :

$$E(S|X) = E(N|X).E(Y_i|X)$$

Critères tarifaires

En assurance automobile, les critères tarifaires utilisés par les assureurs peuvent être classés en trois catégories :

1. les critères propres au véhicule : la puissance, la classe de prix (les réparations d'une voiture chère sont plus onéreuses), l'ancienneté, etc.
2. les critères propres à l'assuré : l'âge, l'ancienneté de permis (les conducteurs dont le permis est récent, en particuliers les jeunes, provoquent plus d'accidents que la

moyenne des conducteurs), la situation matrimoniale, les antécédents de permis (suspension, annulation ou conduite en état d'ivresse), etc.

3. des informations sur la circulation : la zone géographique (les conducteurs qui circulent habituellement dans des zones à faible concentration urbaine ont une probabilité plus faible d'avoir un accident que les autres), l'usage du véhicule (les automobilistes qui utilisent leur voiture pour les besoins de leur profession ont une probabilité plus élevée d'avoir un accident que ceux qui s'en servent uniquement pour la promenade et les déplacements privés), etc.

Tous ces critères de tarification résultent d'analyses statistiques de données historisées provenant de constats de la vie courante des conducteurs. Les assureurs ont pu mettre en évidence certains profils d'assuré et de véhicule, qui disposent d'une fréquence et d'un coût moyen faible ou élevé de sinistres. Aujourd'hui, l'outil principal en tarification est le modèle linéaire généralisé, appelé GLM (Generalized Linear Model).

Chaque assureur est libre de choisir la tarification qu'il souhaite effectuer sur son portefeuille, cependant certains éléments sont fixés ou interdits par la législation. Par exemple, à la suite d'une décision de la Cour de justice de l'Union européenne depuis le 21 décembre 2012, les assureurs ne peuvent plus calculer leur tarif en fonction du sexe de l'assuré.

Le coefficient Réduction-Majoration

Un élément important de la tarification est le coefficient de réduction-majoration (CRM), appelé également coefficient Bonus-Malus. Il s'agit d'un coefficient appliqué à la prime d'assurance à chaque échéance annuelle : la prime est réduite si vous n'avez pas de sinistre responsable, ou au contraire, majorée en fonction du nombre de sinistres enregistrés. Ce système de bonus-malus, pour récompenser «les bons conducteurs», est réglementé par le Code des assurances¹. Donnons un exemple simpliste de ce mécanisme :

Exemple : Si un assuré dispose d'un CRM de 80% et d'une prime pure de 100 euros, sa prime sera de 80 euros. S'il n'a pas d'accident, l'année suivante, il a un bonus (soit 5% de réduction) et son CRM évolue à 76% [cf Table 2.1] et sa nouvelle prime est de 79,80 ($80 \times 0,95$) euros².

Afin de mieux comprendre cette notion de CRM, nous allons maintenant décrire le calcul de ce coefficient :

Lorsqu'on souscrit une assurance auto pour la première fois, le coefficient de départ est de 1. Chaque année, il évolue en fonction des événements déclarés à l'assureur. Le calcul se fait sur la période des 12 mois qui précèdent de 2 mois l'échéance annuelle. Un accident survenu dans les deux derniers mois avant l'échéance sera donc pris en compte dans le calcul de l'année suivante. Son calcul résulte de deux possibilités :

1) L'assuré n'a pas déclaré d'accident responsable : il bénéficie d'un bonus.

Le coefficient se calcule en multipliant le dernier coefficient par 0,95. Il faut attendre 13 ans pour atteindre le coefficient maximal de 0,50.

1. Clause de réduction-majoration : Article A 121-1 du code des assurances.

2. Ce calcul ne tient pas compte des frais, des taxes et d'une éventuelle revalorisation de la prime à la date d'anniversaire du contrat.

Années d'assurance (sans accident responsable)	Coefficient	Bonus/malus
1 an	0,95	5%
2 ans	0,90	10%
3 ans	0,85	15%
4 ans	0,80	20%
5 ans	0,76	24%
6 ans	0,72	28%
7 ans	0,68	32%
8 ans	0,64	36%
9 ans	0,60	40%
10 ans	0,57	43%
11 ans	0,54	46%
12 ans	0,51	49%
13 ans	0,50	50%

TABLE 2.1 – Évolution du coefficient Bonus-Malus sans accident responsable

2) L'assuré a déclaré un ou plusieurs accidents responsables : il subit un malus.

Son coefficient sera majoré en fonction de sa responsabilité dans l'accident. Son nouveau coefficient est calculé en multipliant le dernier bonus malus :

- par 1,125 pour un accident partiellement responsable ;
- ou par 1,25 pour un accident totalement responsable.

Si l'assuré a eu plusieurs accidents responsables, la majoration est multiplicative et non additionnelle. Par exemple, deux accidents responsables la première année entraînent un coefficient de 1,56 la deuxième année ($1 \times 1,25 \times 1,25$) et non de 1,50 ($1 \times (1,25 + 1,25)$). L'assuré subit également un malus si un autre conducteur était au volant, par exemple, si l'assuré prête sa voiture à son enfant : c'est le véhicule responsable de l'accident qui est pris en compte.

Cas particuliers dans le calcul du bonus-malus

- Le malus est limité : le malus maximal est de 250%, soit un coefficient de 3,50 ; au bout de 2 ans sans accident responsable, le coefficient est réduit à 1 s'il était supérieur. Si le conducteur dispose d'un coefficient de 0,50 depuis au moins 3 ans, le premier accident (partiellement ou totalement) responsable n'entraîne pas de malus.
- Un malus supplémentaire est appliqué en cas de circonstances aggravantes : en cas d'accident responsable en état d'imprégnation alcoolique, le CRM est majoré de 150%. En cas de délit de fuite et de fausse déclaration, le CRM est majoré de 100%. En cas d'accident donnant suite à une suspension de permis, le CRM est majoré de 50 à 150%.

Le CRM est transférable d'un assureur à un autre, il est propre à l'assuré en cas de changement d'assureur ou de changement de véhicule, mais il est assigné au véhicule durant la vie du contrat. Le CRM apparaît donc comme un moyen efficace pour inciter les automobilistes à un maximum de prudence.

Chapitre 3

Description de l'étude

3.1 L'acte de concrétisation d'un devis et l'acte de résiliation d'un contrat

Il semble nécessaire d'abord de présenter **l'acte de concrétisation** (ou **de souscription**) d'un devis et **de résiliation** d'un contrat. Ces deux grandes étapes constitueront l'environnement sur lequel nous allons effectuer notre étude.

Le prospect¹ se présente à une agence de la compagnie. L'agent général établit un devis en fonction des caractéristiques du prospect et de son véhicule (l'âge, l'ancienneté du véhicule,...). Ensuite, le prospect décide, selon les garanties et le prix proposés, s'il désire ou non souscrire son contrat d'assurance avec cette compagnie : c'est ce qu'on appelle l'acte de concrétisation. Le **taux de concrétisation** d'un devis, ou taux de transformation, estime la probabilité avec laquelle le prospect devient un client de la compagnie.

L'acte de résiliation, signant la fin de vie d'un contrat, peut s'illustrer de deux façons :

- Renouvellement du contrat : cette étape intervient à la date d'échéance (qui coïncide dans la majorité des cas à l'anniversaire du contrat). En France, en assurance automobile, un contrat est soumis à la *tacite reconduction* : le contrat est automatiquement renouvelé d'une année à l'autre par l'assureur et ce, pour une durée d'un an. Lors d'une éventuelle revalorisation des primes à la date d'échéance : l'assuré peut alors décider de résilier ou de renouveler son contrat.
- Résiliation par la compagnie : Cette résiliation peut aussi avoir lieu suite à un excès de sinistres à l'échéance anniversaire, ou à un non-paiement des primes. La compagnie d'assurance peut également résilier le contrat d'assurance pour fausse déclaration de l'assuré à la souscription, ou de sortie des critères de souscription. Le **taux de renouvellement** donne la proportion de polices d'assurance renouvelées, et à l'inverse, le **taux de résiliation** est la proportion de polices résiliées sur l'ensemble du portefeuille à la fin d'une période donnée. Il existe des cas où la résiliation est effectuée «hors échéance», tels que la perte ou le changement du véhicule, à une date durant l'année de vie du contrat.

3.2 Notions d'élasticité

L'élasticité est une notion souvent utilisée en économie générale ou en microéconomie de l'assurance. En effet, «l'élasticité de la demande» est égale à la variation relative de la

1. Voir Lexique page 97.

demande du bien en fonction de l'augmentation relative du prix. Elle est calculée suivant la formule suivante :

$$\epsilon = \frac{\Delta d/d}{\Delta p/p}$$

où $\frac{\Delta d}{d}$ représente la variation relative de la demande et $\frac{\Delta p}{p}$ représente l'augmentation relative du prix.

Dans notre étude, pour mesurer la sensibilité des assurés au prix, nous utiliserons cette notion en cherchant à calculer l'élasticité du taux de résiliation, ou de concrétisation des devis. Elle correspond donc à la variation relative de ce taux lorsque le prix varie d'une quantité infinitésimale. Son analyse a pour but de déterminer quel impact a une légère modification du prix sur le taux de résiliation, ou de concrétisation. Ainsi, lorsque le prix vaut p , nous pouvons définir l'élasticité $\epsilon(p)$ de la manière suivante :

$$\epsilon(p) = \frac{\partial T(p)/T(p)}{\partial p/p}$$

où $T(p)$ est le taux de résiliation ou de concrétisation de devis en fonction du prix.

Si l'on considère 2 dates t_1 et t_2 , entre lesquelles une variation de tarif a eu lieu, l'élasticité peut s'écrire comme la limite lorsque t_2 tend vers t_1 de :

$$\epsilon(p_{t_1}) = \frac{(T_{t_2} - T_{t_1})/T_{t_1}}{(p_{t_2} - p_{t_1})/p_{t_1}}$$

où $T_{t_i} = T(p_{t_i})$.

Nous remarquons dès à présent que $\epsilon(p)$ sera négative en ce qui concerne l'acte de concrétisation, c'est à dire que le taux de concrétisation diminuera lorsque le prix augmente. Inversement, en ce qui concerne les résiliations, $\epsilon(p)$ sera positive c'est à dire que le taux de résiliation augmente lorsque le prix augmente .

Afin de calculer cette élasticité, nous cherchons donc à expliciter le taux de résiliation et de concrétisation en fonction de variables explicatives, et plus particulièrement du prix payé par l'assuré.

3.3 Périmètre et Démarche

Tout d'abord, il est nécessaire de définir le périmètre sur lequel nous allons travailler : Cette étude s'appuie sur des bases de données des sociétés Aviva Assurances sur le marché de l'Assurance Automobile de Particuliers. Nous utiliserons principalement les données relatives à l'année 2012. En ce qui concerne l'acte de résiliation, nous nous intéresserons aux contrats résiliés à la date d'échéance par l'initiative de l'assuré.

Pour atteindre nos objectifs définis en introduction, nous utiliserons la démarche suivante :

En premier lieu, nous constituerons deux bases de données rassemblant les informations nécessaires pour notre étude : la première pour traiter l'acte de concrétisation et la

deuxième pour traiter l'acte de résiliation. Ces deux bases respecteront le périmètre défini précédemment en prenant soin d'éviter l'introduction de biais liés à la structure même des bases.

Dans un second temps, à l'aide d'une analyse descriptive, nous étudierons les variables de nos bases, ainsi que les taux de concrétisation et de résiliation en fonction de ces variables. Nous pourrons mettre en exergue les individus qui ont tendance à concrétiser un devis et ceux qui ont tendance à résilier leur contrat. Cette étape aura pour objectif de donner une première idée des variables qui entreront dans nos modèles. Nous aurons tendance à retenir les variables ayant les plus forts pouvoirs discriminants sur ces taux et abandonner celles en fonction desquelles les taux sont relativement constants.

Dans une troisième partie, nous chercherons à modéliser, à l'aide de régressions, d'une part le taux de concrétisation et d'autre part le taux de résiliation en fonction de critères concernant l'assuré, son véhicule et particulièrement le tarif. Nous examinerons les résultats obtenus et choisirons pour chaque cas un modèle final fiable.

Enfin, nous calculerons l'élasticité dans l'ensemble du portefeuille et selon différents critères. Nous pourrons ainsi déterminer et commenter l'impact qu'un changement de prix aurait sur la population des assurés.

Cependant, avant d'appliquer cette démarche, nous allons définir les éléments théoriques utilisés pour la modélisation. Tout d'abord, nous commencerons par la présentation du modèle linéaire classique, puis nous enchaînerons avec les modèles linéaires généralisés (GLM) en détaillant la régression logistique qui représente l'outil principal de cette étude. Tous les éléments théoriques décrits par la suite sont issus d'ouvrages, d'articles, de mémoires et de cours (sites internet ou cours de l'EURIA). Toutes ces références sont listées en Bibliographie (page 95).

Deuxième partie

Éléments théoriques

Chapitre 4

Modèle linéaire classique

4.1 Définition et exemples

Le modèle linéaire¹ consiste à établir une relation du type linéaire entre une variable à expliquer et/ou prédire Y et k variables explicatives $X = (X^1, \dots, X^k)$. Il peut s'écrire de la forme suivante :

$$Y = X\beta + \epsilon = \sum_{j=1}^k \beta_j X^j + \epsilon \quad (4.1)$$

- Y est aléatoire ; on suppose que la variance de Y est constante : c'est ce qu'on appelle l'hypothèse d'homoscédasticité ;
- Les k variables $X^1, \dots, X^k$ sont des variables observées et non aléatoires dites déterministes ;
- Les $\beta_j (j = 1, \dots, k)$ sont les paramètres du modèle, non observés et donc à estimer par des techniques statistiques appropriées ;
- ϵ est le terme d'erreur dans le modèle ; c'est une variable aléatoire non observée pour laquelle on pose les hypothèses suivantes :

$$E(\epsilon) = 0 \quad ; \quad Var(\epsilon) = \sigma^2 > 0$$

où σ^2 est un paramètre inconnu, à estimer.

On constate donc que les hypothèses de ϵ impliquent qu'en moyenne Y s'écrit bien comme une combinaison linéaire des X^i . En effet, Y possède les caractéristiques suivantes :

$$E(Y) = \sum_{j=1}^k \beta_j X^j \quad ; \quad Var(Y) = \sigma^2$$

L'estimation des paramètres de ce modèle est basée sur n observations simultanées des variables X^i et Y réalisées sur n individus indépendants. Pour la i -ème observation, les valeurs observées des variables sont notées $y_i; x_i^1, \dots, x_i^k$, de sorte que le modèle s'écrit :

1. Références :

– AILLOT P. (2011) Modèles linéaires. Cours EURIA
– Modèles linéaires <http://www.math.univ-toulouse.fr/~barthe/M1modlin/poly.pdf>

$$y_i = \sum_{j=1}^k \beta_j x_i^j + e_i$$

Nous pouvons écrire le modèle d'une manière matricielle :

$$y = X\beta + e$$

- y le vecteur $\mathbb{R}^n$ composé des valeurs $y_1, \dots, y_n$;
- X la matrice (n,k) de rang k , contenant les valeurs observées des k variables explicatives disposées en colonnes ;
- β le vecteur $\mathbb{R}^k$ contenant les k paramètres du modèle ;
- e le vecteur de $\mathbb{R}^n$ des erreurs du modèle.

On parle de **modèle linéaire gaussien (ou classique)** lorsqu'on ajoute l'hypothèse de normalités des résidus. Ainsi, le modèle $Y = X\beta + \epsilon$ dispose d'une caractéristique en plus :

$$\epsilon \sim N(X\beta, \sigma^2)$$

Cela veut dire que $Y \sim N(X\beta, \sigma^2)$.

Citons deux exemples de modèles linéaires classiques :

a) Le modèle de régression linéaire

On cherche à modéliser une variable quantitative Y en fonction de variables explicatives quantitatives $x^1, x^2, \dots, x^p$. Sous l'hypothèse gaussienne, le modèle de régression linéaire s'écrit :

$$y_i = \theta_0 + \theta_1 x_i^1 + \dots + \theta_p x_i^p + e_i$$

avec $\theta_0, \theta_1, \dots, \theta_p$ inconnus, et $e_1, \dots, e_n$ n observations indépendantes d'une loi $N(0, \sigma^2)$ avec σ^2 inconnue.

b) Le modèle factoriel

On cherche à modéliser une variable quantitative Y en fonction d'une (ou de plusieurs) variable(s) explicative(s) qualitative(s) (appelée facteur). Sous l'hypothèse gaussienne, le modèle à un facteur s'écrit :

$$y_{ij} = \mu_i + e_{ij} \quad i = 1, \dots, I ; j = 1, \dots, n_i$$

avec $\mu_1, \dots, \mu_I$ inconnus, et $e_{11}, \dots, e_{In_I}$ n observations indépendantes d'une loi $N(0, \sigma^2)$ avec σ^2 inconnue.

4.2 Estimation des paramètres du modèle

Méthodes d'estimation

Il existe de nombreuses méthodes d'estimation pour définir les paramètres d'un modèle : la méthode des moments, le maximum de vraisemblance, la méthode des moindres carrés, etc. Par exemple :

1) L'estimation par maximum de vraisemblance (MV) est basée sur la vraisemblance du modèle linéaire gaussien :

$$L(\beta; y) = \prod_{i=1}^n f(y_i; \beta)$$

où $f(y_i; \beta)$ est la densité de la loi Normale sur Y .

Pour obtenir l'estimateur $\hat{\beta}$ du maximum de vraisemblance, on maximise sa log-vraisemblance selon β en résolvant le système d'équations du maximum de vraisemblance :

$$\frac{\partial}{\partial \beta_j} \ln L(\beta_1, \dots, \beta_k; y) = 0 \text{ pour } j = 1, \dots, k$$

dont $\hat{\beta}(y)$ est solution, sous réserve que la condition de second ordre soit vérifiée. On pourra également obtenir l'estimateur du MV de σ^2 en maximisant la log-vraisemblance selon σ^2 .

2) L'estimation par la méthode des moindres carrés consiste à estimer β en minimisant la somme des carrés des résidus (SSR), tel que :

$$\phi(\hat{\beta}(y)) = \min \sum_{i=1}^n (\hat{e}_i)^2 = \min \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Le critère des moindres carrés peut s'écrire aussi de la façon suivante :

$$\|\hat{e}\|^2 = \|y - X\hat{\beta}\|^2 = \inf_{\beta \in R^k} \|y - X\beta\|^2$$

Cette méthode d'estimation ne nécessite pas que l'on pose l'hypothèse de normalité des résidus.

Remarque : Les estimateurs du maximum de vraisemblance de β sont équivalents aux estimateurs des moindres carrés de β . En revanche, certaines propriétés ne sont vraies que sous l'hypothèse de normalité des résidus.

Estimation de β

Si y est la réalisation de Y , l'estimation de β , $\hat{\beta}(y)$, est l'unique élément de R^k tel que :

$$X\hat{\beta}(y) = \hat{y}$$

On a donc :

$$\boxed{\hat{\beta}(y) = (X'X)^{-1}X'y}$$

$\hat{\beta}(y)$ est l'observation de la variable aléatoire $\hat{\beta} = (X'X)^{-1}X'Y$: $\hat{\beta}$ est la transformée de Y par l'expression $(X'X)^{-1}X'$.

$\hat{\beta}$ est un estimateur sans biais de β . Il a pour matrice de variance-covariance : $T_{\hat{\beta}} = \sigma^2(X'X)^{-1}$. Et il suit une loi gaussienne dans R^k . On peut donc écrire :

$$\hat{\beta} \sim N_k(\beta; \sigma^2(X'X)^{-1})$$

Valeurs ajustées et résidus calculés

Les $\hat{y}_i$ s'appellent les valeurs ajustées ou valeurs prédites par le modèle : $\hat{y}_i$ est une valeur approchée de y_i . On estime également les résidus $\hat{e}_i$.

$$\boxed{\hat{y} = X(X'X)^{-1}X'y \quad \text{et} \quad \hat{e} = y - \hat{y}}$$

- $\hat{y} = X\hat{\beta}(y)$ est le vecteur des valeurs ajustées. $\hat{y}$ est l'observation de la v.a $\hat{Y} = X(X'X)^{-1}X'Y$ avec $\hat{Y} \sim N_n(X\beta; \sigma^2 H)$, où $H = X(X'X)^{-1}X'$.
- $\hat{e} = y - \hat{y}$ est le vecteur des résidus calculés. $\hat{e}$ est l'observation de la variable aléatoire $\hat{E} = Y - \hat{Y} = (I_n - H)Y$ avec $\hat{E} \sim N(0; \sigma^2(I_n - H))$.
- Propriétés : $\hat{Y}$ et $\hat{E}$ sont deux variables aléatoires indépendantes ; $\hat{E}$ et $\hat{\beta}$ sont deux variables aléatoires indépendantes.

Estimation de σ^2

Soit σ^2 la vraie valeur théorique de la variance des résidus, $\hat{\sigma}^2$ l'estimateur de σ^2 , et $\hat{\sigma}^2(y)$ la réalisation de la variable aléatoire $\hat{\sigma}^2$ (ou une estimation de σ^2).

L'estimateur σ^2 est :

$$\boxed{\hat{\sigma}^2 = \frac{1}{n-k} \|\hat{E}\|^2 = \frac{1}{n-k} \|Y - \hat{Y}\|^2}$$

L'estimation de σ^2 est donc :

$$\hat{\sigma}^2(y) = \frac{1}{n-k} \|\hat{e}\|^2 = \frac{1}{n-k} \|y - \hat{y}\|^2 = \frac{\|y\|^2 - \|\hat{y}\|^2}{n-k}$$

où

$$\|y\|^2 = \sum_{i=1}^n y_i^2 \text{ et que } \|\hat{y}\|^2 = \hat{\beta}(y)'(X'y)$$

Le dénominateur $(n - k)$ provient du fait que l'on a estimé k paramètres dans le modèle. Voici quelques propriétés :

- $\frac{(n-k)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-k}^2$ (Somme des carrés de n v.a. $N(0,1)$ qui vérifient k relations linéaires).

- $\hat{\sigma}^2$ est un estimateur sans biais de σ^2 et de variance $\frac{2\sigma^4}{n-k}$.
- $\hat{Y}$ et $\hat{\sigma}^2$ sont deux v.a indépendantes ; $\hat{\beta}$ et $\hat{\sigma}^2$ sont deux v.a. indépendantes.

4.3 Limites

Les modèles linéaires classiques ont longtemps été utilisés par les actuaires pour expliquer un phénomène d'intérêt en fonction de différents critères observables de la vie réelle. Cependant, ces modèles, bien que toujours employés aujourd'hui, présentent l'inconvénient d'avoir un champ d'application assez restreint. En effet, ce genre de modèles se base sur des hypothèses qui sont difficiles à valider :

- l'hypothèse de normalité des résidus est la plus restrictive car sous la forme linéaire du modèle elle suppose de vérifier la distribution gaussienne de la variable cible. Cependant, lorsque la variable à modéliser est une variable discrète² ou se présente sous la forme de proportions ou de taux, cette hypothèse ne peut pas être validée.
- l'hypothèse d'homoscédasticité, de la variable à expliquer, se révèle également problématique dans de nombreux cas. En effet, il n'est pas rare de constater une augmentation de la variance à mesure que la moyenne augmente, ce qui ne peut donc pas être modélisé par un modèle linéaire classique.

Au vu de ces limites, il a fallu utiliser de nouveaux outils que nous allons voir par la suite.

2. Une variable quantitative est dite discrète lorsqu'elle prend un nombre fini ou dénombrable de valeurs.

Chapitre 5

Modèles linéaires généralisés

5.1 Définition d'un modèle GLM

Le modèle linéaire généralisé¹ part du même principe que celui du modèle linéaire simple. La différence est qu'au lieu de modéliser la variable à expliquer directement, c'est une fonction de l'espérance de cette variable (appelée fonction de lien) qui est modélisée.

Ce type de modèle tente de relier des variables explicatives $X_1, X_2, \dots, X_p$ à une variable (aléatoire) à expliquer Y de la manière suivante :

- Y suit une loi dont la densité appartient à la famille exponentielle.
- $E(Y)$ dépend de $\eta(X)$ au travers de la fonction de lien g :

$$g(E(Y)) = g(\mu) = \eta(X)$$

g est une fonction inversible.

- On construit une combinaison linéaire des X_j (ie modèle de régression) :

$$\eta(x) = \sum_{j=1}^p \beta_j x_j$$

η est appelé prédicteur, il peut être vu comme une variable «synthétique», un résumé linéaire des variables explicatives ou une direction dans $\mathbb{R}^p$

Les paramètres $(\beta_1, \dots, \beta_n)$ sont les coefficients de régression. Leur estimation se fait par maximum de vraisemblance et le choix de la distribution pour le modèle linéaire se fait en observant les données.

Les principales caractéristiques de cette modélisation sont alors :

- Une transformation de l'espérance conditionnelle par une fonction de lien monotone et dérivable ;

1. Références :

- DE JONG P., ZELLER G. (2008) Generalized Linear Models for Insurance Data. Cambridge University Press.
- GONNET G. (2010) Étude de la tarification et de la segmentation en assurance automobile. ISFA.
- Théorie sur les GLM : <http://www.sites.univ-rennes2.fr/laboratoire-statistique/PAC/doc/glm.pdf>

- La modélisation de cette transformation par une combinaison linéaire des variables explicatives ;
- Des termes d'erreurs non gaussiens ;
- La définition de la loi conditionnelle de Y appartenant à la famille exponentielle.

Distributions de la famille exponentielle naturelle

Dans le cadre particulier des modèles linéaires généralisés, nous allons nous intéresser à une forme particulière de la famille de loi exponentielle. Nous supposons en effet que la variable à expliquer Y possède une densité par rapport à une mesure dominante qui s'écrit :

$$f(y|\theta; \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y; \phi)\right)$$

où :

- $a(\cdot)$ est une fonction non nulle, dérivable sur $\mathbb{R}$;
- $b(\cdot)$ est une fonction trois fois dérivable sur $\mathbb{R}$ et sa dérivée première est inversible ;
- $c(\cdot)$ est une fonction définie sur $\mathbb{R}^2$;
- θ est un paramètre canonique ou paramètre de la moyenne ;
- ϕ est un paramètre de dispersion.

Remarque : $a(\phi) = \phi$ pour des données individuelles sinon $a(\phi) = \frac{\phi}{d}$ pour des données groupées où d est l'effectif du groupe. Dans notre cas, nous fixerons $d = 1$.

Y possède les propriétés suivantes :

- son espérance est : $\mu = E(Y) = b'(\theta)$
- sa variance s'écrit : $Var(Y) = b''(\theta) * a(\phi)$

Le coefficient θ exerce un contrôle sur l'espérance et la variance de la variable à expliquer au moyen des fonctions $b'(\cdot)$ et $b''(\cdot)$. De plus, par définition la fonction $b'(\cdot)$ est inversible, nous disposons ainsi de la relation $\theta = (b')^{-1}(\mu)$. La fonction de lien g évoquée plus haut est alors appelée lien canonique lorsque la relation $g(\mu) = \theta = (b')^{-1}(\mu)$ est vérifiée. Le paramètre de dispersion ϕ n'influence quant à lui la variance de Y qu'à partir de la fonction $a(\cdot)$.

Nous allons à présent évoquer quelques exemples de lois appartenant à la famille exponentielle utilisées en modélisation linéaire généralisée. Nous reprenons les expressions $a(\cdot), b(\cdot), c(\cdot), \theta$ et ϕ :

1) La loi Normale :

Soit Y une variable aléatoire suivant une loi normale $\mathcal{N}(\mu, \sigma^2)$. Y est à valeurs réelles. Sa fonction de densité est :

$$f_{\mu, \sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

Elle peut être mise en forme sous :

$$f_{\mu,\sigma}(y) = \exp\left(\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)}{2}\right)$$

Ainsi, la loi normale $\mathcal{N}(\mu, \sigma^2)$ appartient à la famille exponentielle, avec :

- $\theta = \mu$
- $\phi = \sigma^2$
- $a(\phi) = \phi$
- $b(\theta) = \theta^2/2$
- $c(y, \phi) = -\frac{1}{2}\left(\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)$.

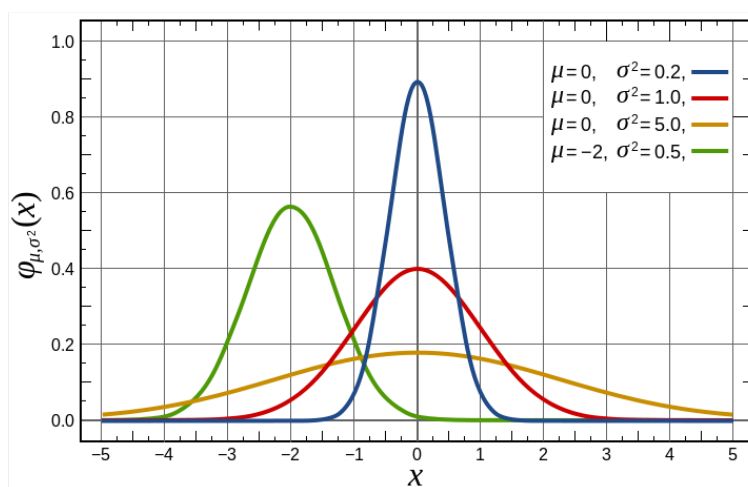


FIGURE 5.1 – Densité de probabilité φ de la loi normale $N(\mu, \sigma^2)$

2) La loi de Poisson :

La loi de Poisson $\mathcal{P}(\lambda)$ appartient à cette famille,

$$f_{\lambda}(y) = \exp(-\lambda) \frac{\lambda^y}{y!} = \exp(y \ln \lambda - \lambda - \ln y!), y \in \mathbb{N}$$

On retrouve donc :

- $\theta = \ln \lambda$
- $\phi = 1$
- $a(\phi) = 1$
- $b(\theta) = \exp(\theta) = \lambda$
- $c(y, \phi) = -\ln y!$

3) La loi Binomiale :

La loi Binomiale $\mathcal{P}(n, p)$ définie par la loi de probabilité suivante :

$$f(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, \dots, n$$

appartient à la famille exponentielle avec :

- $\theta = \ln(p/(1-p))$
- $\phi = 1$
- $a(\phi) = 1$
- $b(\theta) = n \ln(1 + \exp(\theta))$
- $c(y, \theta) = \ln \binom{n}{y}$

4) La loi Gamma :

Soit Y une variable aléatoire suivant une loi Gamma de paramètres ν et α (tous deux strictement positifs). ν est considéré comme un paramètre d'échelle et α est un paramètre de forme. La densité s'écrit :

$$f_{\nu, \alpha}(y) = \frac{1}{\Gamma(\nu)y} \left(\frac{y\nu}{\alpha}\right)^\nu \exp\left(-\frac{y\nu}{\alpha}\right), \quad \text{pour } y > 0$$

où $\Gamma(x) = \int_0^\infty \exp(-u) u^{x-1} du$

La fonction de densité peut également être mise sous la forme :

$$f_{\nu, \alpha}(y) = \exp\left[\frac{y \times \left(-\frac{1}{\alpha} - \left(-\ln\left(\frac{1}{\alpha}\right)\right)\right)}{\frac{1}{\nu}} - \ln(\Gamma(\nu)y) + \nu \ln(y\nu)\right]$$

Ainsi, la loi Gamma appartient à la famille exponentielle avec :

- $\theta = -\frac{1}{\alpha}$
- $\phi = \frac{1}{\nu}$
- $a(\phi) = \frac{1}{\nu}$
- $b(\theta) = -\ln(-\theta)$
- $c(y, \phi) = \ln(\Gamma(\nu)y) + \nu \ln(y\nu)$

Fonctions de lien

Nous avons donc vu ci-dessus qu'il est nécessaire d'effectuer deux choix pour construire un modèle linéaire généralisé. Le premier concerne la loi de la variable à expliquer. Le deuxième choix porte sur la fonction de lien. Une fonction quelconque ne peut pas être une fonction de lien. Il faut que la fonction de lien soit inversible. Un choix particulier qui simplifie les calculs est le choix de la fonction de lien canonique, à savoir $g(.) = (b')^{-1}(.)$. Le tableau ci-dessous donne, pour les lois couramment utilisées dans les GLM, les liens canoniques :

Loi	Nom du lien	Fonction de lien
Bernoulli/Binomiale	lien logit	$g(\mu) = \text{logit}(\mu) = \ln(\mu/1 - \mu)$
Normale	lien identité	$g(\mu) = \mu$
Poisson	lien log	$g(\mu) = \ln(\mu)$
Gamma	lien réciproque	$g(\mu) = -1/\mu$

TABLE 5.1 – Fonctions de liens usuelles

5.2 Exemple d'un GLM : la Régression logistique

La Régression logistique est un cas particulier des modèles linéaires généralisés. Elle permet de modéliser une variable Y binaire en fonction de $X = (X_1, \dots, X_p)$ variables explicatives.

5.2.1 Le modèle LOGIT

Notons Y la variable à prédire et $X = (X_1, X_2, \dots, X_p)$ les variables prédictives (variables explicatives). Notre variable Y prend deux modalités possibles (0,1). Par exemple, Y représente l'acte de concrétisation avec ($Y=1$) l'événement où le prospect a concrétisé le devis et ($Y=0$) l'événement où il a refusé le devis. Y suit donc une loi de Bernoulli. Dans notre étude, nous avons à modéliser une loi de Bernoulli de paramètre $\tau \in [0; 1]$ que nous cherchons à estimer et qui correspond à une fréquence. τ estimera, par exemple, la probabilité de concrétiser un devis. La fonction de lien devra avoir la propriété de conserver l'intervalle de valeurs $[0; 1]$ pour le modèle. La fonction *logit* a cette propriété :

$$g(\tau) = \ln\left(\frac{\tau}{1-\tau}\right) = x'\beta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad , \quad \text{avec } 0 \leq \tau \leq 1$$

Avec $\tau = P(Y = 1|X = x)$ et $1 - \tau = P(Y = 0|X = x)$

où

- g représente notre fonction logit qui est une fonction bijective et dérivable de (0,1) dans $\mathbb{R}$;
- le vecteur $x = (x_1, \dots, x_p)$ représente nos variables explicatives déterministes ;
- le vecteur $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ sont les coefficients de la régression logistique.

Après transformation de l'équation ci-dessus, nous obtenons :

$$\tau = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)}$$

$$\tau = \frac{1}{1 + \exp(-(x'\beta))}$$

L'espérance de Y est donc : $E[Y] = \tau$

La variance de Y nous est donnée par :

$$\text{Var}(Y) = E[Y^2] - E[Y]^2 = \tau - \tau^2$$

$$\text{Var}(Y) = \tau(1 - \tau)$$

5.2.2 Estimation des coefficients

L'estimation s'effectue par la méthode du maximum de vraisemblance à partir d'un échantillon i.i.d de n observations (y_i, x_i) . La vraisemblance L correspond à :

$$\begin{aligned} L(\beta_0, \beta) &= \prod_{i=1}^n \left(\frac{\exp(\beta_0 + \beta' x_i)}{1 + \exp(\beta_0 + \beta' x_i)} \right)^{y_i} \left(1 - \frac{\exp(\beta_0 + \beta' x_i)}{1 + \exp(\beta_0 + \beta' x_i)} \right)^{1-y_i} \\ &= \prod_{i=1}^n \tau(x_i)^{y_i} (1 - \tau(x_i))^{1-y_i} \end{aligned}$$

Pour faciliter les manipulations, on préfère souvent travailler sur la log-vraisemblance :

$$l = \ln L(\beta_0, \beta) = \sum_{i=1}^n y_i \ln \tau(x_i) + (1 - y_i) \ln(1 - \tau(x_i))$$

En annulant les dérivées de la log-vraisemblance, on obtient le système :

$$\begin{cases} \frac{\partial l(\beta)}{\partial \beta_0} = \sum_{i=1}^n (y_i - \tau(x_i)) = 0 \\ \frac{\partial l(\beta)}{\partial \beta_j} = \sum_{i=1}^n x_i^j (y_i - \tau(x_i)) = 0 \quad j = 1, \dots, p \end{cases}$$

Ces équations, non linéaires en β , n'ont pas de solution analytique : leur résolution requiert l'utilisation de méthodes itératives telles que l'algorithme de Newton-Raphson ou la méthode des scores de Fisher.

Par exemple, la méthode de Newton-Raphson est une méthode itérative du gradient s'appuyant sur la relation suivante :

$$\beta^{i+1} = \beta^i - \left(\frac{\partial^2 L}{\partial \beta^2} \right)^{-1} \frac{\partial L}{\partial \beta}$$

- β^i est la solution courante à l'étape i , $\beta^0 = (0, \dots, 0)$ est une initiation possible ;
- $\frac{\partial L}{\partial \beta}$ est le vecteur des dérivées partielles premières de la vraisemblance ;
- $\frac{\partial^2 L}{\partial \beta^2}$ est la matrice des dérivées partielles secondes de la vraisemblance ;
- les itérations sont interrompues lorsque la différence entre deux vecteurs de solutions successifs est négligeable.

Cette dernière matrice, dite matrice hessienne, est intéressante car son inverse représente l'estimation de la matrice de variance-covariance de β .

La méthode de Newton-Raphson peut être appliquée à la log-vraisemblance $l(\beta)$ du modèle décrite ci-dessus, ainsi nous obtenons la matrice, $\hat{V}(\hat{\beta})$, de variance-covariance asymptotique des estimateurs suivante :

$$\hat{V}(\hat{\beta}) = \left[-\frac{\partial^2 l(\beta)}{\partial \beta^2} \right]^{-1} = (X' \hat{V} X)^{-1}$$

où

$$\hat{V} = \begin{bmatrix} \hat{\tau}_1(1 - \hat{\tau}_1) & 0 \\ 0 & \hat{\tau}_n(1 - \hat{\tau}_n) \end{bmatrix}$$

Remarque : L'estimation $\hat{\beta}$ de β dispose des propriétés suivantes :

- Loi de $\hat{\beta}$: $\hat{\beta} \sim N(\beta, \hat{V}(\hat{\beta}))$;
- Intervalle de confiance IC à 95% de $\hat{\beta}$ (appelé Intervalle de confiance de Wald) :

$$IC = [\hat{\beta} - 1,96\sigma(\hat{\beta}); \hat{\beta} + 1,96\sigma(\hat{\beta})]$$

$$\text{où } \sigma(\hat{\beta}) = \sqrt{\hat{V}(\hat{\beta})}$$

Dans la pratique, les logiciels utilisent une procédure approchée pour obtenir une solution satisfaisante, ce qui explique d'ailleurs pourquoi ils ne fournissent pas toujours des coefficients strictement identiques. Les résultats dépendent de l'algorithme utilisé et de la précision adoptée lors du paramétrage du calcul.

