

## Mémoire de fin d'Etudes

Thème :

**Tarification en assurance automobile :  
Modélisation de la prime pure et individualisation du risque  
Cas de la CASH assurances : Garanties dommages**

Présenté et soutenu par :

**AZZOUZI Ziad**

Encadré par :

**Mr. Rassem KTATA**

Etudiant(e) parrainé(e) par :

**Compagnie d'Assurance des Hydrocarbures**

## **Remerciement**

*Au terme de ce travail, je remercie en premier lieu tous les enseignants et toute l'équipe de la direction générale de l'institut de financement du développement du Maghreb arabe.*

*Je tiens à remercier, en particulier, **Mr SLEHEDDINE LOUHICHI** pour sa disponibilité, ses conseils et orientations précieuses tout au long de la période de formation, Notre encadrant **Mr. RASSEM KTATA** pour son suivi et sa disponibilité.*

*Je tiens également à remercier, **Mme. NEHAB TASSADIT** de m'avoir accueilli dans son équipe ainsi que son suivi durant la période de stage. Sans omettre mes remerciements envers toute l'équipe de la direction automobile pour l'intérêt qu'ils m'ont apporté durant la période de stage ainsi que pour leurs coopérations professionnelles.*

*Je remercie également toute l'équipe de la direction de formation.*

*Je remercie **Mme LARBAOUI LYLIA** pour son suivi et son soutien tout au long de la période de formation.*

*Enfin, une pensée particulière pour mes parents et mon épouse, qui m'ont soutenu tout le long de la rédaction de ce mémoire et bien au-delà.*

# **SOMMAIRE**

# Sommaire

## Remerciement

## Sommaire

## Liste des figures

## Liste des tableaux

## Introduction

Enjeux actuels : .....	i
Exposé de la problématique : .....	i
Stratégie du travail : .....	ii

## Partie I : Cadre théorique

### Chapitre 1 | concept des assurances

<b>Section 1</b> : Généralités sur les assurances .....	1
1.1. La notion de risque.....	1
1.1.1. Définition du risque .....	2
1.1.2. La classification du risque .....	2
1.1.3. Le risque assurable .....	4
1.2. Définition du concept d'assurance.....	6
1.2.1 Le contrat d'assurance.....	7
1.2.2. Les branches d'assurance .....	8
1.3. Les fondements de l'assurance .....	9
1.4. Les mécanismes fondamentaux de l'assurance.....	10
1.5. Rôle de l'assurance : .....	13
1.5.1. La fonction réparatrice : .....	13
1.5.2. Facteur dynamisant le progrès social.....	14
1.5.3. Rôle de prévention : .....	15
1.5.4. Rôle financier de collecteur d'Épargne .....	15
<b>Section 2</b> : l'Assurance automobile .....	16
2.1. Définition de l'assurance automobile .....	16
2.2. Le contrat d'assurance automobile .....	17
2.2.1. Le contrat mono véhicule (individuel) .....	17

2.2.2. Le contrat flotte .....	17
2.3. La Responsabilité civile obligatoire.....	17
2.3.1. Définition d'un véhicule terrestre à moteur .....	18
2.3.2. Le conducteur du véhicule assuré .....	18
2.3.3. Les notions d'assuré et de tiers .....	19
2.4. Garantie facultatives .....	19
2.4.1. Les garanties principales .....	19
2.4.2. Les garanties complémentaires .....	21
2.4.3. Les garanties accessoires.....	22
2.4.4. Les garanties optionnelles .....	22

## **Chapitre 2 | La mesure du risque dans l'assurance automobile**

<b>Section 1</b> : principe de tarification et notions de primes .....	27
1.1. Variables de mesures de la sinistralité : .....	27
1.1.1. Le nombre de sinistres .....	27
1.1.2. Le coût de sinistres .....	28
1.1.3. La durée d'exposition au risque .....	28
1.1.4. La Fréquence .....	28
1.2. Principe de tarification et nécessité de segmenter : .....	28
1.3. La tarification de l'assurance automobile .....	30
1.3.1. La tarification a priori .....	30
1.3.2. La tarification a posteriori .....	31
1.4. La détermination des différentes primes ou cotisation .....	33
1.4.1. La cotisation pure (la prime pure) .....	33
1.4.2. Cotisation nette .....	34
1.4.3. Cotisation totale ou commerciale .....	35
1.5. Revue de la littérature .....	36
<b>Section 2</b> : Outils statistiques .....	41
2.1. Le modèle de régression linéaire .....	41
2.2. Principe du modèle linéaire généralisé : .....	42
2.2.1. Les composantes d'un GLM .....	43
2.2.2. Le choix du modèle : .....	44
2.2.3. Estimations des paramètres par maximum de vraisemblance.....	45

2.2.3. Adéquation et validation du modèle : .....	45
---------------------------------------------------	----

## **Partie II : aspect pratique**

### **Chapitre 3 | prérequis pour la modélisation**

<b>Section 1</b> : Présentation de la base et traitement des variables .....	48
1.1. Origine de la base de données.....	48
1.2. Traitement des données.....	51
1.3. Traitement des variables .....	52
<b>Section 2</b> : analyse univariée et bivariée des variables .....	58
2.1. Analyse univariée.....	58
2.2. Analyse bivariée.....	65
2.2.1. Étude de liens entre les variables.....	65
2.2.2. Étude de liens entre variables explicatives et variable à expliquer .....	69

### **Chapitre 4 | Modélisation de la fréquence et le coût de sinistres**

<b>Section 1</b> : Modèle pour la fréquence .....	75
1.1. Choix de la distribution théorique.....	75
1.2. Estimation des coefficients de la régression .....	76
1.3. Sélection des variables .....	78
1.4. Validation du modèle .....	80
<b>Section 2</b> : Modèle pour le coût .....	83
2.1. Choix du modèle .....	83
2.2. Estimation des coefficients de la régression .....	84
2.3. Sélection des variables .....	86
2.4. Validation du modèle .....	88
2.5. Comparaison tarifaire Homme/Femme.....	91

## **Conclusion**

## **Bibliographie**

## **Annexes**

# **LISTE DES FIGURES**

## **Liste des figures :**

<b>Figure 01 :</b> catégories des assurances .....	8
<b>Figure 02:</b> Les différentes formes de la garantie facultative .....	24
<b>Figure 03 :</b> Les différentes cotisations.....	36
<b>Figure 04 :</b> rapport variance inter-classe sur variance totale en fonction du nombre de classes.....	56
<b>Figure 5 :</b> Découpage en classes d'âge de la variable Age. Client. ....	57
<b>Figure 6:</b> montant et nombre de sinistres par année de survenance .....	58
<b>Figure 07:</b> montant moyen de sinistres par année de survenance.....	59
<b>Figure 08:</b> composition du portefeuille et répartition de la charge de sinistres par garantie et par année.....	60
<b>Figure 09:</b> Composition du portefeuille par genre .....	60
<b>Figure 10:</b> fréquence et coût moyen par genre .....	61
<b>Figure 11:</b> Composition du portefeuille par âge du véhicule. ....	61
<b>Figure 12:</b> Répartition de la fréquence et le coût moyen de sinistres par classe d'âge du véhicule .....	62
<b>Figure 13:</b> Composition du portefeuille et répartition de la fréquence et la charge de sinistres par classe de valeur du véhicule.....	62
<b>Figure 14:</b> composition du portefeuille et répartition de la charge sinistre par zone géographique .....	63
<b>Figure 15:</b> Composition du portefeuille et répartition de la fréquence et la charge de sinistres par nombre de chevaux du véhicule.....	64
<b>Figure 16:</b> composition du portefeuille et répartition de la fréquence et la charge sinistre par classe de valeur du véhicule.....	65
<b>Figure 17:</b> tableau de contingence de l'âge de client et l'âge du véhicule .....	67
<b>Figure 18:</b> analyse de comportement de groupes pris deux à deux .....	70

<b>Figure 19:</b> Ajustement de la distribution théorique à une distribution de la loi de poisson	75
<b>Figure 20:</b> estimation des coefficients du modèle initial de fréquence .....	77
<b>Figure 21:</b> processus de sélection des variables .....	78
<b>Figure 22: estimation des coefficients du modèle de fréquence après sélection des variables .....</b>	<b>79</b>
<b>Figure 23:</b> AIC des modèles pour les lois envisagées .....	<b>83</b>
<b>Figure 24: estimation du modèle initial .....</b>	<b>85</b>
<b>Figure 25:</b> processus de sélection des variables .....	<b>86</b>
<b>Figure 26:</b> estimation du modèle après sélection.....	<b>87</b>

# **LISTE DES TABLEAUX**

## Liste des tableaux :

<b>Tableau 01 :</b> Récapitulatif des fonctions de lien des modèles usuels.....	44
<b>Tableau 02 :</b> Les différentes catégories de puissance.....	50
<b>Tableau 03 :</b> Type et nature des variables de la base .....	53
<b>Tableau 04:</b> découpage en classes des variables quantitatives.....	58
<b>Tableau 05:</b> p-value des tests du Khi-deux d'indépendance .....	67
<b>Tableau 06:</b> indices de Cramer.....	68
<b>Tableau 07:</b> significativité de l'influence sur la fréquence et coût de sinistres .....	71
<b>Tableau 08:</b> résultats et Qualité d'ajustement - Synthèse .....	76
<b>Tableau 09:</b> variables intégrées au modèle initial de la fréquence de sinistres.....	77
<b>Tableau 10:</b> la fréquence moyenne par modalité en comparaison avec la fréquence moyenne de référence.....	81
<b>Tableau 11:</b> variables intégrées au modèle initial de la fréquence de sinistres.....	84
<b>Tableau 012:</b> Le coût moyen par modalité en comparaison avec le coût moyen de référence .....	89
<b>Tableau 13:</b> comparaison entre progression des coefficients du modèle et progression des coûts moyens observés .....	89
<b>Tableau 14 :</b> estimation du taux de prime par valeur du véhicule et genre du conducteur	91

# **INTRODUCTION**

## Enjeux actuels

Environ 55 % du chiffre d'affaires du secteur des assurances, qui a atteint 133,9 milliards de dinars en 2016, provenait de l'assurance automobile qui continue de compter énormément sur le produit de responsabilité civile dont le montant est fixé par l'État, sans prospecter suffisamment les autres opportunités. Mais si le nombre de véhicules importés continue de reculer, les compagnies d'assurances, publiques et privées, seront dans l'obligation d'imaginer d'autres stratégies pour proposer aux automobilistes des assurances facultatives.

Le marché de l'assurance automobile est parfois qualifié de marché concurrentiel. De plus, l'assurance automobile constitue généralement un des premiers contacts entre les particuliers et les assureurs, c'est ainsi un produit d'appel. Il est alors important pour les assureurs d'attirer chez eux de nouveaux clients en espérant les fidéliser. En effet, un client satisfait de son assureur pourra souscrire à d'autres produits chez ce dernier.

Cette forte concurrence, avec la nouvelle conjoncture économique, caractérisée par une contraction des recettes en devises issues des exportations des hydrocarbures conduisant à une tendance à la baisse des importations, y compris celle des véhicules, pousse les assureurs à raffiner leurs modèles de tarification, dans le but de se différencier. La maîtrise d'une bonne tarification est donc cruciale.

## Exposé de la problématique :

Quatre-vingt-dix pour cent (90%)<sup>1</sup> des besoins de transports exprimés dans notre pays sont couverts par le transport routier. Il faut toutefois relever que cette activité est ternie par un taux d'accidents alarmant, comme le montrent les statistiques<sup>2</sup> pour l'année 2014 (nombre d'accidents 40101, nombre de blessés 65263, nombre de décès 4812).

---

<sup>1</sup> RIAD Meriem Maitre assistante à Université de Tipaza; Doctorante à l'ENSSEA ; Chercheur Laboratoire LASAP «modèle de tarification optimal en assurance automobile dans le cadre d'un marché règlementé.

<sup>2</sup> Centre national de prévention et sécurité routière <http://www.cnpsr.org.dz/>

Dans ce cas, il est quasiment impossible aux compagnies de savoir avec certitude combien les prestations qu'elles vendent leur coûteront, ainsi, pour fixer le montant de sa prime l'assureur ne peut se baser que sur des études statistiques lui permettant de se faire une idée sur le coût de la prestation en cas de sinistre.

L'enjeu de ce mémoire est de s'assurer que la tarification adoptée par **la CASH assurances** est toujours en adéquation avec la sinistralité constatée de ses assurés. Autrement dit, l'objectif final étant de proposer une tarification sur les profils trop risqués, et parallèlement, d'augmenter la souscription d'affaires nouvelles sur des profils bénéficiaires.

Pour apporter une réponse au problème posé, nous avons extrait un certain nombre d'informations sur la sinistralité des véhicules ayant souscrit au moins une garantie dommage sur la période (2010-2016) à la compagnie CASH Assurances.

Après une analyse descriptive des données nous permettant **de segmenter au mieux le portefeuille** des risques, nous allons construire un modèle de la fréquence moyenne des sinistres en fonction d'un certain nombre de variables en utilisant la technique **GLM**. Ensuite, nous allons estimer le coût relatif moyen des sinistres en fonction également d'un certain nombre de variables. Enfin, il convient de déduire le taux de prime pure des garanties dommages.

### **Stratégie du travail :**

En vue de développer l'ensemble de ces idées, nous avons structuré notre étude en quatre chapitres qui se présentent comme suit :

**Le premier chapitre s'intitulant « concept des assurances »** traitera le risque d'une part et l'assurance d'autre part. Il permet, dans une première partie, d'apprécier l'importance de l'assurance, de connaître les classes du risque et comprendre l'aspect assurable du risque. La deuxième partie sera consacrée à la présentation des différentes notions de l'assurance automobile et les concepts liés à cette matière d'assurance.

**Le deuxième chapitre intitulé «La mesure du risque dans l'assurance automobile »** sera dédié aux notions de mesure de risque automobile, au principe de tarification de ce risque, aux différentes primes ou cotisations de couverture contre ce risque et les outils mathématiques servant à la modélisation de la prime pure par l'approche fréquence-coût.

**Le troisième chapitre s'intitulant « prérequis pour la modélisation »** présentera, dans une première partie, le portefeuille et les différentes étapes de traitements appliqués sur ce portefeuille afin de le rendre exploitable. La deuxième partie, sera consacrée, d'une part, à l'analyse univariée des variables que nous disposons dans notre portefeuille, et d'autre part, à l'étude des liens existants entre ces différentes variables.

**Le quatrième chapitre s'intitulant « modélisation de la fréquence et le coût de sinistres »** a pour objet d'estimer le risque individuel à partir des variables intrinsèques au véhicule, au conducteur et à la police d'assurance. Nous nous intéressons à l'estimation de la fréquence, du coût moyen et de la prime pure.

# **CHAPITRE I**

## **CONCEPT DES ASSURANCES**

Le Développement de l'environnement économique et industriel a intensifié les risques pouvant menacer l'existence des entreprises et empêcher des hommes d'action à accomplir leurs missions. De ce fait, une couverture contre ces risques est primordiale pour préserver la pérennité de ces derniers. Pour cela, il est toujours nécessaire de définir et de mettre en point certains nombres de concepts qui feront comprendre et saisir l'objet et la finalité de ce travail.

Dans le souhait d'atteindre le but et pour mener à bien cette étude, le choix a été de commencer par un chapitre généraliste, intitulé : « le concept des assurances » et qui présente des généralités sur les assurances dans un cadre général et l'assurance automobile en particulier. Il permet de comprendre l'activité de l'assurance, distinguer les risques assurables et d'apprécier les spécificités de cette activité. De ce fait, ce chapitre sera présenté comme suit :

### **Chapitre 1 : le concept des assurances**

- **Section 1** : généralité sur les assurances
- **Section 2** : assurance automobile

---

## **Chapitre 1 : le concept des assurances**

---

### **Section 1 : Généralités sur les assurances**

Cette section traite le risque d'une part et l'assurance d'autre part. Elle permet d'apprécier l'importance de l'assurance, connaître les classes du risque et comprendre l'aspect assurable du risque.

#### **1.1. La notion de risque**

Cette partie traitera la notion du risque en présentant une définition du risque, sa classification et en abordant de manière particulière le risque assurable.

### 1.1.1. Définition du risque

« Un risque est une situation (ensemble d'événements simultanés ou consécutifs) dont l'occurrence est incertaine et dont la réalisation affecte les objectifs de l'entreprise qui le subit »<sup>3</sup>.

Selon l'organisation internationale de normalisation ISO, le risque est « la possibilité d'occurrence d'un événement ayant un impact sur les objectifs, il se mesure en terme de conséquences et de probabilité ».

Une définition donnée par le Committee of Sponsoring Organisations of the Treadway Commission (COSO)<sup>4</sup> dans le cadre de référence du management des risques de l'entreprise est la suivante : « un risque représente la possibilité qu'un événement survienne et nuise à l'atteinte d'objectifs »<sup>5</sup>.

À partir de ces définitions, on peut définir le risque par : un événement dont la réalisation est probable et les conséquences affectent l'atteinte des objectifs.

### 1.1.2. La classification du risque

La classification des différents types de risque peut se faire selon une panoplie de critères. Elle permet d'identifier les caractéristiques techniques, sociales, économiques et juridiques du risque couvert par chaque branche d'assurance.

Le risque peut être classé selon les critères suivants :<sup>6</sup>

➤ **Selon la chose assurée :** on distingue entre l'assurance de dommages et l'assurance de personnes. Le premier type concerne l'assurance de biens (pour les dommages causés aux biens de l'assuré) ou de responsabilité (pour les dommages causés au tiers engageant la responsabilité de l'assuré). Le deuxième type concerne la protection de la vie

<sup>3</sup> B. Barthélémy, Gestion des risques méthode d'optimisation globale, l'organisation novembre 2001, p13

<sup>4</sup> Le management des risques de l'entreprise, Cadre de Référence – Techniques d'application, COSO II, Editions d'Organisation, paru en 2005 est la traduction française de l' « Enterprise Risk Management – Integrated Framework » publié par le Committee of Sponsoring Organisations of the Treadway Commission.

<sup>5</sup> Idem, p 23.

<sup>6</sup> Murielle PAULIN, économie et organisation de l'assurance, Séfi édition, Québec, canada, 2007, p46.

humaine et il englobe les assurances individuelle accident et santé, assurances sur la vie et décès.

➤ **Selon la nature de risque :** On distingue entre :

- Le risque professionnel,
- Le risque particulier,
- Le risque d'entreprise,
- Le risque maritime,
- Le risque de pollution.