5.2.3 Significativité du modèle et des variables

Définition : Soient $X_1, \dots, X_k$ k variables aléatoires indépendantes suivant des lois normales de moyennes respectives μ_i et d'écart-type σ_i ; $Y_i = \frac{X_i - \mu_i}{\sigma_i}$ leurs variables centrées et réduites, alors par définition la variable Z , telle que

$$Z := \sum_{i=1}^k Y_i^2 = \sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

suit une loi du χ^2 (ou Chi-2) à k degrés de liberté. Alors la densité de Z , notée f_Z sera :

$$f_Z(t) = \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} t^{\frac{k}{2}-1} \exp(-\frac{t}{2}), \quad \text{pour tout } t \text{ positif}$$

$$\text{où } \Gamma(x) = \int_0^\infty \exp(-u) u^{x-1} du.$$

Lorsqu'on met en place un modèle de régression, certaines ou l'ensemble des variables explicatives peuvent ne pas avoir d'influence sur les valeurs possibles de la variable Y à expliquer. Tout d'abord, nous voulons juger si le modèle complet est significatif, pour cela nous utiliserons le test global de significativité :

Ce test cherche à valider si l'on accepte ou on rejette l'hypothèse nulle H_0 suivante :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

La statistique utilisée pour tester cette hypothèse est la suivante :

$$T = -2 \ln \left(\frac{\text{Vraisemblance sous } H_0 \text{ (de la constante } \beta_0)}{\text{Vraisemblance du modèle complet}} \right)$$

Sous l'hypothèse nulle H_0 , T suit une loi du χ^2 à k degrés de liberté.

Le modèle est donc jugé significatif si l'hypothèse $\beta = 0$ est rejetée, ce qui se traduit par :

$$P = \text{Prob}(\chi^2(k) \geq T) \leq \alpha$$

où α représente le **seuil de significativité** (il est généralement fixé à 5%).

Si P , appelé aussi p-valeur² du test, est inférieur au niveau α , cela indique clairement une forte valeur de T .

De plus, un modèle peut être significatif même si certaines variables explicatives n'ont pas de lien avec la variable à expliquer. C'est pour cela, que l'on teste aussi la significativité de chaque variable explicative du modèle. On cherche à tester donc l'hypothèse nulle H_0 suivante :

$$H_0 : \exists k \in [1; p] \text{ tq } \beta_k = 0$$

Trois méthodes sont généralement utilisées :

1) Test de Wald :

L'apport d'une variable X est mesuré par la statistique :

$$W = \frac{\hat{\beta}_x}{\sigma(\hat{\beta}_x)}$$

Sous l'hypothèse H_0 , W suit une loi normale $N(0,1)$.

La variable X_k est jugée significative si H_0 est rejetée, ce qui se traduit par :

$$P = Prob(|N(0,1)| \geq W) \leq \alpha$$

Ce test compare l'écart entre le coefficient théorique et sa valeur estimée.

2) Test du rapport de vraisemblance :

L'apport d'une variable X_k dans le modèle est mesuré par la statistique :

$$T = -2 \ln \left(\frac{\text{Vraisemblance sans la variable } X_k}{\text{Vraisemblance avec la variable } X_k} \right)$$

Sous l'hypothèse H_0 , T suit une χ^2 à 1 degré de liberté.

La variable X_k est jugée significative si H_0 est rejetée, ce qui se traduit par :

$$P = Prob(\chi^2(1) \geq T) \leq \alpha$$

3) Test du Score :

$$U = U'(\beta)_{\hat{\beta}_{H_0}} [J(\hat{\beta}_{H_0})]^{-1} U(\beta)_{\hat{\beta}_{H_0}}$$

où $J(\hat{\beta}) = -[\frac{\partial^2 l(\beta)}{\partial \beta^2}]_{\beta=\hat{\beta}}$ et U le vecteur des dérivées partielles de la log-vraisemblance. La statistique U suit une loi du $\chi^2(1)$. La variable X_k est donc jugée significative si :

$$P = Prob(\chi^2(1) \geq U) \leq \alpha$$

2. La p-valeur (ou p-value en anglais) est la probabilité de commettre une erreur de première espèce, c'est-à-dire de rejeter à tort l'hypothèse nulle.

Ce test compare à zéro la pente de la tangente au point théorique.

Ces trois tests donnent des résultats très proches.

5.2.4 Validation du modèle

Un modèle peut être globalement significatif mais cela ne signifie pas qu'il soit bien adapté à nos données. Il est en effet possible que les variables explicatives soient significatives, c'est à dire qu'elles aient un pouvoir discriminant significatif sur les deux populations ($Y=1$) et ($Y=0$), mais que le phénomène ($Y=1$) ne puisse pas être correctement prédit par ces variables.

Dans le cadre du modèle linéaire classique, la validation d'un modèle consiste à s'assurer que les hypothèses de départ sont vérifiées [cf 4.3, page 23]. La formulation plus générale du modèle de régression logistique n'implique pas une telle démarche. En effet, la décomposition de la variable à expliquer en partie déterministe et partie résiduelle n'implique pas de propriété précise quant aux comportements des résidus. De plus, au vu de la forme de la variable à expliquer et de sa loi (Bernoulli), il ne semble pas très intéressant de s'attarder sur les résidus qui sont, par nature eux aussi, compris entre -1 et 1. Il serait donc difficile de s'aider d'une analyse de ce type pour déceler d'éventuelles observations aberrantes.

La validation d'un modèle devient dès lors plus délicate. Dans le cadre de notre étude, nous disposons de plusieurs types de mesures utilisées pour juger de la qualité du modèle global ainsi que de trois statistiques pour tester l'ajustement du modèle.

Les trois statistiques suivantes permettent de tester globalement la qualité d'ajustement du modèle aux données. Plus leur valeur est petite, meilleur est le modèle. Notons L la vraisemblance :

- $-2\ln(L)$
Cette statistique est d'autant plus grande que la vraisemblance est faible.
- Critères d'information d'Akaike
 $AIC = -2\ln(L) + 2m$ où m est le nombre de paramètre.
Ce critère prend en compte le nombre de variables explicatives. Deux modèles ayant la même valeur pour le critère précédent, ne seront pas jugés équivalents pour ce critère si leurs nombres de paramètres diffèrent. Celui qui a le moins de variables explicatives sera préféré.
- Critères de Schwartz
 $SC = -2\ln(L) + m\ln(n)$, où m est le nombre de paramètres et n le nombre d'observations. Ce critère prend en compte le nombre d'observations.

- Corrélations des rangs :

Cette méthode consiste à étudier les corrélations entre fréquences observées de Y et probabilités prédites de $\hat{Y}$ (elle est aussi utilisée pour comparer deux variables ordinales). Nous utiliserons les notations suivantes pour calculer certaines statistiques :

- Nombre de paires concordantes : C ;
- Nombre de paires discordantes : D ;
- Nombre de paires ex-aequo(ou liées) : E ;
- Nombre total de paires : $t = C + D + E$;
- Nombre d'observations : n

Une paire est formée par deux individus (i_1, i_2) ayant une réponse différente : $Y_{i_1} \neq Y_{i_2}$. Il y a concordance si les $\hat{Y}_i$ sont dans le même ordre que les Y_i . Si les ordres sont différents, la paire est discordante. Les paires ex-aequo représentent trois cas :

- (i_1, i_2) est ex-aequo en Y si $Y_{i_1} = Y_{i_2}$ et $\hat{Y}_{i_1} \neq \hat{Y}_{i_2}$. Notons E_Y le nombre de paires ex-aequo en Y .
 - (i_1, i_2) est ex-aequo en $\hat{Y}$ si $Y_{i_1} \neq Y_{i_2}$ et $\hat{Y}_{i_1} = \hat{Y}_{i_2}$. Notons $E_{\hat{Y}}$ le nombre de paires ex-aequo en $\hat{Y}$.
 - (i_1, i_2) est ex-aequo en Y et en $\hat{Y}$ si $Y_{i_1} = Y_{i_2}$ et $\hat{Y}_{i_1} = \hat{Y}_{i_2}$. Notons $E_{Y\hat{Y}}$ le nombre de paires ex-aequo en Y et en $\hat{Y}$.
- Ainsi, le nombre de paires ex-aequo $E = E_Y + E_{\hat{Y}} + E_{Y\hat{Y}}$.

Les mesures suivantes permettent de juger de la corrélation entre la variable cible Y et les variables explicatives. Plus leurs valeurs sont élevées, plus le modèle utilisé permet d'expliquer la variable cible.

- Le $\bar{C}$ varie entre 0 et 1 : $\bar{C} = C + \frac{1}{2} \frac{E}{t}$
- Le $\bar{D}$ de *Sommer* varie entre -1 et 1 : $\bar{D} = \frac{C-D}{t}$
- Le G de *Goodman-Kruskal* varie entre -1 et 1 : $\bar{G} = \frac{C-D}{C+D}$
- Le T de *Kendall* varie entre -1 et 1 : $\bar{T} = 2 \frac{C-D}{n(n-1)}$

Ces statistiques permettent de comparer entre eux plusieurs modèles. Certains résultats d'ordre général permettent de juger d'un modèle sans avoir à le comparer avec d'autres modèles. Par exemple, il est conseillé d'avoir un nombre de paires concordantes supérieur à 60% pour considérer que le modèle est valide. En dessous de ce seuil, la modélisation ne permet pas d'expliquer suffisamment la variable cible. C'est ce seuil que nous essayons de conserver dans la modélisation.

5.3 Sélection des variables

Lors de notre étude, nous avons retenu un certain nombre de variables susceptibles d'expliquer l'acte de concrétisation d'un prospect et l'acte de résiliation de la part de l'assuré. Il n'est cependant pas très pertinent de conserver autant de variables dans la modélisation que nous utiliserons pour prédire la probabilité de concrétisation et de résiliation. Il nous faut donc sélectionner certaines variables que nous jugeons importantes pour l'étude. Pour cela, nous disposons avec le logiciel SAS de 3 méthodes de sélection :

- 1) La méthode ascendante *FORWARD* :

A la première étape, seule la constante est introduite dans le modèle. Puis, à chaque étape suivante, la variable la plus significative selon le critère du chi deux résiduel, c'est-à-dire qui permet d'expliquer le plus la variable cible, parmi celles qui restent, est introduite dans le modèle. Cette procédure se termine lorsqu'il n'y a plus de variable suffisamment significative suivant le seuil que l'on s'est fixé.

2) La méthode descendante *BACKWARD* :

A la première étape, toutes les variables sont introduites dans le modèle. Puis, à chaque étape suivante, la variable la moins significative selon le test du chi deux de Wald est retirée de la modélisation. A nouveau, cette procédure se termine lorsque toutes les variables restant dans le modèle sont significatives selon un seuil fixé d'avance.

3) La méthode pas à pas *STEPWISE* :

Cette méthode est une combinaison des deux précédentes. La première étape consiste à modéliser la variable cible en introduisant uniquement la constante. A chaque étape suivante, comme pour la méthode ascendante, la variable la plus significative est introduite dans le modèle. Dans le même temps, les variables qui constituent le modèle sont testées et, si l'une d'elles n'est plus significative à la suite de l'introduction d'une autre variable, elle est retirée du modèle. Cette dernière méthode est la plus précise dans le sens où elle permet de garder toutes les variables explicatives qui expliquent le phénomène étudié de la variable cible. C'est donc cette méthode que nous utiliserons dans la suite de l'étude.

Ces méthodes de sélection peuvent être aussi réalisées selon d'autres critères jugeant de la significativité des variables tels que le test de déviance ou un critère de choix (AIC, SC, ...).

Troisième partie

Présentation et étude préliminaire du portefeuille

Chapitre 6

Le produit «Assurance Auto»

6.1 Le contrat

Tout d'abord, rappelons qu'un contrat d'assurance est une convention par laquelle l'assureur s'engage à verser à l'assuré une somme d'argent réparant le préjudice subi en cas de survenance d'un sinistre, défini en échange du paiement d'une somme versée dite prime d'assurance, soit à l'origine, soit périodiquement.

Un contrat d'assurance comme tout contrat conclu entre particuliers obéit aux principes généraux énoncés dans le Code Civil, et fait de plus l'objet d'une réglementation particulière¹

Les contrats d'assurance sont constitués de deux éléments bien distincts :

Les «conditions générales» qui doivent être obligatoirement remises à la souscription du contrat, et qui se présentent sous la forme d'un guide expliquant chaque garantie en détail, notamment les exclusions et le fonctionnement du contrat.

Les «conditions particulières» qui sont propres à chaque assuré et regroupent les informations déclarées par celui-ci. Dans notre cas, par exemple, il y apparaît les caractéristiques du véhicule assuré, des conducteurs, le montant de la prime, l'échéance anniversaire et les garanties choisies, etc.

Il existe plusieurs types de contrats en assurance automobile, mais dans le cadre de ce mémoire nous allons nous intéresser seulement aux contrats *mono-véhicules*. Ce type de contrat peut être destiné à l'assurance de particuliers ou d'entreprise pour les voitures, deux-roues, les voitures sans permis, etc. C'est la forme de contrat la plus répandue sur le marché. Le tarif est adapté à l'usage du véhicule : usage privé ou pour rejoindre le lieu de travail, pour les déplacements professionnels simples ou des tournées quotidiennes.

Enfin, le contrat prend effet au jour et à l'heure indiqués aux conditions particulières. Il est conclu pour un an et se renouvelle automatiquement d'année en année tant que l'assuré ou l'assureur n'y a pas mis fin.

1. Un contrat d'assurance est régi par :

- le Code des Assurances pour un particulier qui conclut un contrat d'assurance avec un assureur.
- le Code de la Mutualité pour un particulier qui conclut un contrat d'assurance avec une mutuelle.
- le Code de la Sécurité Sociale pour un particulier qui conclut un contrat d'assurance avec une institution de prévoyance.

6.2 Les garanties

Le produit «Assurance Auto» étudié dans ce mémoire est destiné en majorité aux particuliers mais également aux petits professionnels. La compagnie propose quatre formules qui diffèrent selon les garanties présentes dans le contrat. On peut représenter ces quatre formules selon le tableau suivant :

Garanties	F10	F11	F16	F17
Responsabilité civile	Base	Base	Base	Base
Défense Pénale et Recours Suite à Accident	Base	Base	Base	Base
Protection Juridique	Base	Base	Base	Base
Vol		Base	Base	Base
Incendie-Tempête-Grêle		Base	Base	Base
Catastrophes naturelles		Base	Base	Base
Dommages Accidents par Collision			Base	Base
Dommages Tous Accidents			Base	Base
Bris des glaces	Option	Option	Option	Option
Garanties du Conducteur	Base	Base	Base	Base
Assistance aux personnes	Base	Base	Base	Base
Assistance au véhicule 0 km, sauf panne : 50 km	Base	Base	Base	
Assistance au véhicule 0 km panne	Option	Option	Option	Base
Véhicule de remplacement 3 jours			Base	
Véhicule de remplacement 5, 10 ou 30 jours				Base
Valeur d'achat du véhicule pendant 1 an			Base	
Valeur d'achat du véhicule pendant 2 ans				Base

TABLE 6.1 – Les formules de garanties

Le tableau ci-dessus peut se lire de la manière suivante :

- une garantie est dite de «Base» lorsqu'elle est présente dans la formule, par exemple la Responsabilité civile se trouve dans les quatre formules F10 à F17.
- une garantie est dite d'«Option» lorsqu'elle est considérée comme optionnelle, le choix est laissé à l'assuré. La garantie Bris des Glaces est optionnelle pour les quatre formules.
- lorsqu'une case du tableau n'est pas remplie, cela veut dire qu'une garantie n'est pas présente dans une formule.

Ainsi, nous remarquons bien que la formule F17 est celle la plus complète avec le plus de garanties, donc la plus chère.

6.2.1 Description des garanties

Nous allons décrire les principales garanties présentes dans le tableau 6.1 :

La Responsabilité Civile (Dommages causés au tiers) :

Comme dit précédemment, du fait de son caractère obligatoire, la garantie minimum est la Responsabilité civile [cf le chapitre 1, page 9]. Elle couvre la réparation des dommages ou corporels causés aux autres par l'automobiliste assuré lorsque ces dommages lui sont imputables.

La Défense Pénale et Recours à accident :

Elle a pour but d'assurer la défense du conducteur autorisé lorsqu'il est poursuivi pour infractions aux dispositions régissant la circulation automobile, commises au volant du véhicule assuré, en l'absence de tout dommage à un tiers. De plus, elle permet d'obtenir de la part du responsable la réparation, à l'amiable ou judiciairement, du préjudice subi lors d'un accident impliquant le véhicule assuré.

Cette garantie ne couvre pas les accidents survenant lorsque le conducteur n'est pas titulaire d'un permis de conduire ou lorsqu'il se trouve sous l'emprise de l'alcool ou de la drogue au moment de l'accident.

La Protection juridique :

Elle intervient pour les litiges de la vie privée liés à la consommation, à l'habitation, au droit du travail, etc. L'assureur doit mettre à disposition de l'assuré des conseils, le libre choix d'un avocat et prend notamment en charge au titre de cette garantie les frais de justice, les honoraires d'enquêtes et d'avocat.

Bris de glace :

Elle couvre les dommages subis par le pare-brise et peut aussi s'étendre aux glaces latérales, aux vitres de toit ouvrant, à la lunette arrière, aux blocs optiques de phares et aux rétroviseurs.

Vol :

Cette garantie couvre le vol du véhicule, la tentative de vol, les dommages causés par un malfaiteur (par exemple : serrures endommagées) ainsi que les actes de vandalisme accompagnant le vol ou la tentative de vol.

Incendie-Tempête-Grêle :

Cette garantie entre en jeu lorsque les dommages sont liés à la survenance d'un événement climatique naturel d'intensité anormale (inondation, avalanche, tremblement de terre, etc). Elle intervient après parution au *Journal officiel* de l'arrêté interministériel constatant l'état de catastrophe naturelle.

Dommages Accidents par Collision :

Cette garantie ne s'applique que dans le cas de collision impliquant un tiers identifié, un piéton ou un animal domestique dont le propriétaire est identifié.

Dommages Tous Accidents :

Elle couvre tous les dommages matériels subis par le véhicule, quel que soit le type d'accident ou la faute commise par son conducteur.

L'assistance :

Elle est caractérisée par deux types de prestations :

- l'assistance aux personnes (rapatriement sanitaire, déplacement des membres de la famille, frais de rapatriement du corps, frais médicaux, frais de retour de l'assuré, etc).
- l'assistance aux véhicules (dépannage, remorquage, envoi de pièces détachées, véhicule de remplacement, etc).

L'assistance peut être limitée par un plafond de prise en charge, mais aussi par une franchise kilométrique. En effet, certaines assurances ne prennent en charge

le remorquage que si l'assuré se trouve à plus d'une certaine distance de son lieu d'habitation.

6.2.2 Des garanties limitées

Les garanties proposées par les assurances sont généralement limitées par des plafonds, franchises et exclusions.

L'exclusion de la garantie représente une clause permettant à l'organisme assureur d'exclure de la garantie certains dommages qui seraient provoqués par des circonstances précisées dans la clause, et qui permet ainsi de délimiter les dommages faisant l'objet d'une garantie des dommages n'en faisant pas partie. Deux cas se présentent pour les exclusions : elles sont légales ou contractuelles.

Les plafonds de garantie, ou limites de décaissement, sont les montants maximums sur lesquels la compagnie d'assurance s'engage. Cette clause permet à l'assureur de se prémunir contre un sinistre d'un montant exceptionnel. Elle peut être imposée par le réassureur à la compagnie d'assurance pour limiter son risque.

La franchise correspond à la somme restant à la charge de l'assuré. En général, la franchise est déduite du montant indemnisé par l'assureur. Elle peut être exprimée dans le contrat d'assurance en euros, en pourcentage ou être une combinaison des deux. Suivant les cas, elles permettent à l'assureur de supprimer les petits sinistres qui coûtent plus cher en frais de gestion qu'en indemnisation, de diminuer l'aléa moral, l'assuré continuant à subir une perte en cas de sinistres, de diminuer son risque de ruine, en se protégeant contre un sinistre exceptionnel. Il existe plusieurs types de franchise mais celles les plus répandues sur le marché sont les suivantes :

- Franchise absolue : elle correspond à un montant fixe, quel que soit l'incident survenu et quelle que soit son étendue. Cette franchise est systématiquement déduite de l'indemnisation.
- Franchise proportionnelle : on parle aussi de franchise «variable», elle est exprimée en pourcentage sur la somme globale du préjudice subit. Son montant dépend alors du coût des dommages. Cette franchise est peu utilisée dans les assurances pour les particuliers, mais plus dans les assurances de professionnels.
- Franchise relative : elle indique un montant minimum en dessous duquel l'assureur ne vous indemniser pas. Dans ce cas, les dommages résultant d'un sinistre ne seront totalement indemnisés par votre assureur qu'à partir du moment où leur montant dépasse la franchise relative. S'il est égal, pas d'indemnisation.
- Franchise kilométrique² : cette dernière est applicable dans les contrats d'assurance de véhicules. Elle détermine le nombre de kilomètres à partir desquels l'assurance peut être mise en jeu.

2. Pour le produit étudié dans ce mémoire, cette franchise est proportionnelle, ce qui n'est pas toujours le cas dans les autres compagnies d'assurance.

Chapitre 7

Construction des bases de données

Dans cette partie, nous allons présenter les données sur lesquelles nous nous sommes appuyés pour nos modélisations. En effet, notre étude s'intéresse à deux étapes : l'acte de concrétisation et l'acte de résiliation d'un contrat. Ainsi, deux bases de données ont été créées, une base pour chaque étape.

7.1 Base «Concrétisation» et «Résiliation»

L'étape des traitements des données est une étape primordiale pour préparer les données à la modélisation. En effet, cette étape nous permet de définir le périmètre de l'étude afin d'obtenir des résultats fiables.

a) Les fichiers initiaux

Pour créer nos bases «Concrétisation» et «Résiliations», nous nous sommes appuyés sur les fichiers suivants :

La base «IMAGE» : elle rassemble toutes les polices d'assurance concernant le produit «Assurance Auto» avec toutes les informations concernant l'assuré, son véhicule et ses garanties. Chaque ligne de la base correspond à une image (ou un état) de la vie du contrat à un instant donné. Chaque police dispose d'un ou plusieurs états. Un état de police peut être, par exemple, une affaire nouvelle, un avenant, une police suspendue car le contrat a été résilié, etc. Ainsi, nous comprenons que la base «IMAGE» est riche d'informations et très lourde à manipuler. En effet, elle dispose de 449 variables et 2.562.154 observations.

La base «PJRVECT» : elle rassemble tous les devis proposés par les agents aux prospects. Si un prospect concrétise le devis (décide d'être client d'Aviva), son devis, devenu une affaire nouvelle, sera présent dans la base «IMAGE». PJRVECT a 5.068.643 observations et 230 variables. Elle contient aussi des informations concernant le futur client et son véhicule. Elle nous permettra de récupérer les devis non concrétisés et essayer de comprendre pourquoi le prospect n'a pas désiré choisir une des formules du produit. Était-il sensible au prix proposé par l'agent ?

b) Processus de création des bases

Nous pouvons lister les tâches effectuées durant cette étape de la manière suivante :

1. Découverte des bases de données pour comprendre leur structure.

2. Délimitation du périmètre d'étude : repérer les devis et les contrats ainsi que les variables à conserver.
3. Uniformisation des données si leur provenance diffère.
4. Recodage des variables :
 - les modalités : beaucoup de variables ont des modalités 1,2,3,... ou A,B,C,... donc il a fallu recoder ces modalités afin de leur affecter un libellé permettant une compréhension et une analyse plus rapide et claire des graphiques et des résultats.
 - découpage en classes : par exemple l'âge de l'assuré ou l'ancienneté du véhicule.

Afin de mieux visualiser les tâches exposées précédemment, nous représentons à travers un schéma simplifié la démarche suivie pour créer notre base pour l'acte de concrétisation :

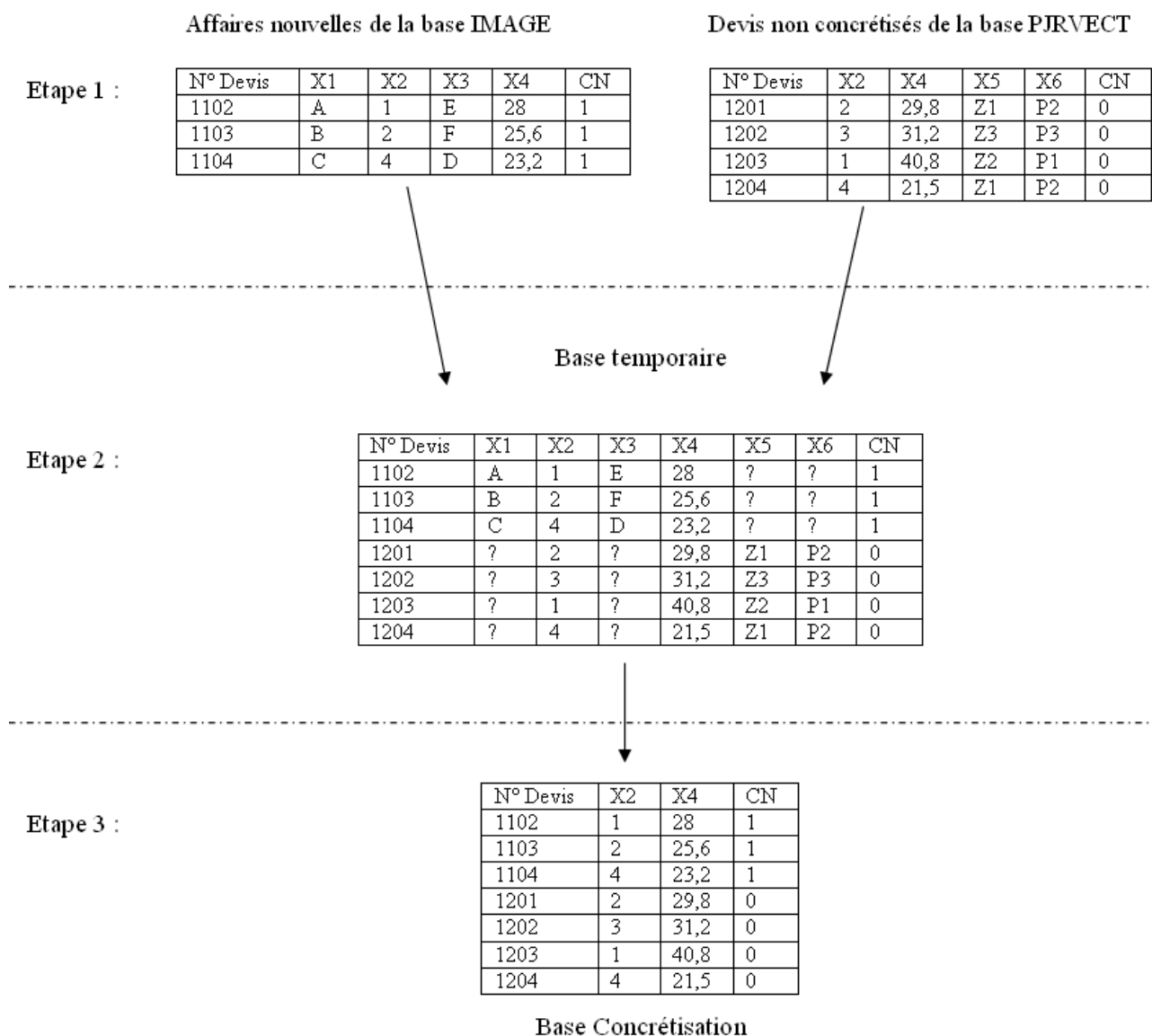


FIGURE 7.1 – Création de la base «Concrétisation»

Ce schéma est divisé en trois étapes :

- Étape 1 : initialement, nous repérons d’une part les contrats de l’année 2012 dont l’image est une affaire nouvelle de la base «IMAGE» et d’autre part les devis non concrétisés parmi tous les devis enregistrés en 2012 de la base «PJRVECT». Nous avons créé une variable «CN» qui est égale à 1 si le devis a été concrétisé et 0 si le devis n’a pas été concrétisé. De plus, nous avons rencontré deux cas : les bases IMAGE et PJRVECT disposent de variables semblables (X2,X4,CN) ou différentes (X1,X3,X5,X6). Un troisième cas (non représenté dans le schéma) est apparu : certaines variables de la base «IMAGE» sont les mêmes que la base «PJRVECT» mais n’ont pas le même nom. Il y a eu donc une étape de recodage des noms variables pour garder un maximum d’informations.
- Étape 2 : Nous créons une base temporaire rassemblant tous les devis (concrétisés ou non) de l’année 2012. Nous constatons des valeurs manquantes « ? » pour les variables non semblables (X1,X3,X5,X6). Ainsi, nous avons dû retirer ces variables car elles ne pourront pas être utilisées pour la modélisation.
- Étape 3 : Nous uniformisons les données. Nous obtenons une base contenant les mêmes variables et modalités pour les affaires nouvelles et les devis non concrétisés.

Quant au processus de création de la base des résiliations, nous nous sommes appuyés seulement sur la base «IMAGE». Nous avons repéré l’image avant l’échéance de chaque contrat. Nous obtenons l’ensemble des polices qui ont connu une échéance en 2012. Les motifs suivants ne seront donc pas considérés dans notre étude comme des résiliations venant de l’assuré :

- Résiliation venant de la compagnie ;
- Destruction ou vente du véhicule ;
- Décès de l’assuré ;
- Suspension ou mise en demeure.

Nous supposons en effet que les résiliations pour ces motifs n’ont pas pour origine un effet de la prime payée par l’assuré. Afin d’éviter d’introduire un biais provenant de ces résiliations, ils ne seront pas considérés comme des résiliations provenant de l’assuré.

Nous avons donc créé une variable binaire «RES» qui est égale à 1 si le contrat est résilié de la part de l’assuré et 0 si le contrat est renouvelé.

c) Synthèse

Au final, nous obtenons deux bases d’étude :

Premièrement, la base «Concrétisation» rassemble tous les devis non concrétisés ainsi que les devis concrétisés. Elle contient 256.250 lignes dont 86.553 affaires nouvelles et 56 variables. On pourra donc calculer un premier taux de concrétisation moyen de la manière suivante :

$$\text{Taux concrétisation moyen} = \frac{\text{Nombre de devis concrétisés}}{\text{Nombre total des devis}}$$

Avec $\text{Nombre total des devis} = \text{Nombre de devis concrétisés} + \text{Nombre de devis non concrétisés}$
 Notre taux moyen de concrétisation est de 33,8%.

Ensuite, la base «Résiliation» se compose des contrats résiliés d'une part et renouvelés d'autre part durant l'année 2012. Elle contient 515.607 lignes dont 17.105 contrats résiliés à leur date d'échéance par l'initiative des assurés et 75 variables. Le taux de résiliation moyen peut être calculé de la manière suivante :

$$\boxed{\text{Taux résiliation moyen} = \frac{\text{Nombre de contrats résiliés}}{\text{Nombre total des contrats}}}$$

Avec $\text{Nombre total des contrats} = \text{Nombre de contrats résiliés} + \text{Nombre de contrats renouvelés}$
 Ce taux est de 3,2%.

7.2 Présentation des variables

Dans cette partie, nous allons exposer les principales variables utilisées pour notre étude. Elle peuvent être classées en trois catégories :

a) Les variables d'identification

Ce sont les variables qui permettent d'identifier le devis, le contrat, le souscripteur ainsi que le portefeuille auquel le devis ou le contrat appartient :

- *Numéro de projet* : il correspond au numéro du devis ;
- *Numéro de police* d'assurance ;
- *Numéro d'image* ;
- *Code VGTA* : chaque véhicule dispose d'un identifiant appelé code VGTA. Cette variable nous a permis de faire le lien avec une autre base de données, appelée *SRA*¹, qui dispose de plusieurs variables donnant une description détaillée du véhicule de l'assuré (la classe de prix, la puissance, le type de véhicule, le carburant, etc.). Certaines de ces variables sont utilisées comme critères tarifaires.

b) Les variables de date

Ces variables nous ont permis de bien cerner le périmètre de l'étude :

- *Date de création du devis* ;
- *Date d'entrée souscripteur* (la date de début du contrat) ;
- *Date d'échéance du contrat* (la date d'anniversaire du contrat) ;
- *Date de fin du contrat* ;
- *Date du début de l'image* ;
- *Date de fin de l'image* ;

1. Dans notre étude, le terme «SRA» est le nom d'une base de données, cependant cette base est conçue par un organisme professionnel, appelé SRA, qui dispose du statut d'association. La compagnie en est adhérente. Pour plus d'informations voir le lexique page 97.

c) Les variables de tarification

Ce sont les variables qui entrent directement dans le calcul du niveau de prime auquel est soumis le contrat :

- *Âge de l'assuré* ;
- *Ancienneté de permis* ;
- *Ancienneté du véhicule* ;
- *Ancienneté du client* dans la compagnie ;
- *Ancienneté de souscription* du contrat «Assurance Auto» ;
- *Coefficient réduction-majoration* (ou coefficient Bonus Malus) ;
- *CRM 50* : indique le coefficient du Bonus Malus de l'assuré en prenant en compte le nombre d'années de l'assuré au niveau de Bonus maximum (50). Il est exprimé de la manière suivante :
 - Si Bonus Malus > 50 alors Bonus Malus 50 = Bonus Malus ;
 - Sinon Bonus Malus 50 = Bonus Malus - Nombre d'année à Bonus 50.
- *Durée de détention du véhicule* : elle indique depuis combien de temps l'assuré possède le véhicule pour lequel le risque est encouru et est exprimé comme suit : Max(durée de détention du véhicule précédent, durée détention véhicule actuel) ;
- *Mode d'achat* : elle précise de quelle façon l'assuré a acquis le véhicule (crédit, comptant, location avec option d'achat, location longue durée) ;
- *Année jeune conducteur* : elle indique le nombre d'années d'assurance en tant que conducteur principal à la souscription (1 an, 2 ans, 3 ans et plus) ;
- *Carburant* du véhicule (diesel, essence, etc.) ;
- *Antécédents de sinistres* (ou Classe de sinistralité noté CLS) : elle mentionne si le prospect a subi un ou plusieurs sinistres au cours des 24 derniers mois. Elle se définit de la manière suivante :
 - CLS=0 si aucun sinistre responsable et au plus un sinistre de type : Incendie, Vol, Bris des glaces, Stationnement ou Non responsable.
 - CLS=1 si un sinistre responsable ou plus d'un sinistre de type : Incendie, Vol, Bris des glaces, Stationnement ou Non responsable.
- *Catégorie socio-professionnel* (CSP) : agriculteur, artisan, commerciale, salarié, fonctionnaire, étudiant, etc. ;
- *Formule* de garanties décrites précédemment ;
- *Mode de garage* : cette variable mentionne l'endroit où est généralement garé le véhicule (terrain privé, parking collectif, voie publique, box fermé) ;
- *Type de véhicule* : précise le type de carrosserie du véhicule (Monospace, Berline, Break, etc.) ;
- *Situation familiale* ;
- *Classe de prix du véhicule* : représente le prix à neuf du véhicule regroupé par classe (A pour le moins cher jusqu'à Z pour le plus cher) ;
- *Puissance du véhicule* : représente la puissance du véhicule regroupée en classe (22 pour le moins puissant jusqu'à 46 pour le plus puissant) ;
- *Clause Forfait Kilométrique* : spécifie si l'assuré a déclaré parcourir moins d'un certain nombre de kilomètres par année (5000, 9000, 15 000 ou 20 000 km), ce qui donne droit à une réduction de prime ;

- *Rachat de franchise* : indique si l'assuré a choisi cette option (voir Lexique 97)
- *Score Assuré* : correspond à un score établi par la compagnie pour noter ses clients. Ce score est calculé en fonction de plusieurs critères : la sinistralité de l'assuré au cours des 12 derniers mois, le nombre de contrats dont il dispose chez la compagnie, le montant de prime payé, etc. Elle dispose de trois modalités : «PREMIUM» pour les très bon clients, «NEMASS» pour les bons clients et «AUTRES» pour le reste des assurés ;
- *Sinistralité Assuré* : indique le type de sinistralité qu'un client de la compagnie a eu au cours des 12 derniers mois : au moins un sinistre responsable (malus), au moins un sinistre non responsable ou pas de sinistres ;
- *Usage du véhicule* : indique quel est l'usage principal qui est fait du véhicule (promenade, trajet régulier, tournée ou usage professionnel) ;
- *Zone RC Dommages* : spécifie quel est le niveau d'urbanisation de l'environnement du risque assuré. Cette classification s'étend de A pour les zones les moins dangereuses à R pour les zones les plus dangereuses ;
- *Zone Vol-Incendie* : de la même manière que la zone RC définit le risque de vol et incendie du véhicule dans la zone de circulation habituelle de l'assuré. Cette classification s'étend de A pour les zones où le risque de vol est faible à R pour les zones les plus exposées à ce risque ;

d) Les variables de prix

Ces variables nous seront utiles pour l'étude car elles nous permettent de savoir quel est le montant de prime dans chaque intervalle de temps qui nous intéressent :

- *Prime du devis* : si un prospect décide de concrétiser le devis, cette prime du devis sera la prime qu'il va payer en tant que nouveau client de la compagnie ;
- *Prime de l'assuré avant échéance* du contrat ;
- *Prime de l'assuré après échéance* du contrat : c'est la nouvelle prime que l'assuré va devoir payer lors du renouvellement du contrat ;

e) Les variables cibles

Ce sont les variables binaires que nous allons modéliser en fonction de différents critères à l'aide de régressions logistiques [cf 5.2.1, 28].

- *CN* : si le devis a été concrétisé par le prospect elle est égale à 1 sinon 0. Elle représente l'acte de concrétisation.
- *RES* : si le contrat a été résilié de la part de l'assuré elle vaut 1 sinon 0. Elle illustre l'acte de résiliation.

f) Les variables supplémentaires

Ce sont les variables concernant l'assuré ou son véhicule qui n'interviennent pas sur la tarification des contrats mais que nous avons jugées pertinentes pour notre étude :

- *Ancienneté de l'agent* : elle exprime le nombre d'années d'exercice de l'agent général ;
- *Qualité de souscription* : elle correspond à une note de l'agent général combinant plusieurs facteurs fixés par la compagnie ;

- *Lien titulaire CG* : elle spécifie le lien existant entre le titulaire de la carte grise et le conducteur (lui-même, conjoint, etc.) ;
- *Code fils* : elle précise si l'assuré a un des ses parents qui possède un contrat Auto dans la compagnie ;
- *Client Vie* : elle permet de savoir si l'assuré dispose d'un contrat d'assurance vie dans la compagnie ;
- *Nombre de contrats* : elle correspond au nombre totale de contrats qu'un assuré possède chez la compagnie ;
- *Origine de la résiliation* : si il y a eu résiliation du contrat, cette variable indique si c'est de l'initiative de l'assuré ou de la compagnie ;
- *Motif de résiliation* ;
- *Motif d'entrée d'une image* ;
- *Motif de sortie d'une image*

Les variables décrites précédemment n'interviennent pas toutes dans la base «Concrétisation» et dans la base «Résiliation». Certaines n'ont été utilisées que dans une des deux bases. En revanche, il se peut que certaines (par exemple, les variables de tarification) soient renseignées dans les deux bases. Un tableau récapitulatif des variables présentes dans chaque base est disponible en annexe 3.

Chapitre 8

Analyse Descriptive

L'étude descriptive de nos bases est une étape obligatoire et préliminaire avant la construction de nos modèles car elle nous permet d'analyser nos variables une par une et aussi de chercher les liens significatifs entre nos variables cibles et les autres variables. Cette partie pourra déjà nous donner certains profils d'individus présents dans notre portefeuille en affaires nouvelles et ceux qui ont le plus tendance à résilier un contrat.

8.1 Profils des individus

8.1.1 La concrétisation

Nous avons réalisé, en premier lieu, une étude descriptive des variables concernant les devis concrétisés afin de pouvoir établir une première description des nouveaux clients.

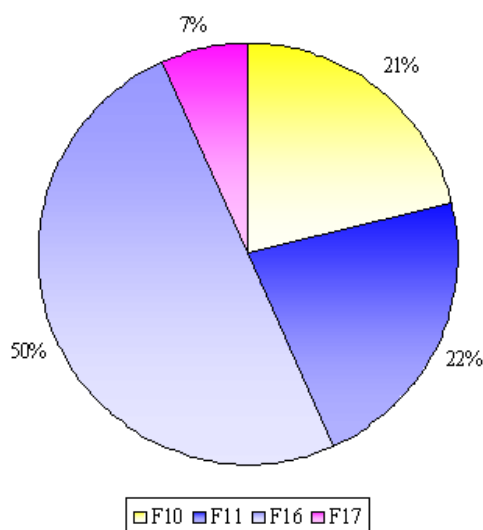


FIGURE 8.1 – Répartition des formules de garanties

Tout d'abord, la formule la plus vendue par les agents est la formule F16 [cf Figure 8.1] : on en déduit que les assurés de notre portefeuille désirent bénéficier d'un maximum de garanties. Seulement 7% des clients ont choisi une formule F17 : cela s'explique par le fait qu'elle est la plus chère et que les garanties supplémentaires à la F16 sont très spécifiques et intéressent peu les particuliers, mais plutôt les petits professionnels.

Variable	Minimum	Moyenne	Médiane	Maximum
Age	18	43,8	43	90
Ancienneté de permis	0	23,1	22	72
Ancienneté de véhicule	0	10,7	10	41
Durée de détention	0	4,1	3	39
CRM	0,5	0,6	0,5	3,04
Prime (TTC)	50	575,7	491	1796

TABLE 8.1 – Statistiques univariées de quelques variables quantitatives

Les affaires nouvelles se caractérisent, entre autres, par les faits suivants :

- l'âge moyen de 43,8 ans ;
- l'ancienneté moyenne du véhicule est de 10,7 ans ;
- les clients sont majoritairement des salariés et mariés (ou en couple) ;
- les véhicules sont majoritairement des berlines et diesels ;
- la plupart des véhicules sont de marque française et possèdent une classe de prix (54,2%) ou une puissance (41,4%) moyenne ;
- les clients ayant déjà un ancien sinistre responsable sont minoritaires ;
- la majorité des clients ont un CRM à 50.

Pour chacune des variables susceptibles d'entrer dans la modélisation, les graphiques représentant la proportion des devis dans la base (noté «% devis»), le taux de concrétisation en fonction des modalités de la variable (noté «Taux CN») ont été tracés. Nous rajoutons à ces graphiques le taux de concrétisation moyen (33,8%) de tous les devis (noté «Taux CN moyen») calculé précédemment :

- $\% \text{ devis} = \frac{\text{Nombre de devis modalité } i}{\text{Nombre total de devis}}$
- $\text{Taux CN} = \frac{\text{Nombre de devis concrétisés modalité } i}{\text{Nombre total de devis modalité } i}$

Par exemple, nous présentons ci-dessous un graphique représentant le taux de concrétisation en fonction de la variable «Âge du conducteur» :

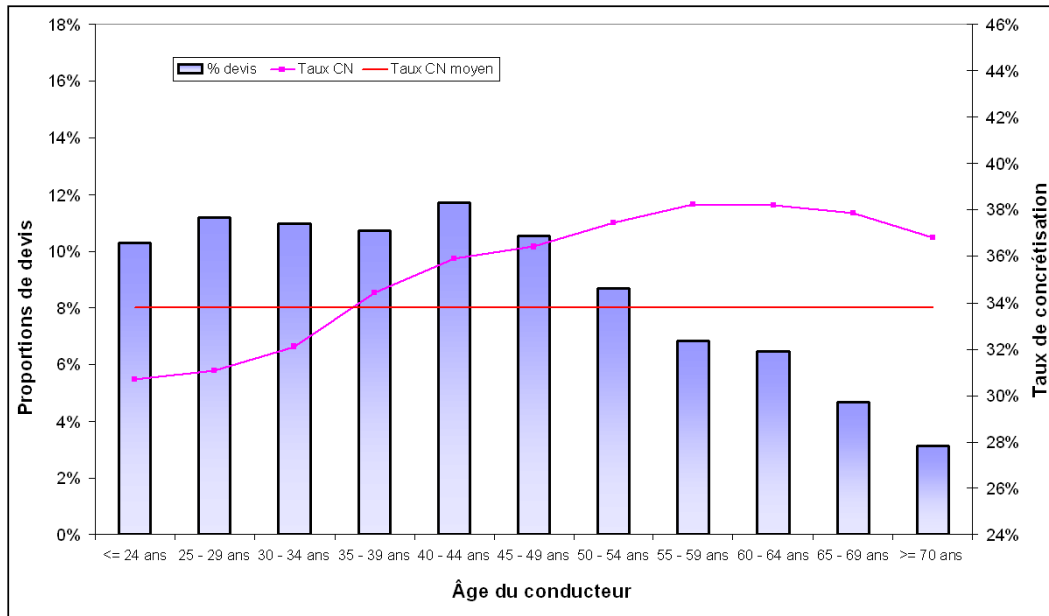


FIGURE 8.2 – Le taux de concrétisation en fonction de l'âge du conducteur

Nous constatons graphiquement que cette variable a un lien avec notre variable cible «CN». En effet, plus l'âge du conducteur augmente plus le taux de concrétisation augmente. Le taux de concrétisation des moins de 24 ans (30,7%) est nettement inférieur au taux de concrétisation moyen de tous les devis. Cela confirme la politique de segmentation faite par les assureurs qui proposent des primes élevées aux jeunes car ils sont moins prudents et se remarquent par une fréquence et une charge de sinistres nettement supérieures à celles des autres conducteurs.

Dans notre cas, la majorité des prospects ayant concrétisé un devis dispose de caractéristiques d'un bon conducteur.

A l'aide des graphiques tels que la figure 8.2, nous pouvons déjà déterminer s'il y a une tendance entre une variable et le taux de concrétisation. Cependant, il faudra effectuer des tests statistiques plus poussés pour déterminer les variables explicatives de nos modèles.

8.1.2 La résiliation

Une étude descriptive a été aussi réalisée sur le type de contrats résiliés de la compagnie. Ces contrats présentent, par exemple, les constats suivants :

- l'âge moyen de l'assuré est de 47,8 ans ;
- l'ancienneté de permis moyenne est de 26,6 ans ;
- la plupart des véhicules sont de marque française, possèdent une ancienneté moyenne de 11,3 ans et 64,7% ont une classe de prix moyenne comprise entre G et M ;
- la majorité des assurés sont des salariés (47,3%) ;
- l'ancienneté de souscription moyenne de ces contrats est de 7 ans ;
- les assurés disposant d'un bon score «PREMIUM» ou «NEMASS» sont minoritaires ;
- l'ancienneté moyenne des clients ayant au moins un contrat (auto, vie, santé, ...) est

de 12,4 ans ;

- les assurés possédant un seul contrat dans la compagnie sont majoritaires.

De la même manière que pour l’acte de concrétisation, pour chacune des variables susceptibles de rentrer dans la modélisation, des graphiques présentant la proportion des contrats dans le portefeuille (noté «% contrats»), le taux de résiliation en fonction des modalités de la variable (noté «Taux RES») ont été tracés. Nous rajoutons à ces graphiques le taux de résiliation moyen (3,2%) de tous les contrats (noté «Taux RES moyen») calculé précédemment.

- $\% \text{ devis} = \frac{\text{Nombre de contrats modalité } i}{\text{Nombre total de contrats}}$
- $\text{Taux RES} = \frac{\text{Nombre de contrats résiliés modalité } i}{\text{Nombre total de contrats modalité } i}$

A l’aide de ces graphiques, nous remarquons des tendances entre les variables et le taux de résiliation.

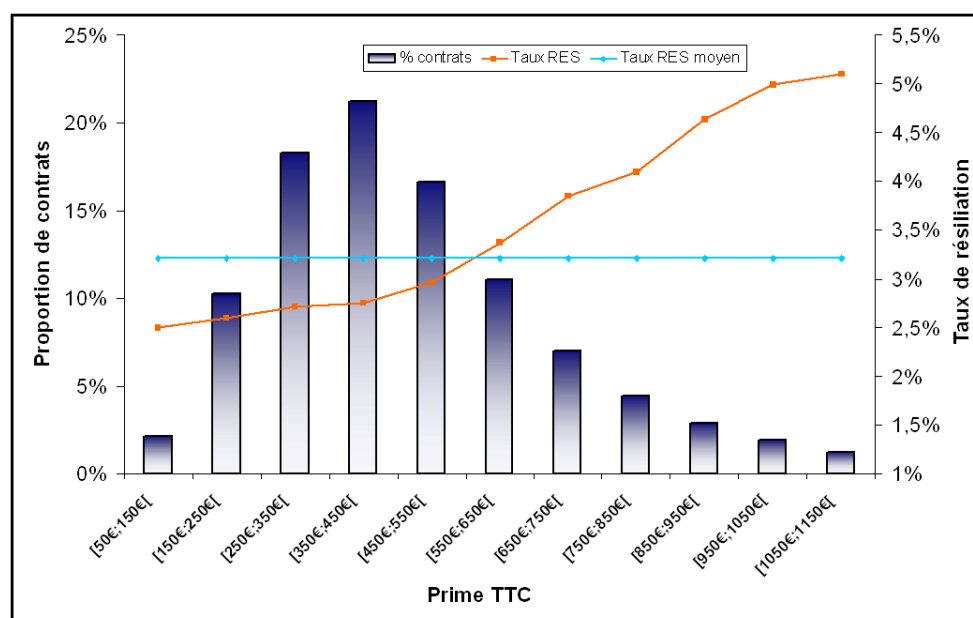


FIGURE 8.3 – Le taux de résiliation en fonction de la prime

Par exemple, la figure 8.3 représente l’évolution du taux de résiliation en fonction du montant de prime payé par l’assuré. Notons que la prime a été découpée en classes de 100€ pour plus de lisibilité. Nous constatons qu’en général plus le montant de prime est élevé plus le taux de résiliation augmente. Ce graphique s’avère intéressant car cette évolution n’est pas strictement linéaire. En effet, pour des montants faibles de prime (entre 50€ et 350€), il y a une faible augmentation, mais au delà de 350€, l’augmentation s’intensifie. L’évolution représentée est alors presque exponentielle. Ainsi, une modélisation du taux de résiliation par une régression linéaire ne peut être envisagée car il est plus probable que l’impact du tarif est d’autant plus élevé que le prix est élevé.

Les graphiques d’analyse descriptive présentant les taux de concrétisation et de résiliation en fonction des autres variables sont listés en annexe. Ces graphiques nous permettent

d'établir les premiers résultats observables sur les deux bases d'étude. Ils permettront aussi de confirmer ou bien, au contraire, de remettre en cause les futurs résultats de la modélisation.

8.2 Corrélations et Mesures d'association

Il convient de retirer de l'étude les variables peu corrélées avec nos variables cibles «CN» et «RES». Afin d'éliminer ces variables non significatives, nous étudions les corrélations entre chaque variable explicative et la variable à expliquer. Pour cela, nous calculerons différentes statistiques et à partir de ces dernières nous effectuerons des tests pour valider la liaison entre une variable explicative X_k et la variable à expliquer Y .

La Statistique du Chi-deux et le test d'indépendance

La méthode consiste à comparer les effectifs réels des croisements des modalités des deux variables qualitatives avec les effectifs théoriques qu'on devrait obtenir dans le cas d'indépendance de ces deux variables. Pour cela, on construit un indice mesurant l'écart constaté entre les effectifs réels et les effectifs théoriques.

$$d = \sum_{i,j} \frac{(n_{ij} - np_{ij})^2}{np_{ij}}$$

où

- n_{ij} = effectif observé des individus possédant la modalité i de la variable à expliquer Y et la modalité j d'une variable explicative X_k ;
- n = effectif total observé ;
- p_{ij} = probabilité d'obtenir une observation possédant la modalité i de Y et la modalité j de X_k lorsqu'elles sont indépendantes ;
- np_{ij} = effectif théorique des individus possédant la modalité i de Y et la modalité j de X_k

Notons que :

$$np_{ij} = \frac{(\sum_j n_{ij})(\sum_i n_{ij})}{n} = \frac{n_{i.}n_{.j}}{n}$$

où $n_{i.}$ est l'effectif des individus possédant la modalité i de Y et $n_{.j}$ l'effectif des individus possédant la modalité j de X_k .

Nous posons les hypothèses suivantes :

$H_0 = \{ \text{la variable cible } Y \text{ et une variable qualitative explicative } X_k \text{ sont indépendantes} \}$
contre $H_1 = \{ Y \text{ et } X_k \text{ sont dépendantes} \}$.

La statistique de test du chi-deux est :

$$D = \sum_{i,j} \frac{(N_{ij} - np_{ij})^2}{np_{ij}}$$

D'après l'expression de D , sous l'hypothèse nulle H_0 :

$\frac{(n_{ij}-np_{ij})}{np_{ij}}$ suit asymptotiquement une loi normale $N(0, 1)$

donc D suit asymptotiquement une loi du χ^2 à $(l-1)(c-1)$ degrés de liberté [cf définition de la loi du χ^2 partie 5.2.3, page 30] où l est le nombre de modalités de Y et c est le nombre de modalités de X_k .

La région critique au niveau asymptotique α s'écrit :

$$W = \{D > \chi^2_{1-\alpha}((l-1)(c-1))\}$$

et la p-valeur associée à ce test est :

$$p - \text{valeur} = P\{\chi^2_{1-\alpha}((l-1)(c-1)) > D\}$$

Ainsi, une valeur de p-valeur inférieure au niveau α indique clairement une forte valeur de D , ce qui signifie d'une liaison significative entre Y et X_k , donc du rejet de l'hypothèse nulle d'indépendance H_0 .

Remarque :

- 1) Pour que l'approximation par la loi du χ^2 soit valable, il est nécessaire que les effectifs théoriques dans chacune des cellules soit au moins égal à 5, ce qui est le cas dans notre étude.
- 2) Ce test s'applique plutôt dans un cadre d'inférence, c'est à dire lorsque l'on dispose d'un échantillon de la population et que l'on souhaite étendre les résultats à l'ensemble de la population. Le test du Chi-deux perd donc de son sens lorsque la population d'étude est constituée de la population totale, car ce test est assez sensible à la taille de l'échantillon.

Nous confirmerons donc systématiquement ce test par d'autres indicateurs tels que :

- Le coefficient de corrélation de Kendall
- Le coefficient de corrélation de Spearman

pour juger de la significativité des liens entre les variable «CN» et «RES» et les variables explicatives.

Le coefficient de corrélation de Kendall

Pour les variables ordinales, c'est à dire lorsque l'on peut ordonner les modalités des variables, nous disposons de statistiques supplémentaires telles que le coefficient tau-b de corrélation de Kendall . En reprenant la définition des «paires concordantes, discordantes et d'ex-æquo» sur l'étude de la corrélation des rangs de la partie 5.2.4 page 32 dans les éléments théoriques, nous pouvons calculer ce coefficient entre notre variable cible Y et une variable X explicative de la manière suivante :

$$\tau_b = \frac{C - D}{[(n^2 - (E_X + E_{XY}))(n^2 - (E_Y + E_{XY}))]^{\frac{1}{2}}}$$

avec

- n est le nombre d'observations
- C le nombre de paires concordantes

- D le nombre de paires discordantes
- E_X le nombre de paires ex-æquo en X
- E_Y le nombre de paires ex-æquo en Y
- E_Y le nombre de paires ex-æquo en Y et en X

On a $-1 \leq \tau_b \leq 1$. Si $\tau_b = 1$ (respectivement $\tau_b = -1$), les classements sont parfaitement identiques (respectivement inverses) et X est une fonction strictement croissante (respectivement décroissante) de Y.

Sous l'hypothèse H_0 d'indépendance de X et Y, r suit une loi Normale de moyenne nulle et de variance $\frac{2(2n+5)}{9n(n-1)}$. La région critique de rejet au niveau α est donc :

$$W = \left\{ \frac{|\tau_b|}{\sqrt{\text{Var}(\tau_b)}} > u_{1-\frac{\alpha}{2}} \right\}$$

où $u_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi Normale centrée réduite (ou $N(0, 1)$).

De la même manière que le test d'indépendance du chi-deux, nous calculons une p-valeur et si elle est inférieure au niveau α alors l'hypothèse H_0 d'indépendance est rejetée.

Le coefficient de corrélation des rangs de Spearman

Si X et Y sont deux variables ordinales observées sur n individus, on peut classer ces derniers en leur attribuant un rang selon la variable X. Ainsi, $r_X = 1$ pour l'individu ayant la plus grande valeur de X. En cas d'ex-æquo, on affecte un rang moyen. On fait la même chose concernant Y. Le coefficient de corrélation de Spearman est le coefficient linéaire entre les variables r_X et r_Y .

$\rho_S(X, Y) = r(r_X, r_Y)$ mesure la ressemblance entre les deux classements. Il est compris entre -1 et 1 et vaut 1 (respectivement -1) si et seulement si les deux classements sont identiques (respectivement inverses l'un de l'autre).

Sous l'hypothèse nulle H_0 d'indépendance de X et Y :

$$T_\rho = \frac{\sqrt{n-2} \times \rho_S}{\sqrt{1-\rho_S^2}} \approx t_{(n-2)}$$

La région critique, au niveau $1 - \alpha$, est donc :

$$W = \{|T_\rho| > t_{1-\frac{\alpha}{2}}(n-2)\}$$

où $t_{1-\frac{\alpha}{2}}(n-2)$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student à n-2 degrés de liberté.

Application sur nos données

Nous calculons donc sur nos deux bases de données, la statistique du chi-deux, le coefficient de corrélation de Kendall et le coefficient de corrélation de Spearman. A l'aide de la p-valeur de chaque test, nous validons ou nous rejetons la liaison entre une variable X_k et une variable cible Y. Voici, par exemple, un tableau rassemblant les résultats pour étudier la liaison entre une variable et la variable «CN» :

Variable	Statistique du CHI-2			Coefficient de KENDALL			Coefficient de SPEARMAN		
	Valeur	P-value	Résultat	Valeur	P-value	Résultat	Valeur	P-value	Résultat
Âge de l'assuré	1862,12	0,000	Liaison	0,08	0,000	Liaison	0,08	0,000	Liaison
Ancienneté de permis	2034,01	0,000	Liaison	0,03	0,000	Liaison	0,03	0,000	Liaison
Ancienneté du véhicule	1819,57	0,000	Liaison	0,07	0,000	Liaison	0,07	0,000	Liaison
CRM	1205,53	0,000	Liaison	-0,05	0,000	Liaison	-0,05	0,000	Liaison
CRM 50	1308,27	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Durée de détention du véhicule	1305,08	0,000	Liaison	0,05	0,000	Liaison	0,06	0,000	Liaison
Mode d'achat	982,76	0,000	Liaison	-0,06	0,000	Liaison	-0,06	0,000	Liaison
Année jeune conducteur	99,06	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Carburant	493,15	0,000	Liaison	0,04	0,000	Liaison	0,04	0,000	Liaison
Antécédents de sinistres	888,92	0,000	Liaison	0,06	0,000	Liaison	0,06	0,000	Liaison
CSP	1057,52	0,000	Liaison	-0,04	0,000	Liaison	-0,05	0,000	Liaison
Mode de garage	1066,52	0,000	Liaison	-0,02	0,000	Liaison	-0,02	0,000	Liaison
Type de véhicule	1698,70	0,000	Liaison	0,26	0,000	Liaison	0,03	0,000	Liaison
Situation familiale	1734,35	0,000	Liaison	0,06	0,000	Liaison	0,06	0,000	Liaison
Classe de prix du véhicule	1492,25	0,000	Liaison	-0,06	0,000	Liaison	-0,07	0,000	Liaison
Puissance du véhicule	1895,70	0,000	Liaison	-0,06	0,000	Liaison	-0,08	0,000	Liaison
Clause Forfait Kilométrique	4957,04	0,000	Liaison	0,01	0,000	Liaison	0,01	0,003	Liaison
Usage du véhicule	2362,70	0,000	Liaison	-0,02	0,000	Liaison	-0,02	0,000	Liaison
Zone RC	818,08	0,000	Liaison	-0,04	0,000	Liaison	-0,05	0,000	Liaison
Zone Vol-Incendie	757,82	0,000	Liaison	-0,04	0,000	Liaison	-0,05	0,000	Liaison
Ancienneté de l'agent	50,16	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Qualité de souscription	0,02	0,881	Indépendance	0,00	0,881	Indépendance	0,00	0,854	Indépendance
Code fils	243,08	0,000	Liaison	0,03	0,000	Liaison	0,03	0,000	Liaison

FIGURE 8.4 – Lien des variables avec la variable «CN»

Ainsi, sur la figure 8.4, nous constatons que toutes les variables excepté la «Qualité de souscription» sont liées à la variable «CN». Les résultats validant la corrélation entre une variable et la variable «RES» sont disponibles en annexe 5.

Remarque : Une étude des corrélations entre les différentes variables explicatives a aussi été réalisée. En effet, deux variables trop corrélées biaiserait les résultats de la modélisation. Nous avons constaté, à l'aide des statistiques décrites précédemment, que les variables ayant une forte corrélation sont :

- l'âge du conducteur et l'ancienneté de permis ;
- l'ancienneté du client et l'ancienneté de souscription ;
- la classe de prix et la puissance du véhicule ;
- Zone RC Dommages et Zone Vol-Incendie.

Ainsi, pour deux variables corrélées, celle qui aura un pouvoir discriminant plus important sur la variable cible, sera préférée à l'autre.

Quatrième partie

Résultats de la modélisation

Chapitre 9

L'acte de concrétisation

Cette partie décrit les étapes et les résultats issus de la modélisation au sein de notre premier environnement de travail : l'acte de concrétisation. Notre premier objectif est de mesurer l'impact du prix sur la souscription ou non d'un devis établi par un agent général d'Aviva Assurances.

Dans un premier temps, nous procéderons à l'analyse d'une variable permettant d'introduire une information liée à la concurrence dans la modélisation du taux de concrétisation. En effet, nous disposons de données concernant l'ancienne prime payée par le prospect chez un concurrent. L'étude d'une telle variable semblerait intéressante pour appréhender des phénomènes de flux de prospects qui peuvent se produire entre différentes compagnies.

Ensuite, il s'agit de déterminer un modèle statistique capable d'estimer la probabilité de concrétisation d'un devis en fonction des caractéristiques du prospect et du prix proposé par l'agent.

Enfin, une fois le modèle mis en place, nous pourrions calculer l'élasticité. Elle nous permettra de visualiser la sensibilité au prix des prospects, de manière globale mais aussi selon certains critères spécifiques intervenant dans la tarification.

9.1 Introduction de la concurrence

Afin d'appréhender les phénomènes de flux qui peuvent se produire entre différentes compagnies dans un environnement concurrentiel, il semble intéressant d'introduire dans nos prochaines étapes une variable faisant appel aux prix pratiqués par la concurrence. Cette variable est renseignée pour environ 35.000 devis (soit 14% de la base «Concrétisation»). A l'aide des caractéristiques de ces devis, notre but est de mettre en place un modèle permettant d'estimer l'ancienne prime d'un prospect chez un concurrent.

La difficulté de cette modélisation consistera à trouver les facteurs permettant une segmentation pour pouvoir se situer sur des cases tarifaires composées d'assurés identiques, tout en conservant des populations suffisamment importantes sur ces segments afin que la modélisation soit robuste.

Notons que plusieurs biais sont d'ores et déjà à prévoir au vu de l'utilisation de cette variable dite «Prime concurrence». En effet, cette dernière provient d'une simple déclaration du prospect lors de l'établissement de son devis :

- Il peut y avoir un décalage entre les assurés déclarant la prime réelle de leur assureur et les assurés qui minorent cette prime dans le but de faire baisser le prix du devis.
- De nombreuses personnes ne connaissent pas le tarif exact de leur assurance, il peut donc souvent s'agir d'approximations.

Choix de la loi exponentielle

Tout d'abord, voici une représentation de la variable «Prime concurrence», découpée en classes de 25€, sur les données observées :

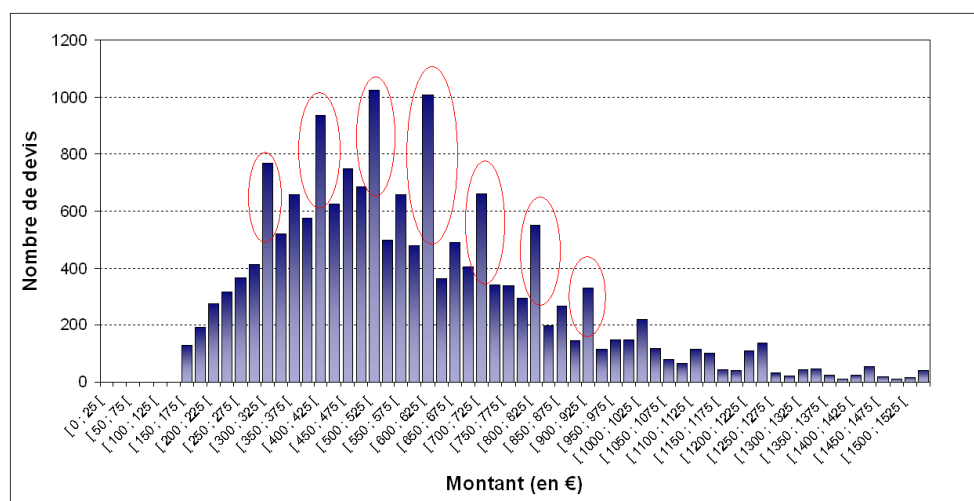


FIGURE 9.1 – Ancienne prime du prospect chez le concurrent

Des pics anormalement élevés sont présents (entourés en rouge) au niveau des classes de montants de centaine d'euros : 300, 400, 500, etc . Ceci confirme l'hypothèse d'approximation du prix donnée par les différents prospects aux agents. Une modélisation de cette variable «brute» ne semble pas envisageable, il est nécessaire d'effectuer un lissage de ces pics. L'objectif est de répartir aléatoirement quelques devis des classes trop volumineuses dans les classes voisines pour obtenir une distribution plus continue et réaliste.

Plusieurs lissages ont été effectués à la suite pour réduire au maximum ces pics, la distribution devient :

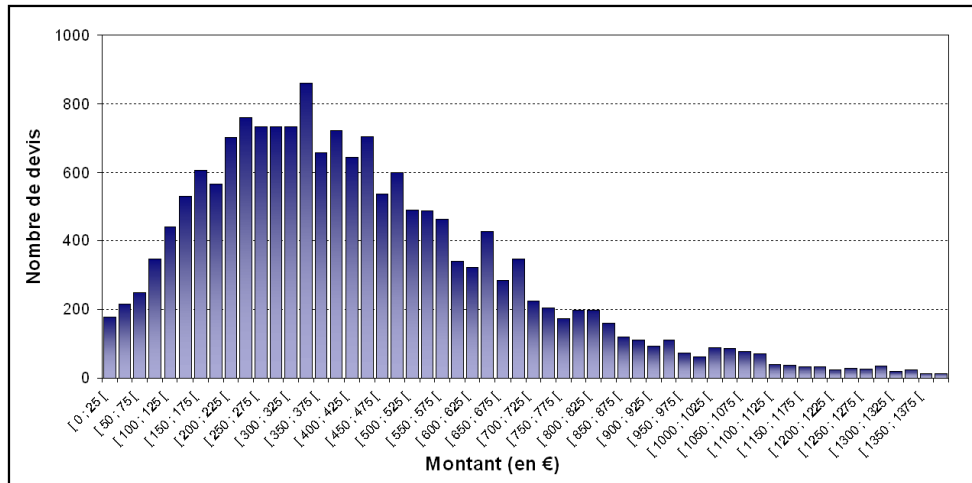


FIGURE 9.2 – Ancienne prime du prospect après lissage

Une modélisation de cette prime par un modèle linéaire généralisé est une possibilité, il faut que la variable à expliquer suive approximativement une loi appartenant à la famille des lois exponentielles. Ensuite, nous choisirons une fonction de lien g permettant la construction du modèle.

Pour ajuster une loi approximative d'une variable continue, le QQ-plot est un très bon indicateur. Ce graphique consiste à comparer les données observées aux données que l'on devrait avoir si elles suivaient « parfaitement » une distribution d'une loi théorique (Normale, Exponentielle, Gamma, ...). Les valeurs observées et « idéales » sont comparées sur un graphe X-Y qui montre une tendance linéaire si la variable à expliquer s'ajuste bien à la loi théorique. Voici deux exemples de QQ-plot :

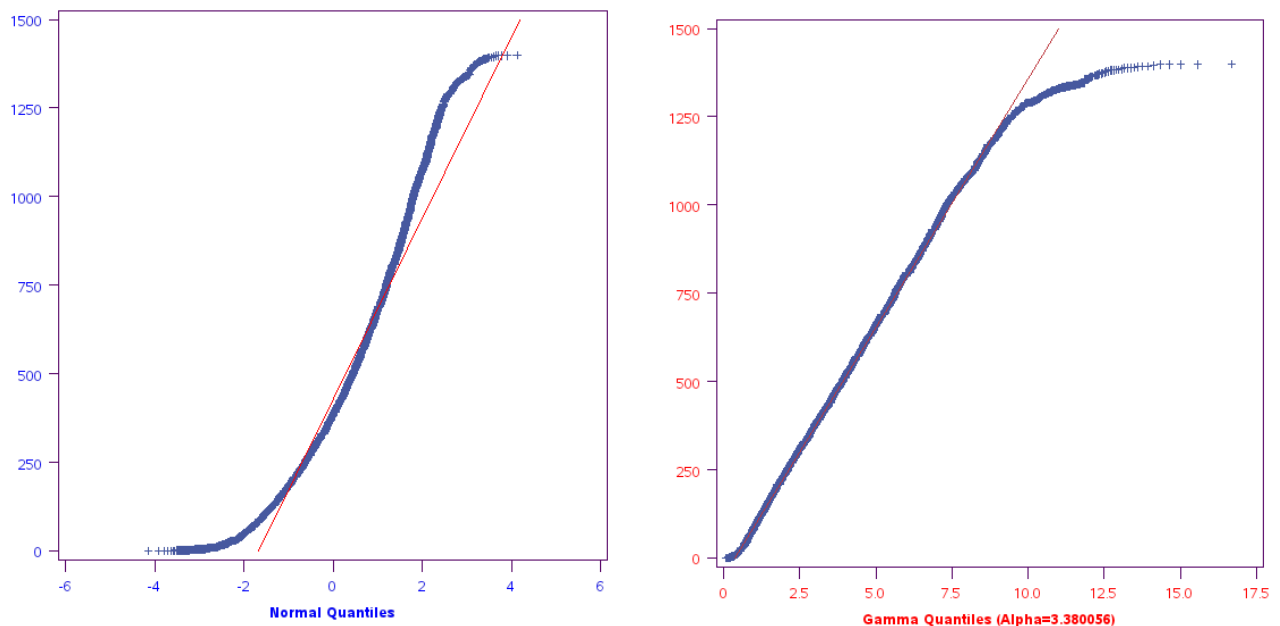


FIGURE 9.3 – QQ-plot d'une loi Normale et QQ-plot d'une loi Gamma

Le premier graphique, à gauche, est un QQ-plot comparant les données observées aux quantiles d'une loi Normale. Les quantiles ne se superposent pas et il n'y a pas de tendance linéaire très claire. Ainsi, il semblerait que les données ne puissent pas être ajustées par une loi Normale. Cependant, sur le deuxième graphique, un QQ-plot comparant nos données aux quantiles d'une loi Gamma, les points sont alignés, sauf pour les montants de prime très élevés qui concernent une minorité. Nous remarquons donc que les données suivent approximativement une loi Gamma.

La représentation graphique suivante de la densité empirique, calculée à partir de nos observations, et de la densité théorique d'une loi gamma ajustée sur les données confirme les constats précédents :

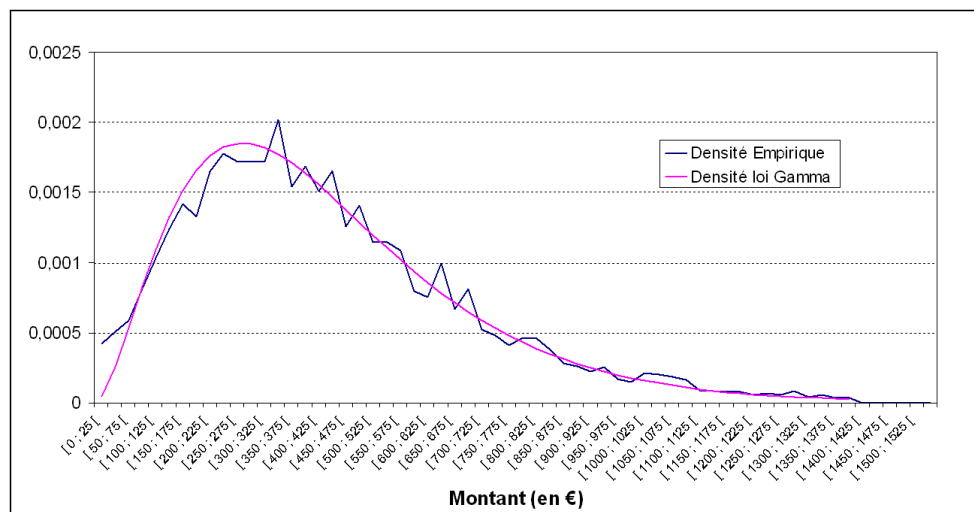


FIGURE 9.4 – Densité empirique et densité théorique

Parmi les autres QQ-plot tracés, nul n'a présenté un meilleur ajustement que celui de la loi Gamma. Nous supposons donc pour la suite que la variable à expliquer «Prime concurrence» suit approximativement une loi Gamma. Maintenant, il nous faut choisir les variables explicatives qui permettront d'estimer cette prime.

Mise en place du modèle

Bien que le lien canonique de la loi Gamma soit la fonction inverse, il est plus fréquent d'utiliser une fonction de lien logarithmique : la forme multiplicative donne des interprétations simples dans le cas des modèles multiples.

Les variables présentant un fort pouvoir discriminant sur l'estimation de la variable à expliquer selon le test de significativité de Wald¹ sont le coefficient de réduction-majoration, la classe de prix du véhicule, la formule de garanties et l'usage du véhicule.

Voici les résultats obtenus à l'aide de la procédure GENMOD² :

1. Il s'agit du même test de significativité décrit pour la régression logistique section 5.2.3, page 31.
2. C'est une procédure du logiciel SAS permettant l'application de la théorie des modèles linéaires généralisés sur des données réelles.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	5,9101	0,0091	5,8923	5,9279	423933	<,0001
CRM >80	1	0,8399	0,0112	0,818	0,8618	5637,24	<,0001
CRM [50,80]	1	0,2348	0,0079	0,2193	0,2503	881,3	<,0001
CRM [60,80]	1	0,5166	0,0069	0,503	0,5302	5528,57	<,0001
Puissance classe SRA faible	1	-0,3361	0,0068	-0,3494	-0,3229	2469,95	<,0001
Puissance classe SRA forte	1	0,3674	0,0071	0,3535	0,3813	2683,57	<,0001
Zone RC grande ville	1	0,1364	0,0068	0,1232	0,1496	408,1	<,0001
Zone RC rurale	1	-0,1008	0,0064	-0,1133	-0,0884	252,03	<,0001
Formule 10	1	-0,7313	0,0077	-0,7464	-0,7162	8986,12	<,0001
Formule 11	1	-0,3633	0,0072	-0,3775	-0,3492	2536,98	<,0001
Formule 17	1	0,0997	0,0104	0,0793	0,1202	91,21	<,0001
CSP ARTI	1	0,0821	0,0116	0,0593	0,1049	49,7	<,0001
CSP AUT	1	-0,0257	0,0103	-0,046	-0,0054	6,17	0,013
CSP FONC	1	-0,0657	0,0091	-0,0836	-0,0478	51,92	<,0001
CSP SANS	1	0,0366	0,0123	0,0125	0,0607	8,87	0,0029
Usage Vie privée/affaires	1	0,1346	0,011	0,1131	0,1561	150,27	<,0001
Usage Vie privée/trajet	1	0,0269	0,0082	0,0108	0,0431	10,68	0,0011
Scale	1	4,153	0,0304	4,0938	4,2131		

FIGURE 9.5 – Estimation des coefficients de régression par maximum de vraisemblance

Afin de bien comprendre la figure précédente, nous allons expliquer les sorties SAS obtenues :

- La colonne «Parameter» indique le nom de la modalité d'une variable explicative. «Intercept» correspond à la constante β_0 . Le paramètre «Scale» correspond au paramètre d'échelle ν de la loi gamma (voir définition 5.1, page 27).
- La colonne «Estimate» fournit les estimations $\hat{\beta}$ (calculées par maximum de vraisemblance) : des coefficients β liés à chaque facteur, de la constante β_0 et du paramètre d'échelle.
- La colonne «Standard Error» indique les écarts types estimés des paramètres associés aux variables explicatives.
- La colonne «Wald 95% Confidence Limits» donne les intervalles de confiance de Wald³ à 95% pour chaque estimation $\hat{\beta}_j$.
- La colonne «Wald Chi-Square» fournit la statistique du chi-deux de Wald. Plus cette statistique est grande, plus le facteur a un pouvoir discriminant élevé sur la variable cible.
- La p-value du test de significativité d'un facteur selon le critère du chi-deux de Wald est donnée dans la colonne «Pr> Chisq». Si elle est inférieure à 5%, ceci indique que le facteur est significatif.