➤ **Selon le mode de gestion :** on distingue entre les assurances par capitalisation et les assurances par répartition. Selon le premier mode de gestion, les primes collectées sont capitalisées selon la méthode d'intérêt durant toute la période du contrat et seront versées à l'assuré ou à ses ayants droit à l'échéance selon les conditions particulières du contrat d'assurance. Selon le second mode de gestion, l'assureur répartit la masse des primes ou des cotisations collectées sur l'ensemble des assurés sinistrés.

➤ **Selon leur caractère :** on distingue entre les assurances obligatoires qui représentent les différents types d'assurance que l'assuré est obligé de souscrire (comme la responsabilité civile en assurance automobile) et les assurances facultatives qui laissent à l'assuré la faculté de la souscription (comme la garantie bris de glace).

➤ **Selon l'importance des risques couverts :** on distingue entre les grands risques et les risques de masse. Ces deux types exigent des approches actuarielles de tarification différente vu la différence du volume, de l'historique des sinistres et de l'importance de l'aléa.

➤ **Selon le mode de souscription :** on distingue entre les assurances individuelles souscrites par un seul assuré et les assurances collectives dont l'assuré est un groupe de personnes (comme les assurances souscrites par une entreprise pour le compte de ses employés : assurance groupe).

➤ **Selon l'événement assuré** : on distingue entre les risques suivants :

- Incendie,
- Dégât des eaux,
- Accident,
- Pertes d'exploitation,
- Vol,
- Vandalisme,
- Catastrophes naturelles,
- Tempête,
- Grêle, neige,
- Bris (de glace, de machine),
- Explosion,
- Litige,
- Etc.

### 1.1.3. Le risque assurable

L'assureur offre aux assurés une couverture contre les risques en s'engageant de les indemniser en cas de survenance du sinistre couvert qui cause des dommages. Toutefois, il existe une certaine catégorie de risques qui ne sont pas assurables. De ce fait, l'assureur accepte de couvrir seulement les risques qui vérifient les critères d'assurabilité.

Le risque n'est assurable que s'il répond à deux types de conditions d'assurabilité : conditions techniques, conditions juridiques.

#### a) Les conditions d'ordre juridique

Le risque assurable revêt les caractères juridiques suivants :<sup>7</sup>

- L'événement doit être futur. Le risque ne doit pas être réalisé (c'est un risque certain),
- L'événement doit être incertain et aléatoire. Cette incertitude réside soit dans la survenance de l'événement (puisque'il y a une possibilité que l'événement ne

---

<sup>7</sup> François COUILBAULT, Constant ELIASHBERG, Les grands principes de l'assurance, l'argus, 9<sup>ème</sup> édition, 2009, p54.

surviennent pas comme l'incendie), soit dans la date de la survenance de l'événement (le décès qui est un événement certain, mais la date de survenance est inconnue).

- La survenance de l'événement ne doit pas dépendre exclusivement de la volonté de l'assuré.

Donc le risque assurable est un événement futur incertain et ne dépend pas exclusivement de la volonté de l'assuré.

### b) Les conditions d'ordre technique

Les différentes conditions que doit vérifier le risque pour qu'il soit assurable sont <sup>8</sup>:

- La possibilité du calcul avec une précision suffisante de la probabilité de réalisation du risque,
- Le classement des risques dans des catégories étroites afin d'assurer l'homogénéité des classes de risques,
- La possibilité de diversifier (risques diversifiables) les risques dans l'espace et dans le temps puisque les risques touchant l'ensemble de la mutualité des assurés en même temps mettent en péril la compagnie d'assurance.

Les conditions juridiques et techniques citées ci-dessus justifient l'inassurabilité de certains risques. Il s'agit de :

- **Le risque nouveau** : ce risque est récemment identifié et n'a pas fait l'objet d'étude statistique.
- **Le risque rare ou les risques spéciaux** : le traitement de ces risques est rare puisque la fréquence de survenance est faible. En conséquence, ces risques restent étranges et leur tarification est difficile.
- **Le risque énorme** : la probabilité de survenance de ce type de risque est importante. Exemple : les séismes dans une zone à risque.

---

<sup>8</sup> Murielle PAULIN, économie et organisation de l'assurance, Séfi édition, Québec, Canada, 2007, P49

- **Le risque isolé** : ce risque concerne un endroit ou un élément bien déterminé et sa fréquence de survenance est faible. L'assureur ne peut construire une mutualité homogène pour assurer ce genre de risque.

## 1.2. Définition du concept d'assurance

L'assurance est définie d'une manière générale comme :

« Une réunion de personnes qui craignant l'arrivée d'un événement dommageable pour elles, se cotisent pour permettre à ceux qui seront frappés par cet événement de faire face à ses conséquences ». <sup>9</sup>

Selon Pedro de SANTAREM, juriste portugais, auteur d'un des premiers ouvrages consacrés à l'assurance maritime, l'assurance est une : « convention par laquelle, le prix d'un risque ayant été convenu, l'un prend pour lui le risque de l'infortune de l'autre ». <sup>10</sup>

En effet, beaucoup d'auteurs ont cherché à définir l'opération d'assurance, ainsi pour Jérôme YEATMAN : « l'assurance est l'organisation rationnelle d'une mutualité de personnes soumises à l'éventualité de réalisation de même risque qui, par leur contribution financière, par ceux d'entre eux qui sont effectivement frappés par le risque ». <sup>11</sup>

On remarque que toutes les définitions précédentes tournent autour d'un ensemble d'idées à savoir une entreprise d'assurance pour couvrir et gérer des risques ou des contrats d'assurance d'un ensemble d'assurés, toutes catégories confondues, à l'aide d'un fonds alimenté par les cotisations collectées au prêt des assurés, et par le biais de cette définition on comprend que l'assurance est une opération financière dont l'objectif est de permettre l'indemnisation des dommages survenus aux personnes.

---

<sup>9</sup>Michel Latrasse, ComstantEliashberg et FrancoisCuilbault, *Les grands principes de l'assurance*, Edition l'Argus, 5<sup>ème</sup> édition, Paris 2002 p : 49.

<sup>10</sup>Frédéric Malaval, *Développement durable, assurance et environnement*, Edition Economica, Paris 1999, p : 20.

<sup>11</sup>JeromeYeatman, *Manuel international de l'assurance*, Edition Economica, Paris 1998, p: 01.

### 1.2.1 Le contrat d'assurance

Le contrat d'assurance est une convention passée entre une entreprise d'assurance et un souscripteur (individu ou collectivité), fixant à l'avance, pour une période déterminée, des charges financières en fonction d'un ensemble bien défini d'événements aléatoires.<sup>12</sup> Il doit comporter les conditions générales qui sont nécessairement complétées par les conditions particulières, ces dernières permettent l'adaptation du contrat à chaque risque et à chaque assuré.

Conformément au droit algérien, les contrats, en Algérie, sont régis par l'ordonnance n° 95-07 du 25 janvier 1995 (modifiée par la loi 06-04) relative aux assurances, ainsi que par les dispositions du Code civil algérien.

Le contrat d'assurance est donc une convention, établie entre l'assuré et la société, qui se matérialise par une police d'assurance qui comprend :<sup>13</sup>

1. Les conditions générales : Ce sont les textes non personnalisés qui définissent les garanties, leurs limites, leurs exclusions, les engagements réciproques des parties, etc., en tenant compte des dispositions légales et réglementaires en vigueur.
2. Les conditions particulières : C'est le document qui précise notamment :
  - les noms et adresse de la personne physique ou morale qui souscrit,
  - la situation de risque,
  - les caractéristiques du risque,
  - les garanties souscrites et le montant des capitaux,
  - la durée de garantie du contrat et sa date d'effet,
  - la prime à payer, le montant de(s) franchise(s) et éventuellement les surprimes et majorations, etc.

<sup>12</sup>François Ewald-Jean HérnéLorenzi, *Encyclopédie d'assurance*, Economica, 1997, p : 432.

<sup>13</sup>KPMG, *Guide des Assurances en Algérie*, Edition 2009, p : 57.

### 1.2.2. Les branches d'assurance

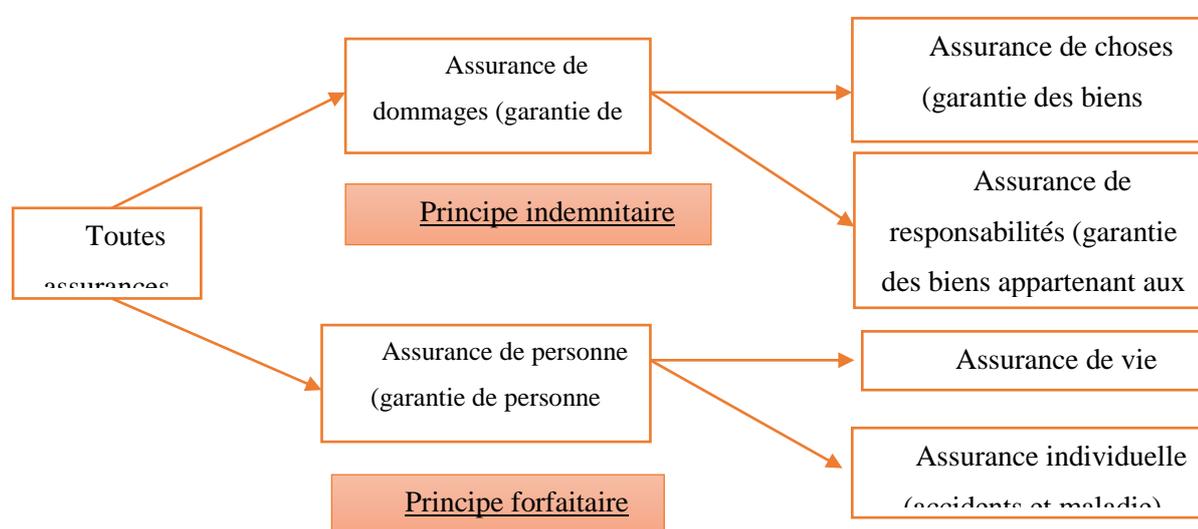
Le secteur des assurances recouvre deux grands domaines d'activité : les assurances de dommages et les assurances de personnes.

#### a) Les assurances de dommages

Regroupent les assurances de biens (assurances relatives aux habitations, aux locaux industriels et commerciaux, aux navires, aéronefs et à leurs cargaisons ....etc.) ainsi que les assurances de responsabilité (responsabilité civile du chef de famille, du chef d'entreprise, du transporteur...etc.).

#### b) Les assurances de personnes

Couvrent, quant à elles, l'ensemble des risques afférents à la personne humaine, elles regroupent l'assurance vie proprement dite ainsi que les assurances relatives aux accidents corporels, à la maladie et à la perte d'emploi.<sup>14</sup>



Source : J.M.Rousseau, T.Blayak et N.Oulmane, *Introduction à la théorie d'assurance*, Édition Dunod, Paris, 2001, p : 42.

**Figure 01** : catégories des assurances

<sup>14</sup> Théodore Corfias, *Assurance vie : technique et produits*, Edition l'Argus, Paris 2003, p : 23.

### 1.3. Les fondements de l'assurance

Les fondements de l'opération d'assurance peuvent être présentés à deux niveaux : le niveau juridique et le niveau technique comme suit :

#### 1.3.1. Les fondements juridiques

De point purement juridique, selon l'article 2 de l'ordonnance n°95/07 du 25 janvier 1995 relative aux assurances modifiée et complétée par la loi n°06/04 du 20 février 2006, «l'assurance est, au sens de l'article 619 du Code civil, un contrat par lequel l'assureur s'oblige, moyennant des primes ou autres versements pécuniaires, à fournir à l'assuré ou au tiers bénéficiaire au profit duquel l'assurance est souscrite, une somme d'argent, une rente ou une autre prestation pécuniaire, en cas de réalisation du risque prévu au contrat ».

D'après cette définition, on peut constater trois éléments caractérisant le contrat d'assurance : le risque, la prime et la prestation de l'assureur.

- **Le risque** : c'est un événement incertain contre lequel l'assuré désire se couvrir en souscrivant le contrat d'assurance. Le risque se définit comme étant un événement incertain, futur, et qui ne dépend pas de la volonté exclusive de l'assuré. En effet, le risque doit répondre à ces trois conditions pour qu'il soit assurable.
- **La prime** : c'est le prix que l'assuré doit payer à l'assureur en contrepartie de la couverture du risque.
- **La prestation de l'assureur** : c'est la contrepartie de la prime. Elle correspond à l'indemnisation que l'assureur doit payer à l'assuré en cas de survenance du sinistre déterminé dans le contrat.

#### 1.3.2. Les fondements techniques

L'activité de l'assurance se distingue par rapport aux autres activités par la mutualité des risques et le recours aux calculs statistiques permettant l'estimation de la probabilité du risque.

### 1.3.3. La mutualité des risques

L'assureur ne couvre un risque que s'il arrive à assurer une bonne mutualisation de ces risques. Cette dernière est définie par la possibilité de compensation entre les bons et les mauvais risques.

La mutualisation des risques n'est que dans la mesure où l'assureur peut constituer un portefeuille respectant les critères suivants<sup>15</sup> :

- L'homogène des risques
- La sélection des risques
- La dispersion des risques
- La division des risques (réassurance, coassurance)
- Le respect des pleins de conservation
- Le renouvellement des risques

### 1.3.4. Le recours aux calculs statistiques

L'assureur gère sa mutualité des risques et recourt aux calculs statistiques lui permettant l'évaluation de la probabilité de la survenance des risques et la mesure du coût des sinistres. L'évaluation de ces deux éléments servira à la détermination de la prime à payer par les assurés.

## 1.4. Les mécanismes fondamentaux de l'assurance

L'assureur perçoit la prime, gère une mutualité de risques qu'il prend en charge afin d'être en mesure d'honorer ses engagements vis-à-vis ses assurés ou bénéficiaires des contrats.

La fixation d'une prime équitable servira à la couverture des sinistres probables futurs que doit payer chaque assuré est effectuée à l'aide des techniques de calcul actuariel. L'exploitation de ces techniques est assurée par des outils, mécanismes fondamentaux adaptés à l'inversion du cycle de production de l'activité des assurances. IL s'agit de : la loi des

---

<sup>15</sup>COUILBAULT François, Constant ELIASHBERG, LATRASSE Michel, Les grands principes de l'assurance, Edition l'Argus, 3<sup>ème</sup> Edition, p 145.

grands nombres et le théorème central limite, les statistiques, la prévision des probabilités futures de survenance des sinistres.<sup>16</sup>

### 1.4.1. La loi des grands nombres et le théorème central limite

Comme c'est présenté dans la première section, le risque n'est assurable que s'il est aléatoire et sa survenance tient du hasard qui, selon les démonstrations mathématiques, est régi par des lois statistiques.

#### a) La loi des grands nombres

Elle a été énoncée par le mathématicien suisse Bernoulli au 18<sup>ème</sup> siècle<sup>17</sup>, elle stipule qu'au fur et à mesure que le nombre des expériences augmente, les écarts absolus augmentent, mais les écarts relatifs diminuent jusqu'à devenir pratiquement nuls pour un nombre très élevé d'expériences.

Cette loi est fondamentale pour les assureurs. Elle permet la tarification des produits d'assurance.

La loi des grands nombres peut être illustrée à l'aide d'un exemple simple qui utilise les résultats fondamentaux de cette loi :

Soit une population composée de  $N$  individus ( $i = 1, 2, \dots, N$ ) identiques, exposés à un risque d'une perte monétaire notée  $S$ , avec une probabilité notée  $P$ . Soit une entreprise d'assurance assure ces individus en contrepartie d'une prime unitaire notée  $\pi$ . Notons  $X_i$ , variables aléatoires représentant l'indemnité payée par l'entreprise à l'individu  $i$ .

Donc  $X_i$  est défini comme suit :

$X_i = S$  si l'individu  $i$  subit un sinistre (probabilité  $P$ )

$X_i = 0$  si non (probabilité  $1-P$ )

---

<sup>16</sup>YEATMAN Jérôme, manuel international de l'assurance, édition ECONOMICA, 1998, p27

<sup>17</sup> Idem

Si les  $X_i$  sont indépendants :  $\lim_{N \rightarrow +\infty} \left( \frac{X_1 + \dots + X_N}{N} \right) = \text{PS}$  avec une probabilité voisine de 1.

Donc, lorsque les risques sont indépendants, et le nombre  $N$  est assez grand le remboursement moyen tend *presque surement* vers *l'espérance mathématique* du risque. Ce remboursement correspond à la prime actuarielle : ( $\pi = \text{PS}$ ).

### b) Le théorème central limite

Reprenant les données de l'exemple précédent. Si les  $X_i$  sont indépendants :

$\frac{X_1 + \dots + X_N - PNS}{\sqrt{P(1-P)} s \sqrt{N}}$  Suit asymptotiquement une loi normale centrale réduite ( $N(0, 1)$ ). Supposons que la compagnie possède des réserves notées  $R$  et applique le tarif actuariel ( $\pi = \text{PS}$ ) pour chaque individu  $i$ .

Dans ce cas, cette compagnie ne peut faire face à ses engagements que si :

$$X_1 + X_2 + \dots + X_N \leq PNS + R$$

Le théorème central limite permet d'évaluer la probabilité de ruine de la compagnie  $P(X_1 + \dots + X_N - PNS - R) > 0$  comme suit :

$$\begin{aligned} P(X_1 + \dots + X_N - PNS - R) > 0 &= 1 - P(X_1 + \dots + X_N - PNS - R) < 0 \\ &= 1 - F \frac{R}{\sqrt{P(1-P)} s \sqrt{N}} \end{aligned}$$

$F$  : la fonction de répartition de la loi normale centrale réduite.

### 1.4.2. Les statistiques

La nature technique du métier de l'assurance exige le recours aux statistiques pour la tarification, la segmentation de la clientèle, l'étude de la solvabilité et éviter la ruine de la compagnie. L'assureur doit rassembler un nombre suffisamment grand pour pouvoir prévoir et tarifier son risque. Les statistiques réunies par l'assureur doivent être établies par :

- Branches d'assurance (construction, incendie, transport...)
- Type de garantie
- Région d'émission
- Groupes d'assurés
- Caractéristiques physiques des biens assurés

En pratique, même les compagnies d'assurance dans les marchés les plus importants ayant un portefeuille d'assurés suffisants, permettant une tarification fiable sont rares. Le portefeuille d'une seule réassurance ne peut être une base statistique qui répond aux exigences de la loi des grands nombres pour la plupart des risques.

La difficulté d'établir une base de données suffisante a poussé les compagnies d'assurance à se regrouper sous forme d'organisations professionnelles pour l'obtention d'une base de données suffisante. Aussi, la nature du métier de la réassurance permet aux réassureurs la possession, d'une base de données assez importante répondant aux exigences de la loi des grands nombres. De ce fait, l'assureur fait appel aux réassureurs pour la tarification des produits d'assurance et le lancement de nouveaux produits.

#### **a) La prévision des probabilités de survenance des sinistres**

Les statistiques utilisées par l'assureur pour la tarification et la souscription de ses garanties portent sur le passé alors que celle-ci est mise en œuvre dans l'avenir. De ce fait, l'assureur doit ajuster le coût moyen et la fréquence des sinistres estimés sur la base de ces statistiques afin de prendre en compte des modifications futures susceptibles d'affecter cette estimation.

### **1.5. Rôle de l'assurance :**

L'assurance joue à divers égards un rôle considérable dans la société contemporaine.

#### **1.5.1. La fonction réparatrice :**

La première fonction de l'assurance est évidemment d'indemniser les préjudices résultant de la réalisation de risques. L'immeuble incendié pourra être reconstruit, l'objet volé sera

remplacé, l'automobile endommagée sera réparée, la dette de responsabilité sera couverte par l'indemnité d'assurance.

Cette fonction paraît moins évidente dans les assurances de personnes, où le capital presté n'est pas absolument calqué sur le montant du dommage subi. Néanmoins, il est clair qu'à travers l'exigence d'un intérêt d'assurance, la fonction réparatrice reste présente. Le capital stipulé sert à pouvoir aux besoins du bénéficiaire après le décès de l'assuré, ou à ceux de l'assuré lui-même s'il survit.

Parfois, le législateur utilise l'assurance dans le souci de fournir une réparation adéquate au préjudice subi non par l'assuré, mais par une tierce victime de celui-ci. C'est notamment le but essentiel de l'assurance de la responsabilité civile en général. L'exemple de l'assurance automobile obligatoire est illustratif à ce propos : si cette assurance est obligatoire, il s'agit moins de protéger le patrimoine des automobilistes contre les actions en responsabilité dont il pourraient faire l'objet, que de donner aux victimes, par la voie de l'action directe, un recours en indemnisation contre un débiteur solvable qui est l'Assureur.

De toute façon, qu'elle soit prise du point de vue de l'assuré ou de celui d'une tierce victime, l'assurance joue en premier lieu une fonction réparatrice qui fait d'elle un très important facteur de sécurité dans la vie de chacun.

### **1.5.2. Facteur dynamisant le progrès social**

Une telle sécurité ne permet pas seulement de supporter les risques ordinaires de l'existence, elle permet aussi d'en assumer des nouveaux : l'assurance se révèle ici un facteur de progrès technique.

De nombreuses activités risquées ne seraient jamais entreprises sans le soutien de l'assurance : Édification de vastes ensembles commerciaux et industriels, utilisation de nouveaux modes de transport (avions géants, superpétroliers), exploitation de nouvelles sources ou formes d'énergie (Plateformes de forage pétrolier, contrôles d'énergie atomique, etc.), ouverture de nouveaux marchés pour commercialiser les produits locaux, etc.

L'assurance favorise ainsi l'innovation en réduisant les risques courus par les entrepreneurs et les financiers de ces initiatives nouvelles.

### 1.5.3. Rôle de prévention :

De plus, l'assurance joue un rôle remarquable en matière de prévention des sinistres. En effet, et pour de réduire le nombre de ses interventions, l'assureur met contractuellement à charge de l'assuré diverses obligations relatives à la prévention. Ces actions des compagnies d'assurance contribuent pour beaucoup à diminuer le nombre de sinistres, notamment en matière d'incendie, de vol (soustraction frauduleuse), de transport ou même de maladies (campagnes dites de vaccination gratuite).

### 1.5.4. Rôle financier de collecteur d'Épargne

Sur un autre plan, et particulièrement en Assurance sur la vie, l'assurance est un mode privilégié de Formation de l'Épargne. Des primes élevées sont payées en vue de la constitution de capitaux importants. Souvent même, comme en assurance mixte (assurance en cas de vie doublée d'une assurance en cas de décès), la prestation de l'assuré sera certainement due, la seule inconnue étant le moment où le capital sera payé (décès prématuré, survie) : la fonction d'épargne joue dans cette hypothèse, un rôle au moins aussi important que la fonction de couverture de risque.

L'Assurance sert encore à garantir la bonne fin de certaines opérations commerciales et financières. Ainsi, le remboursement des emprunts hypothécaires est en général garanti par la souscription d'une assurance vie sur la tête de l'emprunteur qu'il s'agisse d'une assurance du solde restant dû ou d'une assurance servant à reconstituer le prêt.

Enfin, les sommes considérables que les compagnies d'assurances prélèvent sous la forme de primes doivent évidemment être gérées avec prudence, puisqu'elles vont servir au paiement des indemnités et des capitaux.

Les assureurs constituent les provisions adéquates, et veillent à leur placement : acquisition d'immeubles, de fonds publics, d'obligations et d'actions de sociétés, prêts hypothécaires, etc.

Ces placements des compagnies d'assurances sont d'ailleurs strictement réglementés par la législation de contrôle des activités d'assurances.

Par cette voie, l'assurance joue un rôle important en qualité "d'investisseur - institutionnel", dans le financement des investissements<sup>18</sup>.

## Section 2 : l'Assurance automobile

L'assurance automobile appartient au groupe des opérations d'assurance qui n'ont pas pour objet la vie de l'assuré<sup>19</sup>, elle représente d'un côté, une part très importante du patrimoine des individus, d'un autre côté, elle est exposée aux accidents qui causent la mortalité, la chose qui a poussé les états de rendre ce genre d'assurance obligatoire pour sa partie de Responsabilité civile (RC).

Pour cela le produit d'assurance le plus familier du grand public est l'assurance automobile, c'est grâce à ce contrat que la victime d'un accident automobile est indemnisée par la compagnie assurant le responsable de l'accident.

### 2.1. Définition de l'assurance automobile

L'assurance automobile concerne obligatoirement la responsabilité civile du conducteur et, éventuellement, les dommages au véhicule et l'assurance personnelle du conducteur. Elle est obligatoire pour tout propriétaire d'un véhicule terrestre à moteur pour couvrir les dommages causés au tiers.<sup>20</sup>

Cette branche est constituée d'une garantie obligatoire qui est la responsabilité civile (RC) et un ensemble de garanties facultatives. Avec des millions de voitures assurées, l'assurance automobile représente la deuxième activité (sur le plan mondial) des assureurs après l'assurance vie en termes de chiffre d'affaires.

---

<sup>18</sup> IFID Enseignant : « T. Ben Jemia ». « droit des assurances »

<sup>19</sup>Frédéric Planchet, Pierre Therond et Julien Jacquenin, *Modèles financiers en assurance*, Edition Economica, 2005, p : 08.

<sup>20</sup> L'article un (01) de l'ordonnance N° 74-15 du 30 janvier 1974.

## 2.2. Le contrat d'assurance automobile<sup>21</sup>

Le contrat d'assurance automobile est en général un contrat « multirisque » destiné à couvrir des risques aussi divers, deux (02) types de contrats peuvent être distingués :

### 2.2.1. Le contrat mono véhicule (individuel)

Garantissant le véhicule désigné au contrat. Ces contrats concernent les particuliers, les critères pris en considération par l'assureur sont essentiellement les caractéristiques du véhicule et celles associées au conducteur.

### 2.2.2. Le contrat flotte

La technique habituelle pour assurer des véhicules consiste à souscrire un contrat dit « mono véhicule », couvrant un ou plusieurs véhicules, tous nommément désignés dans le contrat. Toute modification de véhicule doit être signalée au préalable à l'assureur, pour éviter la suspension automatique du contrat. Mais lorsqu'une entreprise ou une collectivité utilise de nombreux véhicules pour les besoins de son activité, ce mode de gestion s'avère lourd et inadapté. Afin de faciliter cette gestion, les assureurs proposent d'établir des « polices flottes » prévoyant la couverture de leur parc de véhicules, quel que soit son effectif.

En principe, il est possible de souscrire un contrat d'assurance flotte dès qu'un parc comporte au moins quatre ou cinq véhicules<sup>22</sup>.

## 2.3. La Responsabilité civile obligatoire<sup>23</sup>

L'assurance obligatoire ne porte que sur la garantie responsabilité civile, qui a été instaurée dans l'intérêt des victimes, c'est-à-dire la protection des victimes en cas d'accident. Son rôle est de permettre aux victimes d'accidents ou à leur ayant droit de percevoir des indemnités. L'assurance de responsabilité obligatoire garantit les dommages causés par la circulation des véhicules terrestres à moteur. Autrement dit, cette garantie permet de compenser

---

<sup>21</sup>Michel Latrasse, et al, Comstant Eliashberg et Francois Cuilbault, *Les grands principes de l'assurance*, Edition l'Argus, 5<sup>ème</sup> édition, Paris 2002, p : 192.

<sup>22</sup> IFID, Cours assurance automobile, Mohamed dkhili.

<sup>23</sup> KPMG Guide des Assurances en Algérie, p : 59.

financièrement les dommages matériels et corporels subis par la victime lorsque le véhicule assuré est à l'origine et que la responsabilité de l'assuré et des personnes ayant la qualité d'assuré est engagée.

### 2.3.1. Définition d'un véhicule terrestre à moteur<sup>24</sup>

La notion de véhicule terrestre à moteur englobe de nombreux véhicules :

- Les voitures de tourisme ;
- Les camionnettes et fourgonnettes ;
- Les camions, les bus, les cars ;
- Les voitures à usage particuliers, taxis, ambulances, etc....
- Les tracteurs et autres engins agricoles, les engins de chantier ;

Cette liste n'est pas exhaustive : en fait, les critères qui permettent de déterminer les engins qui répondent à cette définition sont au nombre de trois :

1. Le véhicule doit être muni d'un moteur, quels que soient le mode de propulsion (essence, gazole, électricité...) et la puissance de ce moteur ;
2. Il doit comporter un siège pour le conducteur ;
3. Il est destiné à se déplacer sur la terre ferme sans être guidé par des rails.

### 2.3.2. Le conducteur du véhicule assuré<sup>25</sup>

Le conducteur du véhicule assuré doit être titulaire d'un permis de conduire en état de validité conformément à la réglementation en vigueur, au type du véhicule utilisé et à la nature du transport pour conduire le véhicule assuré. En outre, il doit respecter les conditions restrictives éventuellement mentionnées sur ce permis. Si ces conditions ne sont pas réunies, les garanties souscrites ne sont pas acquises en cas de sinistre.

---

<sup>24</sup> KPMG Guide des Assurances en Algérie, *Idem*, p : 60.

<sup>25</sup> KPMG Guide des Assurances en Algérie, *Op.cit*, p : 60.

L'assureur est tenu d'indemniser la (ou les) victime(s) au titre de la garantie responsabilité civile, mais il exerce contre le (ou les) responsable(s) du sinistre une action en remboursement de toutes les sommes versées.

### **2.3.3. Les notions d'assuré et de tiers<sup>26</sup>**

L'assurance de responsabilité civile obligatoire permet de prendre en charge les conséquences de la responsabilité des personnes qui ont la qualité d'assuré.

Les contrats d'assurance doivent couvrir la responsabilité civile de toute personne ayant la garde ou la conduite, même non autorisée, du véhicule ainsi que la responsabilité civile des passagers du véhicule objet de l'assurance. Les contrats doivent couvrir, en plus de la responsabilité civile des personnes mentionnées ci-dessus, celle du souscripteur du contrat et du propriétaire du véhicule. L'assurance automobile obligatoire couvre la responsabilité civile de l'assuré en raison de dommages subis par des tiers.

Les membres de la famille du conducteur ou de l'assuré sont considérés comme des tiers.

Les dommages causés aux membres de la famille du conducteur, qu'ils soient passagers ou non du véhicule assuré, sont donc couverts dans tous les cas.

## **2.4. Garantie facultatives**

### **2.4.1. Les garanties principales<sup>27</sup>**

#### **a) La garantie dommages accidents**

Les circonstances dans lesquelles cette garantie est appelée à jouer sont minutieusement exposées : choc avec un corps fixe ou mobile, renversement du véhicule, projections ou retombées de substances, chute du chargement, ouverture du capot.

Pour évoquer les circonstances dans lesquelles cette garantie peut être invoquée, on parlera, en recourant à un vocabulaire moins académique, de collision avec un autre véhicule,

<sup>26</sup> KPMG Guide des Assurances en Algérie, *Idem*, p : 61.

<sup>27</sup> Jacques charbonnier « l'assurance du risque automobile » éditions larcier, 2012, page 94

un mur ou un arbre, de dérapage de la voiture finissant par un tonneau, de gravillons projetés sur le pare-brise, de perte d'un bagage mal fixé sur le toit, etc.

#### **b) La garantie incendie/explosion :**

Sont garantis les dommages survenus soit accidentellement soit du fait d'un attentat, d'une émeute ou d'un mouvement populaire, et provenant d'un incendie, d'une explosion, d'une combustion spontanée, ou de la chute de la foudre.

#### **c) La garantie vol**

L'assureur accorde sa garantie dans les cas suivants :

- Disparition du véhicule par : soustraction frauduleuse, menace ou violence à l'encontre du propriétaire ou du gardien, obtention du véhicule par le voleur en échange d'un chèque volé, effraction d'un garage privatif, frais engagés pour la récupération du véhicule;
- Dommages subis par le véhicule volé puis retrouvé, le vol étant établi par la constatation d'une effraction que matérialisent des détériorations matérielles provoquées par exemple par le forçage de la direction ou la détérioration des circuits électriques permettant la mise en route;
- Disparition d'éléments ou d'accessoires du véhicule, survenue dans certaines circonstances (véhicule fermé, forçage d'une serrure, bris de glace);
- Vol ou perte des clés du véhicule.

Soulignons bien à cet égard que les assureurs entendent ainsi se démarquer de la façon dont les modalités de l'infraction sont décrites par la loi pénale et l'existence du vol à des conditions dont l'interprétation s'impose au juge de façon stricte.

**d) La garantie protection des droits de l'assuré :**

- **La garantie défense et recours**

Cette garantie comprend elle-même deux volets :

**Le volet Défense** permet à l'assuré de se faire défendre devant toute juridiction en raison de poursuites ou réclamations susceptibles de mettre en jeu la garantie responsabilité civile.

**Le volet Recours** oblige l'assureur à exercer une réclamation contre un tiers en vue d'obtenir la réparation du préjudice corporel ou matériel subi par l'assuré à la suite d'un accident garanti. Dans les deux cas sont exclus les frais et honoraires d'un mandataire (avocat, expert) saisi sans l'accord de l'assureur, et dans le premier cas seulement sont en outre exclues les amendes et condamnations pénales, ainsi que l'assistance devant la commission de suspension du permis de conduire.

- **La partie protection juridique**

Ce second volet entraîne la fourniture de tous frais ou conseils par l'assureur voire une tentative amiable de sa part ou la prise en charge d'une instance judiciaire, s'il subsiste un désaccord entre l'assuré et le tiers auteur ou responsable d'un accident garanti.

#### **2.4.2. Les garanties complémentaires<sup>28</sup>**

**a) La garantie bris de glace**

Cette garantie permet le remboursement des frais engagés par l'assuré à la suite d'un bris d'éléments tels que pare-brise, glaces latérales, lunette arrière, optiques de phare.

**b) La garantie tempêtes/ grêle**

Cette garantie permet le paiement des réparations nécessitées par les dommages causés par le vent soufflant en tempête ou par la chute de la grêle.

---

<sup>28</sup>Jacques charbonnier « l'assurance du risque automobile » éditions larcier, 2012, page 98

**c) La garantie catastrophes naturelles**

Au titre de cette garantie, l'assuré est indemnisé des dommages subis du fait de l'action d'un élément naturel, tels qu'ouragan, crue, séisme, dont l'intensité anormale a été admise par les pouvoirs publics.

**d) La garantie événements climatiques**

Il s'agit de dommages causés par certains événements d'origine naturelle. Les événements en question sont : un glissement de terrain, une inondation due au débordement d'un cours d'eau ou à un refoulement d'égout, le poids de la neige, la réalité de ces phénomènes étant établis au moyen d'une attestation de la mairie ou d'une coupure de presse.

**2.4.3. Les garanties accessoires<sup>29</sup>**

Enfin, évoquons brièvement les garanties accessoires dont bénéficie l'assuré :

**a) La garantie des frais de remorquage**

Par frais de remorquage, l'assureur entend les frais entraînés par la nécessité de faire remorquer jusqu'au garage le plus proche le véhicule endommagé par un accident, ou à la suite d'un vol ou d'un incendie. Dans les mêmes circonstances, l'assureur rembourse les frais de mise en fourrière (enlèvement et garde).

**b) La garantie insolvabilité des tiers :**

Cette garantie joue en cas de dommages causés au véhicule par un tiers responsable à condition qu'il soit formellement identifié, non assuré et insolvable.

**2.4.4. Les garanties optionnelles**

Ces garanties sont dites optionnelles, car elles doivent être demandées par l'assuré et font l'objet d'une prime séparée.

---

<sup>29</sup>Jacques charbonnier « l'assurance du risque automobile » éditions larcier, 2012, page 99

**a) La valeur majorée du véhicule**

Cette garantie permet à l'assuré victime de la perte totale de son véhicule à la suite d'un événement garanti, d'être indemnisé à concurrence de la valeur d'acquisition et cela pendant les 12 mois suivants la première mise en circulation.

**b) La garantie des accessoires et du contenu privé du véhicule**

Par cette garantie sont couverts les dommages subis par les accessoires du véhicule et par son contenu privé, c'est-à-dire les effets personnels, les objets et les bagages à usage strictement privé, tous objets de la vie courante, ce qui exclut les animaux, les objets précieux, etc.

**c) La garantie véhicule de remplacement**

L'assureur assume la prise en charge des frais de location d'un véhicule de remplacement sur la base de la durée d'immobilisation du véhicule assuré endommagé par un accident, un incendie, un vol, etc.

**d) La garantie assistance**

L'assuré bénéficie de l'aide d'un professionnel dépêché sur place en cas de panne mécanique, de crevaison d'un pneumatique, voire de panne de carburant.