Toutes les variables sont significatives au vu de la p-valeur du test de Wald. Les estimations des paramètres (ou coefficients) disposent d'écart-type très faibles et appartiennent à leurs intervalles de confiance. Pour chaque variable explicative la tendance des coefficients de régression est cohérente. Par exemple, voici une représentation graphique des coefficients de régression estimés et leurs intervalles de confiance à 95% pour la variable Coefficient Réduction-Majoration (CRM) :

3. Une définition de l'intervalle de confiance de Wald est disponible en annexe 2.

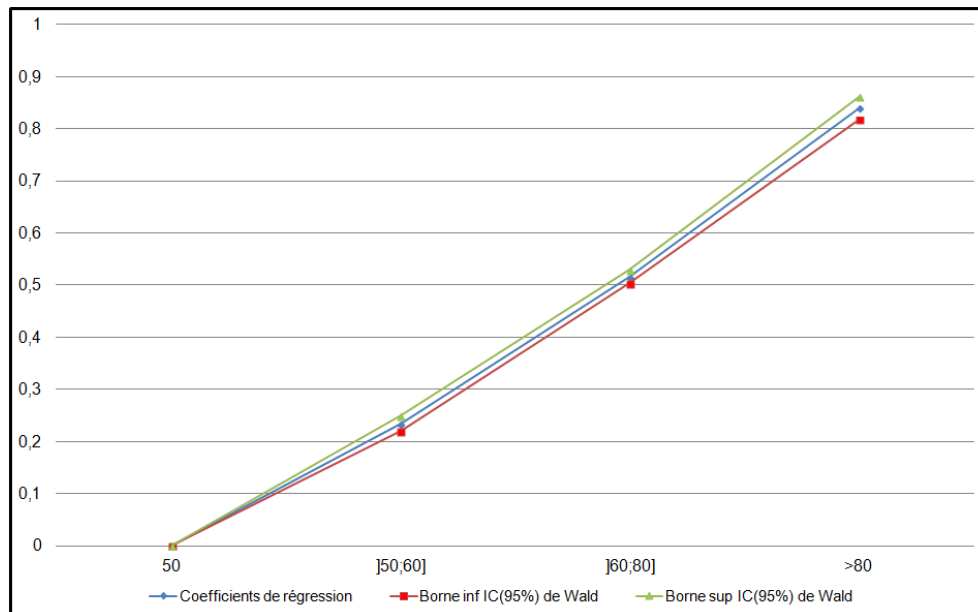


FIGURE 9.6 – Coefficients de régression de la variable CRM

La modalité de référence pour cette variable est la modalité «50», en effet, dans notre base la majorité des prospects ont un CRM de 50%. Le but est d'analyser les estimations des coefficients des autres modalités en fonction de cette modalité où son coefficient de régression est fixé à 0.

Nous observons une tendance croissante des coefficients. Ceci semble cohérent, les données montrent que les conducteurs avec un CRM plus élevé disposent d'une ancienne prime chez le concurrent plus élevée que les autres prospects.

Validation du modèle

Pour juger la qualité du modèle, deux notions vont être utilisées :

- Le rapport de la déviance standardisé sur le nombre de degrés de liberté. Un modèle pertinent présente un rapport inférieur à 1. Ce rapport est sensible au nombre d'observations.
- L'analyse des résidus de la déviance et les résidus de Pearson. Une bonne adéquation du modèle est possible si il n'y a pas de tendance significative des résidus en fonction des valeurs prédites de la variable à expliquer.

La définition théorique de ces dernières notions est disponible en annexe 2. Voici les résultats concernant l'adéquation du modèle :

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	34000	8634	0,25
Scaled Deviance	34000	35856	1,04
Pearson Chi-Square	34000	7782	0,23
Scaled Pearson X2	34000	32318	0,94
Log Likelihood		-226939	
Full Log Likelihood		-226939	
AIC (smaller is better)		453914	
AICC (smaller is better)		453914	
BIC (smaller is better)		454066	

FIGURE 9.7 – Indicateurs sur l'adéquation du modèle

La déviance standardisée («Scaled Deviance») du modèle est de 35 856. Le nombre de degrés de liberté est de 34 000. Ainsi, le rapport de cette déviance sur le nombre de degrés de liberté est de 1,04. Pour l'instant, le modèle ne semble pas assez pertinent mais ce rapport reste très proche de 1. Si le nombre d'observations était plus élevé, ce rapport serait probablement inférieur à 1. Analysons maintenant les résidus :

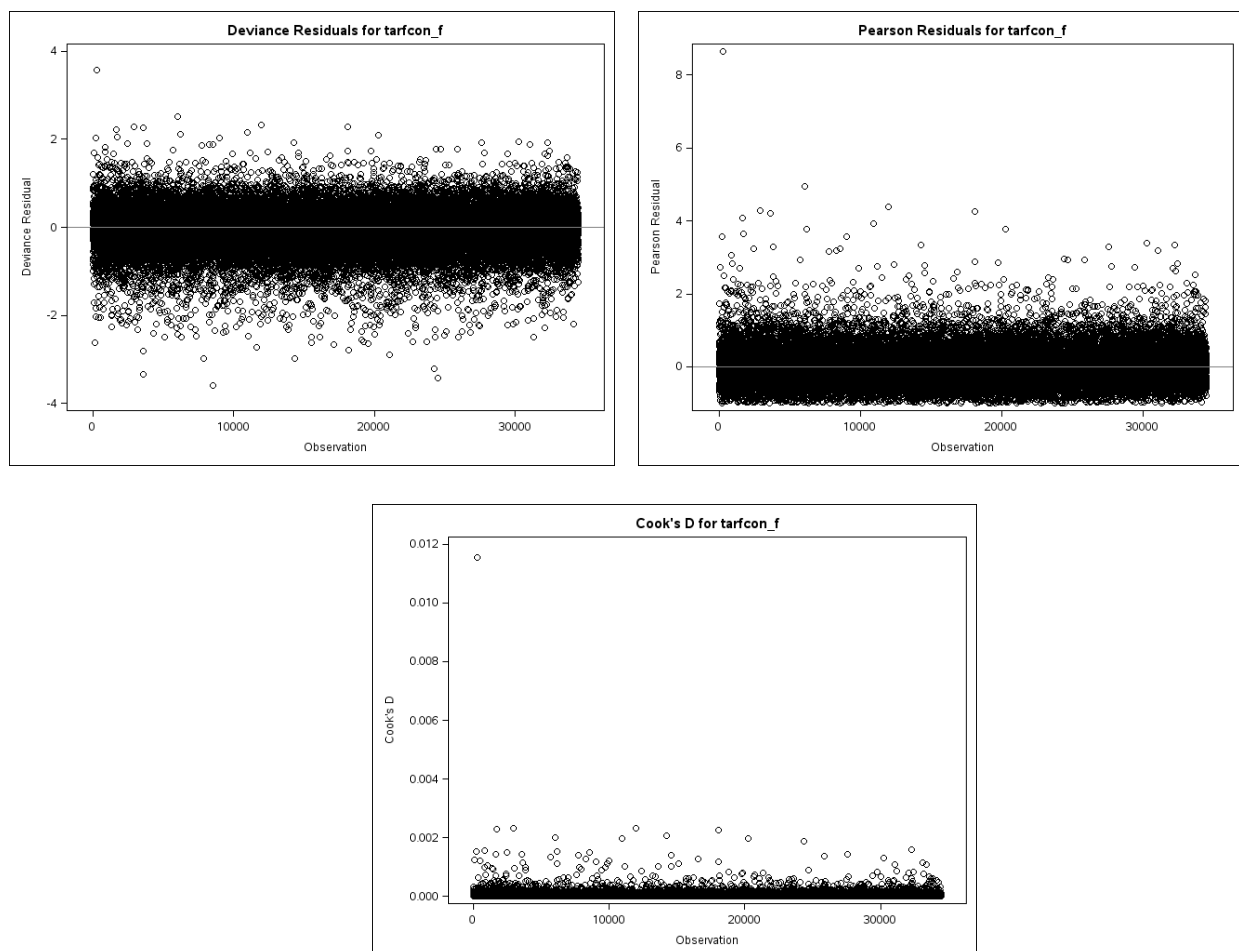


FIGURE 9.8 – Analyse des résidus

Le premier graphique, en haut à gauche, est une représentation des résidus en fonction des valeurs prédites, l'absence de tendance significative et l'équidispersion des points autour de la droite d'équation $y = 0$ indique une bonne adéquation du modèle.

Le second graphique, en haut à droite, est une représentation des résidus de Pearson en fonction des valeurs prédites, il permet aussi de vérifier la présence ou l'absence d'une tendance significative. Nous constatons là aussi une absence de tendance, ceci indique une bonne adéquation du modèle

Enfin, le dernier graphe est celui des distances de Cook. Il permet de connaître l'influence de chaque observation sur le modèle. Une distance supérieure à 1 sera considérée comme anormale. Nous remarquons que toutes les observations de l'étude sont convenables ici et sont donc bien ajustées.

En résumé, l'analyse des résidus met en exergue un bon ajustement du modèle. Nous gardons donc les résultats trouvés et nous allons calculer l'écart tarifaire entre la prime du devis Aviva et l'ancienne prime du prospect chez le concurrent. L'objectif est de visualiser graphiquement une tendance significative et intuitive entre le taux de concrétisation et cet écart tarifaire.

Pour cela, à l'aide des estimations des coefficients de régression, il a été possible d'estimer la «Prime concurrence» des autres devis. Une fois, cette tâche effectuée, pour chaque devis, l'écart tarifaire⁴ (ou «Ecart prix») a été calculé. En reprenant les mêmes notations que l'analyse descriptive [cf. section 8.1.1, page 48], voici un graphique présentant la proportion des devis («% devis»), le taux de concrétisation selon chaque classe de l'écart tarifaire («Taux CN») et le taux de concrétisation moyen («Taux CN moyen») :

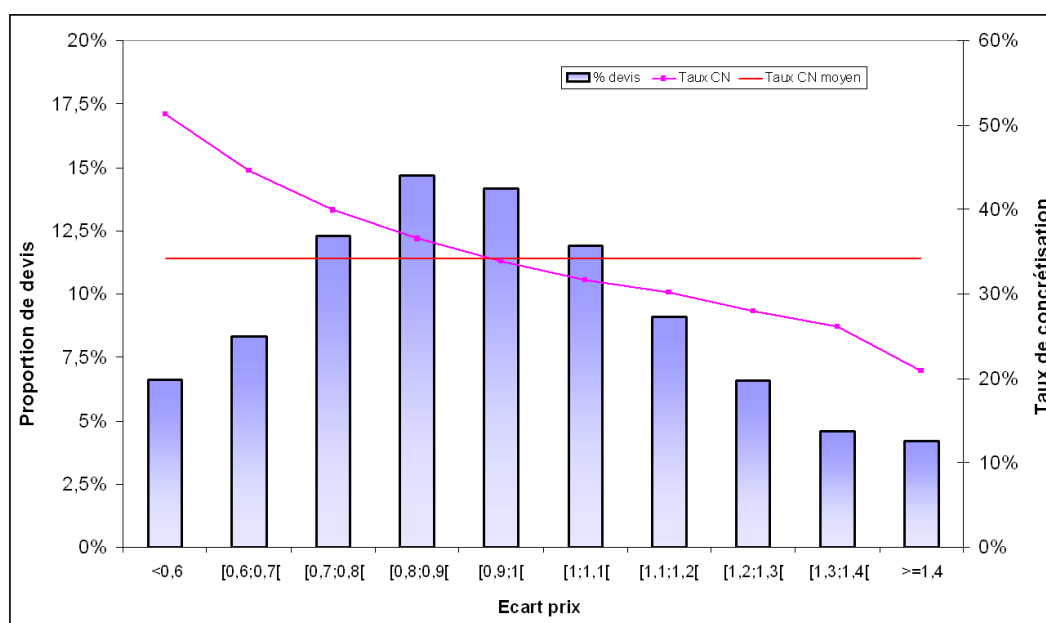


FIGURE 9.9 – Évolution du taux de concrétisation en fonction de l'écart tarifaire

Une tendance quasi-linéaire entre le taux de concrétisation et l'écart tarifaire apparaît sur ce graphique. Nous remarquons que plus l'écart tarifaire augmente, plus le taux diminue. Ceci semble logique car plus la prime d'un devis Aviva est chère par rapport au tarif

4. C'est le rapport entre la prime Aviva et la variable "Prime Concurrence".

d'un concurrent (c'est à dire que l'écart est supérieur à 1), plus le prospect aura tendance à ne pas concrétiser le devis.

Au regard de ces résultats, cette modélisation s'est avéré utile et donne des résultats pertinents. L'ancienne prime du prospect est une information directement liée à la concurrence et donc précieuse. Nous verrons par la suite que l'estimation obtenue de cette prime sera une variable explicative assez significative dans la modélisation du taux de concrétisation.

9.2 Modélisation du taux de concrétisation d'un devis

Suite à l'analyse descriptive [cf 8.2, page 53], les variables retenues de la base «Concrétisation» pour la modélisation logistique du taux de concrétisation sont :

Variables		
Âge du prospect	Code fils	Prime concurrence
Ancienneté de l'agent	CRM 50	Prime du devis
Ancienneté du véhicule	CSP	Puissance du véhicule
Année jeune conducteur	Durée de détention du véhicule	Situation familiale
Antécédents de sinistres	Formule de garanties	Type de véhicule
Carburant	Mode d'achat	Usage du véhicule
Clause Forfait Kilométrique	Mode de garage	Zone RC

FIGURE 9.10 – Liste des variables explicatives pour la modélisation du taux de concrétisation

Nous mettons en place un premier modèle (noté Modèle A) estimant la probabilité de conversion d'un devis d'un prospect en fonction de trois effets dissociables :

- Un effet lié aux caractéristiques du prospect et indépendant du prix.
- Un effet dépendant uniquement du prix quel que soit le segment de population auquel appartient le prospect.
- Un effet dépendant uniquement du tarif de la concurrence estimé lors de la partie précédente quel que soit le segment de population auquel appartient le prospect.

Ce modèle fera intervenir des variables modales explicatives, la prime associée au devis et le tarif de la concurrence estimé. Ces deux dernières variables interviendront de façon continue dans notre modélisation tandis que les variables qualitatives sont découpées en indicatrices pour chaque modalité.

Après une sélection de type «stepwise», le tableau suivant représente les variables toutes suffisamment significatives (au seuil de 95%) présentes dans le modèle A :

Modèle A		
Âge du prospect	CRM 50	Prime du devis
Ancienneté de l'agent	CSP	Puissance du véhicule
Ancienneté du véhicule	Durée de détention du véhicule	Situation familiale
Année jeune conducteur	Formule de garanties	Type de véhicule
Antécédents de sinistres	Mode d'achat	Usage du véhicule
Carburant	Mode de garage	Zone RC
Clause Forfait Kilométrique	Prime concurrence	

FIGURE 9.11 – Variables explicatives du modèle A

Par ailleurs, nous désirons mesurer aussi un impact du prix pour chaque segment de prospect. En effet, suivant le segment auquel il appartient, un prospect n'a pas la même sensibilité au prix pour diverses raisons. Afin de mesurer cet impact supplémentaire, nous introduirons dans le modèle A l'ensemble des variables qualitatives croisées avec le niveau de prime. La variable «Prime concurrence» a été recodifiée en plusieurs modalités représentant des tranches de prime de 100 euros et l'effet croisé avec le niveau de prime proposé sera introduit dans ce nouveau modèle.

Après une sélection de type «stepwise», le tableau suivant représente les variables toutes suffisamment significatives (au seuil de 95%) présentes dans le nouveau modèle (noté Modèle B) :

Modèle B		
<i>Effet de la variable</i>	<i>Direct</i>	<i>Croisé</i>
Âge du prospect	•	•
Ancienneté de l'agent	•	
Ancienneté du véhicule	•	•
Année jeune conducteur	•	
Antécédents de sinistres	•	•
Carburant	•	•
Clause Forfait Kilométrique	•	•
Code fils		
CRM 50	•	•
CSP	•	•
Durée de détention du véhicule	•	•
Formule de garanties	•	•
Mode d'achat	•	•
Mode de garage	•	
Prime concurrence	•	•
Prime du devis	•	
Puissance du véhicule	•	•
Situation familiale	•	•
Type de véhicule	•	•
Usage du véhicule	•	•
Zone RC	•	

FIGURE 9.12 – Variables explicatives du modèle B

Nous représentons certaines statistiques décrivant la qualité des modèles A et B. Elles nous permettront aussi de les comparer :

Modèle	P-valeur	AIC	SC %	Paires concordantes	Somer's D
A	<0,0001	294 759	295 575	68,8%	0,40
B	<0,0001	292 954	294 584	70,3%	0,41

TABLE 9.1 – Indicateurs sur la qualité des modèles A et B

Pour les deux modèles, les p-values associées au test de l'hypothèse nulle, considérant que tous les coefficients associés aux variables explicatives sont nuls, sont inférieures à 1%. Nous rejetons donc cette hypothèse pour les deux modèles et considérons que chaque modèle est significatif à plus de 99%. Le nombre de paires concordantes dépasse largement notre seuil de validité fixé à 60% avec des taux à 68,8% pour le modèle A et 70,3% pour le modèle B.

L'AIC⁵ (ou le SC⁶), pris séparément, est un type d'indicateur très difficile à analyser, mais cela s'avère utile lorsque l'on effectue des comparaisons entre les modèles comportant un nombre différent de variables. Plus la vraisemblance est grande (c'est-à-dire plus on rajoute de paramètres), plus l'AIC diminue, meilleur est le modèle. Dans notre cas, la valeur de ce critère pour le modèle B est plus faible que celle du modèle A. Le modèle B disposant d'un plus grand nombre de variables explicatives peut être jugé comme meilleur que le modèle A.

Cependant, le modèle B présente un nombre de variables explicatives trop élevé et il n'est pas assez simple pour pouvoir être utilisé de façon courante par l'entreprise. Pour obtenir un modèle simplifié, nous allons effectuer une sélection supplémentaire sur ces variables afin de conserver arbitrairement un certain nombre de variables :

- Les variables conservées doivent avoir un impact important sur le tarif ;
- Elles doivent être susceptibles d'intervenir dans la tarification de la concurrence ;
- Elles doivent être assez peu corrélées entre elles ;
- Nous préférons des variables qui font intervenir un faible nombre de modalités avec des effectifs sensiblement égaux par classe.

Ainsi, nous obtenons un troisième modèle (Modèle C) avec les variables explicatives suivantes :

Modèle C		
Âge du prospect	Durée de détention du véhicule	Prime*CSP
Ancienneté du véhicule	Formule	Prime*Formule
Année jeune conducteur	Prime	Prime*Puissance du véhicule
Antécédents de sinistres	Prime concurrence	Prime*Type de véhicule
Clause Forfait Kilométrique	Prime* Prime concurrence	Puissance du véhicule
CRM 50	Prime*Ancienneté du véhicule	Type du véhicule
CSP	Prime*CRM 50	Usage du vehicule

FIGURE 9.13 – Variables explicatives du modèle C

5. C'est le critère d'information d'Akaike, voir définition section 5.2.4, page 32.

6. C'est le critère de Schwartz, voir définition section 5.2.4, page 32.

Voici, les mêmes indicateurs présents dans le tableau 9.1, concernant le modèle C :

Modèle	P-valeur	AIC	SC	% Paires concordantes	Somer's D
C	<0,0001	293 886	294 974	69,1%	0,40

TABLE 9.2 – Indicateurs sur la qualité du modèle C

Les mesures d'associations du modèle C indiquent que lorsque l'on retire des variables de la modélisation, la prédiction de la probabilité de concrétisation est moins fiable : on passe de 70,3% de paires concordantes à 69,1%. Pourtant, au regard des critères de l'AIC et du SC, nous n'avons pas perdu beaucoup d'efficacité entre le modèle B et C. En fait, lorsque l'on retire des variables explicatives du modèle, nous ne nous éloignons pas beaucoup de la valeur réelle du paramètre à estimer, la probabilité de concrétisation. Ce phénomène est dû principalement à la faible variance de la variable cible et à la nature du modèle qui est décomposé en indicatrices. Nous savons, en effet, que, tant que nous garderons un nombre suffisant de variables qualitatives explicatives, la valeur prédite de la variable à expliquer sera très proche de la valeur réelle.

Intéressons-nous maintenant au taux de bons reclassements des individus. Pour prédire la variable cible, il suffit de choisir arbitrairement le seuil de probabilité que l'on considère déterminant pour décider si le contrat est concrétisé ou non. Comme nous considérons équivalents les mauvais reclassements dans les deux sous-populations, nous avons choisi comme seuil le niveau qui nous donnait des taux de mauvais reclassement identiques pour chaque sous population. Ce seuil est de 33,76%.

Nous obtenons un taux de bon reclassement de 68% pour le modèle B et 66% pour le modèle C. Les 2 sous-populations ne sont pas si faciles à différencier avec les variables dont nous disposons. Cependant, le but de cette étude ne s'attache pas particulièrement à expliquer le comportement du prospect en fonction de ses caractéristiques mais plutôt à mesurer l'impact du niveau de prime. Nous accepterons donc ce taux de bon reclassement en admettant que nous ne disposons pas de suffisamment d'information pour l'améliorer.

Ainsi, nous choisirons le modèle C pour la suite de l'étude et nous obtenons une probabilité de concrétisation d'un devis pour un prospect :

$$\hat{\tau} = \frac{1}{1 + \exp[-(a_0 + a_1(i_1)I_{x_1=i_1} + \dots + a_n(i_n)I_{x_n=i_n} + (b_0 + b_1(i_1)I_{x_1=i_1} + \dots + b_m(i_m)I_{x_m=i_m}) \times p]}$$

où

- a_0 et le coefficient associé à la constante ;
- les $a_k(i_k)$ sont les coefficients associés aux variables qualitatives x_k ;
- b_0 est le coefficient associé au niveau de prime p ;
- $b_l(i_l)$ sont les coefficients associés au prix p suivant les modalités des variables qualitatives x_l et du prix pratiqué par la concurrence.

Connaissant les caractéristiques du prospect, nous pouvons regrouper les coefficients précédents sous la forme de deux coefficients :

- Un coefficient ne faisant pas intervenir la prime du devis et qui correspond donc aux caractéristiques du prospect : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)I_{x_k=i_k}$;
- Un coefficient lié aux caractéristiques du prospect et au prix p : $B = B(x_1, \dots, x_m) = b_0 + \sum_{l=1}^m b_l(i_l)I_{x_l=i_l}$;

La probabilité de concrétisation s'écrit alors pour chaque prospect i :

$$\hat{\tau}(i) = \frac{1}{1 + \exp[-(A(i) + B(i)p(i))]}$$

que l'on peut écrire de la manière simplifiée :

$$\hat{\tau}(p) = \frac{1}{1 + \exp[-(A + Bp)]}$$

Les résultats des modèles concernant le test de Wald jugeant la significativité des variables et les mesures d'association sont disponibles en annexe 6.

9.3 Élasticité au prix

Pour mesurer l'impact d'une variation de tarif sur le comportement des prospects, nous utiliserons par la suite, un indicateur couramment utilisé en micro-économie : **l'élasticité au prix**⁷. Elle s'écrit :

$$\hat{\epsilon}(p) = \frac{\partial \hat{\tau}(p)}{\partial p} \times \frac{p}{\hat{\tau}(p)} \quad (9.1)$$

Précédemment, nous avons mis en place un modèle (Modèle C) permettant d'estimer une probabilité de concrétisation. Ainsi, il est possible avec la forme de cette probabilité de calculer une sensibilité de chaque prospect à son niveau de prime. Pour cela, il suffit de dériver la fonction qui permet de calculer cette probabilité en fonction de la prime.

Calcul et propriétés de l'élasticité

D'après la forme de la probabilité à concrétiser un devis, il vient :

$$\hat{\epsilon}(p) = \frac{\partial \hat{\tau}(p)}{\partial p} \times p \times (1 + e^{-(A+Bp)})$$

avec

$$\frac{\partial \hat{\tau}(p)}{\partial p} = \frac{-Be^{-(A+Bp)}}{(1 + e^{-(A+Bp)})^2}$$

D'où

$$\hat{\epsilon}(p) = \frac{Bpe^{-(A+Bp)}}{1 + e^{-(A+Bp)}} = Bp(1 - \hat{\tau}(p))$$

La prime est toujours définie positive. Le taux de refus du devis (qui est le complémentaire au taux de concrétisation) est compris entre 0 et 1. Le signe de l'élasticité est

7. Références :

– HENRIET D., ROCHET J.-C. (1991) Microéconomie de l'assurance. Economica.
– Notions d'élasticité et des applications : <http://iml.univ-mrs.fr/~lafont/licence/CHAP2.pdf>

donc uniquement déterminé par celui du coefficient B .

Nous considérerons dans l'étude aux limites qui suit que le coefficient B est négatif, ce qui sera globalement le cas, comme nous le verrons par la suite :

- Lorsque le prix tend vers 0, $\hat{\epsilon}(p) \approx \frac{Bp}{(1+e^A)} \xrightarrow{p \rightarrow 0} 0$
- Lorsque le prix tend vers l'infini, $\hat{\epsilon}(p) \approx (e^{-A}) \times Bpe^{-Bp} \xrightarrow{p \rightarrow \infty} 0$

Nous pouvons donc prédire le comportement de l'assuré en fonction du niveau de prime auquel il sera soumis. La variation de probabilité de concrétisation des devis en fonction de la variation de prime d'une année sur l'autre, pour une variation infinitésimale, s'écrit :

$$\boxed{\frac{\Delta \hat{\tau}(p)}{\hat{\tau}(p)} = \frac{\Delta p}{p} \times \hat{\epsilon}(p)} \quad (9.2)$$

Une élasticité presque nulle indique donc que l'assuré n'est absolument pas sensible au prix qu'il paye. Si son tarif varie, sa probabilité de concrétiser son devis n'en sera pas affectée. A l'inverse, si l'élasticité est élevée en valeur absolue, l'assuré est très sensible au niveau de tarif qu'il paye.

Remarque : Il est techniquement possible d'observer des élasticités positives de la probabilité de concrétiser un devis en fonction du niveau de prime. Cela signifierait que l'assuré préférerait avoir une prime plus élevée. Ce phénomène n'est pas très courant et apparaît le plus souvent pour des produits de luxe pour lesquels un prix élevé est gage de qualité ou de prestige. Le produit d'assurance ne peut pas être considéré comme un produit de luxe.

L'observation de valeurs positives d'élasticité indiquerait plus probablement un biais qu'il n'a pas été possible de prendre en compte dans l'étude.

Résultats

Sur l'ensemble des devis nous obtenons les résultats suivants :

Le taux de concrétisation moyen estimé est de 33,77%. Il est égal au taux de concrétisation moyen observé. Comme nous l'avons déjà vu précédemment, ce résultat est conforme à nos attentes puisque l'estimé est égale à l'observé.

L'élasticité moyenne estimée est de -2,1. Le niveau de prime moyen proposé lors des devis sur l'année 2012 étant de 672 €, ces affaires nouvelles génèrent un chiffre d'affaires de 58,2 M€.

D'après l'équation précédente (9.2) , nous pouvons effectuer certains scénarios de hausse ou de baisse des tarifs en supposant que les tarifs de la concurrence sont inchangés :

Scénario		Prime TTC	Variation du taux	Nouveau taux	CA	Variation du CA
Hausse	1%	679 €	-0,71%	33,09%	57 551 138 €	-1,10%
	2%	685 €	-1,42%	32,38%	56 874 230 €	-2,26%
	3%	692 €	-2,13%	31,67%	56 172 876 €	-3,46%
Baisse	-1%	665 €	0,71%	34,51%	58 831 617 €	1,10%
	-2%	659 €	1,42%	35,22%	59 435 188 €	2,14%
	-3%	652 €	2,13%	35,93%	60 014 314 €	3,14%

FIGURE 9.14 – Scénarios de hausse ou de baisse des tarifs

Par exemple, si l'on avait effectué une hausse de 1% sur les tarifs, le niveau de prime moyen aurait été de 679 €. Le taux de concrétisation aurait été de 33,1%, soit une baisse de 0,71%, et le chiffre d'affaires généré aurait été de 57,6 M€, soit une perte de 1,1%. Cependant, si l'on avait effectué une baisse de 1% sur les tarifs, le niveau de prime moyen aurait été de 665 €. Le taux de concrétisation aurait été de 34,5% et le chiffre d'affaires généré aurait été de 58,8 M€, soit une augmentation de 1,1%.

L'analyse des résultats peut être effectuée par variable pour obtenir des informations plus précises sur l'impact du tarif parmi les différents segments de prospects. Nous constaterons que certains prospects sont plus sensibles au prix que d'autres. Nous nous intéresserons principalement dans cette partie aux variables «Âge du conducteur», «Catégorie Socio-Professionnelle» et «Classe de prix du véhicule».

Lorsque l'on compare deux élasticités ou deux taux de concrétisation différents pour des modalités différentes, il faut aussi comparer les niveaux de prime auxquels sont soumis les prospects de ces deux catégories. En effet, le taux de concrétisation et l'élasticité sont dépendants du niveau de prime appliqué sur chaque modalité.

Les tableaux suivants [Figure 9.15, 9.17, 9.19] reprennent pour chaque variable, l'effectif, le niveau de prime moyen, le taux de concrétisation moyen estimé, l'élasticité estimée par modalité et le chiffre d'affaires généré par les devis concrétisés. Les graphiques⁸ [Figure 9.16, 9.18, 9.20] représentent l'écart à la moyenne de la prime et de l'élasticité pour chaque modalité. Ces graphiques nous permettront de comparer le degré de sensibilité de chaque segment au prix.

1) L'âge du conducteur

Âge du conducteur	Effectif	Elasticité	Prime TTC	Taux de concrétisation	CA
moins de 24 ans	32350	-2,68	1 082 €	27,51%	9 630 902 €
25 - 29 ans	29568	-2,50	903 €	27,35%	7 299 293 €
30 - 34 ans	28948	-2,14	713 €	32,68%	6 790 359 €
35 - 39 ans	28295	-1,99	615 €	33,29%	5 790 398 €
40 - 44 ans	30922	-1,92	583 €	35,73%	6 445 874 €
45 - 49 ans	27814	-1,91	570 €	36,32%	5 755 833 €
50 - 54 ans	22870	-1,85	530 €	37,37%	4 532 923 €
55 - 59 ans	17956	-1,80	496 €	37,96%	3 383 023 €
60 - 64 ans	17058	-1,77	482 €	37,91%	3 118 010 €
65 - 69 ans	12250	-1,75	464 €	37,53%	2 133 627 €
70 ans et plus	8219	-1,76	450 €	37,11%	1 373 081 €
Portefeuille	256250	-2,08	672 €	33,78%	58 188 876 €

FIGURE 9.15 – Résultats sur le critère de l'âge

8. Les résultats concernant d'autres variables sont disponibles en annexe 7.

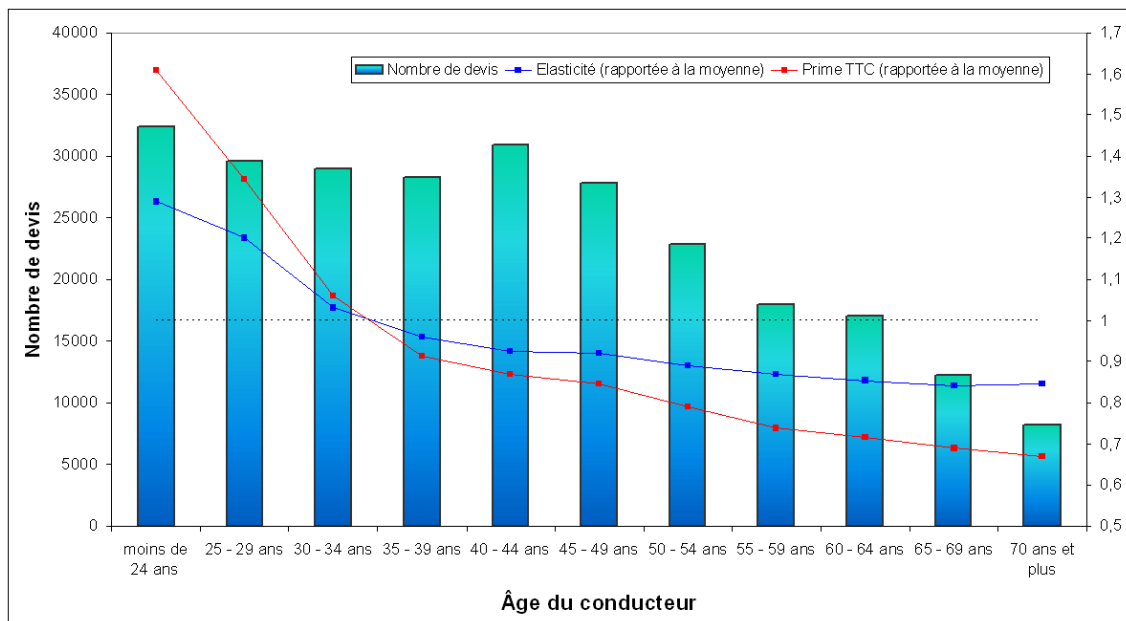


FIGURE 9.16 – Écart à la moyenne de l'élasticité et de la prime pour l'âge

Sur le graphique ci-dessus, l'élasticité évolue de la même manière que la prime moyenne. En effet, plus le prospect est âgé plus le montant de prime diminue et plus l'élasticité diminue. Ceci peut se justifier par les études de sinistralité et la politique de tarification appliquée par la compagnie sur ce critère : les jeunes conducteurs sont considérés comme «des conducteurs inexpérimentés» avec une fréquence et d'un coût moyen de sinistres nettement supérieur à ceux des autres conducteurs. Ainsi, la prime du devis proposée par l'agent est élevée et un jeune conducteur sera donc plus sensible au prix que les autres prospects. Par exemple, dans notre cas, les moins de 24 ans possèdent une prime de 1082€ et une élasticité de -2,68 nettement supérieur à l'élasticité moyenne. Cependant, pour les classes d'âge au dessus de 35 ans, les conducteurs sont assez responsables et considérés comme de bons risques par la compagnie. Ainsi leur montant de prime diminue, ce qui induit une augmentation du taux de concrétisation et donc une baisse de l'élasticité.

La sinistralité n'est pas la seule cause pour les valeurs d'élasticité décrite dans le tableau ci-dessus. En effet, le pouvoir d'achat est à prendre en compte dans cette analyse. En général, les jeunes ont des moyens financiers plus faibles que les autres conducteurs et certains d'entre eux dépendent encore financièrement de leurs parents. Ainsi, la prime élevée qui leur est proposée ne fait qu'augmenter leur sensibilité.

2) La catégorie Socio-Professionnelle

CSP	Effectif	Elasticité	Prime TTC	Taux de concrétisation	CA
Agriculteur	8063	-1,71	523 €	40,57%	1 709 538 €
Artisan, Ent	35758	-1,90	716 €	38,10%	9 756 097 €
Autres (Retraités,...)	30169	-1,75	449 €	36,68%	4 965 423 €
Etudiant	14319	-2,67	1 040 €	26,78%	3 988 597 €
Fonctionnaire	20526	-2,23	590 €	29,95%	3 629 243 €
Salarié	138207	-2,13	693 €	33,33%	31 924 279 €
Demandeur d'emploi	9208	-2,05	666 €	27,60%	1 692 501 €
Portefeuille	256250	-2,08	672 €	33,78%	58 188 876 €

FIGURE 9.17 – Résultats sur le critère de la catégorie Socio-Professionnelle

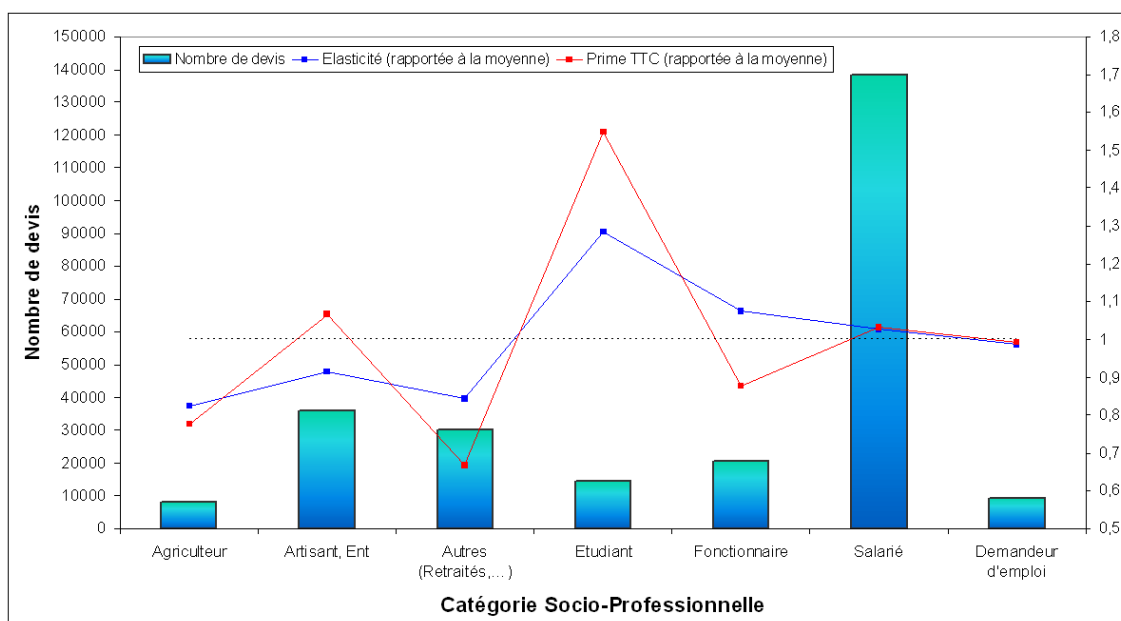


FIGURE 9.18 – Écart à la moyenne de l'élasticité et de la prime pour la CSP

En ce qui concerne l'élasticité, on peut remarquer de grandes disparités selon les différentes catégories socio-professionnelles définies.

Les agriculteurs circulent habituellement dans des zones rurales et ont une probabilité plus faible d'avoir un accident que les autres. Ceci implique donc une prime moins élevée, d'où un fort taux de concrétisation et donc une sensibilité plus faible que celles des autres prospects.

Les résultats trouvés sur les prospects actuellement étudiants (une élasticité de -2,67) rejoignent ceux trouvés pour les conducteurs de moins de 24 ans. Un pouvoir d'achat faible et une sinistralité importante confirment un niveau de prime élevé et donc un taux de concrétisation faible pour ces prospects.

Les fonctionnaires se remarquent avec une élasticité supérieure à la moyenne, cependant le niveau de prime (590€) associé à ces prospects est inférieur à la moyenne. Les

artisans et commerçants apparaissent moins sensibles au prix que les autres catégories. Ceci est en contradiction avec une prime moyenne sur ces prospects relativement élevée.

Ces résultats indiquent que la compagnie aurait peut-être intérêt à minorer le niveau de prime des contrats sur certaines catégories afin d'attirer de nouveaux clients. En contrepartie, la compagnie pourrait majorer, par exemple, le niveau de prime des contrats concernant les artisans et commerçants afin d'accroître son chiffre d'affaires sans perdre trop d'assurés.

3) La classe de prix du véhicule

Classe de prix du véhicule	Effectif	Elasticité	Prime TTC	Taux de concrétisation	CA
<= C	7605	-1,63	372 €	45,96%	1 298 394 €
D	9030	-1,76	459 €	42,12%	1 747 212 €
E	10965	-1,82	498 €	40,18%	2 193 971 €
F	15619	-1,89	529 €	38,98%	3 218 546 €
G	19125	-2,02	591 €	36,56%	4 130 062 €
H	23068	-2,14	648 €	34,26%	5 118 323 €
I	23709	-2,15	667 €	33,05%	5 225 053 €
J	23736	-2,11	661 €	33,21%	5 207 918 €
K	24203	-2,11	658 €	32,38%	5 158 713 €
L	25844	-2,15	688 €	31,90%	5 668 288 €
M	22169	-2,10	688 €	31,83%	4 855 496 €
N	17891	-2,09	727 €	31,57%	4 107 813 €
O	10988	-2,13	781 €	30,27%	2 596 488 €
P	7111	-2,17	849 €	29,18%	1 762 103 €
Q	4365	-2,26	950 €	27,63%	1 146 240 €
R	3854	-2,35	1 042 €	26,63%	1 069 192 €
S	2777	-2,39	1 195 €	26,01%	863 281 €
>= T	4191	-2,53	1 379 €	24,11%	1 393 789 €
Portefeuille	256250	-2,08	672 €	33,78%	58 188 876 €

FIGURE 9.19 – Résultats sur le critère de la Classe de prix du véhicule

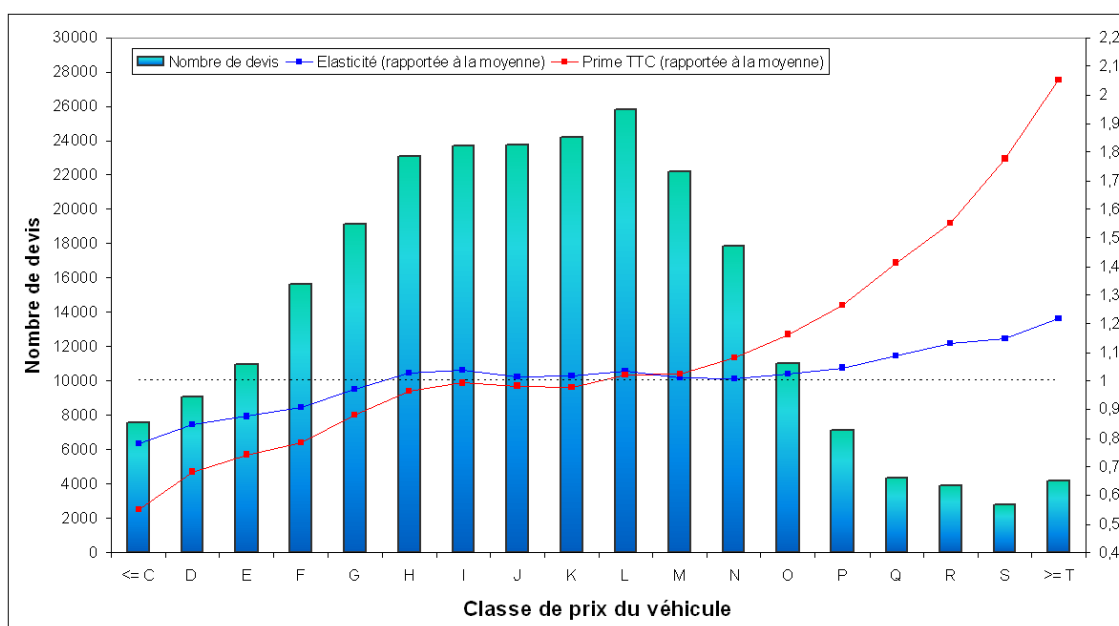


FIGURE 9.20 – Écart à la moyenne de l'élasticité et de la prime pour la Classe de prix

L'étude de l'élasticité montre que la sensibilité au prix des prospects semble presque constante en fonction de la valeur (à neuf) du véhicule, ce qui est en contradiction avec l'augmentation du niveau de prime qui s'intensifie devenant exponentielle à partir de la modalité M. Ceci peut s'expliquer par les frais d'indemnisation, en cas de dommages, des véhicules de classe de prix forte (au dessus de N) qui coûtent assez chers à la compagnie.

De plus, cette variable est à mettre en parallèle avec la puissance du véhicule, ces véhicules chers sont majoritairement des véhicules très puissants. Ils sont statistiquement plus impliqués dans des accidents de la route à cause d'une vitesse excessive.

Il semblerait donc que la politique de tarification adoptée par la compagnie en ce qui concerne la classe de prix du véhicule soit relativement bien adaptée à la sensibilité au prix des prospects.

Chapitre 10

L'acte de résiliation

Cette dernière partie décrit les étapes et les résultats obtenus de la modélisation durant notre deuxième environnement de travail : l'acte de résiliation. Notre deuxième objectif est d'analyser l'impact d'une variation de tarif sur le comportement des clients.

En premier lieu, nous mettrons en place, comme pour l'acte de concrétisation, un modèle logistique permettant d'estimer la probabilité qu'un assuré résilie à la date d'échéance de son contrat en fonction de ses caractéristiques et de son montant de prime. Puis, une fois cette tâche effectuée, un calcul d'élasticité sera réalisé pour mesurer la sensibilité au prix de chaque client présent dans le portefeuille.

Cependant, nous verrons par la suite, qu'avec l'introduction d'une nouvelle variable explicative, que l'élasticité au prix de chaque assuré ne sera pas équivalente à celle calculée à l'aide du premier modèle. La prime payée par l'assuré ne sera plus le seul indicateur de prix.

10.1 Modélisation du taux de résiliation en fonction du niveau de prime

Suite aux résultats de l'analyse descriptive, les variables utilisées pour la modélisation du taux de résiliation sont les suivantes :

Variables		
Âge de l'assuré	Client Vie	Prime TTC
Ancienneté de l'agent	Code Fils	Puissance du véhicule
Ancienneté du client	CRM 50	Score Assuré
Ancienneté du véhicule	Durée de détention du véhicule	Sinistralité
Année jeune conducteur	Garantie BDG	Situation familiale
Carburant	Mode d'achat	Type de véhicule
Catégorie Socio-Professionnelle	Mode de garage	Usage du véhicule
Clause Forfait Kilométrique	Nombre de contrats	Zone RC

FIGURE 10.1 – Liste des variables explicatives pour la modélisation du taux de résiliation

Nous souhaitons obtenir un premier modèle estimant la probabilité de résiliation de la part de l'assuré en fonction de deux effets dissociables :

- Un effet lié aux caractéristiques de l'assuré et indépendant du prix.

- Un effet dépendant uniquement du prix quel que soit le segment de population auquel appartient l'assuré.

De plus, au cours de l'étude de l'acte de concrétisation, nous avons vu l'importance de croiser les variables qualitatives avec le niveau de prime pour le calcul de l'élasticité permettant de mesurer l'impact du prix sur chaque segment d'assuré. Nous introduirons donc dans ce modèle ce croisement.

Après une sélection de type «stepwise», certaines variables toutes suffisamment significatives (au seuil de 95%) ont été gardées dans le modèle (noté Modèle 1) :

Modèle 1		
Effet de la variable	Direct	Croisé
Âge de l'assuré	•	•
Ancienneté du client		•
Ancienneté de l'agent	•	
Carburant		•
Catégorie Socio-Professionnelle	•	•
Clause Forfait Kilométrique		•
CRM 50	•	•
Durée de détention du véhicule	•	•
Formule		•
Garantie BDG	•	•
Nombre de contrats		•
Prime	•	
Puissance du véhicule	•	
Score Assuré	•	•
Sinistralité		•
Type de véhicule	•	•
Zone RC	•	

FIGURE 10.2 – Variables explicatives du modèle 1

Pour juger de la qualité de ce modèle, nous disposons des statistiques suivantes :

P-valeur	AIC	SC	% Paires Concordantes	Somer's D
<0,0001	144 069	145 199	66,9%	0,372

TABLE 10.1 – Indicateurs sur la qualité du modèle 1

Nous rejetons l'hypothèse nulle considérant que tous les coefficients associés aux variables explicatives sont nuls car le p-value associé à ce test est inférieur à 1%. Le pourcentage de paires concordantes est de 66,9%, nous dépassons le seuil fixé à 60%.

Nous avons vu que pour la modélisation du taux de concrétisation, une sélection supplémentaire a été effectuée sur le nombre de variables explicatives (le passage du modèle B au modèle C). Dans ce cas, cette sélection n'a pas besoin d'être effectuée. Nous préférons garder cette efficacité afin de ne pas diminuer la vraisemblance et le pourcentage de paires

concordantes qui semble correct.

Le taux de bon reclassement est de 64%. En considérant équivalents les mauvais reclassements dans les deux-sous populations (contrats résiliés ou renouvelés), le seuil donnant des taux de mauvais reclassement identiques pour chaque sous-population est de 3,19%. Nous acceptons ce taux même s'il s'avère difficile de différencier les 2 sous-populations. Comme dit précédemment pour la modélisation du taux de concrétisation, le but de l'étude est de mesurer l'impact du niveau de prime sur le comportement des assurés.

Les résultats du modèle 1 concernant le test de Wald jugeant la significativité des variables et les mesures d'association sont disponibles en annexe 8.

10.2 Calcul de l'élasticité pour le modèle 1

Comme pour la probabilité de concrétisation d'un devis, nous écrivons la probabilité de résiliation d'un contrat à sa date d'échéance pour un assuré :

$$\hat{\tau}_1(i) = \frac{1}{1 + \exp[-(A(i) + B(i)p(i))]}$$

que l'on peut écrire de la manière simplifiée :

$$\hat{\tau}_1(p) = \frac{1}{1 + \exp[-(A + Bp)]}$$

où nous avons :

- Un coefficient ne faisant pas intervenir le prix et qui correspond donc aux caractéristiques de l'assuré : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)I_{x_k=i_k}$;
- Un coefficient lié aux caractéristiques de l'assuré et au prix : $B = B(x_1, \dots, x_m) = b_0 + \sum_{l=1}^m b_l(i_l)I_{x_l=i_l}$;

Tout comme le taux de concrétisation d'un devis [cf. section 9.3, page 68], il s'avère que l'élasticité au prix pour ce modèle 1 équivaut à :

$$\hat{\epsilon}_1(p) = Bp(1 - \hat{\tau}_1(p))$$

La prime est toujours définie positive. Le taux de renouvellement du contrat (qui est le complémentaire au taux de résiliation) est compris entre 0 et 1. Le signe de l'élasticité est donc toujours déterminé par celui du coefficient B.

Nous nous attendons cette fois à observer des élasticités positives, ce qui signifie que pour une augmentation de la prime, le taux de résiliation des contrats va augmenter. Pour chaque assuré, tout comme pour notre précédente analyse, nous avons donc estimé la probabilité de résiliation d'un contrat à son terme ainsi que l'élasticité.

Sur l'ensemble du portefeuille, nous obtenons les résultats suivants :

Le taux de résiliation moyen estimé est de 3,21%. Il est égal au taux de résiliation moyen observé de 3,21%. Comme nous l'avons déjà vu précédemment, ce résultat est conforme à nos attentes puisque l'estimé est égal à l'observé.

L'élasticité moyenne observée est de 0,65. Le niveau de prime moyen est de 498 €. Le chiffre d'affaires généré par ces polices est de 265,3 M €. Notons CA(1) ce chiffre d'affaires.

Scénario		Prime TTC	Variation du taux	Nouveau taux
Hausse	1%	503 €	0,02%	3,23%
	2%	508 €	0,04%	3,25%
	3%	513 €	0,06%	3,27%
Baisse	-1%	493 €	-0,02%	3,19%
	-2%	488 €	-0,04%	3,17%
	-3%	483 €	-0,06%	3,15%

FIGURE 10.3 – Scénarios de hausse ou de baisse des tarifs

Selon les scénarios ci-dessus, si l'on avait opéré une hausse de 1% sur les tarifs, le niveau de prime moyen aurait été de 503 €. Le taux de résiliation assuré aurait été de 3,23%, soit une variation de 0,02%. En revanche, si l'on avait effectué une baisse de 1% sur les tarifs, le niveau de prime moyen aurait été de 493 €. Le taux de résiliation assuré aurait été de 3,19%, soit une baisse de 0,02%.

Cette élasticité semble assez faible puisque l'assuré n'apparaît pas très sensible à son niveau de tarif. D'après les scénarios ci-dessus, la compagnie pourrait augmenter ses tarifs pour accroître son chiffre d'affaires puisqu'elle ne perdra pas beaucoup d'assurés.

10.3 Modélisation du taux de résiliation en fonction de la variation de prime

Dans le modèle 1 utilisé jusqu'à présent pour estimer la probabilité de résiliation, la variable liée au prix est le montant de prime payé par l'assuré. L'élasticité moyenne calculée est assez faible.

Cependant, quelques mois avant la date d'échéance du contrat, la compagnie effectue une revalorisation de la prime de chaque assuré. Ceci a pour conséquence, pour la majorité des cas, d'engendrer une majoration de cette prime.

Les compagnies d'assurance sont conscientes des lourdes conséquences engendrées par les mauvais comportements du conducteur. C'est pour cette raison que la majoration de prime a été mise place. Les majorations de prime sont destinées à freiner les conducteurs les plus audacieux ayant une attitude à risque. L'assureur peut ainsi augmenter la prime de ses assurés en fonction de la sinistralité observée et la rentabilité attendue.

De plus, depuis le 1er août 2005 avec la loi Chatel, les assureurs sont tenus d'informer l'assuré, au plus tôt 3 mois et au plus tard 15 jours avant la date limite de résiliation, de

la possibilité de ne pas reconduire le contrat.

Ainsi, un assuré avant la date d'échéance connaît la future prime qu'il devra payer s'il renouvelle son contrat. L'assuré peut résilier en invoquant le motif de «hausse de prime» à condition que cette dernière ne résulte pas d'un malus.

Cette information semble alors très pertinente pour estimer la probabilité de résiliation en ajoutant la variation de prime comme une nouvelle variable explicative dans notre modélisation. Nous définissons cette variation de la manière suivante :

$$\text{Variation de prime} = \frac{p_1 - p_0}{p_0}$$

où p_0 est la prime actuelle de l'assuré (celle utilisée précédemment pour le modèle 1) et p_1 est la nouvelle prime pour la nouvelle échéance du contrat. Pour simplification, nous nommerons p_0 la «prime avant échéance» et p_1 la «prime après échéance».

Nous mettons en place alors un deuxième modèle, cette fois-ci, en utilisant la variation de prime comme indicateur du prix. Comme pour le modèle 1, la variation de prime a été croisée avec les variables caractéristiques puisque c'est cette variable qui fait maintenant directement intervenir le prix. La variable «prime avant échéance» devient alors une variable caractéristique que la compagnie connaît et sur laquelle elle n'a plus de pouvoir décisionnel. Seule la variation de prime peut être modifiée. Ainsi, nous retirons l'effet dépendant uniquement de la prime avant échéance pour éviter la corrélation avec celui dépendant de la variation de prime.

Après une sélection de type «stepwise», voici les variables toutes suffisamment significatives (au seuil de 95%) qui ont été gardées dans ce nouveau modèle (noté Modèle 2) :

Modèle 2		
Effet de la variable	Direct	Croisé
Âge de l'assuré	•	
Ancienneté de l'agent	•	
Ancienneté du client		•
Ancienneté du véhicule	•	
Carburant	•	
Catégorie Socio-Professionnelle	•	•
Classe de prix du véhicule	•	
Clause Forfait Kilométrique	•	
CRM 50	•	•
Durée de détention du véhicule	•	
Formule	•	
Garantie BDG	•	
Nombre de contrats		•
Score Assuré	•	•
Sinistralité		•
Type de véhicule	•	
Variation de prime	•	•
Zone RC	•	

FIGURE 10.4 – Variables explicatives du modèle 2

Pour juger de la qualité de ce modèle, nous utilisons les mêmes statistiques que pour le modèle 1 :

P-valeur	AIC	SC	% Paires Concordantes	Somer's D
<0,0001	145 707	146 836	65,6%	0,366

TABLE 10.2 – Indicateurs sur la qualité du modèle 2

La p-value associée au test, où nous considérons que tous les coefficients associés aux variables explicatives sont nuls, est inférieur à 1%. Le pourcentage de paires concordantes est de 65,6%, et reste supérieur au seuil fixé à 60%. L'efficacité du modèle 2 est très proche de celle du modèle 1 au regard des critères d'information d'Akaike et de Schwartz.

Selon les mêmes principes que le modèle 1, le taux de bon reclassement est de 63%. Le seuil donnant des taux de mauvais reclassement identiques pour chaque sous-population est de 3,19%. Nous acceptons ce taux même s'il semble difficile de différencier les contrats résiliés des contrats renouvelés avec les variables dont nous disposons. L'objectif final est de calculer une nouvelle élasticité, noté $\hat{e}_2(p_1)$, qui dépendra dorénavant de la «prime après échéance».

Les résultats du modèle 2 concernant le test de Wald jugeant la significativité des variables et les mesures d'association sont disponibles en annexe 8.

10.4 Calcul de l'élasticité pour le modèle 2

Suite aux résultats du modèle 2, pour chaque assuré, la probabilité de résiliation devient :

$$\hat{\tau}_2(i) = \frac{1}{1 + \exp[-(A(i) + C(\frac{p_1(i) - p_0(i)}{p_0(i)})])]$$

que l'on peut écrire de la manière simplifiée :

$$\hat{\tau}_2(p_1) = \frac{1}{1 + \exp[-(A + C(\frac{p_1 - p_0}{p_0})))]$$

où nous avons :

- Un coefficient associé aux variables caractéristiques qualitatives : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)I_{x_k=i_k}$;
- Un coefficient associé à la variation de prime : $C = C(x_1, \dots, x_m) = c_0 + \sum_{l=1}^m c_l(i_l)I_{x_l=i_l}$;

Cette formulation peut être transformée pour obtenir un calcul se rapprochant de la forme obtenue pour le modèle 1 :

$$\begin{aligned}\hat{\tau}_2(p_1) &= \frac{1}{1 + \exp[-(A + C(\frac{p_1 - p_0}{p_0})))]} = \frac{1}{1 + \exp[-(A + C\frac{p_1}{p_0} - C)]} \\ \hat{\tau}_2(p_1) &= \frac{1}{1 + \exp[-((A - C) + (\frac{C}{p_0})p_1)]} = \frac{1}{1 + \exp[-(\alpha + \beta p_1)]}\end{aligned}$$

avec : $\alpha = A - C$ et $\beta = \frac{C}{p_0}$

- α est un coefficient ne faisant intervenir que des variables explicatives fixes, sur lesquelles il est impossible d'intervenir.
- β est un coefficient lié à l'impact du nouveau prix p_1 appliqué par la compagnie en fonction des caractéristiques de l'assuré.

La probabilité de résiliation de l'assuré s'écrit alors sous cette forme :

$$\boxed{\hat{\tau}_2(p_1) = \frac{1}{1 + \exp[-(\alpha + \beta p_1)]}}$$

Et nous pouvons effectuer la même démarche que pour le modèle 1 et estimer une élasticité de la probabilité de résiliation au prix.

L'élasticité s'écrit donc pour le modèle 2 :

$$\boxed{\hat{\epsilon}_2(p_1) = \beta p_1 (1 - \hat{\tau}_2(p_1))}$$

Le signe de l'élasticité est déterminé par celui du coefficient β . En effet, la prime «après

«échéance» est définie positive et le taux de renouvellement est compris entre 0 et 1. L'élasticité $\hat{\epsilon}_2(p_1)$ estimée sera donc positive.

Sur l'ensemble du portefeuille, les résultats sont les suivants :

A nouveau, le taux de résiliation moyen estimé est de 3,21%. Il est égal au taux de résiliation moyen observé de 3,21%. La variation moyenne de prime est de 4,5%. La nouvelle prime moyenne est de 520€. Les polices renouvelées génèrent un chiffre d'affaires de 268 M€. Notons CA(2) ce chiffre d'affaires. Il est plus élevé de 1,1% par rapport au chiffre d'affaires CA(1)¹. Il s'avère que même si 3,21% des clients ont décidé de résilier leur contrat, la majoration de prime effectuée par la compagnie lui permet de ne pas réaliser de pertes.

L'élasticité moyenne observée est de 2,4. Selon les mêmes principes que pour le premier modèle, nous pouvons effectuer certains scénarios de hausse ou de baisse concernant cette fois-ci la nouvelle prime moyenne :

Scénario		Prime TTC après échéance	Variation du taux	Nouveau Taux	CA(2)	Variation du CA
Hausse	1%	525 €	0,08%	3,29%	270 500 282 €	2,03%
	2%	530 €	0,15%	3,37%	272 960 827 €	2,96%
	3%	535 €	0,23%	3,44%	275 417 105 €	3,89%
Baisse	-1%	515 €	-0,08%	3,13%	265 566 386 €	0,17%
	-2%	509 €	-0,15%	3,06%	263 093 035 €	-0,76%
	-3%	504 €	-0,23%	2,98%	260 615 417 €	-1,69%

FIGURE 10.5 – Scénarios de hausse ou de baisse de la «prime après échéance»

Si l'on avait opéré une hausse de 1% sur les tarifs après échéance, le nouvelle prime moyenne aurait été de 525 €. Le taux de résiliation assuré aurait été de 3,29%, soit une variation de 0,08%. Les polices renouvelés aurait généré un nouveau chiffre d'affaire CA(2) de 270,5 M€, soit une augmentation de 2,03%.

Si l'on avait effectué une baisse de 1%, cette prime aurait été de 515 €. Le taux de résiliation assuré aurait été de 3,13%. Nous attendions à une baisse de CA(2), au contraire, il aurait été de 265,6 M€, soit une augmentation de 0,17% par rapport à CA(1). Il aurait fallu effectuer une baisse à partir de 2%, pour constater une perte du chiffre d'affaires.

Cette valeur d'élasticité est radicalement différente de la valeur trouvée pour le modèle précédent. Elle implique des variations du taux de résiliation plus élevées. Il semble donc que l'assuré soit beaucoup plus sensible à la variation de sa prime plutôt qu'à son niveau. Ceci montre que l'assuré a un repère de prix qui est sa prime «avant échéance», et que la variation de prime, s'il s'agit d'une forte majoration, peut attirer son attention sur les prix proposés par la concurrence.

Au vu de la modélisation effectuée, nous avons pu estimer, pour chaque assuré, son élasticité ainsi que sa probabilité de résilier à échéance son contrat. Il est donc possible d'étudier les résultats par segment de population suivant la ou les variables que l'on juge intéressantes. Nous nous intéresserons aux variables «Ancienneté du véhicule» et «An-

1. défini précédemment section 10.2, page 78.

cienneté de souscription».

Nous rappelons que lorsque l'on compare deux élasticités ou deux taux de résiliation différents pour des modalités différentes, il faut aussi comparer les niveaux de prime auxquels sont soumis les assurés de ces deux catégories.

Les tableaux suivants [Figure 10.6, 10.8] reprennent pour chaque variable : l'effectif, la prime moyenne après échéance, la variation moyenne de prime, le taux de résiliation moyen estimé, l'élasticité estimée par modalité. Les graphiques² [Figure 10.7, 10.9] représentent l'écart à la moyenne de la prime après échéance et celui de l'élasticité pour chaque modalité.

1) Ancienneté du véhicule :

Ancienneté du véhicule	Effectif	Elasticité estimée	Prime TTC moyenne après échéance	Variation moyenne de prime	Taux résiliation estimé
0 - 1 an	35134	2,63	652 €	4,51%	1,80%
2 - 3 ans	64584	2,41	603 €	4,57%	2,19%
4 - 5 ans	64700	2,41	628 €	4,72%	2,87%
6 - 7 ans	63264	2,44	609 €	4,75%	3,12%
8 - 9 ans	58152	2,36	573 €	4,62%	3,11%
10 - 11 ans	56453	2,42	525 €	4,51%	3,51%
12 - 13 ans	49988	2,38	468 €	4,30%	3,56%
14 - 15 ans	35669	2,32	422 €	4,12%	3,85%
16 - 17 ans	31030	2,36	383 €	4,07%	4,21%
18 - 19 ans	12570	2,36	364 €	4,02%	4,25%
19 - 20 ans	18330	2,26	347 €	4,02%	4,20%
21 ans et plus	42838	2,31	276 €	4,18%	3,90%
Total	532712	2,40	520 €	4,46%	3,21%

FIGURE 10.6 – Résultats sur le critère de l'ancienneté du véhicule

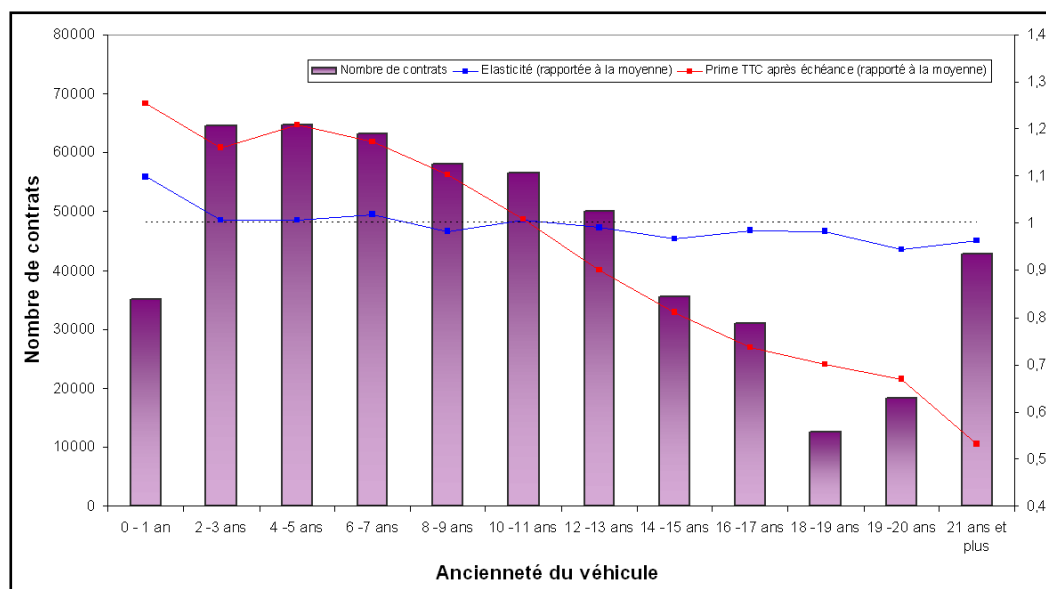


FIGURE 10.7 – Écart à la moyenne de l'élasticité et de la prime pour l'ancienneté du véhicule

2. Les résultats concernant d'autres variables sont disponibles en annexe 9.

La nouvelle prime appliquée à l'assuré semble diminuer en fonction de l'ancienneté du véhicule. Selon le tableau ci-dessus, l'évolution du taux de résiliation estimé est visiblement croissant en fonction de ce critère si l'on excepte les véhicules de plus 21 ans. Ce phénomène est dû à la corrélation naturelle entre l'ancienneté du véhicule et son état général. En effet, plus le véhicule est ancien, plus son état général se dégrade et plus le conducteur est susceptible de s'en séparer.

Globalement, nous remarquons une très légère baisse de l'élasticité. Selon le graphique, le degré de sensibilité semble presque le même pour les différents segments d'assuré. Seuls les véhicules de moins d'un an se détachent avec une élasticité de 2,63.

Ainsi, il semblerait que les mesures tarifaires adoptées par la compagnie en ce qui concerne l'ancienneté du véhicule soit relativement bien adaptée à la sensibilité au prix des assurés. Bien sûr, il est toujours possible d'affiner ce résultat en baissant les tarifs pour les assurés sensibles et en augmentant ceux des assurés qui le sont moins, mais il semble difficile de dégager une tendance très nette de l'élasticité concernant cette caractéristique.

2) Ancienneté de souscription :

Ancienneté de souscription	Effectif	Elasticité estimée	Prime TTC moyenne après échéance	Variation moyenne de prime	Taux résiliation estimé
moins d'un an	2273	2,74	576 €	0,70%	3,20%
1 an	77220	3,06	433 €	3,88%	3,44%
2 ans	63934	1,32	479 €	4,42%	3,86%
3 - 4 ans	89440	1,32	462 €	4,46%	3,72%
5 - 10 ans	131901	2,41	578 €	4,51%	3,48%
10 - 14 ans	73361	2,56	561 €	4,52%	2,66%
15 - 19 ans	38261	2,62	546 €	4,72%	2,48%
20 ans et plus	56322	2,63	527 €	5,08%	1,95%
Total	532712	2,40	520 €	4,46%	3,21%

FIGURE 10.8 – Résultats sur le critère de l'ancienneté de souscription

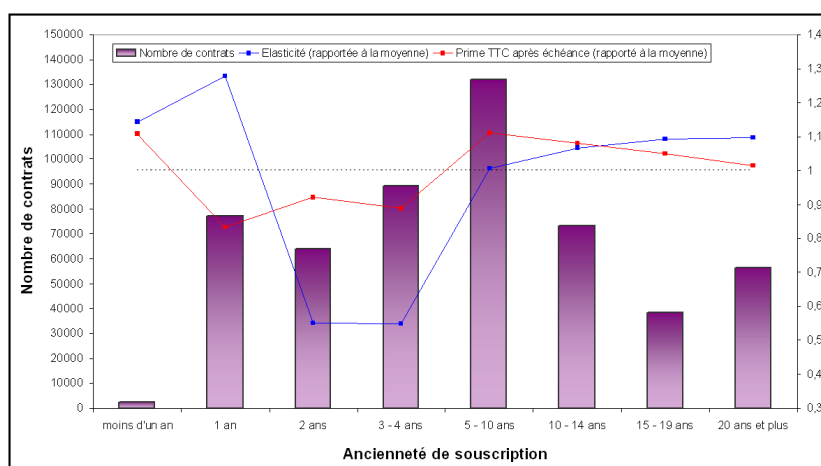


FIGURE 10.9 – Écart à la moyenne de l'élasticité et de la prime pour l'ancienneté de souscription

Nous constatons que le taux de résiliation est de 3,20% pour les contrats de moins d'un an et de 3,86% pour ceux d'une durée de deux ans. En revanche, il diminue à partir de 5 ans de souscription, pour parvenir à un taux 2% sur les contrats de plus de 20 ans.

A partir de 5 ans d'ancienneté, le niveau de prime baisse pour des raisons liées à l'âge et d'évolution du bonus-malus où les conducteurs deviennent de plus en plus expérimentés.

Nous nous attendions à une prime élevée pour les contrats récents. Cependant, les contrats d'un an bénéficient d'une certaine souplesse tarifaire sous l'hypothèse que l'assuré n'a pas subi de sinistres.

A partir de 2 ans d'ancienneté, l'élasticité chute à 1,32 pour augmenter à partir de 5 ans. Ce phénomène est dû à la fidélisation des clients.

Notons que le produit d'assurance automobile est un produit d'appel : il s'agit, dans la majorité des cas, du premier contrat souscrit par un client. Par exemple, chez Aviva, ce produit permet aux assureurs de proposer des réductions tarifaires sur d'autres produits, si le client décide de souscrire d'autres contrats d'assurance tels que l'assurance multi-risque habitation. Cette politique commerciale débute dès la concrétisation du devis.

Ainsi, grâce aux réductions proposées par l'agent, le client peut passer rapidement d'un statut mono-détenteur³ à celui de multi-détenteur, et sa probabilité de résiliation s'affaiblit alors avec le temps et le nombre total de contrats souscrits. Comme le montre le tableau suivant :

Nombre de contrats souscrits	Effectif	Taux résiliation estimé
1	68192	5,75%
2 à 3	171486	3,29%
4 à 5	114825	2,69%
6 et plus	178209	2,49%

FIGURE 10.10 – Taux de résiliation estimé en fonction du nombre total de contrats souscrits

Les assurés ayant un seul contrat dans la compagnie résilient presque deux fois plus que les autres clients.

Après étude de ces résultats, nous en déduisons que le lien entre l'évolution du taux de résiliation et celle du niveau de prime ne semble pas aussi évident que celui que nous avons établi lors des constats précédents : nous comprenons que plusieurs mesures tarifaires ont été appliquées indirectement sur le portefeuille par des actions commerciales.

10.5 Autres approches sur l'étude de l'élasticité-prix

Avant de conclure ce mémoire, nous allons évoquer et commenter certains points de l'article « The customer, the insurer and the market » écrit par Dutang⁴. Cet article présente d'autres approches pour l'étude de la sensibilité au prix. L'auteur s'intéresse à l'analyse de l'élasticité au prix dans le cadre d'un renouvellement de contrat d'assurance non-vie (produit assurance automobile) en explorant la relation triangulaire assuré - assureur - marché.

3. Voir Lexique page 97.

4. Pour tout détail sur la démarche, les données, les outils théoriques et les résultats des modélisations, se référer à l'article : DUTANG C. (2012) The customer, the insurer and the market. Bulletin français d'Actuariat, Vol. 12, n°24, pp. 35-85.

Tout d'abord, le premier point important est celui d'une autre définition de l'élasticité qui peut être définie comme suit :

$$\Delta_p(dp) = r(p + dp) - r(p)$$

où $r(p)$ est le taux de résiliation en fonction d'un indicateur de prix p . Cette approche semble pertinente car elle nous permet aussi d'identifier quelle impact a une légère modification du prix (dp) sur le taux de résiliation $r(p)$. Pour modéliser le taux de résiliation, une régression de type LOGIT⁵ est utilisée où l'auteur estime la probabilité moyenne de résiliation $\hat{\pi}_n(p)$ de la manière suivante :

$$\hat{\pi}_n(p) = \frac{1}{n} \sum_{i=1}^n g^{-1}(\hat{\beta}_0 + x_i(p)^T \hat{\beta}_{-p} + z_i(p)^T \hat{\beta}_{+p} * p)$$

où

- $x_i(p) = (x_{1i}, \dots, x_{ki})$ sont des variables explicatives non croisées avec l'indicateur de prix p ;
- $z_i(p) = (z_{1i}, \dots, z_{ki})$ sont des variables explicatives croisées avec l'indicateur de prix p ;
- $\hat{\beta}_0$ est l'estimation de la constante ;
- $\hat{\beta}_{-p}$ rassemble les coefficients de régression associés à $x_i(p)$;
- $\hat{\beta}_{+p}$ rassemble les coefficients de régression associés à $z_i(p)$;
- $p = \frac{\text{Prime de l'année étudiée}}{\text{Prime de l'année précédente}}$;
- g est la fonction de lien logit où $g^{-1}(x) = \frac{1}{1+e^{-x}}$

Cette modélisation est équivalente à celle effectué dans notre étude pour estimer le taux de concrétisation d'une part et le taux de résiliation d'autre part. Une probabilité de concrétisation pour chaque prospect i et une probabilité de résiliation pour chaque assuré i ont été calculées en fonction de leurs caractéristiques et du prix (l'aide d'une régression LOGIT), ensuite la moyenne de ces probabilités estime les taux moyens de concrétisation et de résiliation. Cependant, un terme de l'équation ci-dessus diffère, c'est la définition de l'indicateur du prix p . La probabilité $\hat{\pi}_n(p)$ applique un ratio de prix constant pour toutes les polices.

Notons que $\hat{\pi}_n(1)$ est un taux moyen de résiliation, appelé le taux central de résiliation, si la prime reste constante par rapport à celle de l'an dernier pour tous les clients. En s'appuyant sur $\hat{\pi}_n(1)$, le delta du taux de résiliation peut se définir comme suit :

$$\Delta_{1-}(\delta) = \hat{\pi}_n(1 - \delta) - \hat{\pi}_n(1) \quad \text{et} \quad \Delta_{1+}(\delta) = \hat{\pi}_n(1 + \delta) - \hat{\pi}_n(1)$$

où δ représente la variation de prime, par exemple 5%. Dutang s'appuie surtout sur les résultats trouvés $\hat{\pi}_n(1)$, Δ_{1-} et Δ_{1+} pour l'étude de la sensibilité des clients au prix. L'étude de différentes valeurs de p est possible mais cela demande un traitement informatique considérable.

L'auteur distingue trois types de résiliations : résiliation par la compagnie, celle par l'assuré et enfin un défaut de paiement. Dans son étude, les polices résiliées par la compagnie n'ont pas été gardées tandis que celles résiliées par l'assuré et celles suite à un défaut de paiement ont été conservées. Ce dernier motif est à considérer avec précaution car il pourrait représenter une décision cachée venant de l'assuré. Il peut résulter d'une prime

5. Voir présentation du modèle LOGIT section 5.2.1 , page 28.

trop élevée que le client ne peut pas se permettre.

En se basant sur notre périmètre d'étude pour traiter l'acte de résiliation, cette méthode peut être appliquée en posant :

$$p = \frac{\text{Prime après échéance } p_1}{\text{Prime avant échéance } p_0} = 1$$

où p_0 est la prime de l'assuré de l'année étudiée et p_1 est la nouvelle prime pour la nouvelle échéance du contrat en cas de renouvellement.

Une fois le taux de résiliation estimé en fonction de p , afin d'obtenir $\hat{\pi}_n(1)$, nous pourrions calculer $\Delta_{1-}(\delta)$ et $\Delta_{1+}(\delta)$. Mais, nous rappelons que notre objectif principal est l'étude de la sensibilité au prix des assurés pour le produit «Assurance Automobile» d'Aviva Assurances, ainsi nous incluons dans la modélisation du taux de résiliation la variation de prime⁶ observée propre de chaque assuré suite à la sinistralité constatée et la rentabilité attendue par la compagnie, qui permet d'obtenir pour chaque assuré sa propre probabilité de résilier son contrat. Un rapport de prix p constant pour toutes les polices ne serait pas adapté.

Le deuxième point important à souligner est d'étudier la sensibilité au prix des clients selon le réseau de ventes (ou de distribution) des contrats et le type de couverture de garanties (ou formule de garanties). Dans notre étude, nous nous sommes restreint à un réseau d'agents généraux, il aurait été intéressant d'obtenir des résultats sur le taux de résiliation et d'élasticité sur d'autres réseaux de ventes de contrats : mutuelles, courtiers ou assurances directes (appelé aussi assurances en ligne), etc. « De toute évidence, la façon dont nous vendons des produits d'assurance joue un rôle majeur dans la décision du client de renouveler ou de résilier », souligne Dutang. Il est possible qu'une augmentation (ou une baisse) de 5% des prix n'a pas le même impact sur la décision de renouveler ou de résilier un contrat des clients d'un courtier, d'un agent ou d'un réseau direct.