Figure 02: Les différentes formes de la garantie facultative

## Conclusion

L'activité assurantielle a pour objectif de préserver la richesse créée et de soutenir son développement par la couverture des risques susceptibles de l'atteindre. Bien que l'assurance vise à couvrir tous les risques afférents à l'activité de l'entreprise, les compagnies ne peuvent assurer tous les risques du fait de l'inassurabilité de certains risques. En effet, certains risques sont considérés comme des risques non assurables puisqu'ils ne répondent pas aux critères d'assurabilité techniques ou juridiques.

L'activité des assurances est une activité particulière. Elle se distingue par rapport aux autres activités économiques par ces fondements et ces mécanismes fondamentaux.

# **CHAPITRE II**

**LA MESURE DU RISQUE DANS  
L'ASSURANCE AUTOMOBILE**

Le métier de l'assureur est d'assurer les risques, en réduisant au minimum les marges de manœuvre du hasard.

Le présent chapitre est dédié aux notions de mesure de risque automobile, principe de tarification de ce risque, les différentes primes ou cotisations de couverture contre ce risque et les outils mathématiques servant à la modélisation de la prime pure par l'approche fréquence-coût. De ce fait, ce chapitre sera présenté comme suit :

### **Chapitre 2 : La mesure du risque dans l'assurance automobile**

- **Section 1** : principe de tarification et notions de primes
- **Section 2** : outils mathématiques

---

## **Chapitre 2 : La mesure du risque dans l'assurance automobile**

---

### **Section 1 : principe de tarification et notions de primes**

La mesure du risque de la branche automobile est une opération importante, dans la mesure où elle aboutit à la détermination de la prime à payer par l'assuré qui est la contrepartie du service d'assurance, et c'est cette contrepartie qui permet d'indemniser un assuré en cas de sinistre.

#### **1.1. Variables de mesures de la sinistralité :**

- **Le nombre de sinistres**

C'est le nombre de sinistres déclarés par l'assuré. Cette information est connue avec précision et est donc extrêmement fiable pour des analyses statistiques. Ainsi, cette variable sera le support de notre étude.

- **Le coût de sinistres**

C'est le coût total mis à la charge de la compagnie pour un sinistre donné. Lors de l'étude des coûts, il convient de tenir compte du nombre de sinistres engendrés. On analysera le plus souvent le coût moyen des sinistres. Cette valeur peut être connue très longtemps après la survenance d'un sinistre.

- **La durée d'exposition au risque**

C'est le nombre de jours où la police a été exposée à un risque sous-jacent particulier (nombre de jours entre le début et la fin d'une période donnée). Cette variable permet de quantifier l'exposition du portefeuille à un risque donné sur une période donnée.

- **La Fréquence**

Le nombre de sinistres divisé par l'exposition. Cette variable permet de bien mesurer l'occurrence du sinistre, car un individu qui aura eu un sinistre en quatre mois d'exposition aura le même risque qu'un individu qui a eu trois sinistres en un an d'exposition.

### **1.2.Principe de tarification et nécessité de segmenter :**

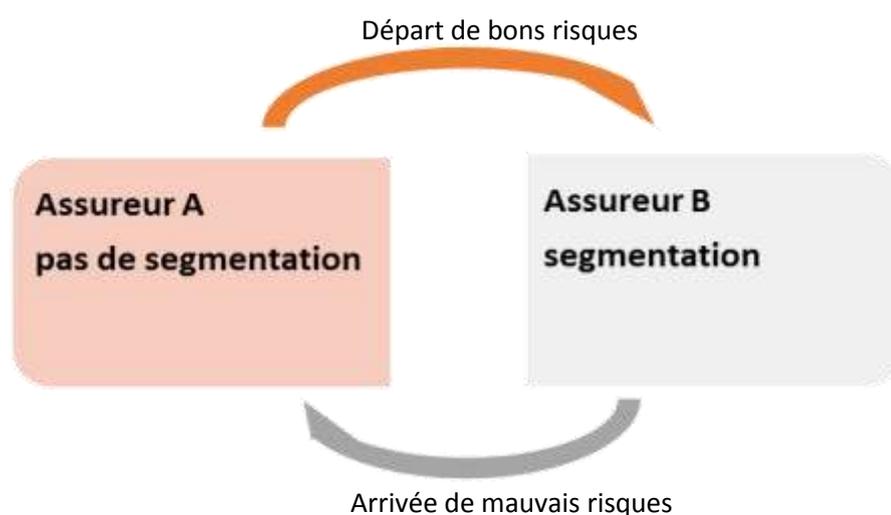
Dans un contexte de marché concurrentiel, chaque assureur adopte une tarification qui lui est propre et qui se base sur des considérations techniques et commerciales lui permet de se différencier et d'adapter le prix de risque au profil de son portefeuille d'assurés. Pour ce faire, la segmentation des risques est nécessaire permettant de différencier les risques et les assurés qu'il porte à sa charge.

Ainsi donc, la segmentation nous permet d'obtenir différentes catégories de risques selon les caractéristiques de l'assuré et les garanties souscrites par ce dernier. Un tarif propre à chaque catégorie sera attribué en adéquation avec le risque associé.

La segmentation consiste à adapter les primes aux sinistres selon des classes de risques homogènes, de façon à en tirer des conséquences du point de vue technique. La segmentation

permettra à l'assureur de prendre le recul nécessaire et des mesures adéquates à chacun des niveaux de segmentation, tant en tarification, qu'en souscription.

Un exemple illustratif nous permet de mettre en évidence la nécessité de segmenter, tout<sup>30</sup> particulièrement en assurance automobile. Prenons un assureur A qui décide de n'effectuer aucune segmentation. Cela veut dire 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 segmenter son portefeuille d'assurés, autrement dit, de faire payer plus cher les conducteurs à risque, déterminés à partir de variables telles que l'âge ou les antécédents de sinistres. En contrepartie, l'assureur B réduit, à un niveau correspondant au risque assuré, la prime demandée des assurés à faible sinistralité. Suite à cette segmentation, les assurés prudents, ayant une faible probabilité d'avoir un sinistre et que l'on qualifiera de bons risques, décideront de s'assurer auprès de l'assureur B, proposant des primes plus faibles. L'assureur A n'ayant pas effectué une segmentation en proposant un tarif unique, constituera un portefeuille, à forte sinistralité, comportant la plupart des personnes ayant un comportement à risque. En effet, le tarif proposé par l'assureur A sera plus attractif que celui de l'assureur B car les bons assurés vont payer pour les mauvais. En conclusion, l'assureur A se trouve face à une sinistralité très élevée qu'il ne pourra pas supporter étant donné l'inadéquation des primes demandées aux assurés avec le profil de risque, ce qui va déséquilibrer son bilan, alors que l'assureur B dégage du profit grâce à la segmentation qu'elle a effectuée .



<sup>30</sup>Jean-pierredaniel préface de jacques vandier « l'assurance automobile aujourd'hui », 2003.

### 1.3. La tarification de l'assurance automobile

Le tarif d'assurance automobile comprend :<sup>31</sup>

**Une information a priori** sur l'assuré et l'usage qu'il fera du véhicule, son lieu de garage habituel et, d'autre part, sur la voiture, son modèle, son état d'entretien ...etc. ;

**Une information a posteriori** sur les sinistres déclarés et sur la manière dont le conducteur se comporte (entretien du véhicule, infractions au Code de la route), relative à la période écoulée depuis la souscription du contrat.

#### 1.3.1. La tarification a priori<sup>32</sup>

La tarification a priori est le calcul des primes pour des risques nouvellement assurés, lorsque les antécédents avec le client en matière de sinistre n'existent pas.

- **Modélisation du tarif a priori**

Ce système de classification des risques est basé sur des variables facilement observables, il consiste à utiliser une méthode de régression linéaire pour identifier les variables qui affectent de façon significative le nombre de sinistres dans le cas de la tarification individuelle.

Le modèle est de la forme suivante :

$$Y = a_0 + a_1X_{1i} + a_2X_{2i} + a_3X_{3i} + a_4X_{4i} + a_5X_{5i} + a_6X_{6i} + \varepsilon_i$$

$Y$  : est la variable à expliquer, c'est le nombre de sinistre durant la période d'échantillonnage.

$a_k$  : Les paramètres du modèle à estimer (coefficient des variables explicatives)

$k = 1 \dots 8$

$i = 1 \dots n$  (le nombre d'observations)

<sup>31</sup>P.J. Delaporte, *Les Mathématiques de l'assurance Automobile*, THE ASTIN BULLETIN, VOL. VI, PART (03), Editor H. G. Verbeek, Amsterdam, MAY 1972, p: 187.

<sup>32</sup>Jean Luc Besson et Christian Partrat, *Assurance non vie, modélisation et simulation*, Edition Economica, Paris 2005, p : 31.

Les variables explicatives sont :

- $X_{1i}$  : La zone de circulation
- $X_{2i}$  : Le sexe du conducteur ;
- $X_{3i}$  : La catégorie socioprofessionnelle ;
- $X_{4i}$  : L'âge de l'individu ;
- $X_{5i}$  : L'âge du véhicule ;
- $X_{6i}$  : Puissance du véhicule ;
- $X_{7i}$  : L'usage du véhicule ;
- $X_{8i}$  : Le genre du véhicule ;
- $\epsilon_i$  : L'erreur de modèle.

### 1.3.2. La tarification a posteriori

Il s'agit de la correction de la tarification a priori, en prenant en compte les observations des sinistres des assurés dans le cadre des risques individuels

#### ➤ Les critères de tarification de la Responsabilité civile (RC)

En Algérie, la prime en RC est règlementée (administrée) par les autorités publiques qui sont la tutelle, représentées par le Ministère des Finances, les compagnies d'assurance ne peuvent déterminer librement leurs tarifs de RC, qui est basée sur les critères suivants :

- **Le Genre du véhicule**

Il permet d'identifier le type du véhicule, l'ensemble des genres est présenté en détail dans l'annexe.

- **La Zone de circulation**

La tarification diffère selon les zones dans lesquelles le véhicule est amené à circuler, on comprend aisément que le fait de circuler majoritairement à la campagne ou majoritairement en ville n'expose pas l'automobiliste aux mêmes risques. La fréquence est forte dans une zone à densité urbaine très élevée.

L'Algérie est subdivisée en deux (02) zones de circulation, la zone Nord et la zone sud, ainsi la zone sud subie de taux de prime moins élevés que celle du Nord. Les deux (02) zones avec les Wilayas associées sont présentées en détail en annexe.

- **L'Usage du véhicule**

Plus le véhicule circule plus il est exposé aux accidents, donc le risque d'être impliqué dans un accident augmente. Les fréquences des accidents sont proportionnelles au kilométrage parcouru. Les codes usages sont présentés dans l'annexe

- **La Puissance fiscale du véhicule**

La puissance du véhicule entre également dans la tarification de l'assurance. Les différents codes des puissances sont présentés en détail dans l'annexe.

En dehors de la garantie Responsabilité civile, puisque elle est obligatoire, les autres garanties restent déréglementées et chaque compagnie a la latitude de fixer ses tarifs en fonction de sa propre politique

- **La tarification des garanties facultatives**

La tarification des garanties facultatives est fixée par la compagnie d'assurance, et conformément au code des assurances, les assureurs sont tenus de les communiquer au Ministère des Finances.

En Europe, les critères de tarification sont très nombreux, leur nombre et leur degré de sophistication ont tendance à augmenter. Aujourd'hui, les sociétés qui activent sur le marché automobile ont établi des modèles intégrant un nombre très élevé de critères pour traduire la probabilité de survenance d'un sinistre pour un couple conducteur/véhicule donné.<sup>33</sup>

---

<sup>33</sup> Jean Pierre Daniel, préface de Jacques Vandier « l'assurance automobile aujourd'hui », 2003, p : 84.

## 1.4. La détermination des différentes primes ou cotisation

### 1.4.1. La cotisation pure<sup>34</sup> (la prime pure)

Elle est représentative de l'opération d'assurance, abstraction faite de tout autre élément. C'est le montant que doit verser chaque assuré pour permettre à l'assureur de régler l'ensemble des sinistres frappant la mutualité qu'il a constitué.

En assurances de biens ou de responsabilité, la cotisation pure est égale à la fréquence du risque multipliée par le coût d'un sinistre.

**Cotisation pure = fréquence moyenne x coût moyen**

Soit un portefeuille comprenant un nombre  $N$  de risque, et soit  $n$  le nombre de sinistre pour un exercice donné.

La fréquence des sinistres est égale à :  $f = \frac{n}{N}$

Soit  $C$ , le coût moyen d'un sinistre, le coût total des sinistres à la charge de l'assurance est  $S$ , telle que :  $S = C \times n$

Cette charge  $S$  doit être supportée par tout le portefeuille, c'est-à-dire :

$$\frac{S}{N} = \frac{C \times n}{N} = \frac{n}{N} \times C = f \times C = P$$

$P$  ; est appelée prime pure, c'est le prix de revient pris en charge par l'assureur, et pour l'évaluer, il faut estimer préalablement  $f$  et  $C$ .

Toutefois, les connaissances statistiques laissent toujours subsister une incertitude dénommée « écart-type », qui doit être comptabilisée par un chargement de sécurité

<sup>34</sup> Francis Noël, Francis Noël, *Économie et Organisation de l'Assurance*, BTS Assurances, p : 114.

(exemple : +5%). Plus le nombre d'assurés est important, moins le chargement de sécurité est important, ce qui implique que l'entreprise améliore sa compétitivité.

**Cotisation pure = (fréquence + chargement de sécurité) x coût moyen**

Les contrats d'assurance automobile sont des contrats multirisques (la responsabilité civile, les dommages accidentels, le bris de glace, le vol..... etc.), et chaque risque voit sa cotisation pure définie par un calcul de cette nature. Le regroupement de garanties pour un même bien économise sur les coûts de gestion.

#### 1.4.2. Cotisation nette<sup>35</sup>

**Cotisation nette = cotisation pure + chargements (frais d'acquisition et les frais de fonctionnement de la société d'assurance).**

Pour acquérir, gérer des contrats d'assurance, l'entreprise supporte un certain nombre de frais qui vont s'ajouter à la cotisation pure.

#### 1.4.3. Les frais d'acquisition

##### a) Les frais de publicité

Les sociétés d'assurances anonymes ou mutuelles ont recours à la publicité pour inciter les prospects à s'assurer auprès d'elles. Les formes sont multiples : journaux, télévision, affichages, etc.

##### b) Les frais de distribution

Les entreprises employant des intermédiaires versent des commissions. Souvent pour encourager ceux-ci à développer une clientèle nouvelle, des bonifications sont accordées et il n'est pas rare de rencontrer des gratifications de l'ordre de 20% à 30%.

---

<sup>35</sup> Francis Noël, *idem*, p : 116.

### c) Les frais de gestion

Outre les commissions annuelles versées par les entreprises qui font appel à des intermédiaires, toutes doivent assumer des frais liés au personnel, aux investissements immobiliers, des frais d'impression des contrats, des frais informatiques, des frais téléphoniques, des coûts de formation et des taxes diverses (taxe professionnelle, taxe d'apprentissage).

### d) Les charges financières

- L'alimentation de la marge de solvabilité correspondant à l'accroissement du chiffre d'affaires.
- Dividende destiné aux actionnaires pour les sociétés d'assurance anonymes.

### 1.4.3. Cotisation totale ou commerciale<sup>36</sup>

C'est le montant à payer par le client. Elle est constituée de la cotisation nette augmentée des frais et accessoires, l'ensemble étant majoré des taxes.

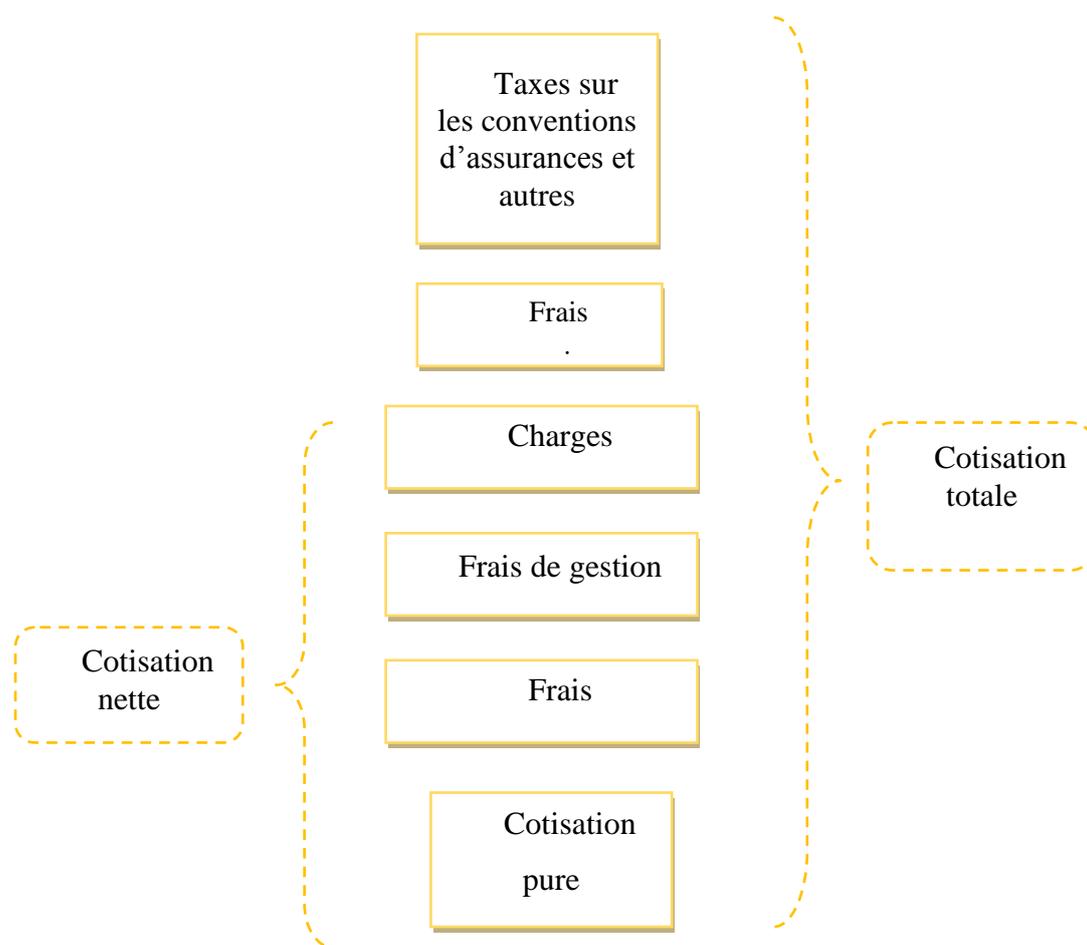
**Cotisation totale = cotisation nette + frais et accessoire + taxes**

- **Les frais et accessoires** correspondent au coût de l'envoi de l'appel de cotisation et du coût de l'encaissement. L'informatisation tend à les réduire, voire à les faire disparaître.
- **Les taxes** : sont des impôts indirects calculés sur la cotisation nette et les frais accessoires, elles sont réservées à l'État. Cette taxe n'est pas récupérable par les entreprises individuelles ou collectives assujetties à la TVA.

---

<sup>36</sup> Francis Noël, *Op.cit.*, p : 117.

Le schéma suivant résume les différentes cotisations :



Source : Francis Noël, *Économie et Organisation de l'Assurance*, BTS Assurances, 2007, p117.

**Figure 03** : Les différentes cotisations

### 1.5. Revue de la littérature

Nous présenterons une vue d'ensemble des résultats de la littérature qui porte sur ce volet, les études choisies ayant traité l'assurance automobile contiennent les travaux suivants :

Aldjia MEZAIE [2007, Algérie], Olfa N. GHALI [2001, Tunisie], Meriem MAATIG [2010, France], OkbaRimi, AbdelouhabLatreche et RiadhRimi [2015, Algérie], ainsi que d'autres travaux statistiques élaborés par le Conseil National des Assurances CNA [2001].

- La recherche menée par Aldjia MEZAIE<sup>37</sup> dont l'objet porte sur la tarification a posteriori en assurance automobile s'est effectuée pendant la période 2000-2003, et les modèles d'estimations ont été appliqués sur un échantillon de 1500 assurés de la CAAT<sup>38</sup>.
- Cette recherche propose de modéliser l'événement de la survenance des accidents par les lois de probabilités (Poisson-Mélange, Binomiale Négative) et le système Bonus-Malus en introduisant un critère qui est le nombre des accidents provoqués durant la période du contrat d'assurance dont le conducteur est responsable.

Les résultats obtenus sont trois (03) systèmes Bonus-Malus optimaux basés exclusivement sur le nombre d'accidents passés avec les caractéristiques suivantes :

- Le système Bonus-Malus optimal du modèle Binomial Négatif engendre des Malus élevés inapplicables en réalité ;
  - Le système Bonus-Malus optimal par le principe de l'utilité nulle diminue le Malus et augmente le Bonus, l'inconvénient de cette méthode est que ce système n'est financièrement pas équilibré ;
  - Le système Bonus-Malus optimal par le principe de maximisation de l'utilité espérée ajuste la prime selon le nombre d'accidents et l'effectif des groupes, donc les assurés ont bénéficié de la collectivité tout en maximisant l'utilité espérée de l'assureur.
- Les études menées et élaborées par le Conseil National des Assurances CNA dans l'optique de proposer un choix de facteurs de risques à utiliser pour la tarification et son introduction dans le calcul de la prime. Ex. : l'étude statistique relative au tarif de la Responsabilité civile automobile,<sup>39</sup> qui a porté sur le traitement de données statistiques relatives à la sinistralité matérielle et corporelle observée sur la période 1993-2000. L'objectif était de vérifier la pertinence du modèle de tarification qui existait et de réviser ou actualiser la grille tarifaire en matière d'assurance RC automobile.

---

<sup>37</sup>AldjiaMezaie, *Tarification a posteriori en assurance automobile par le principe de maximisation de l'utilité espérée*, mémoire de magister, option Finance et Actuariat, INPS 2006/2007.

<sup>38</sup> CAAT : Compagnie Algérienne d'Assurance Total.

<sup>39</sup>Conseil National des Assurances EDITIONS & DOCUMENTATION, Responsabilité Civile automobile, Tarif, Indice & Barème d'indemnisation, 2001.

Sur la base des observations collectées auprès d'un échantillon d'agences représentatives de la distribution géographique de la sinistralité dans la région centre.

Dans la seconde phase, la collecte a été étendue à l'ensemble du territoire. Après organisation des données, la taille de l'échantillon global, incluant les régions Centre, Est, Ouest et Sud, atteint 719.711 contrats contenant 37.836 sinistres, dont 31.940 sinistres matériels et 5.896 sinistres corporels.

Les travaux de la deuxième phase ont porté sur les tâches suivantes :

- Étude statistique exploratoire de la distribution des coûts de la sinistralité par région ;
- Validation des résultats obtenus lors de la première phase sur ce nouvel échantillon ainsi étendu : la capacité explicative de ces facteurs a été confirmée ;
- Étude statistique comparative du zonage. L'étude de facteur zone a été menée en utilisant des tests graphiques et non paramétriques. Ces tests rapportés aux fréquences et aux coûts moyens de la sinistralité estimée ont permis de déterminer les niveaux
- Statistiques de la similarité entre les différentes régions (Est ; Ouest ; Centre ; Sud) ;
- Mise en œuvre d'une approche algorithmique pour l'ajustement des classes de risques et la simulation de la sinistralité.

Il a été confirmé lors de cette étude que les facteurs de risques de tarif, pris conjointement, expliquent significativement la sinistralité, avec pour facteur dominant la puissance fiscale du véhicule. L'étude confiée a abouti à valider le modèle de tarification en confirmant les critères de tarification utilisés.

Mais ce que ces études n'expliquent pas ou ne montrent pas, c'est la liaison qui existe entre d'autres critères propres aux conducteurs ou aux véhicules qui ne sont pas pris en considération lors de la détermination des primes d'assurance, est la probabilité de provoquer un accident.

- Dans le même cadre de la présente étude, Olfa N. GHALI a présenté un article pour analyser le système de tarification de l'assurance automobile tunisien à l'aide d'un modèle de tarification optimal.<sup>40</sup>

Le système de tarification automobile utilisé en Tunisie, instauré en 1993, se base essentiellement sur la puissance et l'usage de l'automobile, et d'un système Bonus-Malus pour la responsabilité civile. Il est réservé à l'usage privé et appliqué selon la même règle à tous les assurés par toutes les compagnies d'assurances.

Le principal problème de l'assurance automobile en Tunisie est le faible niveau des primes, déterminées par le Ministère des Finances, pour les différentes catégories de véhicules en fonction de la hausse croissante des coûts.

Cette étude propose de formaliser un modèle optimal de tarification basé sur les caractéristiques des assurés (un modèle a priori) et sur le nombre d'accidents passés des individus (un modèle a posteriori), de sorte que chaque assuré paye une prime proportionnelle à sa fréquence d'accident et que l'assureur soit financièrement équilibré.

Pour ce faire, l'auteur a utilisé un modèle de tarification basé sur les caractéristiques des individus et leurs nombres d'accidents passés, la base de données s'étale sur cinq ans (1990-1995) et contient 46 337 observations.

Cette base a permis d'estimer, à partir de données annuelles et à l'aide des modèles de comptage (poisson et binomiale négative), l'importance relative des facteurs qui expliquent le nombre d'accidents durant une période et de construire des tables bonus-malus optimales.

L'étude a pu démontrer que le système de tarification automobile tunisien pour l'usage privé n'est pas efficace, et elle est arrivée à la conclusion que d'autres variables que la puissance et l'usage des véhicules (qui sont utilisées comme seuls critères de tarification dans le régime actuel) sont significatives pour expliquer le nombre d'accidents (région de résidence de l'assuré, les garanties auxquelles il souscrit, la marque et l'âge de l'automobile) et que, par ailleurs, la table Bonus-Malus adoptée par le ministère des Finances n'est pas optimale.

---

<sup>40</sup>Olfa N. Ghali, *Un modèle de tarification optimal pour l'assurance automobile dans le cadre d'un marché réglementé : application à la Tunisie*, Cahier de recherche 01-09, Décembre 2001

- Le sujet de cette étude se rapproche de la recherche réalisée par Meriem MAATIG<sup>41</sup>, intitulée « *Modélisation multi variée des comportements à risque des conducteurs d'automobile* », l'objet de cette étude était d'avoir une spécification économétrique qui modélise simultanément les comportements à risque des conducteurs, d'où la provocation des accidents.

L'étude s'est appuyée sur des données collectées à partir d'une enquête, réalisée par questionnaire sur un échantillon représentatif de la région parisienne en termes de sexe du conducteur et son âge.

Un modèle probit a été effectué pour tester l'existence de lien de causalité entre les comportements suivants :

L'utilisation des places de parking réservées aux handicapés, la conduite sous l'emprise de l'alcool et l'utilisation du téléphone portable au volant, car la présence d'informations inobservées par l'assureur dans une classe peut venir invalider l'hypothèse d'homogénéité.

Cette étude a révélé d'importantes informations sur le lien existant entre ces trois comportements du conducteur, l'auteur a constaté que les coefficients de corrélation des erreurs du probit sont significativement différents de zéro, donc il y a interdépendance entre ces comportements. Ces coefficients étant négatifs, les caractéristiques individuelles qui expliquent la conduite sous l'emprise de l'alcool et l'utilisation des places de parking réservées aux handicapés expliquent aussi négativement la probabilité de l'utilisation du téléphone portable au volant.

**Okba Rimi, Abdelouhab Latreche et Riadh Rimiont** présentaient, 30 décembre 2015, un article intitulé « *Une évaluation empirique de la tarification de l'assurance automobile en Algérie* » portant le principal problème de l'assurance automobile en Algérie est le faible niveau des primes, déterminées par le ministère des Finances. Ils ont proposé, dans cette étude, de formaliser un modèle de tarification basé sur les caractéristiques des assurés (tarification a priori) et sur le nombre d'accidents passés des individus (tarification a

---

<sup>41</sup>MérimMaatig, *Modélisation multivariée des comportements à risque des conducteurs d'automobile*, publié dans "42<sup>èmes</sup> Journées de Statistique, 24 juin 2010

posteriori) à l'aide d'un système Bonus-Malus. La méthodologie utilisée pour calibrer un système Bonus-Malus à la française, est celle proposée par (PITREBOIS, et al, 2003c).

La comparaison entre le système de tarification appliqué par le ministère des Finances et le système théorique obtenu dans cette étude a permis de conclure que le système de tarification appliqué par le ministère des Finances n'est pas optimal.

## Section 2 : Outils statistiques

### 2.1. Le modèle de régression linéaire

Soit n individus observés pour lesquels nous disposons :

- d'une **variable réponse**  $(Y_t)_{t=1, \dots, n}$ , aussi appelée **variable à expliquer**
- de p variables explicatives  $(x_{1t}, \dots, x_{pt})_{t=1, \dots, n}$

L'objectif est d'établir une relation entre la variable à expliquer  $Y_i$  et les p variables explicatives  $(x_{1t}, \dots, x_{pt})$ . Supposons l'existence d'une relation linéaire entre la variable à expliquer et les variables explicatives du type

$$y_t = a_0 + a_1 x_{1t} + \dots + a_k x_{kt} + \varepsilon_t, \text{ Pour } t=1, \dots, n$$

Le modèle, sous forme matricielle :

$$Y_{(n,1)} = X_{(n,K+1)} a_{(K+1,1)} + \varepsilon_{(n,1)}$$

Avec :

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \\ \vdots \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1t} & x_{2t} & \dots & x_{kt} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}; \quad a = \begin{pmatrix} a_0 \\ \vdots \\ a_t \\ \vdots \\ a_k \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

### 2.1.1. Estimation des coefficients de régression :

Pour obtenir le meilleur ajustement, la première idée consiste à minimiser les écarts entre la droite de régression et la variable endogène observée. Cependant, ces écarts pouvant être de signes opposés et donc se compenser. Pour ce faire, la méthode des moindres carrés ordinaire cherche à minimiser la somme des carrés des écarts, appelée somme des carrés des résidus et notée SCR.

La méthode des moindres carrés ordinaires (**MCO**) consiste à minimiser la somme des carrés des erreurs on trouve :

$$\hat{a} = (X'X)^{-1} X'Y$$

Cette solution est réalisable si la matrice  $X'X$  de dimension  $(k+1, k+1)$  est inversible

Le modèle estimé s'écrit :

Avec :

$$y_t = \hat{a}_0 + \hat{a}_1 x_{1t} + \dots + \hat{a}_k x_{kt} + e_t$$

$e_t = y_t - \hat{y}_t$ ,  $e_t$ , est le résidu, c'est-à-dire l'écart entre la valeur observée de la variable à expliquer et sa valeur estimée (ajustée).

**Le modèle linéaire est facile à mettre en œuvre, mais n'est pas entièrement adapté à la tarification en assurance Non-Vie : il suppose que la moyenne est une fonction linéaire des variables. Pour pallier à ces restrictions, le modèle linéaire généralisé (GLM) est une généralisation de la régression linéaire. Il permet d'étudier la liaison entre une variable réponse  $Y$  et un ensemble de variables explicatives.**

### 2.2. Principe du modèle linéaire généralisé :

Le modèle linéaire généralisé est une extension du modèle linéaire. Il a été introduit par John Nelder et Robert Wedderburn en 1972 et permet d'étudier le lien entre une variable à expliquer et un ensemble de variables explicatives qualitatives ou quantitatives. Dans notre

cas de figure, nous cherchons à expliquer les montants de sinistre par un ensemble de variables explicatives.

### Quand utilise-t-on un GLM plutôt qu'un modèle linéaire ?

Dans le cas où la variable réponse n'est pas continue, le modèle linéaire n'est pas approprié. Une autre raison expliquant le fait qu'un modèle linéaire n'est pas approprié est tout simplement le cas où la relation entre la variable endogène et la (les) variable(s) exogène(s) n'est pas linéaire. Une fonction de lien adéquatement choisie permet de mieux modéliser l'effet des variables exogènes sur la variable endogène dans un GLM.

#### 2.2.1. Les composantes d'un GLM

##### 1) La variable réponse, composante aléatoire

La variable réponse est la variable aléatoire que l'on cherche à expliquer et dont on souhaite déterminer l'espérance. Dans notre cas, nous considérons le montant de sinistres comme une variable aléatoire dont on cherche l'espérance pour proposer un prix.

On dispose de  $n$  observations  $Y_1, \dots, Y_n$  considérées comme des réalisations de la variable  $Y$  (binaire, discrète ou continue). La loi de  $Y$  doit appartenir à la famille exponentielle, c'est-à-dire sa densité s'écrit sous la forme :

$$\exp\left[\left(\frac{y\theta - b(\theta)}{a(\phi)}\right) + c(y, \phi)\right] \quad \text{Où } \phi \text{ et } \theta \text{ sont des paramètres, } a, b \text{ et } c \text{ des}$$

fonctions souvent  $a(\phi) = \phi$ ,  $\phi$  est un paramètre d'échelle.  $\theta$  est appelé le paramètre naturel car il est lié aux deux premiers moments de la loi.

##### 2) Les variables explicatives, composante déterministe

Comme pour le modèle linéaire classique, on considère une combinaison linéaire des variables explicatives :

$$\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

### 3) la fonction de lien :

Elle contient le lien entre l'espérance de la variable réponse et la composante déterministe.

Au lieu de modéliser la moyenne directement, nous allons introduire une fonction de lien monotone et dérivable. Cette fonction matérialise la relation entre l'espérance de notre variable réponse et la composante déterministe.

$$g(E(y)) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Voici les fonctions de lien des lois exponentielles les plus courantes :

Nom du lien	Fonction de lien
Lien identité	$g(\mu) = \mu$
Lien log	$g(\mu) = \ln(\mu)$
Lien cloglog	$g(\mu) = \ln(-\ln(1-\mu))$
Lien logit	$g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$
Lien probit	$g(\mu) = \Phi(\mu)$ $\Phi$ fonction inverse de la fonction de répartition d'une loi $\mathcal{N}(0,1)$
Lien réciproque	$g(\mu) = -1/\mu$
Lien puissance	$g(\mu) = \mu^\gamma$ avec $\gamma \neq 0$ $g(\mu) = \ln(\mu)$ avec $\gamma = 0$
Aranda Ordaz (asymétrique)	$g(\mu) = \ln\left(\frac{(1-\mu)^{-\lambda} - 1}{\lambda}\right)$

Tableau 01 : Récapitulatif des fonctions de lien des modèles usuels

#### 2.2.2. Le choix du modèle :

Le plus souvent, le choix de la loi de probabilité de la fonction de réponse découle naturellement de la nature du problème étudié. On peut alors choisir comme fonction de lien la fonction de lien canonique associée à la loi de probabilité de la fonction de réponse étudiée.

Il est toujours possible d'utiliser d'autre fonction de lien par exemple : l'identité, le Logit, Probit, Puissance et Logarithme. **Ce choix dépend également de la fonction de réponse (Y).**

### 2.2.3. Estimations des paramètres par maximum de vraisemblance.

L'estimation des paramètres du *GLM* se fait par maximum de vraisemblance dont nous rappelons ici la méthode.

Rappelons tout d'abord que la vraisemblance est, par définition, un produit de fonctions de densité. Pour en déterminer le maximum, il suffit de déterminer la valeur du paramètre de la fonction qui l'annule tout en gardant la dérivée seconde négative. Pour des raisons pratiques, on préfère dériver le logarithme de la vraisemblance. On dérive alors une somme plutôt qu'un produit. De plus, comme la fonction logarithme est strictement croissante, maximiser le logarithme de la fonction équivaut à maximiser la fonction.

Concrètement, prenons  $n$  variables indépendantes  $Y_i$  de loi appartenant à la famille exponentielle, de paramètre  $\theta_i$  et de densité  $f$ . Soient  $(y_1, y_2, \dots, y_n)$   $n$  observations de cette variable.

La vraisemblance s'écrit :

$$l(y; \theta; \emptyset) = \prod_{i=1}^n f(y; \theta; \emptyset)$$

Et la log-vraisemblance s'écrit :

$$L(y; \theta; \emptyset) = \ln l(y; \theta; \emptyset) = \sum_{i=1}^n \ln f(y; \theta; \emptyset)$$

### 2.2.3. Adéquation et validation du modèle :

La validation du modèle doit être guidée par la théorie économique. Cela revient à vérifier le signe des coefficients estimés ainsi que leurs ordres de grandeur en examinant les écarts entre les valeurs théoriques (obtenues avec le modèle) et les valeurs observées. Pius, il convient de contrôler la légitimité de ce modèle par la déviance du modèle.