Une modélisation du taux de concrétisation ou de résiliation par formules de garanties a été envisagée dans notre cas, cependant les résultats des modélisations GLM par formule de garanties n'étaient pas robustes à cause du peu de données pour certaines formule de garanties (seule la formule F16, près de 50% du portefeuille, présentait de bons résultats et une possibilité de calculer l'élasticité). Au final, Dutang exploite neuf sous-groupes d'individus se répartissant selon le type de couverture⁷ (Couverture complète, couverture partiellement complète et couverture minimum de garanties ou "standard") et selon le réseau de distributions (Agent, Courtier, Direct). Une modélisation GLM du taux de résiliation pour chaque sous-groupes (neuf régressions) permet d'obtenir des résultats permettant une comparaison des clients sur une même échelle.

Le troisième point concerne certaines variables clés à prendre en considération pour l'étude de la sensibilité des clients. Suite à des analyses descriptives et de modélisations GLM, des comparaisons de taux de résiliation et d'élasticité (suite à une hausse ou une baisse du tarif de 5%) selon le réseau de distribution et selon le type de couverture sont effectuées. Ces variables ont eu un impact important sur l'estimation du taux de résiliation. Pour chaque régression réalisée (dans chaque sous-groupe), elles sont significatives :

6. Voir section 10.3, page 78.

7. Full Comprehensive (FC), Partial Comprehensive (PC), Third-Part Liability (TPL).

- **Le niveau de remise** : représentant le montant de remise accordée au client par l’agent, le courtier ou un gestionnaire de clientèle. Il diffère selon le réseau de ventes et selon la formule de garanties. Le réseau de distribution ayant la plus grande proportion de clients disposants d’une remise est celui des courtiers. Ils devancent les agents et les assureurs directs. Un assureur direct ne fournit généralement pas de remises car la prime est souvent plus faible que celles des autres réseaux de distributions. De plus, moins la couverture choisie par le client contient de garanties, plus il est difficile d’obtenir une remise. Ceci est remarqué par Dutang, la proportion de clients d’une couverture standard (Third-Part Liability) ayant une remise est la plus faible.

Des estimations de coefficients de régression négatives sont établies pour cette variable, ceci montre que le niveau de remise a une influence négative sur la résiliation. Plus le niveau de remise est élevé plus le taux de résiliation est faible, ceci semble logique.

Les résultats trouvés ne concernent pas seulement le comportement des assurés mais révèlent aussi certains faits sur le comportement des vendeurs de contrats. Par exemple, pour le réseau des courtiers, l’élasticité estimée est nettement plus élevée par rapport à celle trouvée sans l’ajout de cette variable. Ce fait représente le travail des courtiers vis à vis de leurs clients qui cherchent le tarif le moins cher.

- **Market premium proxy** : c’est un indicateur de prime de marché. Il est défini comme la dixième prime la plus faible parmi les primes des concurrents d’un produit d’assurance dommage standard d’un tiers pour lequel il n’y a pas de franchise. Une telle évaluation est faite sur une base de données des primes de marchés complétées par tous les assureurs du marché.

Les résultats conclus sur l’étude de cette variable ne concernent que deux sous-groupes : les clients d’agent et ceux d’assurance directe ayant une couverture standard de garanties. Il décide d’insérer cette variable dans la régression via une différence relative par rapport à la prime proposée :

$$m_i = \frac{market_i - premium_i}{premium_i}$$

où $market_i$ et $premium_i$ représentent, respectivement, la prime du marché et la prime proposée pour la i ème police.

Une telle variable permet d’introduire une notion de concurrence lié au marché. Dans un environnement concurrentiel, un assureur se trouve dans l’obligation de diversifier ses produits d’assurances et d’offrir une prestation de service adéquate afin de garder sa position sur le marché et de chercher à conquérir voire à améliorer sa position en deuxième lieu. Une amélioration du positionnement offre à l’assureur la possibilité d’accroître sa rentabilité et de limiter l’effet de la concurrence qui affecte sa marge.

L’introduction d’une notion de concurrence n’est pas seulement intéressante dans le cadre d’un renouvellement d’un contrat mais aussi dans le cadre de concrétisation d’un devis. Lors de l’étude de l’acte de concrétisation, nous avons remarqué l’importance d’insérer une notion de concurrence dans la modélisation du taux de concrétisation. Cette variable est l’ancienne prime du prospect qu’il payait chez un concurrent (noté Prime concurrence). Elle est l’une des variables les plus significatives avec la prime du devis⁸.

De plus, à l’aide de ce type de variable, des scénarios de hausses ou de baisses de

8. Voir annexe 6 page 112.

prix de marché et de la prime proposée peuvent être réalisés. Ensuite, une comparaison de l'évolution du taux de résiliation selon le réseau de distribution peut être effectuée. D'après son jeu de données, Dutang constate que la réaction des clients, suite à une hausse ou une baisse des tarifs diffèrent selon le réseau de distributions : par exemple, les clients d'agent généraux semblent moins sensibles à une variation des tarifs du marché que les clients d'assurances direct. Selon le type de scénarios, l'écart constaté entre les taux de résiliation concernant les clients d'agents généraux est nettement plus faible que celui des clients d'assurances direct. Il est possible que les clients d'agents généraux soient prêts à payer un prix plus élevé leur permettant une facilité et une rapidité de gestion de leurs contrats.

Le quatrième point important est l'effet de l'asymétrie d'information. On parle d'asymétrie d'information lors d'un échange quand certains des participants disposent d'informations pertinentes que d'autres n'ont pas. C'est un phénomène connu des praticiens de l'assurance. Son identification par les économistes a permis de réelles contributions en matière de compréhension du fonctionnement des marchés d'assurance. Cette asymétrie d'information entre l'assureur et l'assuré se manifeste par deux phénomènes bien connus, l'aléa moral et l'antisélection.

L'aléa moral concerne l'influence du contrat d'assurance sur le comportement de l'assuré. Il peut se définir comme le risque qu'un assuré se comporte différemment que s'il était totalement exposé au risque. L'antisélection (ou sélection adverse) est un terme utilisé à l'origine par les compagnies d'assurance pour désigner le fait que les individus à hauts risques ont tendance à s'assurer plus que les autres, de sorte qu'une « sélection » se fait parmi les assurés, mais dans un sens défavorable pour l'assureur. Généralement, le client devra choisir une franchise⁹ en sa faveur sur la base de sa propre évaluation des risques. Ainsi, les individus à haut risque auront tendance à choisir une franchise plus faible.

Un exemple classique est fourni en matière de fixation de primes d'assurances : si un assureur fixe un prix non adapté, la prime peut apparaître trop chère à ceux dont les risques sont les moins élevés (et qui, de fait, ne s'assurent pas) et intéressante pour ceux qui présentent le risque. L'assureur sélectionne ainsi les agents qui présentent le plus de risques (antisélection).

Afin de repérer une présence d'asymétrie d'information pouvant influencer sur le comportement des clients, Dutang teste la présence ou non d'antisélection sur son jeu de données, particulièrement sur le produit d'assurance contenant un maximum de garanties (Full Comprehensive coverage product) pour les trois réseaux de ventes (agent, courtier et direct), à l'aide de la théorie des GLM. Il s'inspire des travaux de Dionne et al. (2001) :

Soit X un vecteur de variables exogène¹⁰, N une variable endogène¹¹ et Z une variable de décision. L'absence de sélection adverse est équivalente à la prédiction de Z basée sur la distribution conjointe de X et N coïncidant avec celle basée sur la distribution de X seul. Cette caractérisation indirecte mène à

$$l(Z|X, N) = l(Z|X)$$

9. Voir définition section 6.2.2, page 39.

10. Voir Lexique page 97.

11. Voir Lexique page 97.

où $l(\cdot|\cdot,\cdot)$ représente la fonction de densité conditionnelle de probabilité. Une approche simple consiste à effectuer une dépendance conditionnelle dans le cadre paramétrique GLM, tels que le modèle est définie comme

$$l(Z|X, N) = l(Z|aX + bN)$$

. Tester l'indépendance conditionnelle de Z par rapport à N et X est réalisée par régression de la variable Z en fonction de X et N, afin de vérifier si le coefficient de N est significatif. Cependant, cette approche peut conduire à des conclusions fausses en raison d'effets non linéaires entre X et N. Dionne et al. (2001) recommandent d'utiliser le modèle économétrique suivant :

$$l(Z|X, N) = l(Z|aX + bN + c\hat{E}(N|X))$$

où $\hat{E}(N|X)$ est l'espérance conditionnelle de N sachant X, elle sera estimée par un modèle de régression initialement. L'introduction de l'espérance estimée $\hat{E}(N|X)$ permet de prendre en compte les effets non linéaires entre X et N, et pourtant pas les effets non linéaires avec Z.

Les principales étapes pour détecter le risque d'antisélection peuvent se résumer de la manière suivante :

- Étape 1 : Une première régression de N sur X pour obtenir $\hat{E}(N|X)$;
- Étape 2 : Une deuxième régression de la variable décision Z sur X, N et $\hat{E}(N|X)$;
- Étape 3 : Vérification de la significativité du coefficient de régression lié à N. Si c'est le cas dans la deuxième régression alors le risque de sélection adverse est détecté.

Le choix pertinent pour Z est le choix de la franchise par l'assuré, avec des facteurs X (caractérisant l'assuré) et N le nombre des sinistres observés. $\hat{E}(N|X)$ est estimée à l'aide d'une modélisation GLM où la variable à expliquer N est modélisée par une loi de Poisson¹² ou d'autres lois plus sophistiquées. De plus, une fois que le modèle GLM estimant la probabilité de choisir une franchise (150€, 200€, 300€, ...) Z en fonction de variables toutes significatives est mis en place, l'auteur insère cette probabilité (comme variable explicative) dans ses précédents modèles de régressions estimant le taux de résiliation. Une application concrète est donnée dans l'article. Le but final est de vérifier que le choix de franchise incluant un effet d'antisélection a un impact sur le renouvellement d'un contrat.

Au final, l'auteur ne détecte pas de sélection adverse dans son jeu de données car le coefficient de régression concernant N (la fréquence de sinistres) n'est pas significatif (il n'a aucun impact sur le choix de franchise). Ainsi, il sera difficile de détecter les individus à haut risque qui auront tendance à choisir une franchise plus faible. En outre, il constate finalement que le choix d'une franchise n'a pas d'impact important sur l'estimation du taux de résiliation $\hat{\pi}_n(1)$ et les valeurs d'élasticité $\Delta_{1-}(\delta)$ et $\Delta_{1+}(\delta)$. Cependant, il reste persuadé que le choix de franchise par le client joue un rôle majeur sur la probabilité de résiliation lors du renouvellement de sa police. « Les personnes à faible risque, c'est à dire avec des franchises élevées, sont susceptibles d'être les clients les plus sensibles, par rapport aux individus à haut risque ».

Enfin, nous allons finir cette section par une présentation brève d'autres modèles de régressions pouvant modéliser le taux de résiliation (ou le taux de concrétisation). Durant cette étude, des régressions de type GLM ont été utilisées, cependant d'autres modèles

12. Voir définition section 5.1, page 26.

de régression existent et peuvent être plus performants : les modèles de type GAM (the Generalized Additive Models).

Le modèle additif généralisé (ou GAM) est une extension du GLM. En fait, une grande partie de la définition du GLM vaut aussi pour lui. Ce qui les différencie c'est le prédicteur. Il est linéaire dans le GLM, il est dit additif dans le GAM, c'est-à-dire qu'il n'est pas obligé d'être linéaire. De plus, il est composé d'une somme de fonctions qui ne sont pas forcément paramétriques. Le GAM suppose que la fonction non paramétrique de n variables $X_1, X_2, \dots, X_n$ dont l'estimation et la représentation sont complexes puisse être approchée par une combinaison linéaire de fonctions non paramétriques :

$$S(x_1, \dots, x_n) \sim f_1(x_1) + \dots + f_n(x_n)$$

. Le modèle GAM se présente sous la même forme qu'un modèle GLM¹³, seul le prédicteur change :

$$\begin{aligned} Y_i &\sim L_{exp} & \text{et} & & \mu_i &= E[Y_i] \\ \eta_i &= g(\mu_i) \\ \eta_i &= \alpha + \sum_{j=1}^p f_j(x_{ij}) \end{aligned}$$

où Y (la variable à expliquer) suit une loi dont la densité appartient à la famille exponentielle¹⁴, g est une fonction de lien inversible (tels que $E(Y_i) = g^{-1}(\eta_i)$), les x_i sont les variables explicatives et η le prédicteur. Les fonctions f_j sont des fonctions quelconques d'une ou plusieurs variables. Elles peuvent être paramétriques (polynomiales, trigonométriques, etc.) ou non paramétriques comme les fonctions splines non paramétriques (ou splines de lissage). Ces modèles ont été introduits par Trevor Hastie et Rob Tibshirani en 1990. L'idée est de garder une forme linéaire, sauf que l'on s'autorise à avoir des transformations non-linéaires de variables explicatives.

Remarque : les modèles linéaires (et GLM) sont un cas particulier des modèles additifs (et GAM) avec $f_j(x) = \beta_j x$.

Pour déterminer les fonctions f_j et estimer les coefficients de régression liés à ces fonctions, plusieurs méthodes existent, mais elles ne seront pas détaillées ici. Pour plus d'informations, consulter les ouvrages¹⁵ d'Hastie et Tibshirani (1990), Venables et Ripley (2002) ou Wood (2008).

L'auteur applique une modélisation de type GAM pour estimer le taux de résiliation selon le réseau de ventes et selon le type de couverture, et constate certaines similitudes et différences par rapport à une modélisation GLM. Voici la nouvelle estimation de la probabilité moyenne de résiliation $\hat{\pi}_n(p)$ effectué par Dutang :

$$\hat{\pi}_n(p) = \frac{1}{n} \sum_{i=1}^n g^{-1}(\hat{\beta}_0 + x_i(p)^T \hat{\beta}_{-p} + z_i(p)^T \hat{\beta}_{+p} * p + \sum_{j=1}^p \hat{f}_j(\tilde{z}_i(p), p))$$

où $(\hat{\mu}, \hat{\beta}_{-p}, \hat{\beta}_{+p})$ sont les estimations des paramètres de régression, $\hat{f}_j$ sont les fonctions f_j ajustées, $(x_i, z_i, \tilde{z}_i)$ sont les variables explicatives pour chaque individu i et g est la

13. Voir définition et forme d'un modèle GLM section 5.1, page 24.

14. Voir définition section 5.1, page 25.

15. Voir Bibliographie 95.

fonction de lien logit.

Les modèles GAM permettent d'estimer une probabilité de résiliation de l'assuré en fonction de ses caractéristiques et d'un indicateur de prix et de permettre ainsi le calcul d'une élasticité permettant d'analyser le comportement des clients. Cependant, les modèles GAM ont besoin de plus de temps pour ajuster les données et exigent une meilleure puissance de calcul par rapport aux modèles GLM. Enfin, un jugement de l'utilisateur est nécessaire pour sélectionner, linéariser ou rejeter les variables explicatives afin d'obtenir le modèle GAM final.

Conclusion

Le but de ce mémoire était de mesurer et d'analyser l'impact du tarif sur le comportement des prospects lors de l'acte de concrétisation de devis et des assurés lors de l'acte de résiliation auprès des agents généraux d'Aviva Assurances.

Tout d'abord, nous nous sommes aperçus de l'importance de certaines étapes, notamment la construction des bases de données. Il s'agissait de bien comprendre leur structure, délimiter le périmètre d'étude, détecter les anomalies et les valeurs aberrantes. Ce travail fut primordial pour ne pas obtenir de résultats biaisés.

L'analyse statistique s'est, elle aussi, avérée très utile. Elle permet d'avoir une vision d'ensemble et d'obtenir les profils des prospects et des assurés. De plus, à l'aide de certains graphiques et de tests statistiques, elle nous a permis d'analyser nos variables une par une et également de chercher les liens significatifs entre nos variables cibles et les autres variables susceptibles d'entrer dans la modélisation.

Ensuite, une modélisation de type logistique nous a permis d'expliquer la probabilité de concrétisation et celle de résiliation en fonction des caractéristiques des individus et du prix. Ainsi, à l'aide de calcul de dérivées, une estimation de l'élasticité a été réalisée en fonction du prix. A l'aide de cette élasticité, nous avons été en mesure de déterminer la sensibilité au prix pour chaque individu. L'étude de cette sensibilité a permis d'établir des comparaisons entre certains profils d'assurés selon leur propension à concrétiser un devis ou résilier un contrat. Cet outil est donc particulièrement adapté comme aide à la décision dans un contexte qualitatif.

Durant l'analyse de l'acte de concrétisation, nous avons constaté que certains devis disposaient de l'ancienne prime payée par le prospect chez un concurrent. Cette variable est très précieuse dans ce contexte économique où ce type d'information est difficile à obtenir. Afin d'estimer cette prime pour le reste des devis, une modélisation a été réalisée à l'aide d'un modèle linéaire généralisé. Une étude descriptive de l'estimation de cette prime a révélé que cette information a un lien assez significatif sur l'évolution du taux de concrétisation. De plus, ceci a été confirmé lors de la modélisation du taux de concrétisation où l'estimation de cette prime était parmi les variables les plus significatives. Ainsi, cette variable nous a permis d'intégrer dans notre étude une notion directement liée à la concurrence pour le calcul de la sensibilité des prospects au prix.

Concernant l'analyse de l'acte de résiliation, deux modèles ont été mis en place donnant des résultats d'élasticité relativement différents. Le premier est construit à partir du niveau de prime payé par l'assuré et reflète, dans un sens, le montant que l'assuré est prêt à consacrer au contrat d'assurance. En revanche, le second modèle prend en compte la variation de tarif subie par l'assuré. L'élasticité obtenue du modèle 1 est plus faible que celle du modèle 2. Ceci peut s'expliquer par le fait que l'assuré a maintenant un repère de

tarif qui était son ancien niveau de prime. La variation, surtout si c'est une hausse, peut attirer l'attention de l'assuré sur les prix proposés par la concurrence. Il est donc, ce qui semble logique, beaucoup plus sensible à la variation du tarif de son contrat.

Par l'interprétation des résultats selon les différents segments d'assurés, cette étude a mis en avant plusieurs phénomènes généraux . Ils interviennent d'une façon directe ou indirecte, sur la variation du niveau de prime, et donc sur les valeurs d'élasticité. Ces phénomènes peuvent être classés de la manière suivante :

- les études de sinistralité dévoilant les profils à risque ;
- les actions commerciales effectuées par la compagnie ;
- le pouvoir d'achat des assurés : cette sensibilité au prix dépend de la conjoncture économique actuelle.

Cette étude révèle donc une situation complexe où la compagnie cherche à maîtriser le risque auquel elle doit faire face, tout en essayant de proposer des prix plus attractifs que la concurrence afin d'attirer de nouveaux clients et de fidéliser ceux déjà en portefeuille.

D'après des scénarios de hausse ou de baisse de prime effectués suite aux élasticités estimées, nous pouvons nous demander, en tenant compte de la sensibilité de chaque assuré, s'il existe un niveau de tarif optimal pour lequel les critères stratégiques de la compagnie sont satisfaits. Les résultats établis guideront la compagnie vers la possibilité de répondre à cette dernière question.

Par ailleurs, cette étude a été effectuée sur des données issues d'un réseau d'agents généraux. Il se peut alors que l'acte de concrétisation et de résiliation soient liés au prix mais aussi à d'autres facteurs exogènes indisponibles dans nos bases de données. Il est possible que les clients soient prêts à payer un prix leur permettant une facilité et une rapidité de gestion de leurs contrats. Ainsi, il peut s'avérer que les valeurs d'élasticité trouvées soient plus faibles que celles d'autres acteurs du marché comme par exemple les réseaux directs d'assurance. Cependant, dans un tel contexte de concurrence, ces informations sont difficiles à obtenir.

Pour conclure, cette étude démontre que le rôle de l'actuaire au sein de l'entreprise ne s'arrête pas au seul établissement de résultats techniques, mais s'inscrit dans une logique de contrôle, de communication et d'innovation constante au sein de l'entreprise.

Bibliographie

Articles

- DIONNE G., GOURIEROUX C., VANASSE C. (2001) Testing for evidence of adverse selection in the automobile insurance market : A comment, *Journal of Political Economy* 109(2), 444-453.
- DUTANG C. (2012) The customer, the insurer and the market. *Bulletin français d'Actuariat*, Vol. 12, n°24, pp. 35-85.
- PAGLIA A., PHELIPPE-GUINVARC'H M. V. (2011) Tarification des risques en assurance non-vie, une approche par modèle d'apprentissage statistique. *Bulletin français d'Actuariat*, Vol. 11, n°22, pp. 49-81.
- THIRY G. (2000) Assurance automobile études actuarielles et politique tarifaire en 2001. *Bulletin français d'Actuariat*, Vol. 4, N°7, pp. 61-82.
- WOOD S. N. (2010) Fast stable direct fitting and smoothness selection for generalized additive models, *Journal of the Royal Statistical Society : Series B* 70(3).

Cours EURIA

- AILLOT P. (2011) Modèles linéaires.
- LENCA P. (2012) Aide à la décision.
- PICARD C. (2012) Mathématiques des assurances IARD.
- VERMET F. (2010) Régression linéaire.

Ouvrages

- ALLISON P. D. (1999) *Logistic Regression Using SAS : Theory and Application*. Cary NC : SAS Institute Inc.
- DENUIT M., CHARPENTIER A. (2005) *Mathématiques de l'assurance non vie, tome 2 : Tarification et provisionnement*. Economica.
- DE JONG P., ZELLER G. (2008) *Generalized Linear Models for Insurance Data*. Cambridge University Press.
- HASTIE T. J., TIBSHIRANI R. J. (1990) *Generalized additive Models*. Chapman and Hall.
- HENRIET D., ROCHET J.-C. (1991) *Microéconomie de l'assurance*. Economica.
- KAAS R., GOOVAERTS M., DHAENE J., DENUIT M. (2009) *Modern Actuarial Risk Theory*. Springer Verlag.
- VENABLES W. N., RIPLEY B. D. (2002) *Modern Applied Statistics with S*, 4th edn. Springer.

Mémoires

- BERUBE V. (2007) Modèles Avancés en Régression appliqués à la tarification IARD. Faculté des Sciences et de Génie, Université LAVAL, QUEBEC.
- CHARBONIER F., GRUET P. (2012) Étude de la sensibilité et de la déformation d'un portefeuille de prévoyance. ENSAE.
- GONNET G. (2010) Étude de la tarification et de la segmentation en assurance automobile. ISFA.
- STUMPF C. (2000) Élasticité de la demande au prix. ULP.

Sites internet

- Site des mémoires d'actuariat : <http://www.ressources-actuarielles.net/bfa>
- Site de la Fédération Française des Sociétés d'Assurance :
 - Assurance Automobile : http://www.ffsa.fr/sites/jcms/c_47603/fr/comment-sont-etabl
 - Rapport Annuel 2011 Assurances de bien et de Responsabilité : <http://www.ffsa.fr/sites/upload/docs/application/pdf/2013-01/assurancesdebiensresponsabili>
[donneescles2011ffsa.pdf](http://www.ffsa.fr/sites/upload/docs/application/pdf/2013-01/assurancesdebiensresponsabili)
- Site de la loi Chatel : <http://www.loichatel.com/>
- Site de recherche d'ouvrage : <http://books.google.fr/>
- Site du logiciel SAS : SAS/STAT(R) 9.2 User's Guide, Second Edition :
 - Procédure GENMOD : http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#genmod_toc.htm
 - Procédure LOGISTIC : http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#logistic_toc.htm
- Site de l'association Sécurité et Réparation Automobiles : <http://www.sra.asso.fr/>

Autres références

- Élasticité de la demande : <http://www.lescoursdecogestion.fr/Outils/ELASTICITE%20DE%20LA%20DEMANDE.pdf>
- Notions d'élasticité et des applications : <http://iml.univ-mrs.fr/~lafont/licence/CHAP2.pdf>
- Modèles linéaires : <http://www.math.univ-toulouse.fr/~barthe/M1modlin/poly.pdf>
- Théorie sur les GLM : <http://www.sites.univ-rennes2.fr/laboratoire-statistique/PAC/doc/glm.pdf>
- TENENHAUS M. Modèle linéaire généralisé (Proc Genmod) : https://studies2.hec.fr/jahia/webdav/site/hec/shared/sites/tenenhaus/acces_anonyme/home/doctorat/statistique_avancee/GENMOD.PPT
- TENENHAUS M. Régression logistique : https://studies2.hec.fr/jahia/webdav/site/hec/shared/sites/tenenhaus/acces_anonyme/home/fichier_ppt/regression_logistique.ppt#312,1,Diapositive1

Lexique

- **Assurance de dommages** : L'assurance de dommages comprend les assurances de biens qui garantissent les biens appartenant à l'assuré et les assurances de responsabilité qui prennent en charge les conséquences financières des dommages que l'assuré peut causer à des tiers.
- **Avenant** : Un avenant est une nouvelle édition des conditions particulières du contrat d'assurance qui est faite systématiquement lorsqu'une modification leur est apportée (changement de véhicule, nouvelles garanties, etc). L'avenant remplace ou vient compléter les conditions particulières précédentes.
- **Endogène** : Une variable est endogène dans un modèle économique, lorsque ses valeurs sont déterminées par le modèle lui-même, par exemple par les décisions économiques des acteurs, ou l'influence d'autres variables économiques. Une variable endogène est à la fois explicative et expliquée.
- **Exogène** : Une variable est exogène dans un modèle économique, lorsque ses valeurs sont tirées de l'observation, ou résultent d'hypothèses spéciales, non économiques. Une variable exogène est explicative, mais non expliquée par le modèle.
- **IARD** : C'est une abréviation utilisée dans le monde de l'assurance signifiant «Incendie, Accidents et Risques Divers», qui s'oppose à l'assurance de personnes.
- **Mono-détenteur** : Ce terme commerciale désigne les clients de la compagnie ayant un seul contrat. Il s'oppose au terme «Multi-détenteur» représentant les assurés ayant plus d'un contrat.
- **Multirisques** : Le contrat multirisques est un contrat garantissant plusieurs risques. Par exemple, la multirisques Habitation comprend essentiellement les garanties catastrophes naturelles, explosions et incendies, actes de terrorisme, vol, dégât des eaux, bris de glace et responsabilité civile familiale. L'assuré doit bien vérifier dans les conditions particulières de son contrat les garanties prises en charge et les exclusions.
- **Mutualisation** : La mutualisation est le principe fondateur de l'assurance, basé sur le fait qu'un danger peut survenir. Le principe de la mutualisation consiste à répartir le coût de la réalisation du danger entre les membres d'un groupe soumis potentiellement au même danger et qui pourrait frapper certains d'entre eux.
- **Prospect** : Un prospect est le terme commercial utilisé pour définir un client probable de la compagnie.

- **Rachat de franchise** : Cette option, définie dans le contrat d'assurance, permet la prise en charge des franchises normalement dues en cas de sinistres. Ces franchises ne seront donc pas déduites du remboursement lors de sinistres.
- **Segmentation** : Opération par laquelle l'assureur distingue les risques qu'il prend en charge, afin de les classer dans des catégories de risques homogènes et de leur appliquer un traitement adéquat en ce qui concerne le tarif et la garantie.
- **SRA** : ou Sécurité et Réparations Automobiles est un organisme professionnel, créé en 1977. Il a le statut d'association loi 1901. Toutes les entreprises d'assurances automobiles sont adhérentes.
 La vocation de SRA est de promouvoir, au sein de la profession et avec les acteurs de l'automobile, toutes études et de mettre en oeuvre tous moyens utiles à la réalisation des actions pouvant contribuer à la limitation du nombre et du coût des sinistres dans l'intérêt des assurés.
 L'une des principales missions de SRA est de diffuser aux adhérents de l'information sur les véhicules par l'intermédiaire de ses bases de données, SRA met à disposition de ses adhérents toutes les caractéristiques techniques et commerciales des véhicules (4, 3 et 2 roues) afin de permettre leur identification.
 Les bases concernant les véhicules particuliers et utilitaires inférieurs à 3,5 T ainsi que les véhicules 2/3 roues et quads motorisés comportent également un système de classification, déterminé à partir de données objectives et uniquement liées à la conception du véhicule.
 Les entreprises adhérentes peuvent ainsi librement croiser dans leurs propres nomenclatures les données techniques et commerciales des véhicules avec celles concernant le conducteur et l'usage.

Cinquième partie

Annexes

Annexe 1 : Articles du Code des Assurances

Article A121-1

Modifié par *Arrêté 2007-07-19 art. 1 1°, art. 2 1° 2° JORF 21 juillet 2007*

Les contrats d'assurance relevant des branches mentionnées au 3 et au 10 de l'article R. 321-1 du code des assurances et concernant des véhicules terrestres à moteur doivent comporter la clause de réduction ou de majoration des primes ou cotisations annexée au présent article.

Sauf convention contraire, la clause visée au premier alinéa n'est pas applicable aux contrats garantissant les véhicules, appareils ou matériels désignés par les termes ci-après, tels que définis à l'article R. 311-1 du code de la route : cyclomoteur, engin de service hivernal, engin spécial, motocyclette légère, quadricycle léger à moteur, quadricycle lourd à moteur, véhicule de collection, véhicule d'intérêt général, véhicule d'intérêt général prioritaire, véhicule d'intérêt général bénéficiant de facilités de passage, véhicule et matériel agricoles, matériel forestier, matériel de travaux publics.

Article Annexe à l'article A121-1

Créé par *Arrêté du 31 octobre 2003 - art. Annexe, v. init.*

Art. 1er -Lors de chaque échéance annuelle du contrat, la prime due par l'assuré est déterminée en multipliant le montant de la prime de référence, telle qu'elle est définie à l'article 2, par un coefficient dit coefficient de réduction-majoration », fixé conformément aux articles 4 et 5 suivants. Le coefficient d'origine est de 1.

Art. 2. -La prime de référence est la prime établie par l'assureur pour le risque présentant les mêmes caractéristiques techniques que celles présentées par l'assuré et figurant au tarif communiqué par l'assureur au ministre chargé de l'Économie et des Finances dans les conditions prévues à l'article R. 310-6. Les caractéristiques techniques concernent le véhicule, la zone géographique de circulation ou de garage, l'usage socioprofessionnel ou le kilométrage parcouru, éventuellement la conduite exclusive du véhicule, ainsi que les réductions éventuelles figurant au tarif des entreprises d'assurance. Cette prime de référence ne comprend pas les majorations éventuellement prévues pour les circonstances aggravantes énumérées à l'article A. 335-9-2 du Code des assurances. En revanche, pour l'application des dispositions de la clause, cette prime de référence comprend la surprime éventuellement prévue pour les conducteurs novices à l'article A. 335-9-1 du Code des assurances ainsi que les réductions éventuelles mentionnées à l'article A. 335-9-3.

Art. 3. -La prime sur laquelle s'applique le coefficient de réduction-majoration est la prime de référence définie à l'article précédent, pour la garantie des risques de responsabilité civile, de dommages au véhicule, de vol, d'incendie, de bris de glaces et de catastrophes naturelles.

Art. 4. -Après chaque période annuelle d'assurance sans sinistre, le coefficient applicable est celui utilisé à la précédente échéance réduit de 5 %, arrêté à la deuxième décimale et arrondi par défaut ; toutefois, lorsque le contrat garantit un véhicule utilisé pour un usage Tournées » ou Tous Déplacements », la réduction est égale à 7%. Le coefficient de réduction-majoration ne peut être inférieur à 0,50. Aucune majoration n'est appliquée pour le premier sinistre survenu après une première période d'au moins trois ans au cours de laquelle le coefficient de réduction-majoration a été égal à 0,50.

Art. 5. -Un sinistre survenu au cours de la période annuelle d'assurance majore le coefficient de 25 % ; un second sinistre majore le coefficient obtenu de 25%, et il en est de même pour chaque sinistre supplémentaire. Le coefficient obtenu est arrêté à la deuxième décimale et arrondi par défaut. Si le véhicule assuré est utilisé pour un usage Tournées » ou Tous Déplacements », la majoration est égale à 20 % par sinistre. La majoration est, toutefois, réduite de moitié lorsque la responsabilité du conducteur n'est que partiellement engagée notamment lors d'un accident mettant en cause un piéton ou un cycliste. En aucun cas le coefficient de réduction-majoration ne peut être supérieur à 3,50.

Après deux années consécutives sans sinistre, le coefficient applicable ne peut être supérieur à 1.

Art. 6. -Ne sont pas à prendre en considération pour l'application d'une majoration les sinistres devant donner lieu ou non à une indemnisation, lorsque :

1-l'auteur de l'accident conduit le véhicule à l'insu du propriétaire ou de l'un des conducteurs désignés, sauf s'il vit habituellement au foyer de l'un de ceux-ci ;

2-la cause de l'accident est un événement, non imputable à l'assuré, ayant les caractéristiques de la force majeure ;

3-la cause de l'accident est entièrement imputable à la victime ou à un tiers.

Art. 7. -Le sinistre survenu à un véhicule en stationnement par le fait d'un tiers non identifié alors que la responsabilité de l'assuré n'est engagée à aucun titre, ou lorsque le sinistre mettant en jeu uniquement l'une des garanties suivantes : vol, incendie, bris de glace, n'entraîne pas l'application de la majoration prévue à l'article 5 et ne fait pas obstacle à la réduction visée à l'article 4.

Art. 8. -Lorsqu'il est constaté qu'un sinistre ne correspond pas à la qualification qui lui avait été donnée initialement, la rectification de la prime peut être opérée soit par le moyen d'une quittance complémentaire, soit à l'occasion de l'échéance annuelle suivant cette constatation.

Aucune rectification de prime ne sera, toutefois, effectuée si la constatation est faite au-delà d'un délai de deux ans suivant l'échéance annuelle postérieure à ce sinistre.

Art. 9. -La période annuelle prise en compte pour l'application des dispositions de la présente clause est la période de douze mois consécutifs précédant de deux mois l'échéance annuelle du contrat.

Si le contrat est interrompu ou suspendu pour quelque cause que ce soit, le taux de réduction ou de majoration appliqué à l'échéance précédente reste acquis à l'assuré mais aucune réduction nouvelle n'est appliquée, sauf si l'interruption ou la suspension est au plus égale à trois mois.

Par exception aux dispositions précédentes, la première période d'assurance prise en compte peut être comprise entre neuf et douze mois.

Art. 10. -Le coefficient de réduction-majoration acquis au titre du véhicule désigné au contrat est automatiquement transféré en cas de remplacement de ce véhicule ou en cas d'acquisition d'un ou plusieurs véhicules supplémentaires.

Toutefois, le transfert de la réduction n'est applicable que si le ou les conducteurs habituels du ou des véhicules désignés aux conditions particulières du contrat demeurent les mêmes, sauf en cas de réduction du nombre des conducteurs.

Art. 11. -Si le contrat concerne un véhicule précédemment garanti par un autre assureur, le coefficient de réduction-majoration applicable à la première prime est calculé en tenant compte des indications qui figurent sur le relevé d'informations mentionné à l'article 12 ci-dessous, et des déclarations complémentaires de l'assuré.

Art. 12. -L'assureur délivre au souscripteur un relevé d'informations lors de la résiliation du contrat par l'une des parties et dans les quinze jours à compter d'une demande expresse du souscripteur.

Ce relevé comporte notamment » les indications suivantes :

-date de souscription du contrat ;

-numéro d'immatriculation du véhicule ;

-nom, prénom, date de naissance, numéro et date de délivrance du permis de conduire du souscripteur et de chacun des conducteurs désignés au contrat ;

-nombre, nature, date de survenance et conducteur responsable des sinistres survenus au cours des cinq périodes annuelles précédant l'établissement du relevé d'informations, ainsi que la part de responsabilité retenue ;

-le coefficient de réduction-majoration appliqué à la dernière échéance annuelle ;

-la date à laquelle les informations ci-dessus ont été arrêtées.

Art. 13. -Le conducteur qui désire être assuré auprès d'un nouvel assureur s'engage à fournir à celui-ci le relevé d'informations délivré par l'assureur du contrat qui le garantissait précédemment, au souscripteur de ce contrat.

Art. 14. -L'assureur doit indiquer sur l'avis d'échéance ou la quittance de prime remis à l'assuré :

- le montant de la prime de référence ;
- le coefficient de réduction-majoration prévu à l'article A.121-1 du Code des assurances ;
- la prime nette après application de ce coefficient ;
- la ou les majorations éventuellement appliquées conformément à l'article A335-9-2 du Code des assurances ;
- la ou les réductions éventuellement appliquées conformément à l'article A. 335-9-3 du Code des assurances.

Article L211-1

Modifié par *LOI n°2007-1774 du 17 décembre 2007 - art. 1*

Toute personne physique ou toute personne morale autre que l'État, dont la responsabilité civile peut être engagée en raison de dommages subis par des tiers résultant d'atteintes aux personnes ou aux biens dans la réalisation desquels un véhicule est impliqué, doit, pour faire circuler celui-ci, être couverte par une assurance garantissant cette responsabilité, dans les conditions fixées par décret en Conseil d'État. Pour l'application du présent article, on entend par «véhicule» tout véhicule terrestre à moteur, c'est-à-dire tout véhicule automoteur destiné à circuler sur le sol et qui peut être actionné par une force mécanique sans être lié à une voie ferrée, ainsi que toute remorque, même non attelée.

Les contrats d'assurance couvrant la responsabilité mentionnée au premier alinéa du présent article doivent également couvrir la responsabilité civile de toute personne ayant la garde ou la conduite, même non autorisée, du véhicule, à l'exception des professionnels de la réparation, de la vente et du contrôle de l'automobile, ainsi que la responsabilité civile des passagers du véhicule objet de l'assurance. Toutefois, en cas de vol d'un véhicule, ces contrats ne couvrent pas la réparation des dommages subis par les auteurs, coauteurs ou complices du vol.

L'assureur est subrogé dans les droits que possède le créancier de l'indemnité contre la personne responsable de l'accident lorsque la garde ou la conduite du véhicule a été obtenue contre le gré du propriétaire.

Ces contrats doivent être souscrits auprès d'une entreprise d'assurance agréée pour pratiquer les opérations d'assurance contre les accidents résultant de l'emploi de véhicules automobiles.

Les membres de la famille du conducteur ou de l'assuré, ainsi que les élèves d'un établissement d'enseignement de la conduite des véhicules terrestres à moteur agréé, en cours de formation ou d'examen, sont considérés comme des tiers au sens du premier alinéa du présent article.

Annexe 2 : Notions sur les modèles linéaires généralisés

Lors de la description des éléments théoriques, nous nous sommes particulièrement intéressé à la régression logistique. Cette partie met en avant des notions plus générales applicables à un modèle linéaire généralisé.

Nous reprenons les notions décrites précédemment :

$$f(y_i|\theta; \phi) = \exp\left(\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i; \phi)\right)$$

où :

- $a(\cdot)$ est une fonction non nulle, dérivable sur $\mathbb{R}$;
- $b(\cdot)$ est une fonction trois fois dérivable sur $\mathbb{R}$ et sa dérivée première est inversible ;
- $c(\cdot)$ est une fonction définie sur $\mathbb{R}^2$;
- θ est un paramètre canonique ou paramètre de la moyenne ;
- ϕ est un paramètre de dispersion.