### La déviance

Une mesure<sup>42</sup> d'adéquation appelée la déviance  $D = 2(L_{max} - l_0)$ , qui représente le double du logarithme d'un rapport de vraisemblance. Il est toujours possible de construire un modèle comprend autant de paramètres que d'observations. Ce modèle comprend autant de paramètres que d'observations distinctes (modèle dit saturé) et puisqu'il ne permet pas de les résumer. Cependant, on peut le considérer comme une référence : aucun modèle ne s'ajuste mieux et sa log-vraisemblance  $L_{max}$  est la plus élevée parmi tous ceux de la famille considérée. La log-vraisemblance  $l_0$  du modèle dont on veut mesurer l'adéquation est inférieure, mais si la différence n'est pas trop élevée on pourra affirmer qu'il s'ajuste bien aux données. Sous l'hypothèse que le modèle considéré est correct, on montre que  $2(l_{max} - l_0)$  suit asymptotiquement une distribution de chi-2 dont le nombre de degrés de liberté est égal à la différence entre le nombre de paramètres des 2 modèles.

En pratique, l'usage est souvent de comparer la déviance standardisée (divisée sur le coefficient de dispersion) avec le nombre de degrés de liberté (ddl) donné par :

$$ddl = n - p$$

Un modèle est jugé satisfaisant si  $DS / ddl < 1$

### **Conclusion**

Chaque assureur adopte une tarification qui lui est propre et qui se base sur des considérations techniques et commerciales lui permet de se différencier et d'adapter le prix de risque au profil de son portefeuille d'assurés. Pour ce faire, la segmentation des risques est nécessaire permettant de différencier les risques et les assurés qu'il porte à sa charge.

Par ailleurs, la segmentation permettra à l'assureur de prendre le recul nécessaire et des mesures adéquates à chacun des niveaux de segmentation, tant en tarification, qu'en souscription.

---

<sup>42</sup>Estimation du modèle linéaire généralisé et application, Malika CHIKHI, Michel CHAVANCE, Sciences & Technologie A – N°35, Juin 2012, 13-21, page 19.

# **CHAPITRE III**

## **PRÉREQUIS POUR LA MODÉLISATION**

Ce chapitre sera réservé à l'étude du portefeuille ainsi que les principales étapes de traitements affectées pour le rendre exploitable. Par la suite, une analyse univariée et bivariée des variables recueillies sera menée et qui constitue une étape indispensable à la mise en place d'un modèle de tarification.

Pour développer davantage la logique d'idées, ce chapitre sera structuré comme suit :

### Chapitre 3 : prérequis pour la modélisation

- **Section 1** : présentation de la base et traitement des variables
- **Section 2** : analyse univariée et bivariée des variables

---

## Chapitre 3 : prérequis pour la modélisation

---

### Section 1 : Présentation de la base et traitement des variables

#### 1.1. Origine de la base de données

Notre étude est basée sur des données provenant du portefeuille d'assurance automobile de la compagnie **CASH assurances**. Ces données ont été fournies **sur** quatre fichiers Excel différents ;

- trois fichiers, relatif à la production, qui contiennent respectivement des informations sur le contrat, le véhicule assuré, et le conducteur du véhicule
- Un fichier sinistre contient tout sinistre ayant eu lieu dans la période d'étude, à savoir de 2010 à 2016 (le sinistre est rattaché à la date de survenance et non pas à la date de règlement. .

Le fichier production contient 154 996 lignes, chaque ligne correspond à un véhicule ayant souscrit une garantie dommages durant 6 ans, à savoir de 2010 à 2016. Pour chaque

ligne, nous disposons de variables qui caractérisent l'assuré, son contrat ainsi que son véhicule. À savoir :

### 1. Liste des variables intrinsèques à la police :

- Numéro de police de base.
- Date d'effet et date d'échéance de la police
- Formule de garantie dommages (garantie souscrite)

Nous disposons de 5 garanties :

1. Dommages avec ou sans collision (DASC)
2. Dommage-collision
3. Bris de glace
4. Vol-incendie
5. Assistance automobile

### 2. Liste des variables intrinsèques au conducteur :

- Date de naissance (**Âge. Client**)
- Genre (M : homme, F : femme)
- La zone de circulation : L'assureur distingue entre la zone nord et la zone sud. Les wilayas de la zone nord sont indiquées en annexe.

### 3. Liste des variables intrinsèques au véhicule

- Marque
- La puissance fiscale du véhicule : elle est exprimée en fonction du nombre de CV ou du tonnage. Les différentes catégories de puissance sont présentées comme suit :

Code puissance	Désignation
0	2 CV
1	3 à 4 CV
2	5 à 6 CV
3	7 à 10 CV
4	11 à 14 CV
5	15 à 23 CV
6	24 CV et plus
7	50 CM3
8	125 CM3
9	175 CM3

**Tableau 02** : Les différentes catégories de puissance

- Date de mise en circulation (Âge du véhicule)
- genre du véhicule : Les différentes classes du genre de véhicules sont présentées comme suit : (véhicules particuliers sans remorque, scooters...), ces différentes classes sont présentées en annexe. Nous nous sommes intéressés à la classe véhicules particuliers sans remorque
- Valeur du véhicule

Le **fichier sinistre**, pour sa part, est constitué de 18 922 lignes et comporte les sinistres du portefeuille durant la période 2010-2016. Nous disposons comme variables :

- Numéro de police de base
- Année de survenance du sinistre
- Charge de sinistre

Nous avons, également, créé les variables suivantes :

- **Nbr.sinistre** : le nombre de sinistres observés. Cette variable prend ses valeurs dans {0, 1, 2,3}

- **Exposition** : durée de couverture au prorata temporise de l'exercice. En effet, un individu qui aura eu un sinistre en quatre mois d'exposition aura le même risque qu'un individu qui a eu trois sinistres en un an d'exposition.
- **Fréquence** : le nombre de sinistres divisé par l'exposition. Cette variable prend ces valeurs dans  $[0 - 7,2]$
- **Charge cumulée** : le montant cumulé des sinistres observés.
- **coût unitaire** :

$$\left\{ \begin{array}{ll} \text{Charge cumulée} / \text{Nbr.sinistre} & \text{si nbr} > 0 \\ 0 & \text{sinon} \end{array} \right.$$

Il important de noter qu'après un travail de traitement et de vérification des données, nous disposons d'un tableau de production constitué de 76 086 observations (ce qui correspond à 6421 sinistres) et 11 variables. En effet, ce travail préliminaire qui est très déterminant pour la qualité de l'étude nous a couté la moitié de notre période de stage.

## 1.2. Traitement des données

Avant d'entamer toute démarche de modélisation, il est indispensable de veiller à la qualité des données en traitant les valeurs erronées, oublie de renseignement, faute de frappe..., etc.

### 1. Valeurs erronées et données exclues :

Tout d'abord, tout contrat autre que celui relatif au véhicules particuliers sans remorque a été retiré de la base de données, et ce dans un souci d'avoir une population le plus homogène possible. De plus, les variables possédaient des valeurs tout à fait aberrantes, telle que la variable genre de véhicule ont été exclues de notre base de données.

De même, nous supprimons également les contrats dont la variable :

- Âge de véhicule est supérieur à 40 ans à la date de prise d'effet de la garantie
- Âge conducteur est inférieur à 18 ans et supérieur à 90 ans.
- Valeur du véhicule est inférieure à 10 000 DA et supérieure à 35000000 DA.

La garantie Dommages-collision est également exclue de notre étude. Cette garantie compte des sous garantie dont la valeur assurée est plafonnée selon l'option choisie (ex : Dommages-collision à 10000 DA). De ce fait, pour pouvoir la tarifier nous devons disposer des données relatives à chacune de ces sous garantie ce qui n'était pas le cas pour notre base de données, faute de disponibilité.

De même, la garantie Assistance a été retirée de l'étude, car elle n'affiche que 4 sinistres dans notre base de données après la phase de traitement de cette dernière.

Finalement, tout contrat résilié, suspendu ou ayant un avenant de changement de garantie, changement de véhicule...etc., a été retiré de la base pour éviter de baisser les informations disponibles.

## 2. Données complétées :

La variable **genre** compte beaucoup de valeurs manquantes (non renseignée). Compte tenu de l'information que porte cette variable, nous avons jugé utile de ne pas la supprimer. Nous avons choisi de la laisser dans l'état actuel et de fixer sa valeur à « ND » (non disponible), le non-remplissage peut être considéré comme une information en lui-même.

### 1.3.Traitement des variables

Chaque classe de risque pour chacune des variables servant à la tarification se verra attribuée le même montant de prime. Ainsi dans une logique de marketing, mais aussi actuarielle, il est judicieux de répartir un ensemble de données en k classes homogènes ou de regrouper un ensemble de modalités dans une classe affichant les mêmes caractéristiques pour ne pas se trouver avec un nombre important de tarifs différents.

De même, rappelons que certains outils statistiques ou modèles, notamment GLM, exigent que les variables quantitatives continues soient transformées en classes pour pouvoir les exploiter.

Il existe deux grandes familles de méthodes de classification :

Les méthodes de partitionnement, auxquelles on doit fournir le nombre de groupes a priori et qui trouvent la meilleure façon de créer de tels groupes.

- Les méthodes hiérarchiques, qui n'ont pas d'a priori sur le nombre de groupes et aboutissent à un arbre (de type phylogénétique) appelé dendrogramme.

En fait, le choix entre ces deux grandes familles se fait en fonction de type de variable et de la finalité de classification.

### 1.3.1. Type de variables disponibles :

variable	Type	Nature	Nombre de modalités
<b>Âge. Client</b>	Quantitative	Continue	-
<b>Âge. véh</b>	Quantitative	Continue	-
<b>V.véh</b>	Quantitative	Continue	-
<b>Zone</b>	Qualitative	Nominale	2
<b>Nb.cheveux</b>	Qualitative	Ordinale	10
<b>Type. Client</b>	Qualitative	Nominale	9
<b>Sexe</b>	Qualitative	Nominale	3
<b>garanties</b>	Qualitative	Nominale	3

Tableau 03 : Type et nature des variables de la base

À partir de ce tableau, nous constatons la nécessité de transformation des variables :

1. **Âge. Client**
2. **Âge. Véh**
3. **V.véh**

Puisque les variables sont quantitatives, leur transformation en classe se fera par la méthode de partitionnement et en particulier par l'algorithme **Kmeans**.

**Clustering par K-means : regroupement d'observations autour de centres mobiles**

Le data clustering est une méthode statistique qui permet de diviser un ensemble de données en groupes homogènes. Cette méthode peut être utilisée pour classer une variable quantitative continue. Nous comptons de nombreux algorithmes permettant le partitionnement des données. Dans notre cas de figure, nous allons partitionner nos données avec l'algorithme des K-means car il est le plus simple et le plus rapide.

Dans cet algorithme chaque classe est représentée par la moyenne. k-Means est un algorithme itératif. Il commence avec un ensemble de k individus de référence choisis de façon aléatoire. Les individus de données sont ainsi partitionnés dans k classes ; un individu appartient à une classe si le centre de cette classe est le plus proche de lui (en terme de distance). La mise à jour des moyennes et l'affectation des individus de données aux classes sont réalisées pendant les itérations successives.

**1.3.2. Aspect théorique de la méthode de partitionnement<sup>43</sup> :****Définition**

La distance sur un ensemble E est une application d définie sur le produit  $E^2 = E \times E$  à valeurs dans l'ensemble  $\mathbb{R}_+$  des réels positifs. Il existe notamment :

- La distance euclidienne :

$$d(x, y) = \sqrt{\sum_{j=1}^p (x^j - y^j)^2}$$

- La distance de Manhattan :

$$d(x, y) = \sum_{j=1}^p |x^j - y^j|$$

---

<sup>43</sup>Master actuariat de dauphine, Mise en place de modèles de tarification alternatifs face à la suppression réglementaire d'une variable tarifaire en automobile, ATIA Rachel, page 65

**Définition :**

L'inertie d'un nuage de  $n$  points est la moyenne des carrés des distances de ces points à leur centre de gravité noté  $g$ .

Si  $d(x_i, g)$  est la distance du point  $i$  au centre de gravité, alors l'inertie totale vaut :

$$I_T = \frac{1}{n} \sum_{i=1}^n d^2(x_i, g) \text{ où } g = \frac{1}{n} \sum_{i=1}^n x_i$$

**Définition :**

**Un cluster** (une classe)  $k$  est un ensemble non vide d'observations.

Notons  $\pi_k = \{ x_i \in \text{cluster } k \}$ .

**Définition :**

**Le centre de gravité** du cluster  $K$  est  $g_k = \frac{1}{n_k} \sum_{i \in \pi_k} x_i$  où  $n_k = \text{Card } \pi_k$

**Définition :**

Supposons que nous ayons  $K$  classes, la  $k^{\text{ième}}$  classe  $\pi_k$  de centre de gravité  $g_k$  est formée de  $n_k$  ( $k = 1, \dots, K$ ) observations, son inertie vaut :

$$I_K = \frac{1}{n_k} \sum_{i \in \pi_k} d^2(x_i, g_k)$$

Elle est d'autant plus homogène que ses éléments sont proches de son centre de gravité, son inertie est alors faible.

Une mesure globale de l'homogénéité des classes appelée inertie intra-classes (d'où le symbole  $W$  comme Within) est donc la moyenne des inerties des  $K$  classes, chacune pondérée par son importance relative :

$$I_W = \sum_{k=1}^K \frac{n_k}{n} I_K = \frac{1}{n} \sum_{k=1}^K \sum_{i \in \pi_k} d^2(x_i, g_k)$$

L'inertie du nuage des  $K$  centres de gravité, appelée inertie inter-classes (d'où le symbole  $B$  comme Between), est :

$$I_B = \sum_{k=1}^K \frac{n_k}{n} d^2(g_k, g)$$

Le théorème de Huygens nous permet d'écrire :

$$I_T = I_W + I_B$$

Le critère de classification choisi est alors celui qui rend l'inertie intra-classes minimale ou l'inertie inter-classes maximale. Le problème se résout par itérations successives.

### 1.3.3. Application : Partitionnement de l'âge du client

Dans ce qui suit, nous mettrons en évidence la démarche de transformation de la variable Age.client en classe de valeur.

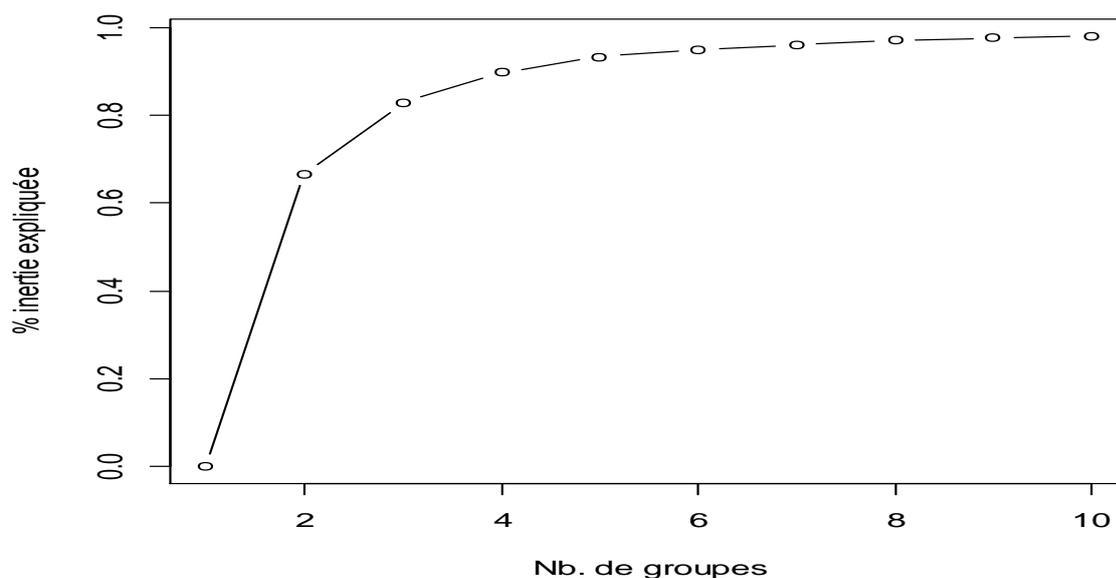
Il est à noter que la méthode des K-means est utilisée en se plaçant dans le cas où  $p=1$  puisque seul l'âge du client caractérise ici l'individu.

Rappelons les statistiques de nos données :

```
> summary(donnee$Age.client)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 18.00  34.00   42.00   43.61  51.00   90.00
```

Nous avons  $n = 76015$  individus dont l'âge varie de 18 ans à 90 ans. La différence ou l'écart entre la moyenne et la médiane témoigne d'une distribution d'âge symétrique et une population équilibrée.

Le choix du nombre de classes se fait en fonction du nombre de classe qui maximise l'inertie inter classe. Le graphique ci-dessous laisse apparaître l'évolution du rapport variance inter classe sur variance totale en fonction du nombre de classes

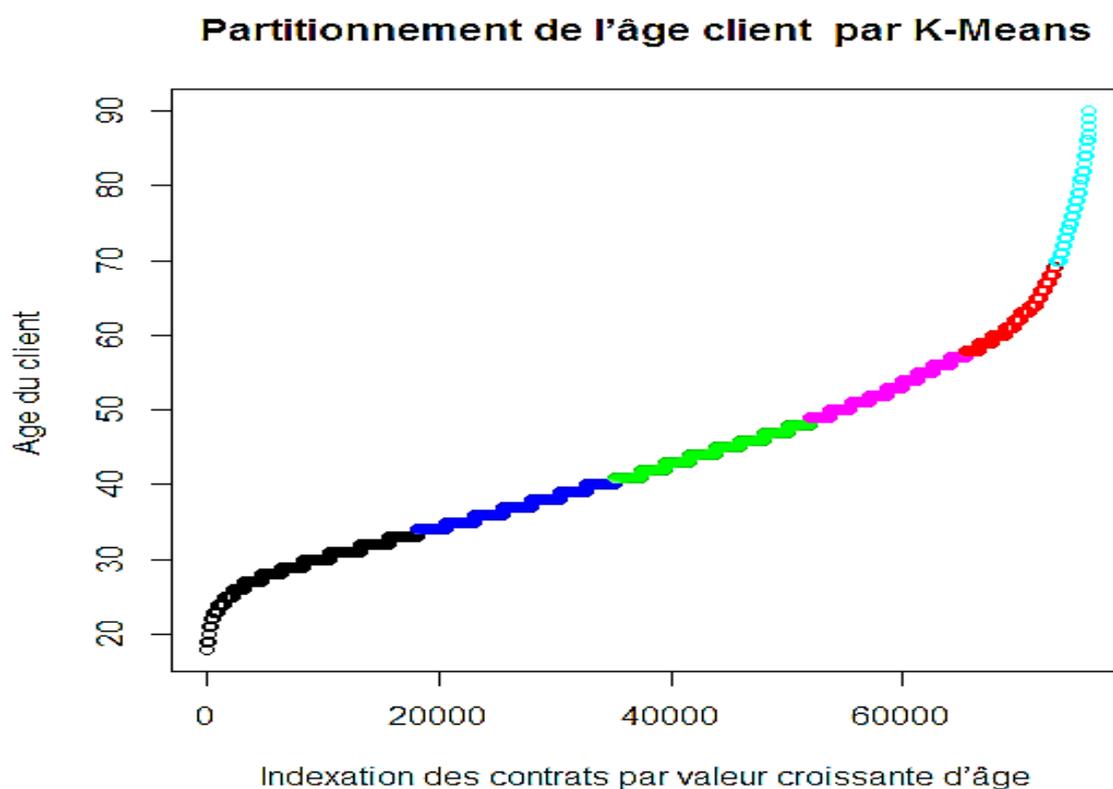


**Figure 04** : rapport variance inter-classe sur variance totale en fonction du nombre de classes.

Nous remarquons, à partir de ce graphique, que l'inertie expliquée commence à se stabiliser à partir de  $k=6$ , ce qui explique que l'adjonction d'un groupe, à partir de  $k=4$ , n'augmente pas significativement l'inertie expliquée par la partition.

Six classes sont retenues ce qui correspond à un rapport de la variance inter-classe sur la variance totale de l'ordre de 95.2%.

À cette étape, il nous reste qu'à chercher la délimitation des classes de l'âge de client à l'aide de la fonction `kmeans()` sous **R**. dans notre cas, l'algorithme `kmeans` est lancé **25** fois pour assurer sa robustesse à l'aide de l'argument `nstart`.



**Figure 5** : Découpage en classes d'âge de la variable Age. Client.

La même démarche s'est imposée pour les autres variables. Les classes retenues ainsi que leurs bornes (les illustrations y afférentes sont en annexe) sont détaillées dans le tableau ci-dessous:

Numéro de classe	Age.clinet	Effectif	Part du port	Age.véh	effectif	Part du port	V.véh en (mda)	effectif	Part du port
1	[18-33]	18098	29,8 %	[0-1]	28604	37,63%	] 0-655]	13755	18,10%
2	[34-40]	17029	28,04 %	[2-2]	11339	14,92%	[656-907]	18840	24,78%
3	[41-48]	16756	27,59 %	[3-5]	22132	29,12%	[908-1152]	14472	19,04%
4	[49-57]	13473	22,19%	[6-40]	13939	18,34%	[1153-1420]	10352	13,62%
5	[58-90]	10658	17,54 %	-	-	-	[1421-1733]	7393	9,73%
6	-	-	-	-	-	-	[1734-2065]	4948	6,51%
7	-	-	-	-	-	-	[2067-35000[	6254	8,23%
<b>Total</b>	-	<b>60 714</b>	<b>100 %</b>	-	<b>60 714</b>	<b>100%</b>	-	<b>76 014</b>	<b>100%</b>

Tableau 04: découpage en classes des variables quantitatives

Il est important de noter que certaines classes ont été regroupées. En effet, chaque classe devrait comporter un nombre considérable d'individus, ce qui nous a amené à regrouper toute classe représente moins de 5% du portefeuille. Pour plus de détail sur le partitionnement des classes initiales, nous invitons le lecteur à consulter l'annexe.

## Section 2 : analyse univariée et bivariée des variables

### 2.1. Analyse univariée

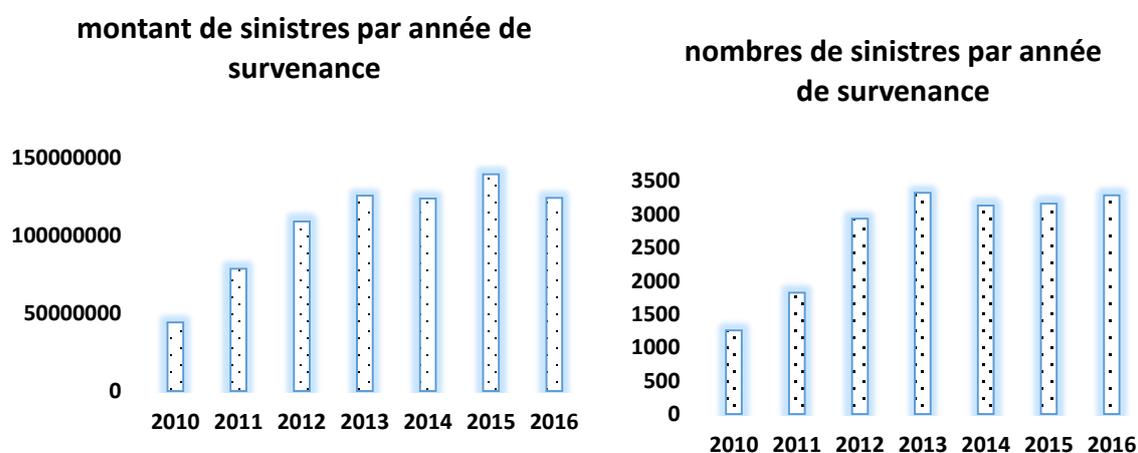


Figure 6: montant et nombre de sinistres par année de survenance

Le nombre de sinistres et leurs montants dans la globalité suivent la même allure qui se traduit par une tendance haussière tout au long de la période d'étude. Cela s'explique par une augmentation de la sinistralité.

montant moyen de sinistres par année de survenance

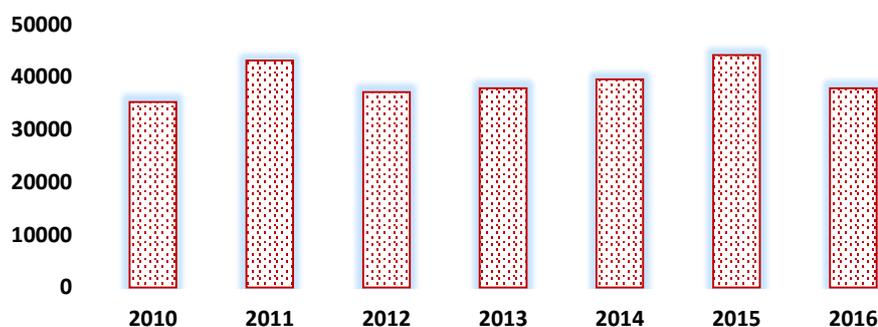
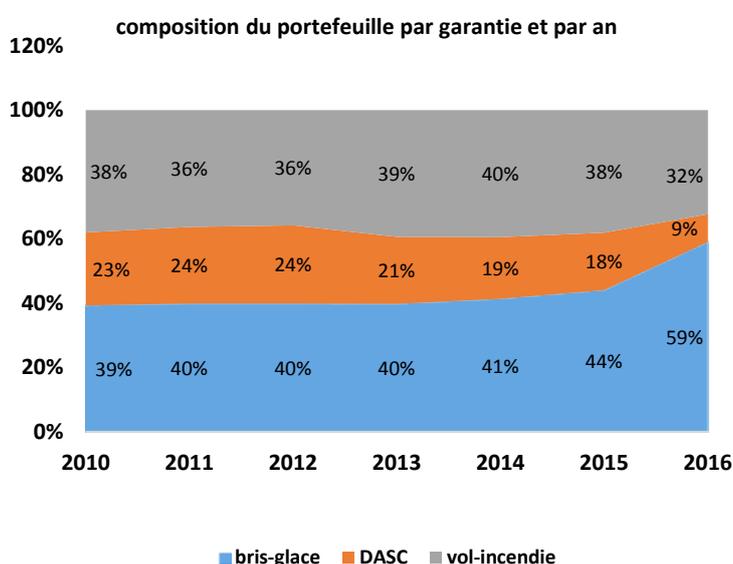


Figure 07: montant moyen de sinistres par année de survenance

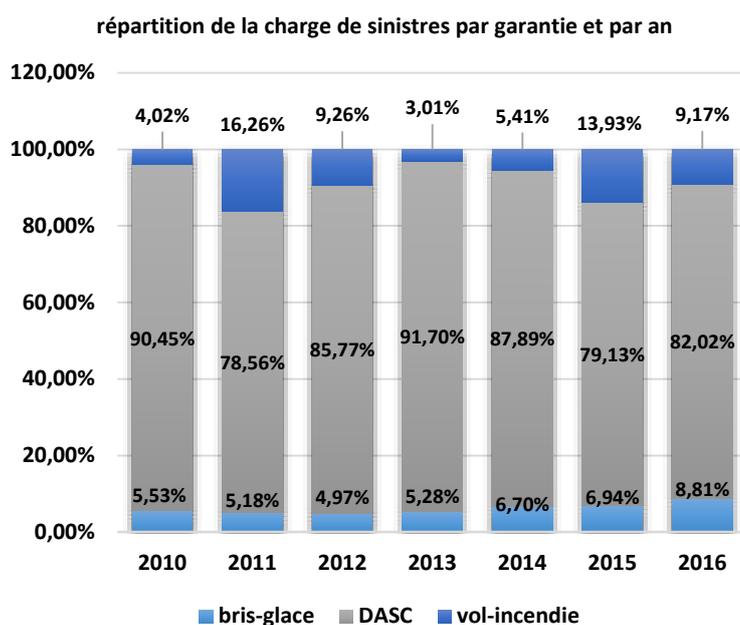
Remarquons tout d'abord que le montant moyen des sinistres matériels suit une augmentation légère, mais constante d'année en année. Cependant, l'année 2016, pour sa part, témoigne d'une baisse de sinistralité. Cette constatation s'aligne avec la situation du marché automobile en 2016. En effet, les chiffres publiés par « **Revue de l'assurance N°17** » annoncent que les garanties facultatives suivent également une tendance baissière de 1,8 % comparativement à 2015.

**Garanties :**



Nous observons sur ces dernières années une augmentation de la proportion de la garantie bris de glace et une diminution de la proportion de la garantie tous risques. Il est important de noter que l'année 2016 ne compte que les contrats dont la date d'échéance se rapporte à 2016, et ce pour ne pas fausser la modélisation.

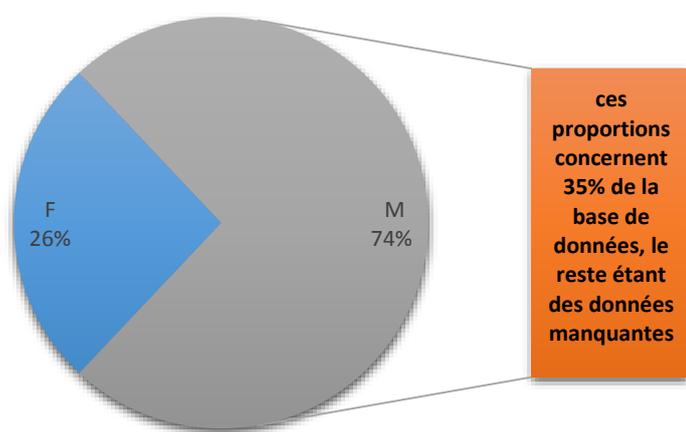
La garantie vol-incendie est quant à elle relativement stable pour les trois premières années avant d'augmenter les deux années suivantes, puis une baisse significative pour le reste de la période d'étude.



Pour ce qui est de la sinistralité, la proportion de la garantie bris de glace affiche une tendance à la hausse, ce qui est dû à l'augmentation de sa part dans le portefeuille. La garantie tous risques, malgré la diminution de sa part dans le portefeuille, la sinistralité y afférente continue d'augmenter et plus particulièrement en 2016.

Figure 08: composition du portefeuille et répartition de la charge de sinistres par garantie et par année

### Genre



Le portefeuille est à dominance masculine.

Il y a une majorité écrasante du genre masculin.

Figure 09: Composition du portefeuille par genre

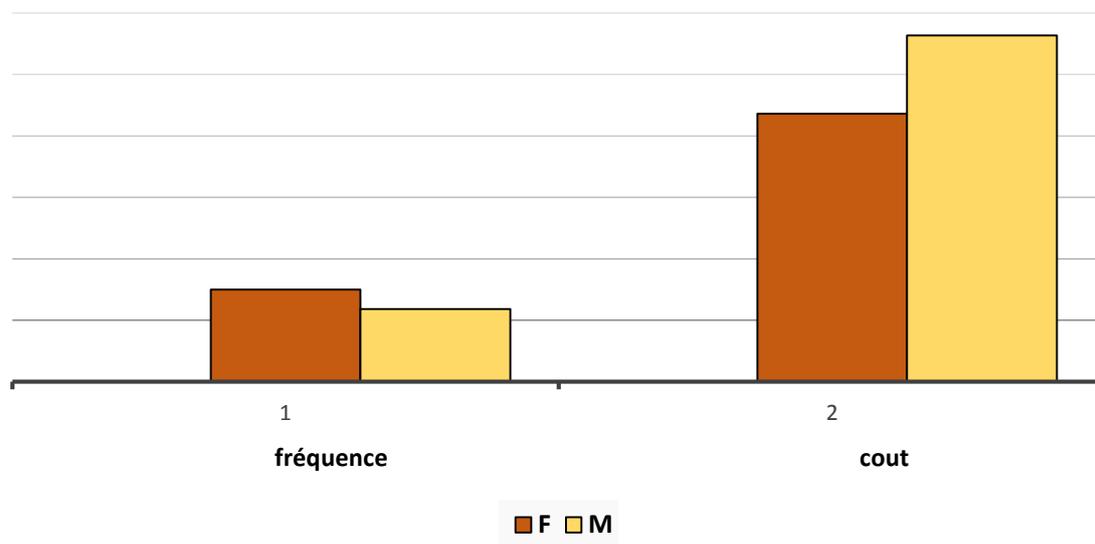
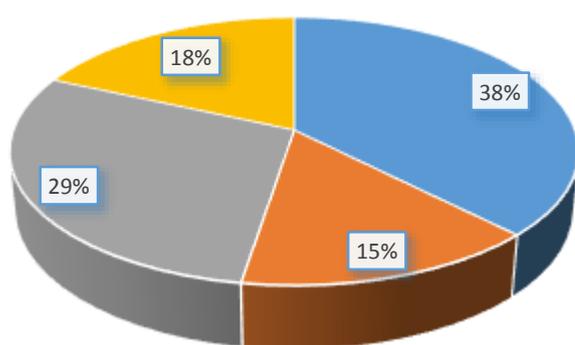


Figure 10: fréquence et coût moyen par genre

Ce graphique laisse apparaitre que la fréquence moyenne des femmes est plus élevée à celle des hommes. Cependant, ce constat n'est pas valable pour le coût moyen de sinistres, car en termes de sévérité, les sinistres causés par les hommes sont plus importants relativement à ceux causés par les femmes.

**Âge du véhicule :**



Un poids presque équitable des quatre classes dans le portefeuille, avec toutefois une prédominance des véhicules dont l'âge se situe entre 0 et 1 an.

Figure 11: Composition du portefeuille par âge du véhicule.

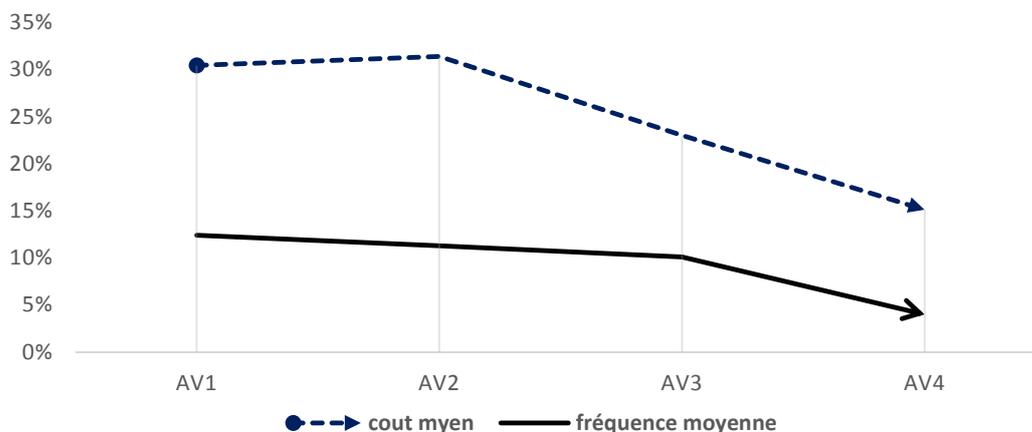


Figure 12: Répartition de la fréquence et coût moyen de sinistres par classe d'âge du véhicule

Nous constatons que la fréquence et le coût moyen de sinistre diminue au fur et à mesure avec le vieillissement du véhicule assuré. Ainsi donc, plus le véhicule vieillit moins il est sinistré. De ce fait cette variable a une influence significative sur la sinistralité observée.