Remarque : $a(\phi) = \phi$ pour des données individuelles sinon $a(\phi) = \frac{\phi}{d}$ pour des données groupées où d est l'effectif du groupe. Dans notre cas, nous fixerons $d = 1$.

Le modèle linéaire généralisé s'écrit :

$$g(E[Y_i]) = g(\mu_i) = \sum_j \beta_j x_{ji}$$

Estimation des β_j par maximum de vraisemblance

La loi de Y_i :

$$f(y_i|\theta; \phi) = \exp\left(\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i; \phi)\right)$$

peut s'écrire en fonction de $\beta_1, \dots, \beta_p$ en remplaçant θ_i en remplaçant θ_i par :

$$\boxed{\theta_i = g_c(\mu_i) = g_c(g^{-1}(\sum_j \beta_j x_{ji}))} \quad (10.1)$$

puisque $g(\mu_i) = \sum_j \beta_j x_{ij}$

Résultats de la maximisation de la vraisemblance

De l'équation précédente (10.1), on en déduit la vraisemblance L :

$$L = \prod_{i=1}^n f(y_i; \theta_i; \phi) = \prod_{i=1}^n f(y_i; g_c(g^{-1}(\sum_j \beta_j x_{ji})); \phi)$$

On définit la log-vraisemblance l :

$$l = \text{Log}(L) = \text{Log}\left(\prod_{i=1}^n f(y_i; g_c(\mu_i); \phi)\right) = \text{Log}(L(y, \mu))$$

où $\mu_i = g^{-1}(x_i'\beta) = g^{-1}(\sum_j \beta_j x_{ji})$

On obtient $\hat{\beta}$ en annulant le vecteur Score $U = \frac{\partial l}{\partial \beta}$:

$$U = \sum_{i=1}^n \frac{y_i - \mu_i}{Var(Y_i)} \times \frac{\partial \mu_i}{\partial \beta} = 0$$

Déviance normalisé D* (ou déviance standardisé) et Déviance D

La déviance normalisée est donnée par la formule suivante :

$$D* = 2[l(y, y) - l(y, \hat{\mu})]$$

où

- $l(y, y) = \text{Log}(L(y, y))$ est la log-vraisemblance du modèle saturé, celui qui ajuste exactement toutes les valeurs de la variable Y.
- $l(y, \hat{\mu}) = \text{Log}(L(y, \hat{\mu}))$ est la log-vraisemblance du modèle étudié .

Elle suit une loi du χ^2 de degrés de liberté (n-p), où n est le nombre d'observations et p le nombre de paramètres.

La déviance D est égale à :

$$D = \phi \times D*$$

Une fois ϕ fixé, le maximum de vraisemblance conduit à minimiser la déviance D. La figure suivante représente la déviance D de certaines lois standards (définies dans les éléments théoriques) :

Loi	ϕ	Déviance
Normale	σ^2	$\sum (y_i - \hat{\mu}_i)^2$
Poisson	1	$2 \sum \left[(y_i \text{Log} \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i)) \right]$
Binomiale (proportion $y_i = r_i/m_i$)	1	$2 \sum \left[(y_i \text{Log} \frac{y_i}{\hat{\mu}_i} + (1 - y_i) \text{Log} \frac{1 - y_i}{1 - \hat{\mu}_i}) \right]$
Gamma	v^{-1}	$2 \sum \left[-\log \frac{y_i}{\hat{\mu}_i} + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$
Gauss inverse	σ^2	$2 \sum \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^2 y_i}$

FIGURE 10.11 – Déviance D des lois standards

Intervalles de confiance et test de significativité

L'estimation $\hat{\beta}$ de β dispose des propriétés suivantes :

- Loi de $\hat{\beta}$: $\hat{\beta} \sim N(\beta, J^{-1})$; où $J = -E[\frac{\partial^2 \text{Log}(L)}{\partial \beta^2}]_{\beta=\hat{\beta}}$
- Intervalle de confiance IC à 95% de $\hat{\beta}$ (appelé **Intervalle de confiance de Wald**) :

$$IC = [\hat{\beta} - 1,96\sqrt{\text{Var}(\hat{\beta}_j)} \ ; \ \hat{\beta} + 1,96\sqrt{\text{Var}(\hat{\beta}_j)}]$$

A l'aide des estimations $\hat{\beta}$, on obtient $\mu_i = g^{-1}(x'_i \hat{\beta})$ et on en déduit l'intervalle de μ_i à 95% :

$$IC = g^{-1}[x'_i \hat{\beta} - 1,96\sqrt{x'_i \text{Var}(\hat{\beta}) x_i} \ ; \ x'_i \hat{\beta} + 1,96\sqrt{x'_i \text{Var}(\hat{\beta}) x_i}]$$

Les tests permettant de juger de la significativité des paramètres sont les mêmes que ceux décrits pour la régression logistique : Test de Wald, Test de vraisemblance et celui du Score.

Analyse des résidus

Les résidus observés se définissent de la manière suivante :

$$r_i = y_i - \hat{\mu}_i$$

Nous définissons les résidus de Pearson et Déviance :

- Résidus de Pearson :

$$r_{P_i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{Var}(\hat{Y}_i)}}$$

- Résidus de Déviance : De $D = \sum_i d_i$ on déduit :

$$r_{D_i} = \text{signe}(y_i - \hat{\mu}_i) \sqrt{d_i}$$

Les résidus de Pearson et de Déviance sont normalisés en les divisant par leurs écarts-types.

Les résidus observés parfois appelés résidus bruts sont difficiles à analyser, les résidus de Pearson et de Déviance leurs sont souvent préférés permettant de vérifier une bonne adéquation du modèle linéaire généralisé utilisé.

Annexe 3 : Liste des variables

Type de variables	Nom des variables	Base de données		
		Concrétisation	Résiliation	Nombre de modalités
<i>Identifiant</i>	Numéro de projet	•		-
	Numéro de police	•	•	-
	Numéro d'image	•	•	-
	Code VGTA	•	•	-
<i>Date</i>	Date de création du devis	•		-
	Date d'entrée souscripteur	•	•	-
	Date d'échéance du contrat		•	-
	Date de fin du contrat		•	-
	Date du début d'image	•	•	-
	Date de fin d'image		•	-
<i>Tarifaire</i>	Age de l'assuré	•	•	11
	Ancienneté de permis	•	•	8
	Ancienneté de souscription		•	8
	Ancienneté du client		•	10
	Ancienneté du véhicule	•	•	12
	CRM	•	•	-
	CRM 50	•	•	9
	Durée de détention du véhicule	•	•	9
	Mode d'achat	•	•	5
	Année jeune conducteur	•	•	3
	Carburant	•	•	3
	Antécédents de sinistres	•		2
	CSP	•	•	7
	Formule	•	•	4
	Mode de garage	•	•	4
	Type de véhicule	•	•	7
	Situation familiale	•	•	5
	Classe de prix du véhicule	•	•	18
	Puissance du véhicule	•	•	11
	Clause Forfait Kilométrique	•	•	4
	Usage du véhicule	•	•	3
	Rachat de franchise	•	•	2
	Sinistralité Assuré		•	3
	Score Assuré		•	3
	Zone RC	•	•	9
	Zone Vol	•	•	9
<i>Prix</i>	Prime du devis	•		-
	Prime avant échéance		•	-
	Prime après échéance		•	-
<i>Cible</i>	CN	•		2
	RES		•	2
<i>Supplémentaires</i>	Ancienneté de l'agent	•	•	4
	Qualité de souscription	•	•	4
	Lien titulaire CG	•	•	3
	Nombre de contrats		•	4
	Code fils	•	•	2
	Client Vie		•	2
	Origine résiliation		•	2
	Motif résiliation		•	17
	Motif d'entrée d'une image		•	5
	Motif de sortie d'une image		•	20

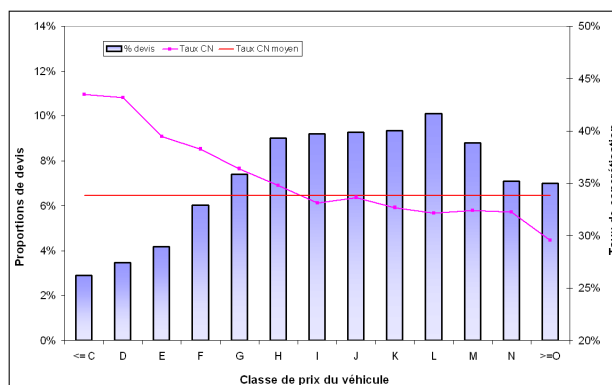
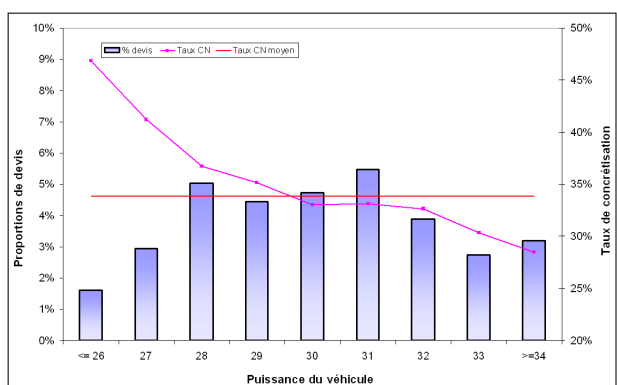
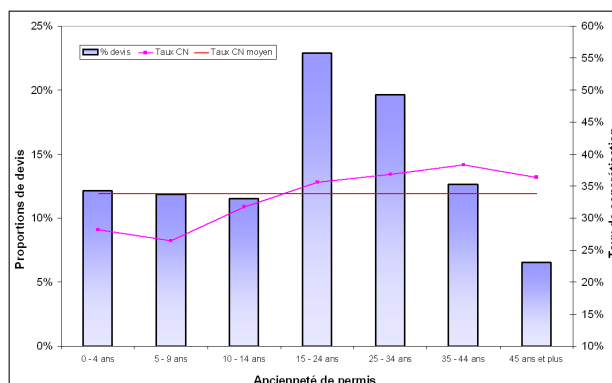
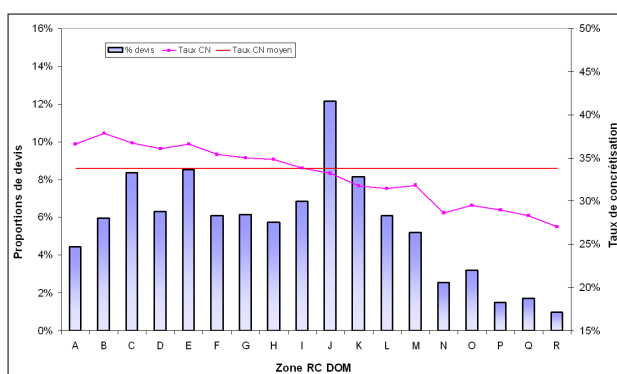
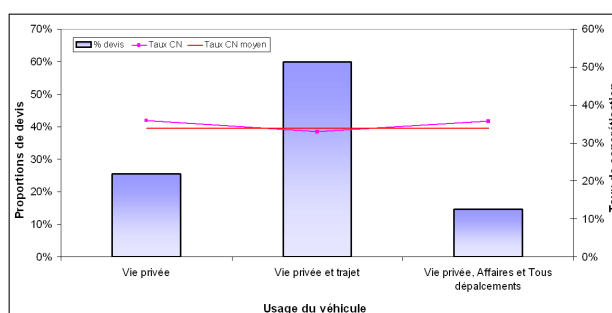
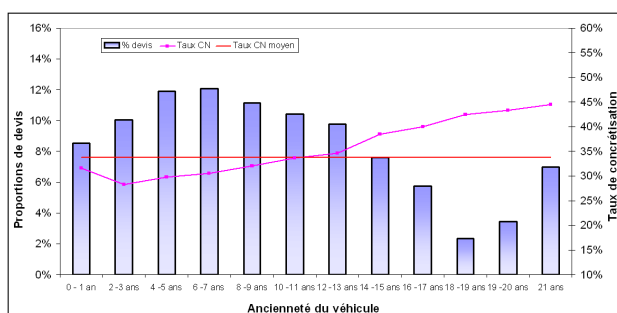
FIGURE 10.12 – Tableau récapitulatif des variables

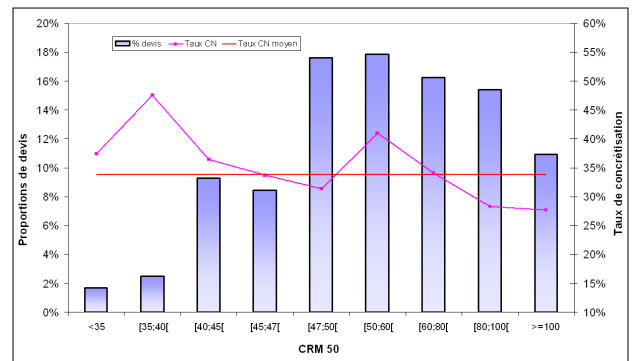
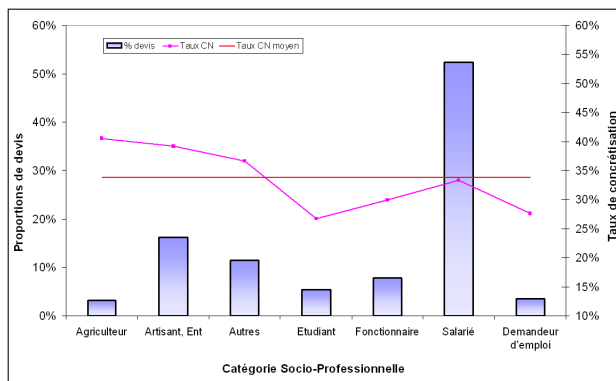
Annexe 4 : Analyse descriptive sur l'acte de concrétisation

Nous rappelons les notions suivantes :

- $\% \text{ devis} = \frac{\text{Nombre de devis modalité } i}{\text{Nombre total de devis}}$
- $\text{Taux CN} = \frac{\text{Nombre de devis concrétisés modalité } i}{\text{Nombre total de devis modalité } i}$
- $\text{Taux CN moyen} = \frac{\text{Nombre de devis concrétisés}}{\text{Nombre total des devis}}$

Le taux de concrétisation moyen de tous les devis est de 33,8%. Le nombre de total de devis est de 256 250. Ainsi, voici quelques graphiques représentant «%devis», «Taux CN» et «Taux CN moyen» en fonction de certaines variables susceptibles d'entrer dans la modélisation :



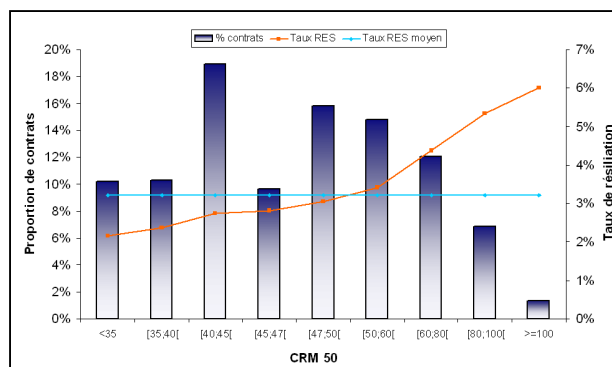
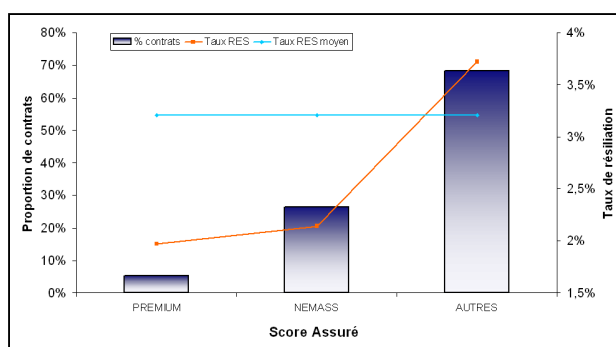
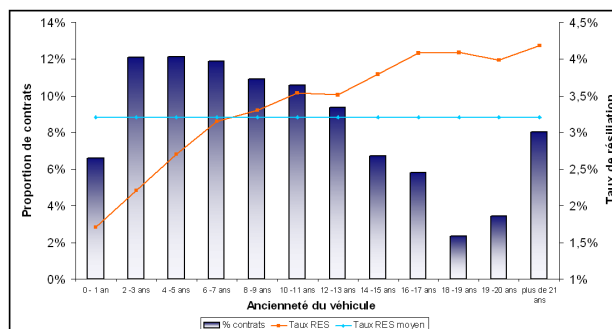
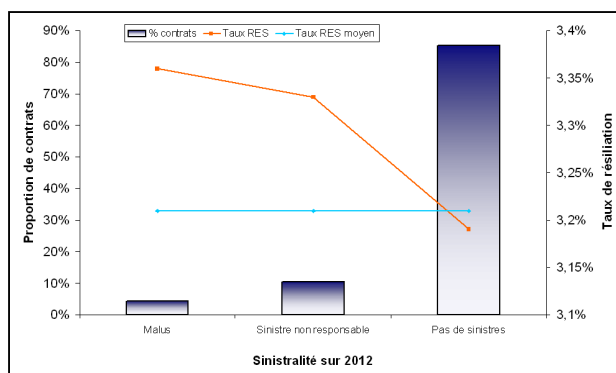
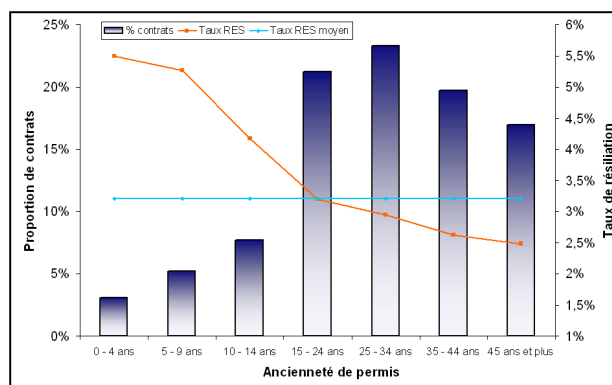
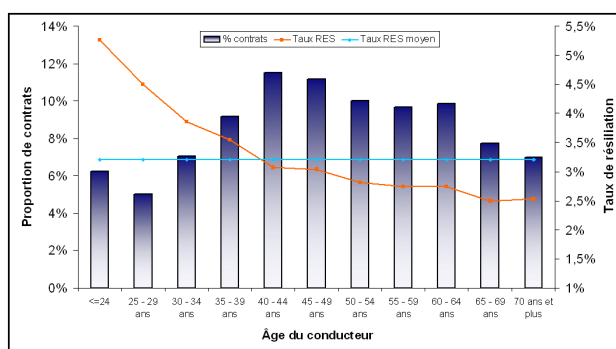


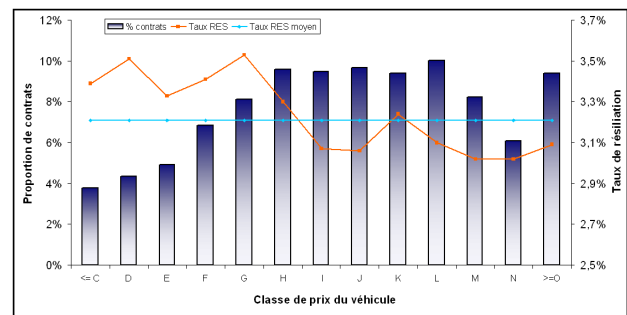
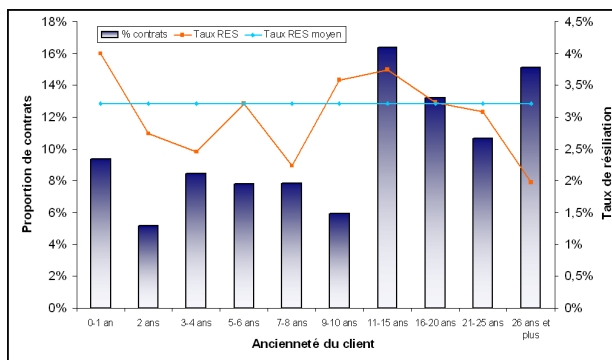
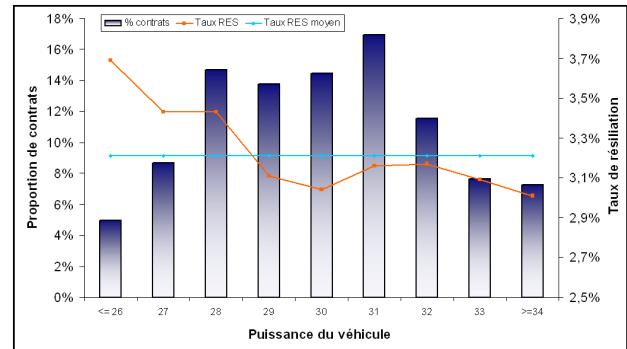
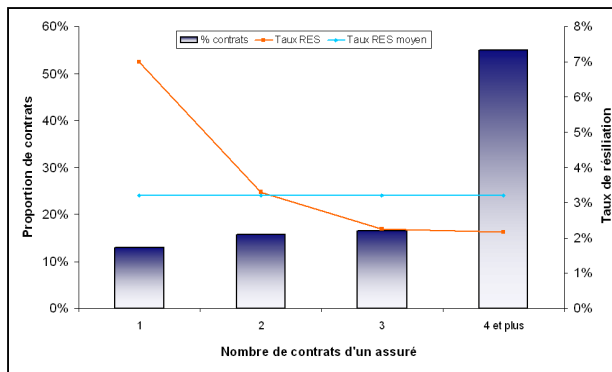
Annexe 5 : Analyse descriptive sur l'acte de résiliation

Nous rappelons les notions suivantes :

- $\% \text{ devis} = \frac{\text{Nombre de contrats modalité } i}{\text{Nombre total de contrats}}$
- $\text{Taux RES} = \frac{\text{Nombre de contrats résiliés modalité } i}{\text{Nombre total de contrats modalité } i}$
- $\text{Taux RES moyen} = \frac{\text{Nombre de contrats résiliés}}{\text{Nombre total des contrats}}$

Notre taux de résiliation moyen est de 3,2%. Le nombre de total de contrats est de 515 607. Ainsi, voici quelques graphiques représentant «%devis», «Taux RES» et «TAUX RES moyen» en fonction de certaines variables susceptibles d'entrer dans la modélisation :





Les résultats du test d'indépendance du chi-deux, du coefficient de corrélation de Kendall et du coefficient de corrélation de Spearman validant la corrélation entre une variable et la variable «RES» sont résumés dans le tableau suivant :

Variable	Statistique du CHI-2			Coefficient de KENDALL			Coefficient de SPEARMAN		
	Valeur	P-value	Résultat	Valeur	P-value	Résultat	Valeur	P-value	Résultat
Âge de l'assuré	1028	0,000	Liaison	-0,03	0,000	Liaison	-0,03	0,000	Liaison
Ancienneté de l'agent	27	0,000	Liaison	-0,01	0,000	Liaison	-0,01	0,000	Liaison
Ancienneté de permis	1192	0,000	Liaison	0,00	0,000	Liaison	0,00	0,000	Liaison
Ancienneté de souscription	391	0,000	Liaison	-0,02	0,000	Liaison	-0,02	0,000	Liaison
Ancienneté du client	1765	0,000	Liaison	0,04	0,000	Liaison	0,04	0,000	Liaison
Ancienneté du véhicule	847	0,000	Liaison	0,02	0,000	Liaison	0,03	0,000	Liaison
Année jeune conducteur	110	0,000	Liaison	-0,01	0,000	Liaison	-0,01	0,000	Liaison
Carburant	26	0,000	Liaison	-0,01	0,000	Liaison	-0,01	0,000	Liaison
Catégorie Socio-Professionnelle	351	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Classe de prix du véhicule	51	0,000	Liaison	0,00	0,000	Liaison	-0,01	0,000	Liaison
Clause Forfait Kilométrique	107	0,000	Liaison	0,01	0,000	Liaison	0,01	0,003	Liaison
Client Vie	231	0,000	Liaison	-0,02	0,000	Liaison	-0,02	0,000	Liaison
Code fils	187	0,000	Liaison	0,02	0,000	Liaison	0,02	0,000	Liaison
CRM 50	1345	0,000	Liaison	0,02	0,000	Liaison	0,03	0,000	Liaison
Durée de détention du véhicule	337	0,000	Liaison	-0,02	0,000	Liaison	-0,02	0,000	Liaison
Garantie BDG	476	0,000	Liaison	-0,01	0,000	Liaison	-0,01	0,000	Liaison
Lien titulaire CG	17	0,005	Liaison	0,00	0,650	Indépendance	0,00	0,720	Indépendance
Mode d'achat	15	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Mode de garage	450	0,000	Liaison	0,03	0,000	Liaison	0,03	0,000	Liaison
Nombre de contrats	4932	0,000	Liaison	-0,08	0,000	Liaison	-0,08	0,000	Liaison
Prime TTC	949	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Puissance du véhicule	57	0,000	Liaison	0,00	0,009	Liaison	0,00	0,009	Liaison
Rachat de franchise	6	0,065	Indépendance	0,00	0,080	Indépendance	0,01	0,086	Indépendance
Score Assuré	964	0,000	Liaison	-0,04	0,000	Liaison	-0,04	0,000	Liaison
Sinistralité	101	0,000	Liaison	-0,03	0,000	Liaison	-0,03	0,000	Liaison
Situation familiale	142	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Type de véhicule	89	0,000	Liaison	-0,01	0,000	Liaison	-0,01	0,000	Liaison
Usage du véhicule	101	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison
Zone RC	625	0,000	Liaison	0,02	0,000	Liaison	0,02	0,000	Liaison
Zone Vol-Incendie	718	0,000	Liaison	0,01	0,000	Liaison	0,01	0,000	Liaison

Annexe 6 : Résultats de la modélisation du taux de concrétisation

Résultats du modèle A

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Âge du prospect	3	346	<,0001
Ancienneté du véhicule	4	142	<,0001
CRM 50	8	1354	<,0001
Durée de détention du véhicule	3	1231	<,0001
Type du véhicule	6	400	<,0001
CSP	6	696	<,0001
Année jeune conducteur	2	732	<,0001
Antécédents de sinistres	1	1438	<,0001
Mode d'achat	1	307	<,0001
Mode de garage	3	112	<,0001
Formule de garanties	3	4216	<,0001
Puissance du véhicule	10	257	<,0001
Usage du véhicule	2	50	<,0001
Carburant	1	4	0,0399
Clause Forfait Kilométrique	3	1570	<,0001
Zone RC	8	18	0,0195
Situation familiale	1	295	<,0001
Ancienneté de l'agent	3	31	<,0001
Prime	1	3776	<,0001
Prime concurrence	7	589	<,0001

FIGURE 10.13 – Significativité des variables explicatives du modèle A

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	68,6	Somers' D	0,400
Percent Discordant	30,9	Gamma	0,401
Percent Tied	0,4	Tau-a	0,179
Pairs	1,4688E+10	c	0,686

FIGURE 10.14 – Mesures d'association du modèle A

Résultats du modèle B

Type 3 Analysis of Effects			
Effect	DF	Wald	Pr > ChiSq
		Chi-Square	
Âge du prospect	3	68	<,0001
Ancienneté du véhicule	4	440	<,0001
CRM 50	8	594	<,0001
Durée de détention du véhicule	3	390	<,0001
Type de véhicule	6	36	<,0001
CSP	6	132	<,0001
Année jeune conducteur	2	718	<,0001
Antécédents de sinistres	1	559	<,0001
Mode d'achat	1	20	<,0001
Mode de garage	3	74	<,0001
Formule de garanties	3	2063	<,0001
Puissance du véhicule	10	21	0,0189
Usage du véhicule	2	33	<,0001
Carburant	1	0	0,711
Clause Forfait Kilométrique	3	215	<,0001
Zone RC	8	18	0,0244
Situation familiale	1	24	<,0001
Ancienneté de l'agent	3	30	<,0001
Prime	1	1100	<,0001
Prime concurrence	7	269	<,0001
Prime * Âge du prospect	3	18	0,0004
Prime * Ancienneté du véhicule	4	315	<,0001
Prime * CRM50	8	149	<,0001
Prime * Durée de détention du véhicule	3	46	<,0001
Prime * Type de véhicule	6	147	<,0001
Prime * CSP	6	67	<,0001
Prime * Antécédents de sinistres	1	31	<,0001
Prime * Mode d'achat	1	12	0,0006
Prime * Puissance du véhicule	10	180	<,0001
Prime * Usage du véhicule	2	8	0,0184
Prime * Carburant	1	18	<,0001
Prime * Clause Forfait Kilométrique	3	30	<,0001
Prime * Formule de garanties	3	534	<,0001
Prime * Situation Familiale	1	8	0,0055
Prime * Prime concurrence	7	824	<,0001

FIGURE 10.15 – Significativité des variables explicatives du modèle B

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	70,3	Somers' D	0,409
Percent Discordant	29,4	Gamma	0,411
Percent Tied	0,3	Tau-a	0,183
Pairs	14687784441	c	0,705

FIGURE 10.16 – Mesures d'association du modèle B

Résultats du modèle C

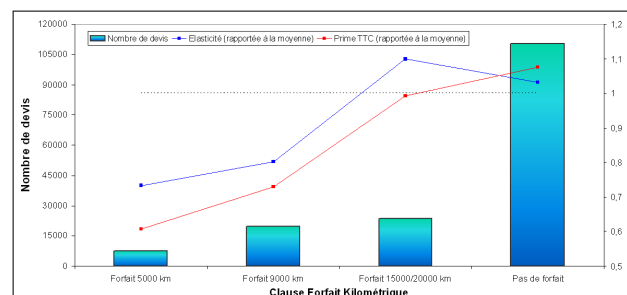
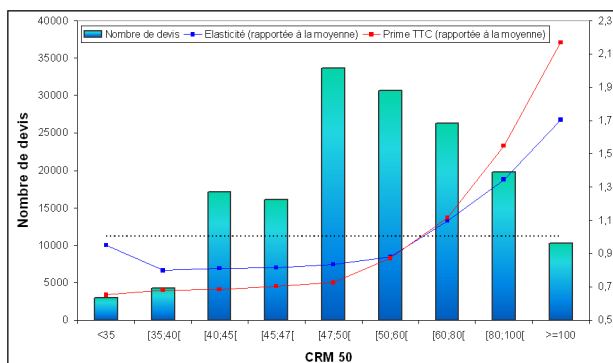
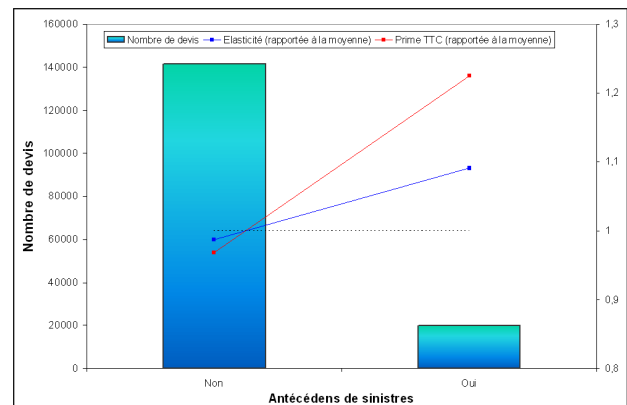
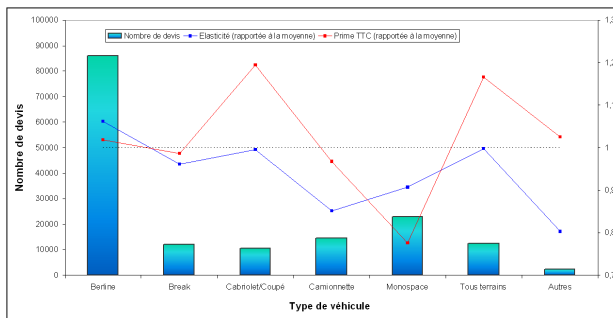
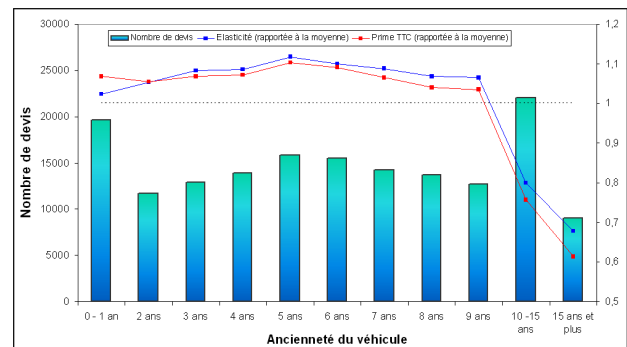
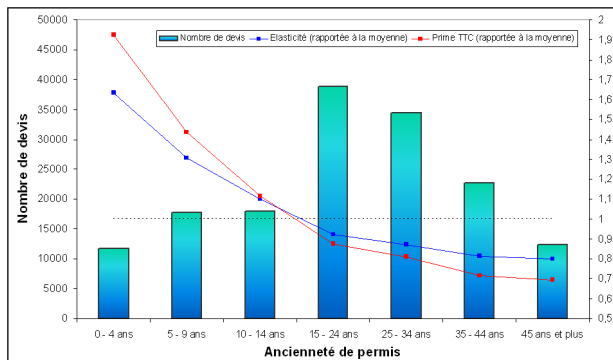
Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Âge du prospect	3	364	<,0001
Ancienneté du véhicule	4	590	<,0001
CRM 50	8	387	<,0001
Durée de détention du véhicule	3	1335	<,0001
Type de carrosserie	6	40	<,0001
CSP	6	181	<,0001
Année jeune conducteur	2	715	<,0001
Antécédents de sinistres	1	1460	<,0001
Puissance du véhicule	10	43	<,0001
Usage du véhicule	2	53	<,0001
Clause Forfait Kilométrique	3	1395	<,0001
Formule de garantie	3	1923	<,0001
Prime concurrence	7	121	<,0001
Prime	1	982	<,0001
Prime * Ancienneté du véhicule	4	230	<,0001
Prime * CRM 50	8	176	<,0001
Prime * Type de véhicule	6	178	<,0001
Prime * CSP	6	71	<,0001
Prime * Puissance du véhicule	10	229	<,0001
Prime * Formule de garanties	3	378	<,0001
Prime * Prime concurrence	7	132	<,0001

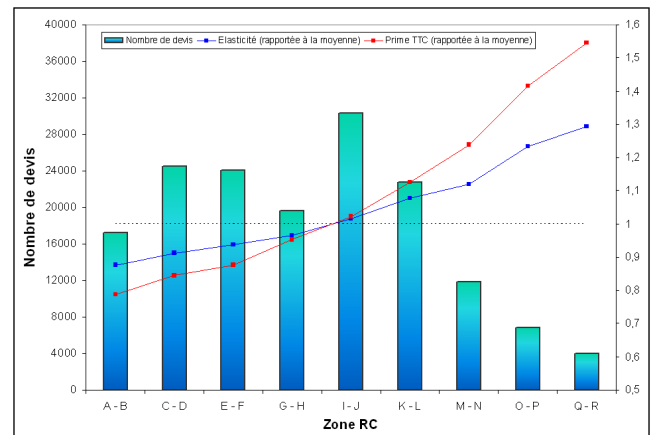
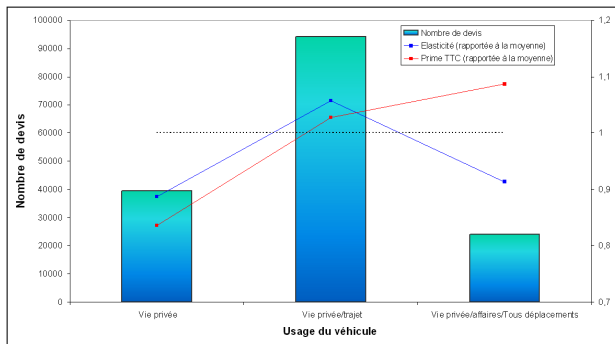
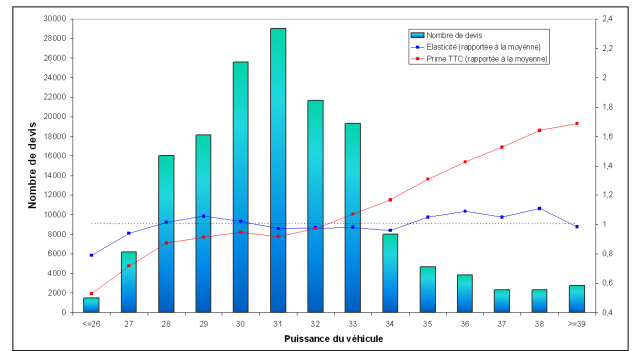
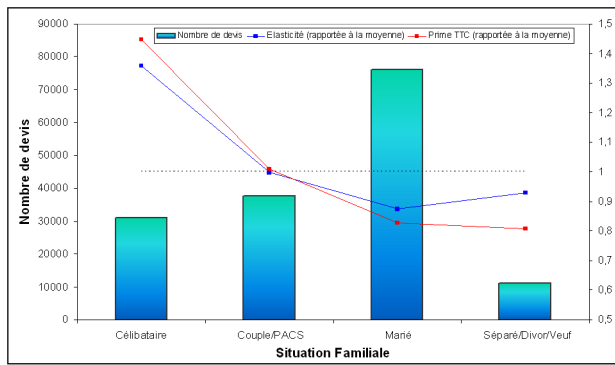
FIGURE 10.17 – Significativité des variables explicatives du modèle C

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	69,1	Somers' D	0,403
Percent Discordant	30,6	Gamma	0,404
Percent Tied	0,3	Tau-a	0,18
Pairs	14687784441	c	0,692

FIGURE 10.18 – Mesures d'association du modèle C

Annexe 7 : Résultats d'élasticité lors de l'acte de concrétisation





Annexe 8 : Résultats de la modélisation du taux de résiliation

Résultats Modèle 1

Type 3 Analysis of Effects			
Effect	DF	Wald	Pr > ChiSq
		Chi-Square	
Catégorie Socio-Professionnelle	6	51,7788	<,0001
CRM 50	8	15,7059	0,0468
Durée de détention du véhicule	4	16,5636	0,0023
Âge de l'assuré	4	41,4083	<,0001
Garantie BDG	3	108,6006	<,0001
Clause Forfait Kilométrique	3	32,3126	<,0001
Zone RC	7	82,3911	<,0001
Puissance du véhicule	8	25,2608	0,0014
Ancienneté de l'agent	3	9,2959	0,0256
Type de véhicule	3	15,2191	0,0016
Score assuré	2	305,865	<,0001
Prime	1	12,9365	0,0003
Prime * Carburant	1	36,4511	<,0001
Prime * Catégorie Socio-Professionnelle	6	32,0784	<,0001
Prime * Formule de garanties	3	741,6963	<,0001
Prime * Sinistralité	2	13,0061	0,0015
Prime * CRM 50	8	70,6517	<,0001
Prime * Durée de détention du véhicule	4	13,7407	0,0082
Prime * Âge de l'assuré	4	12,8074	0,0123
Prime * Garantie BDG	3	14,9997	0,0018
Prime * Clause Forfait Kilométrique	3	23,0141	<,0001
Prime * Puissance du véhicule	8	51,223	<,0001
Prime * Type de véhicule	3	12,8111	0,0051
Prime * Score assuré	2	165,0339	<,0001
Prime * Nombre de contrats	3	1521,3617	<,0001
Prime * Ancienneté du client	4	384,5174	<,0001

FIGURE 10.19 – Significativité des variables explicatives du modèle 1

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	66.9	Somers' D	0.372
Percent Discordant	29.7	Gamma	0.385
Percent Tied	3.4	Tau-a	0.023
Pairs	8819457735	c	0.686

FIGURE 10.20 – Mesures d'association du modèle 1

Résultats Modèle 2

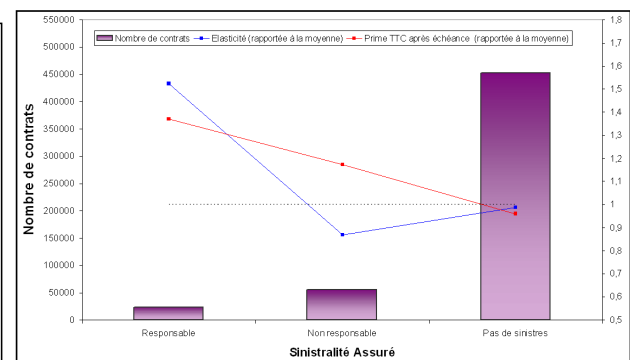
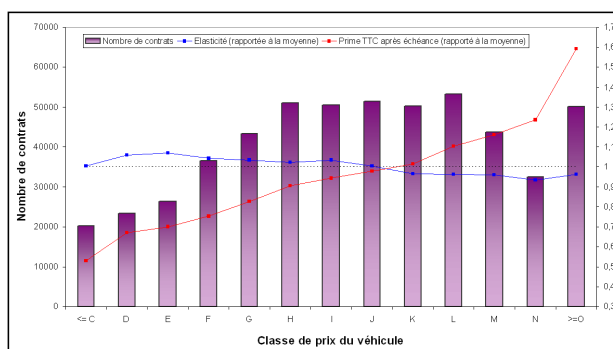
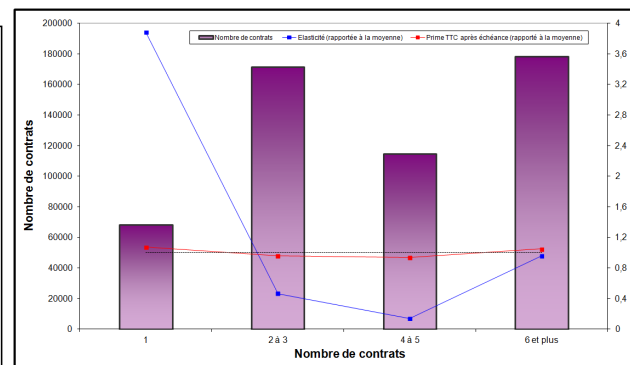
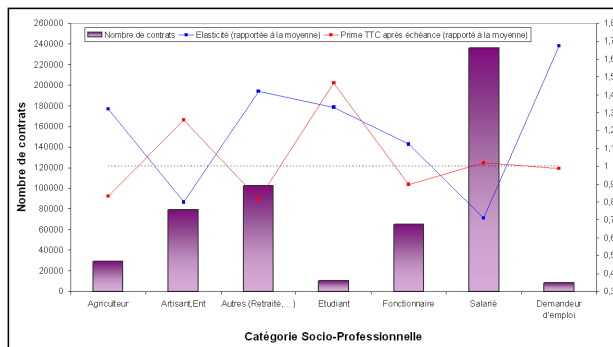
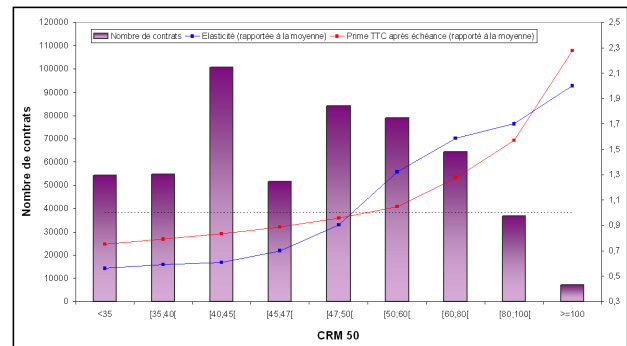
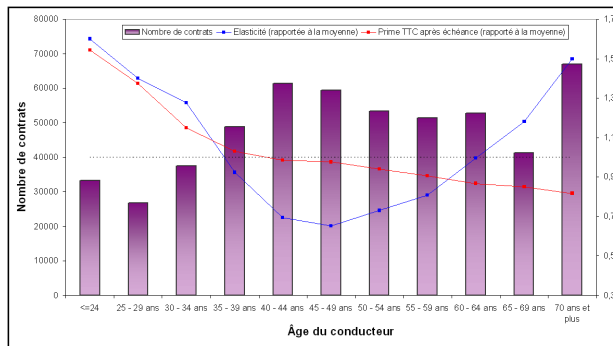
Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Carburant	1	57	<,0001
Catégorie Socio-Professionnelle	6	33	<,0001
CRM 50	8	23	0,0032
Durée de détention du véhicule	4	31	<,0001
Âge de l'assuré	4	160	<,0001
Formule de garanties	3	212	<,0001
Garantie BDG	3	74	<,0001
Ancienneté du véhicule	4	362	<,0001
Clause Forfait Kilométrique	3	22	<,0001
Zone RC	7	32	<,0001
Classe de prix du véhicule	12	99	<,0001
Ancienneté de l'agent	3	38	<,0001
Type de véhicule	3	30	<,0001
Score Assuré	2	622	<,0001
Variation de prime	1	2	0,1311
Variation de prime * Catégorie Socio-Professionnelle	6	21	0,0017
Variation de prime * Sinistralité	2	59	<,0001
Variation de prime * CRM 50	8	24	0,0021
Variation de prime * Zone RC	7	16	0,0229
Variation de prime * Score	2	123	<,0001
Variation de prime * Nombre de contrats	3	729	<,0001
Variation de prime * Ancienneté du client	4	244	<,0001

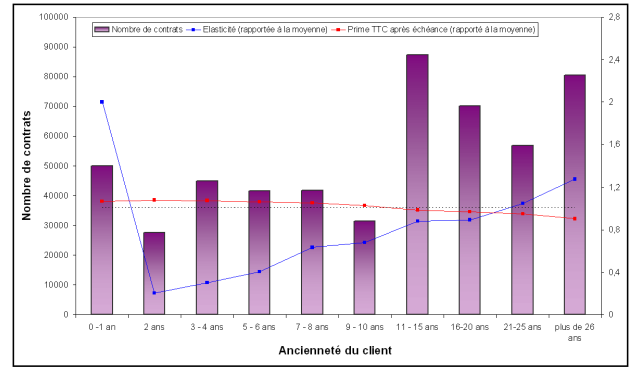
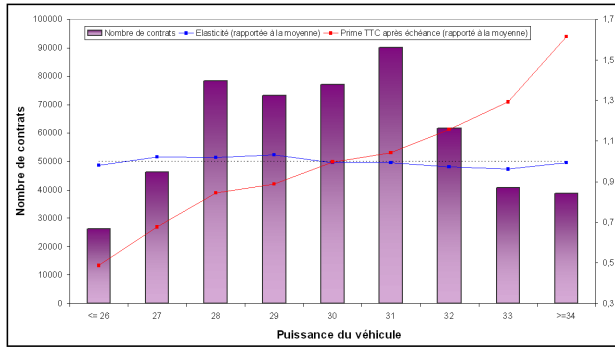
FIGURE 10.21 – Significativité des variables explicatives du modèle 2

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	65,6	Somers' D	0,366
Percent Discordant	31,5	Gamma	0,377
Percent Tied	2,9	Tau-a	0,023
Pairs	8819457735	c	0,657

FIGURE 10.22 – Mesures d'association du modèle 2

Annexe 9 : Résultats d'élasticité lors de l'acte de résiliation





Liste des tableaux

1.1	Montants et variation des cotisations des différents secteurs de l'assurance dommages (<i>Source FFSA</i>)	9
2.1	Évolution du coefficient Bonus-Malus sans accident responsable	14
5.1	Fonctions de liens usuelles	28
6.1	Les formules de garanties	37
8.1	Statistiques univariées de quelques variables quantitatives	48
9.1	Indicateurs sur la qualité des modèles A et B	66
9.2	Indicateurs sur la qualité du modèle C	67
10.1	Indicateurs sur la qualité du modèle 1	76
10.2	Indicateurs sur la qualité du modèle 2	80

Table des figures

1.1	Assurance française en 2011 (<i>Source FFSA</i>)	8
1.2	Répartition de la charge totale selon le type de sinistre Auto en 2011 (<i>Source FFSA - GEMA</i>)	9
1.3	Cotisations perçues par les principaux acteurs du marché automobile (<i>Source FFSA</i>)	10
5.1	Densité de probabilité φ de la loi normale $N(\mu, \sigma^2)$	26
7.1	Création de la base «Concrétisation»	41
8.1	Répartition des formules de garanties	47
8.2	Le taux de concrétisation en fonction de l'âge du conducteur	49
8.3	Le taux de résiliation en fonction de la prime	50
8.4	Lien des variables avec la variable «CN»	54
9.1	Ancienne prime du prospect chez le concurrent	57
9.2	Ancienne prime du prospect après lissage	58
9.3	QQ-plot d'une loi Normale et QQ-plot d'une loi Gamma	58
9.4	Densité empirique et densité théorique	59
9.5	Estimation des coefficients de régression par maximum de vraisemblance	60
9.6	Coefficients de régression de la variable CRM	61
9.7	Indicateurs sur l'adéquation du modèle	62
9.8	Analyse des résidus	62
9.9	Évolution du taux de concrétisation en fonction de l'écart tarifaire	63
9.10	Liste des variables explicatives pour la modélisation du taux de concrétisation	64
9.11	Variables explicatives du modèle A	65
9.12	Variables explicatives du modèle B	65
9.13	Variables explicatives du modèle C	66
9.14	Scénarios de hausse ou de baisse des tarifs	70
9.15	Résultats sur le critère de l'âge	70
9.16	Écart à la moyenne de l'élasticité et de la prime pour l'âge	71
9.17	Résultats sur le critère de la catégorie Socio-Professionnelle	72
9.18	Écart à la moyenne de l'élasticité et de la prime pour la CSP	72
9.19	Résultats sur le critère de la Classe de prix du véhicule	73
9.20	Écart à la moyenne de l'élasticité et de la prime pour la Classe de prix	73
10.1	Liste des variables explicatives pour la modélisation du taux de résiliation	75
10.2	Variables explicatives du modèle 1	76
10.3	Scénarios de hausse ou de baisse des tarifs	78
10.4	Variables explicatives du modèle 2	80
10.5	Scénarios de hausse ou de baisse de la «prime après échéance»	82
10.6	Résultats sur le critère de l'ancienneté du véhicule	83
10.7	Écart à la moyenne de l'élasticité et de la prime pour l'ancienneté du véhicule	83

10.8 Résultats sur le critère de l'ancienneté de souscription	84
10.9 Écart à la moyenne de l'élasticité et de la prime pour l'ancienneté de sous- cription	84
10.10 Taux de résiliation estimé en fonction du nombre total de contrats souscrits	85
10.11 Déviance D des lois standards	104
10.12 Tableau récapitulatif des variables	106
10.13 Significativité des variables explicatives du modèle A	112
10.14 Mesures d'association du modèle A	112
10.15 Significativité des variables explicatives du modèle B	113
10.16 Mesures d'association du modèle B	113
10.17 Significativité des variables explicatives du modèle C	114
10.18 Mesures d'association du modèle C	114
10.19 Significativité des variables explicatives du modèle 1	117
10.20 Mesures d'association du modèle 1	117
10.21 Significativité des variables explicatives du modèle 2	118
10.22 Mesures d'association du modèle 2	118

Synthèse

Mots clés : *Assurance automobile, tarification, segmentation, sensibilité au prix, prime commerciale, taux de concrétisation, taux de résiliation, modèles linéaires généralisés.*

En France, l'assurance automobile est le plus important marché¹⁶ de l'assurance dommages ; principalement à cause de l'obligation de s'assurer et du volume du parc automobile. Le nombre de clients est donc élevé, de même que les cotisations perçues par les assureurs. Les chiffres de la Fédération Française des Sociétés d'Assurance en témoignent : 19,5 milliards d'euros, soit 40% des cotisations des assurances de biens et de responsabilités, sont apportés par le marché de l'assurance automobile, le rendant vital pour les compagnies.

Les différents acteurs¹⁷ se livrent une guerre de prix pour conquérir les parts de marché. En parallèle, avec le développement des nouvelles technologies, des comparateurs d'assurance publient les évaluations des différents contrats selon plusieurs critères (tarif, garanties, qualité des prestations fournies, etc.) ; accentuant la pression concurrentielle.

L'actuaire se retrouve confronté à deux problématiques tarifaires dont les objectifs divergent : il s'agit tout d'abord de segmenter avec précision les assurés par niveau de risque encouru et ensuite de proposer un prix attractif. Respectivement, les objectifs sont de maîtriser le risque et d'attirer les prospects.

L'étude de la sensibilité d'un assuré au tarif semble alors nécessaire pour chaque assureur. Un prix trop élevé par rapport à la concurrence sera un obstacle à la souscription ou bien une incitation à la résiliation. Le comportement des consommateurs est difficile à prévoir, chaque individu accordant une importance différente à chaque critère.

L'objectif de ce mémoire est d'analyser l'impact d'une variation de tarif sur le comportement de l'assuré, en ce qui concerne l'acte de souscription et l'acte de résiliation. Pour mesurer la sensibilité des assurés au prix, nous utiliserons un indicateur de micro-économie, l'élasticité-prix $\epsilon(p)$, évaluée par la variation relative du taux de concrétisation ou de résiliation lorsque le prix varie d'une quantité infinitésimale :

$$\epsilon(p) = \frac{\partial T(p)/T(p)}{\partial p/p}$$

où $T(p)$ est le taux de résiliation ou de concrétisation de devis en fonction du prix p .

Son analyse a pour but de déterminer l'impact d'une légère évolution du prix sur le taux de concrétisation ou de résiliation.

16. en terme de nombre de contrats effectifs.

17. sociétés d'assurance, mutuelles et bancassureurs.

L'acte de concrétisation décrit la succession d'évènements :

1. Le prospect se présente à une agence de la compagnie et fait part de son intérêt pour un produit.
2. L'agent général établit un devis en fonction des caractéristiques du prospect et de son véhicule (âge, ancienneté du véhicule,...).
3. Le prospect décide, selon les garanties et le prix proposés, s'il désire ou non souscrire son contrat d'assurance avec cette compagnie.

Le taux de concrétisation d'un devis, ou taux de transformation, est défini comme la probabilité avec laquelle le prospect devient un client de la compagnie.

L'acte de résiliation, quant à lui, englobe :

- La résiliation par l'assuré qui intervient à la date d'échéance (elle coïncide dans la majorité des cas à l'anniversaire du contrat). Lors d'une éventuelle revalorisation des primes, l'assuré décide de résilier ou de renouveler son contrat.
- La résiliation par la compagnie qui peut avoir lieu à la date d'échéance suite à un excès de sinistres ou à tout moment pour non-paiement des primes, fausse déclaration de l'assuré à la souscription ou sortie des critères de souscription.
- La résiliation hors échéance en cas de perte ou de changement de véhicule.

Le taux de résiliation est la proportion de polices résiliées sur l'ensemble du portefeuille à la fin d'une période donnée.

Avant d'aborder la modélisation à proprement parler, tous les éléments théoriques nécessaires sont expliqués, la compréhension des méthodes étant cruciale à l'interprétation des résultats. L'outil principal est la régression logistique, membre de la famille des modèles linéaires généralisés. Différents indicateurs de qualité d'ajustement et de prédiction de modèles sont utilisés¹⁸.

Le périmètre choisi est délimité dans le temps, seules les données relatives à l'année 2012 sont utilisées. Des bases sont construites pour chaque environnement de l'étude (souscription/résiliation). En ce qui concerne l'acte de résiliation, nous traitons uniquement les contrats résiliés à la date d'échéance par l'initiative de l'assuré. Les variables sont traitées et nettoyées. Cette étape est primordiale pour préparer les données à la modélisation afin d'obtenir des résultats fiables.

L'analyse descriptive, a dévoilé les profils des prospects et des assurés mettant en avant différents critères tarifaires qui se regroupent en trois catégories liées :

1. au conducteur : âge, ancienneté de permis, situation matrimoniale, antécédents de permis, etc.
2. au véhicule : puissance, classe de prix, ancienneté, etc.
3. à la circulation : zone géographique, usage du véhicule, etc.

Des études bivariées, à l'aide de graphiques de distribution des taux¹⁹ en fonction des modalités et de tests de corrélation, ont permis de trouver des variables significatives pour la modélisation. Les variables étant majoritairement qualitatives (et ordinales), le

18. AIC, critère de Schwartz, paires concordantes, taux de bon reclassement.

19. concrétisation et résiliation

test d'indépendance du chi-deux a été utilisé. Ce test présente l'inconvénient d'être sensible au nombre d'observations. Pour confirmer les potentielles dépendances, deux autres tests ont été mis en oeuvre : les tests de corrélation de Kendall et de Spearman.

Ces étapes représentent la base de l'étude, il est maintenant possible d'entrer dans la phase de modélisation qui fait l'objet de la quatrième partie.

Le premier objectif est de mesurer l'impact du prix sur la souscription ou non d'un devis établi par un agent général. L'information liée à la concurrence est capitale pour la concrétisation ou non du prospect, elle doit être incorporée au modèle. La base inclut l'ancienne prime payée par certains prospects chez un concurrent. Cette variable est très précieuse dans ce contexte économique où ce type d'information est difficile à obtenir, elle doit toutefois être considérée avec prudence, étant exposée à certains biais. En effet, ce sont les clients eux-mêmes qui la rapportent, ainsi :

- certains assurés peuvent minorer cette prime dans le but de faire baisser le prix du devis,
- d'autres donnent un tarif approximatif, le tarif exact étant difficile à retenir.

Dans ce contexte, pour avoir une distribution continue et réaliste, les primes ont été lissées (plusieurs fois).

D'après les comparaisons de QQ-Plot et de test d'ajustement, la variable ancienne prime chez le concurrent suit approximativement une loi Gamma. Elle est modélisée grâce à un modèle linéaire généralisé, permettant d'explicitier cette prime pour le reste des devis.

Ensuite, il s'agit de déterminer un modèle statistique capable d'estimer la probabilité de concrétisation d'un devis en fonction des caractéristiques du prospect et du prix proposé par l'agent.

Un premier modèle a été mis en place composé de trois effets dissociables, liés :

- aux caractéristiques du prospect et indépendant du prix ;
- uniquement au prix quel que soit le segment de population auquel appartient le prospect ;
- au tarif de la concurrence estimé lors de la partie précédente quel que soit le segment de population auquel appartient le prospect.

Désireux de mesurer l'impact du prix par segment de la population, nous avons mis en place un deuxième modèle en introduisant l'ensemble des variables qualitatives du premier modèle croisées avec le niveau de prime.

Une comparaison des indicateurs gageant de la qualité des modèles²⁰ a montré que le deuxième modèle était plus pertinent que le premier.

Cependant, le nombre important de variables empêche l'application opérationnelle du deuxième modèle au quotidien. Il est nécessaire d'effectuer une sélection supplémentaire afin de conserver les variables :

- qui ont un impact important sur le tarif ;
- qui sont susceptibles d'intervenir dans la tarification de la concurrence ;
- qui sont peu corrélées entre elles ;

20. AIC, critère de Schwartz, paires concordantes, taux de bon reclassement.

- qui font intervenir un faible nombre de modalités avec des effectifs sensiblement égaux par classe.

A l'issue de ces sélections, qui font passer le nombre de variables de trente quatre à vingt, l'efficacité du modèle est conservée même si son pouvoir prédictif sur la variable à expliquer semble plus faible. La probabilité de concrétisation en fonction des caractéristiques du prospect et de son niveau prime est :

$$\hat{\tau}(i) = \frac{1}{1 + \exp[-(A(i) + B(i)p(i))]}$$

où nous disposons de deux coefficients A et B :

- Un coefficient ne faisant pas intervenir la prime du devis et qui correspond donc aux caractéristiques du prospect : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)I_{x_k=i_k}$;
- Un coefficient lié aux caractéristiques du prospect et au prix p : $B = B(x_1, \dots, x_m) = b_0 + \sum_{l=1}^m b_l(i_l)I_{x_l=i_l}$;

que l'on peut écrire de la manière simplifiée :

$$\hat{\tau}(p) = \frac{1}{1 + \exp[-(A + Bp)]}$$

Cette probabilité permet de calculer la sensibilité de chaque prospect à son niveau de prime, il suffit de dériver la fonction $\hat{\tau}(p)$ par rapport à la prime :

$$\hat{\epsilon}(p) = \frac{\partial \hat{\tau}(p)}{\partial p} \times \frac{p}{\hat{\tau}(p)}$$

$$\hat{\epsilon}(p) = Bp(1 - \hat{\tau}(p))$$

La prime est toujours définie positive. Le taux de refus du devis (qui est le complémentaire au taux de concrétisation) est compris entre 0 et 1. Le signe de l'élasticité est donc uniquement déterminé par celui du coefficient B.

L'élasticité de l'acte de concrétisation est négative, c'est-à-dire que le taux de concrétisation diminue lorsque le prix augmente. Une élasticité presque nulle indique que l'assuré n'est absolument pas sensible au prix qu'il paye. Si son tarif varie, sa probabilité de concrétiser son devis n'en sera pas affectée. A l'inverse, si l'élasticité est élevée en valeur absolue, cela signifie que l'assuré est très sensible au niveau de tarif qu'il paye.

La formule de l'élasticité rend possible l'étude des résultats sur l'ensemble des devis ou par segment de population suivant les variables jugées intéressantes.

Le deuxième objectif est d'analyser l'impact d'une variation de tarif sur le comportement des clients lors de l'acte de résiliation. Comme pour l'acte de concrétisation, une modélisation logistique a permis d'estimer la probabilité qu'un assuré résilie à la date d'échéance son contrat en fonction de ses caractéristiques et de son montant de prime

Considérant l'importance de croiser le niveau de prime avec les autres variables explicatives dans le calcul d'élasticité, le modèle contient trois effets respectivement liés :

- aux caractéristiques de l'assuré et indépendant de la prime ;
- uniquement à la prime quel que soit le segment de population auquel appartient l'assuré ;
- à l'effet croisé des caractéristiques de l'assuré en fonction du niveau de prime.

Une fois la modélisation validée (par les mêmes statistiques que pour le premier environnement d'étude), la probabilité de résiliation d'un contrat à sa date d'échéance pour un assuré s'écrit de la même manière que celle de concrétisation où nous avons aussi :

- Un coefficient ne faisant pas intervenir le prix et qui correspond donc aux caractéristiques de l'assuré : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)I_{x_k=i_k}$;
- Un coefficient lié aux caractéristiques de l'assuré et au prix : $B = B(x_1, \dots, x_m) = b_0 + \sum_{l=1}^m b_l(i_l)I_{x_l=i_l}$;

Pour les résiliations, l'élasticité est positive c'est-à-dire que le taux de résiliation augmente lorsque le prix augmente. Pour chaque assuré, tout comme dans l'analyse précédente, nous avons estimé la probabilité de résiliation d'un contrat à son terme ainsi que l'élasticité.

Quels que soient les scénarios, hausse ou baisse de prime, cette valeur d'élasticité engendre une variation de taux très faible.

Avant la date d'échéance du contrat, la compagnie effectue une revalorisation de la prime de chaque assuré qui a pour conséquence, dans la majorité des cas, une majoration de cette prime. Ainsi, avant de renouveler son contrat, un assuré connaît la future prime qu'il devra payer. Il peut résilier en invoquant le motif de «hausse de prime» à condition que cette dernière ne résulte pas d'un malus. Cette information est très pertinente pour estimer la probabilité de résiliation par l'ajout d'une nouvelle variable explicative : la variation de prime.

Un deuxième modèle a été mis en place, en utilisant la variation de prime comme indicateur du prix. Comme pour le premier modèle, la variation de prime a été croisée avec les variables caractéristiques car elle fait directement intervenir le prix. La variable prime avant échéance devient alors une variable que la compagnie connaît et sur laquelle elle n'a plus de pouvoir décisionnel. Seule la variation de prime peut être modifiée.

L'élasticité trouvée à l'aide de ce deuxième modèle est supérieure à celle du modèle précédent. Dans ce cas, selon les scénarios, hausse ou baisse de prime, cette valeur d'élasticité engendre une variation du taux de résiliation plus élevée : l'assuré est plus sensible à la variation de sa prime qu'à son niveau de prime. Il semble que l'assuré a un repère de prix qui est sa prime «avant échéance», et que la variation de prime, en cas de forte majoration, peut attirer l'attention sur les prix proposés par la concurrence.

L'analyse des résultats obtenus sur les différents segments d'assurés a mis en avant plusieurs phénomènes généraux qui interviennent d'une façon directe ou indirecte, sur la variation du niveau de prime, et donc sur les valeurs d'élasticité :

– **Les études de sinistralité dévoilent les profils à risque :**

La segmentation des primes engendre une différence de tarif entre les différents conducteurs selon leurs caractéristiques. Par exemple, les résultats établis sur le cri-

rière de l'âge ont révélé que plus le conducteur est âgé plus le montant de prime diminue tout comme l'élasticité. En effet, les jeunes conducteurs sont des conducteurs inexpérimentés avec une fréquence et un coût moyen de sinistres nettement supérieurs. Ainsi, leurs primes sont élevées, le jeune conducteur sera donc plus sensible au prix que les autres prospects. A l'opposé, les classes d'âge au dessus de 35 ans sont des conducteurs responsables et considérés comme de bons risques par la compagnie. Leur montant de prime est minoré, ce qui induit une augmentation du taux de concrétisation et donc une baisse de l'élasticité.

- **Les actions commerciales effectuées par la compagnie fidélisent les clients :**
Le produit d'assurance automobile est un produit d'appel ; il s'agit, dans la majorité des cas, du premier contrat souscrit par un client, donnant droit à des réductions tarifaires sur d'autres produits (en cas de souscription à ces produits). Ainsi, grâce aux réductions proposées par l'agent, le client peut passer rapidement d'un statut mono-détenteur à celui de multi-détenteur, avec une probabilité de résiliation qui s'affaiblit avec le temps et le nombre total de contrats souscrits.
- **Le pouvoir d'achat des assurés doit être pris en compte :**
La sensibilité au prix dépend de la conjoncture économique actuelle. En général, les jeunes ont des moyens financiers plus faibles que les autres conducteurs. Ainsi, la prime élevée qui leur est proposée ne fait qu'augmenter leur sensibilité.

En conclusion, cette étude révèle une situation complexe où la compagnie cherche à maîtriser le risque auquel elle doit faire face, en proposant des tarifs par segment, tout en essayant de proposer des prix plus attractifs que la concurrence afin d'attirer de nouveaux clients et de fidéliser les assurés actuels.

Par ailleurs, cette étude a été effectuée sur des données issues d'un réseau d'agents généraux. Il se peut que l'acte de concrétisation et de résiliation soient liés au prix mais aussi à des facteurs exogènes qui n'apparaissent pas dans les bases de données. Certains clients peuvent être prêts à payer un prix majoré contre une facilité et une rapidité de gestion. Ainsi, il peut s'avérer que les valeurs d'élasticité trouvées soient plus faibles que celles d'autres acteurs du marché comme par exemple les réseaux directs d'assurance.

Synthesis

Key words : *Car pricing, pricing, segmentation, price sensitivity, market premium, subscription rate, cancellation rate, generalized linear models.*

In France, car insurance is the most important non-life insurance market²¹, primarily because of the need to ensure and the number of vehicles. As there are a many customers, premiums collected by insurers are a lot of money. Figures from the French Federation of Insurance Companies show that 19,5 billion, or 40 % of the insurance premiums of property and liability, are provided by the auto insurance market, making it vital for companies.

Different actors²² are involved in a price war to win market shares. In parallel, with the development of new technologies, insurance comparators publish evaluations of individual contracts based on several criteria (price, guarantees, quality of services provided, etc.); accentuating the competitive pressure.

The actuary is confronted with two tariff issues whose objectives diverge : first segmentation by level of risk and then attractiveness of the price. Respectively, the objectives are to control the risk and attract prospects.

The study of the sensitivity of an insured to the quote seems necessary for each insurer. Too high prices compared to the competition will be an obstacle to the purchase or incitement to terminate the contract. Consumer behavior is difficult to predict, individual giving different weights to each criteria.

The objective of this paper is to analyze the tariff impact on the insured behavior regarding the act of subscription and the act of termination. To measure the sensitivity of the insured price, we use a micro-economic indicator, price elasticity $\epsilon(p)$, measured by the relative change in the rate of subscription or termination when the price varies from infinitesimal amount :

$$\epsilon(p) = \frac{\partial T(p)/T(p)}{\partial p/p}$$

where $T(p)$ is the rate of subscription or termination of policies based on the price p .

The aim of this analysis is to determine the impact of a small change in the price of the rate of subscription or termination.

The act of realization describes the succession of events :

1. The prospect comes to an agency of the company and shares his interest in a product.

21. in terms of number of contract.

22. insurance, bank-insurance, and mutual societies.

2. The sales representative establishes an estimate based on the characteristics of the prospect and his vehicle (age, age of the vehicle, ...).
3. The prospect decides, according to the guarantees and the price offered, whether or not to purchase the insurance contract with this company.

The success rate of a quote or conversion rate is defined as the probability that the prospect becomes a customer of the company.

Meanwhile, the act of termination, includes :

- termination by the insured at the maturity date (it coincides in most cases with the anniversary of the contract), after a possible increase of the premium, the insured decides to terminate or renew the contract.
- termination by the company that can take place at the due date because of excessive claims or at any time for non-payment of premiums, fake information at subscription or change of criteria.
- termination for loss or change of vehicle.

The termination rate is the proportion of terminated policies across the portfolio at the end of a given period.

Before discussing the actual modeling, all the necessary theoretical elements are explained, understanding the methods is crucial for the interpretation of results. The main tool is the logistic regression, member of the family of generalized linear models. Various indicators of goodness of fit and prediction models are used²³.

The perimeter of the study is bounded in time, only the data of the year 2012 is used. Databases are constructed for each study environment (subscription / termination). Regarding the act of termination, we only deal with contracts terminated on the date of maturity by the initiative of the insured. The variables are treated and cleaned. This step is essential to prepare the data for modeling in order to obtain reliable results.

The descriptive analysis revealed the profiles of prospects and policyholders highlighting different tariff criteria falling into three categories linked to :

1. the prospect/policyholder : driver age, seniority permits, marital status, history of permits, etc.
2. the vehicle : power, price range, length, etc.
3. circulation : geographical area, vehicle use, etc.

Bivariate studies, using graphics distribution of the rate²⁴ based on the terms and correlation tests show the significant variables for modeling. The variables being predominantly qualitative (and ordinal), the independence chi-square test is used. This test has the disadvantage of being sensitive to the number of observations. To confirm the potential dependencies, two tests are additionally applied : the correlation Kendall and Spearman tests.

These steps form the basis of the study, it is now possible to enter the modeling phase which is the subject of Part IV.

The first objective is to measure the impact of price on the subscription, or not, of a policy offer made by a sales representative. Information related to competition is crucial

23. AIC, Schwartz criterion, concordant pairs, good rate reclassification.

24. implementation and termination

for the subscription of the prospect or not, it must be incorporated to the model. The database includes the premium paid by some former policyholders to a competitor. This variable is very valuable in this economic environment where this type of information is difficult to obtain, however, it must be considered with caution, being exposed to certain biases. Indeed, it is the customers themselves who relate it so :

- some policyholders may reduce the premium in order to bring down the price of the quote,
- others give an approximate price, the exact rate being difficult to remember.

In this context, for a continuous and realistic distribution, premiums were smoothed (several times).

Based on comparisons of QQ-Plot and fit tests, the variable former premium to the competitor has approximately a gamma distribution. It is modeled with a generalized linear model, allowing to explicit the premium for the rest of the policyholders.

Then it comes to the determination of a statistical model capable of estimating the probability of a subscription based on the characteristics of the prospect and the price offered by the sales representative.

The first model was developed taking into account :

- the characteristics of the prospect, independently of the price ;
- price without distinction of population segment ;
- premium paid to a competitor (previously estimated)without distinction of population segment.

Interested in measuring the impact of the price segment of the population, we have developed a second model by introducing all the qualitative variables of the first model cross-leveled with the premium.

A comparison of the key quality indicators²⁵ showed that the second model is more effective than the first.

However, the large number of variables of the model prevents the operational application of the second model in everyday life. It is necessary to make an additional selection to lower the number of variables keeping the ones that :

- have a significant impact on the price ;
- may intervene in the pricing of the competition ;
- are not correlated ;
- involve a small number of modalities with approximately equal individuals per class.

After these selections, which decrease the number of variables from thirty four to twenty, the effectiveness of the model is retained even if the predictive power of the dependent variable seems weaker. The probability of occurrence based on the characteristics of the prospect and level premium is :

$$\hat{\tau}(i) = \frac{1}{1 + \exp[-(A(i) + B(i)p(i))]}$$

25. AIC criterion Schwartz, concordant pairs, good rate reclassification.

where we have two coefficients A and B :

- A is a coefficient which is independant of the premium quote and therefore corresponds to the characteristics of the prospect : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)i_{x_k=i_k}$;
- B is a coefficient related to the characteristics of the prospect and the premium $B = B(x_1, \dots, x_m) = b_0 + \sum_{l=1}^m b_l(i_l)i_{x_l=i_l}$;

Which can be simplified :

$$\hat{\tau}(p) = \frac{1}{1 + \exp[-(A + Bp)]}$$

This probability allows to calculate the sensitivity of each prospect's level of premium by deriving the $\hat{\tau}(p)$ function in regards to the premium :

$$\hat{\epsilon}(p) = \frac{\partial \hat{\tau}(p)}{\partial p} \times \frac{p}{\hat{\tau}(p)}$$

$$\hat{\epsilon}(p) = Bp(1 - \hat{\tau}(p))$$

The premium is always positive. The refusal rate for the estimate (which is the complement of the subscription rate) is between 0 and 1. The sign of the elasticity is therefore determined solely by the coefficient B.

The elasticity of the act of realization is negative, that is to say, the success rate decreases as the price increase. An inexistant elasticity indicates that the insured is not sensitive to the price he pays. If the rate varies, the probability of subscription will not be affected. Conversely, if the elasticity is high in absolute terms, this means that the insured is very sensitive to the level of price he pays.

The formula for elasticity makes it possible to study the results of all quotations or segment of the population according to the variables considered interesting.

The second objective is to analyze the impact of a tariff on customer's behavior during the act of termination. As for the act of subcription, a logistic model was used to estimate the probability that an insured person terminates his contract at the date of maturity according to its characteristics and the price of the premium.

Considering the importance of crossing the premium level with the other explanatory variables in the calculation of elasticity, the model contains three effects :

- the characteristics of the insured, independently of the premium ;
- the premium regardless of the segment of the population of the policyholder ;
- the cross-leveled effect of the characteristics of the insured based on the level of premium.

Once the model validated (by the same statistics as the first environmental study), the probability of termination of a contract at the due date for an insured is similar than the one for realization :

- A a coefficient not involving price and thus corresponding to the characteristics of the insured : $A = A(x_1, \dots, x_n) = a_0 + \sum_{k=1}^n a_k(i_k)i_{x_k=i_k}$;

- B a coefficient related to the characteristics of the insured and the price : $B = B(x_1, \dots, x_m) = b_0 + \sum_{l=1}^m b_l(i_l)i_l = x_{i_l}$;

For terminations, the elasticity is positive that is to say that cancellation increases when the price increases. For each insured, as in the previous analysis, we estimate the probability of termination of a contract as well as elasticity.

Whatever the scenario, increase or decrease in premiums, the value of elasticity causes a very low variation of rates.

Before the expiry of the contract, the company performs a revaluation of the premium of each insured which results in most cases, in an increase of the premium. Thus, before renewing his contract, an insured knows his future premium. He may terminate based on grounds of "premium increase" provided that it does not result from a penalty. This information is relevant to estimate the probability of termination by the addition of a new explanatory variable : the change in premium.

A second model was developed, using the change of premium as an indicator of price premium. As the first model, the change in premium was cross-leveled with variables because it is directly involved the price. The variable premium prior to maturity becomes a fixed known variable ; only the change in premium can be modified.

Elasticity found using this second model is greater than that of the previous model. In this case, depending on the scenario, increase or decrease in premiums, the value of elasticity causes a higher variation of cancellation rate : the insured is more sensitive to changes in the premium than to the premium level. It seems that the insured has a benchmark which is its price premium "before maturity" and that the variation of premium, if it is high can draw attention to the prices offered by competitors.

Analysis of the results obtained in the different segments of insured highlighted several general phenomena involved directly or indirectly, on changes in the level of premium, and thus the elasticity values :

- **Claim studies reveal risk profiles :**

Segmentation generates a different premium price for each driver depending on its characteristics. For example, the results established on the grounds of age revealed that the more a driver ages, the more the premium and the elasticity decrease. Indeed, young drivers are inexperienced drivers with a frequency and average cost of claim significantly higher. Thus, with higher premiums, the young driver will be more sensitive to price than other prospects. In contrast, the age groups above 35 years are responsible drivers and considered good risks by the company. Their premiums are therefore reduced, which leads to an increase in the rate of subscription and thus a decrease in elasticity.

- **The commercial activities carried out by the company gain the loyalty of customers :**

The auto insurance product is a loss leader, it is, in most cases, the first contract signed by a client, making him entitled to tariff reductions on other products (in case of subscription to these products). Thus, thanks to the cuts proposed by the

sales representative, the customer can quickly switch from the mono-holder to the multi-holder status, with a weaker probability of termination over time and the total number of contracts signed.

– **The purchasing power of the insured must be taken into account :**

The price sensitivity depends on the current economic environment. In general, young people are less fortunate than other drivers. Thus, the high premiums they are offered increase their sensitivity.

In conclusion, this study reveals the complex situation where the company seeks to control the risk it faces by offering rates by segment, while trying to offer more attractive premium than the competitors to attract new customers and retain current policyholders.

Furthermore, this study was performed on data from a network of sales representatives. It may be that the act of subscription and termination are related to price but also to exogenous factors that do not appear in the database. Some customers may be willing to pay a higher premium for an ease and speed of policy management. Thus, it may be that the elasticity values found are lower than those of other market participants such as direct insurance networks.