**Valeur du véhicule (VV) :**

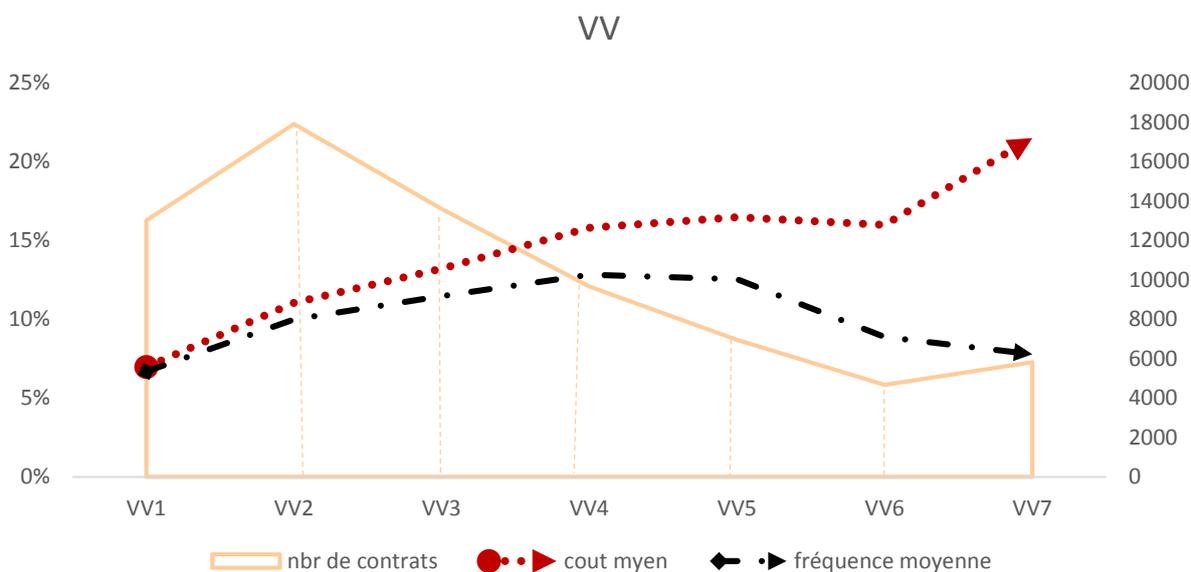
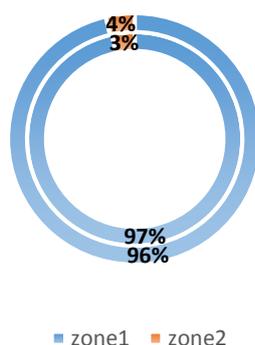


Figure 13: Composition du portefeuille et répartition de la fréquence et la charge de sinistres par classe de valeur du véhicule

Les quatre premières classes composent 76% du portefeuille avec une prédominance de la deuxième classe [656-907] KDA. Cette prédominance s'expliquerait par l'attitude des assurés qui sous-assurent leurs véhicules pour payer moins de la prime effective.

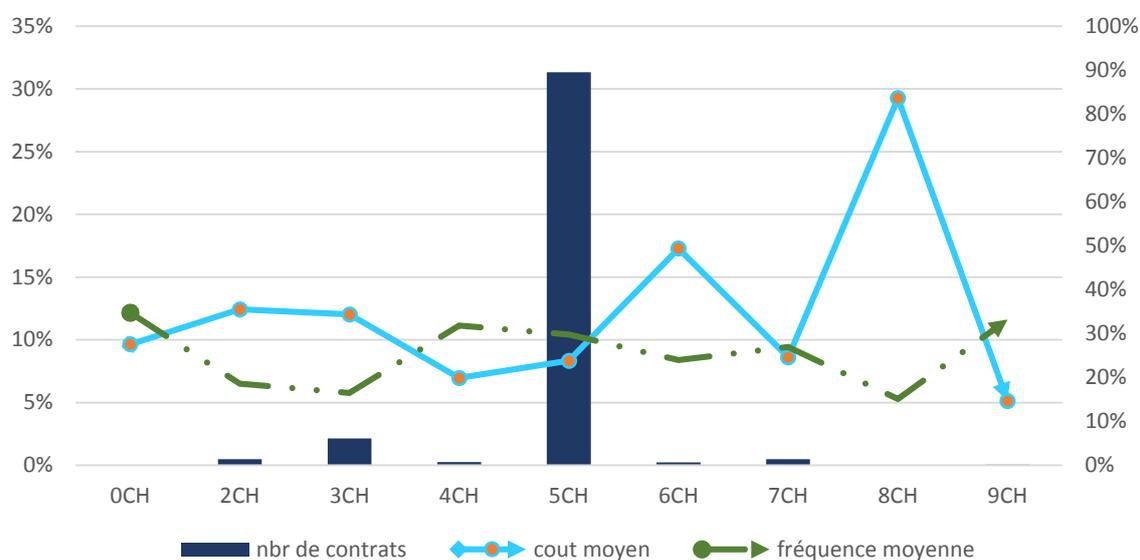
Les classes de valeur du véhicule sont caractérisées par une fréquence moyenne croissante jusqu'à la classe 5 où la fréquence commence à prendre une tendance baissière. Par ailleurs, le coût moyen est toujours en augmentation avec l'accroissement de la valeur du véhicule. Ce constat paraît logique, car plus la valeur du véhicule augmente plus le montant du sinistre sera conséquent. Donc, nous attendons une présence d'influence de cette variable sur la sinistralité.

### Zone



Le premier cercle, celui qui est à l'extérieur, représente la composition du portefeuille par zone. Nous remarquons qu'il y a une majorité écrasante de la classe zone1. Le deuxième cercle met en évidence la part de chaque classe dans la charge de sinistres. Ce constat pourrait s'expliquer par deux facteurs. Le premier est la disparité en termes de population au sein des zones en Algérie. Le deuxième pourrait s'expliquer par le fait que le taux de pénétration de la zone2 par le réseau commercial de l'entreprise demeure faible. Dans tous les cas de figure, cette variable n'apporte ni de l'information ni de la segmentation. Ainsi donc, elle ne sera pas prise comme une variable de tarification.

**Figure 14:** composition du portefeuille et répartition de la charge sinistre par zone géographique

**Nombre de chevaux :**

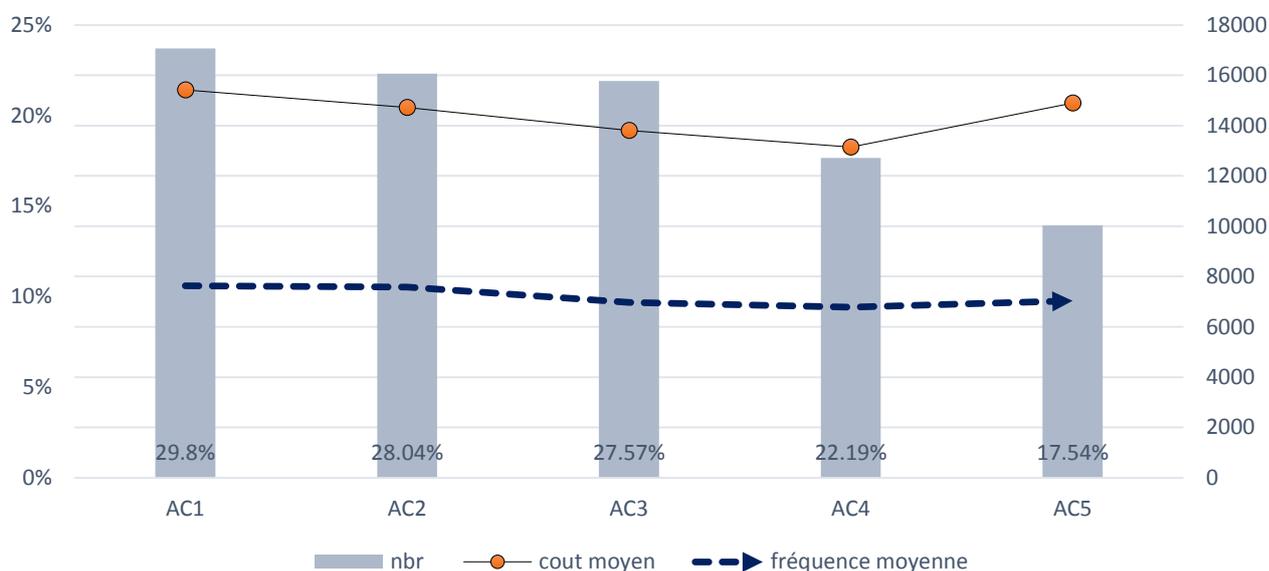
**Figure 15:** Composition du portefeuille et répartition de la fréquence et la charge de sinistres par nombre de chevaux du véhicule

Nous observons à partir de ce graphique que la classe 5 représente près de 90% du portefeuille. Quant aux autres classes, elles ne dépassent pas les 7% si nous prenons par exemple la classe 3 qui se positionne en deuxième place.

Par ailleurs, le graphe par ligne met en évidence que la charge moyenne de sinistres diffère d'une classe à une autre. En effet, en moyenne, nous constatons que la classe 6 renvoie la sinistralité la plus élevée et vient juste après la classe 8 et la classe 1.

En ce qui concerne la fréquence moyenne, elle fluctue d'une manière aléatoire d'une classe à l'autre. Nous observons que les classes 0, 4, 5, 6, 7, et 9 sont marquées par une fréquence élevée par rapport aux autres classes.

Âge du client



**Figure 16:** composition du portefeuille et répartition de la fréquence et la charge sinistre par classe de valeur du véhicule

Nous remarquons que les parts des classes d’âge dans le portefeuille sont plus ou moins équilibrées. Le graphe en ligne indique que la part moyenne de la charge de sinistre par classe diminue légèrement au fur et à mesure avec l’augmentation d’âge des assurés avant de reprendre son niveau initial à partir de la 4ème classe ce qui correspond à l’âge de 58. Pour ce qui est de la fréquence de sinistres, elle reste stable dans les différentes classes d’âge. Ainsi donc, nous attendons une absence d’influence de la variable âge du client sur la sinistralité.

**2.2. Analyse bivariée**

**2.2.1. Étude de liens entre les variables**

Avant de procéder à la construction des modèles ajustés à nos données, il est indispensable d’étudier les liens existant entre nos variables. Pour ce faire, nous allons procéder en deux étapes.

Une première consiste à repérer, parmi les variables explicatives candidates, celles qui sont corrélées entre elles. Les variables sont testées deux à deux, si la corrélation est élevée nous procéderons par l'élimination d'une des variables.

Dans la deuxième étape, nous repérerons les variables qui sont les plus corrélées à la variable à expliquer. Généralement, si la corrélation est élevée nous garderons la variable, sinon elle sera éliminée.

### 1.1 Étude de liens entre variables explicatives :

Les tests permettant de déceler la corrélation entre variables diffèrent selon la nature des variables à tester. Dans notre cas de figure, nous avons fait recours au test d'indépendance du  $\chi^2$  et V de Cramer, car nous ne disposons que des variables qualitatives. Rappelons que les variables quantitatives ont été transformées en classe (qualitative) dans la première partie de ce chapitre.

C'est donc le test du  $\chi^2$  d'indépendance qui est adapté à l'étude des corrélations entre ces variables. Pour chaque paire de variables sur lesquelles le test est réalisé, nous mettons les hypothèses suivantes :

- Hypothèse nulle ( $H_0$ ): les deux variables sont indépendantes
- Hypothèse alternative ( $H_1$ ) : les deux variables sont corrélées

Nous construisons le tableau de contingence des deux variables à étudier, puis nous obtenons, à l'aide de la fonction `chisq.test()` sur R, la valeur de la statistique  $\chi^2$  et la p-value associée au test. Si la p-value est inférieure à 0.05 nous acceptons  $H_0$ .

Dans ce qui suit, nous allons compter sur un exemple d'application entre la variable âge du client et âge du véhicule pour mettre en évidence les étapes menant à l'application de ce test. Pour les autres variables, leurs résultats sont synthétisés dans le tableau n°(05).

Étape 1 : la réalisation du tableau de contingence

```
> table(donneep$Age.C, donneep$Age.V)

      AV1  AV2  AV3  AV4
AC1 6957 2675 5221 3245
AC2 6416 2573 4991 3049
AC3 6257 2556 4799 3144
AC4 4921 2005 4008 2539
AC5 4053 1530 3113 1962
```

Figure 17: tableau de contingence de l'âge de client et l'âge du véhicule

Étape 2 : calcul de la statistique de  $\chi^2$  et la p-value y

```
> test<-chisq.test(table_c, correc=F)
> test

Pearson's Chi-squared test

data:  table_c
X-squared = 23.354, df = 12, p-value = 0.02486
```

Étape 3 : prise de décision

Nous constatons que la p-value est inférieure à 0.05, ce constat nous amène à rejeter l'hypothèse H0 qui suppose l'indépendance entre l'âge du véhicule et celui du client.

De la même manière nous procédons pour l'étude des autres variables dont les résultats sont récapitulés dans tableau ci-après :

	Age.V	VV	nb.CH	Type. Client	Genre	garantie
Age.C	0.02486	0.3941	0.008248	0.0001568	0.00008441	5.798e-12
Age.V		< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16
VV			< 2.2e-16	2.2e-16	< 2.2e-16	< 2.2e-16
nb.CH				< 2.2e-16	< 2.2e-16	< 2.2e-16
Type. Client					< 2.2e-16	< 2.2e-16
genre						< 2.2e-16

Tableau 05: p-value des tests du Khi-deux d'indépendance

Les résultats fournis annoncent l'existence de lien entre toutes les variables prises deux à deux à l'exception de la variable âge du véhicule (**Age.V**) et sa valeur (**VV**) avec une  $p\text{-value}=0.3941$  qui est supérieure à 0.05. De ce fait, nous devrions mesurer le degré d'association des variables à l'aide de **l'indice de Cramer** pour repérer les variables fortement corrélées et les éliminer.

L'indice de Cramer prend ces valeurs dans l'intervalle  $[0,1]$ , plus il est proche de 1 plus les variables sont corrélées.

	Age.C	Age.V	VV	nb.CH	Type. Client	genre	garantie
Age.C	1	0.01011991	0.009106585	0.01336235	0.01506892	0.01456215	0.02140252
Age.V		1	0.3511626	0.08428243	0.06934559	0.06526616	0.1399684
VV			1	1	0.1327095	0.1105993	0.1119466
nb.CH				1	0.1751836	0.09854519	0.03037835
Type. Client					1	0.2102952	0.06033574
Genre						1	0.0348821
garantie							1

**Tableau 06:** indices de Cramer.

À première lecture, nous remarquons que les variables prises deux à deux ne sont pas fortement corrélées exception faite pour la variable valeur du véhicule et ses nombres de chevaux ( $V=1$ ). En effet, ces deux variables sont positivement fortement corrélées. Ceci paraît intuitif puisqu'une voiture puissante aura tendance à se situer dans une classe de valeur élevée.

De plus, il existe un certain lien ( $V=0.35$ ) entre l'âge du véhicule et sa valeur. Ceci est aussi légitime, car plus l'âge du véhicule augmente plus sa valeur diminue.

D'après l'étude d'association, la variable **nb.CH** sera éliminée pour la suite de l'étude en raison de l'importance du lien existant avec la variable **VV**, **et ce pour ne pas paramétrer nos modèles par des variables ayant un double effet.**

### 2.2.2. Étude de liens entre variables explicatives et variables à expliquer

À présent, nous verrons de près l'influence des variables explicatives sur les variables à expliquer (la fréquence et le coût de sinistres). L'étude de lien se fera à l'aide de deux tests statistiques : le test de Student pour groupes indépendants et le test de Kruskal-Wallis. Le choix du test est fonction de nombres de modalités de la variable explicative. Nous réalisons donc :

- le test de Kruskal-Wallis car toutes les variables que nous disposons admettent de trois à sept modalités

Rappelons que la variable zone a été éliminée et la variable genre compte 3 modalités du fait de la non-disponibilité de l'information. Pour plus de détail, nous invitons le lecteur à consulter la partie relative au traitement de la base de données.

#### Déroulement du test :

Pour chaque variable observée, une modalité sera considérée comme un groupe. L'objectif, par la suite, est de vérifier si l'un des groupes (modalités) se comporte différemment des autres en termes de fréquence ou de coût de sinistre, cela revient à vérifier s'il existe une influence de la variable étudiée sur la sinistralité.

Pour ce faire, les hypothèses du test sont :

- **Hypothèse nulle ( $H_0$ )** : la variable  $x$  n'a aucune influence sur la fréquence (coût) de sinistre.
- **Hypothèse alternative ( $H_1$ )** : la variable  $x$  influe sur la fréquence (coût) de sinistre.

La décision sera fonction de la p-value associée au test. Si la p-value est inférieure à 0.05 nous rejetons  $H_0$ .

Compte tenu du nombre important de tests à effectuer. Nous allons, comme nous avons fait précédemment, nous appuyer sur un exemple pour tester l'influence de l'âge client sur la sinistralité afin de mettre en évidence les étapes menant à l'application de ce test. Pour les autres variables, leurs résultats sont synthétisés dans le tableau n°(07).

### Exemple d'application : Corrélacion entre la fréquence de sinistre et l'âge du véhicule

En pratique, le teste se fait à l'aide de la fonction `kruskal.test()` sous **R**. les résultats retournés sont les suivants :

```
> kruskal.test(donnees$fréquence2~donnees$Age.V)

Kruskal-Wallis rank sum test

data:  donnees$fréquence2 by donnees$Age.V
Kruskal-Wallis chi-squared = 653.47, df = 3, p-value < 2.2e-16
```

La p-value est inférieure à 0.05 ce qui nous amène à rejeter  $H_0$  au profit de  $H_1$ . Ainsi donc, la variable âge du véhicule a une influence significative sur la fréquence de sinistres.

Nous pouvons même aller plus loin en identifiant, à l'aide de logiciel R, les groupes qui se comportent différemment des autres. Cette analyse se fait entre les groupes deux à deux à l'aide de la fonction `kruskalmc()`.

```
> kruskalmc(donnees$charge.total~donnees$Age.V)
Multiple comparison test after Kruskal-Wallis
p.value: 0.05
Comparisons
```

	obs.dif	critical.dif	difference
AV1-AV2	235.8547	642.4547	FALSE
AV1-AV3	723.1724	518.2715	TRUE
AV1-AV4	2739.3146	598.0091	TRUE
AV2-AV3	487.3177	668.5896	FALSE
AV2-AV4	2503.4599	732.1349	TRUE
AV3-AV4	2016.1422	626.0023	TRUE

Figure 18: analyse de comportement de groupes pris deux à deux

Nous remarquons que la différence se trouve entre AV1-AV3, AV1-AV4, AV2-AV4 et AV3-AV4. Nous n'avons pas relevé une différence entre AV1-AV2 ce qui paraît logique étant donné la succession des deux classes

Les résultats du test pour l'ensemble des variables sont synthétisés dans le tableau ci-dessous :

	fréquence	coût
Age.C	p-value = 0.0008544	p-value = 0.0008893
Age.V	p-value < 2.2e-16	p-value < 2.2e-16
VV	p-value < 2.2e-16	p-value < 2.2e-16
nb.CH	p-value = 2.652e-14	p-value = 1.515e-13
Type. Client	p-value = 8.112e-10	p-value = 0.000000002552
Genre	p-value < 2.2e-16	p-value < 2.2e-16
Garantie	p-value < 2.2e-16	p-value < 2.2e-16

**Tableau 07:** significativité de l'influence sur la fréquence et coût de sinistres

La première observation que nous pouvons faire est que toutes nos variables ont une influence significative sur la fréquence et le coût de sinistres (p-value < 0.05), bien que la variable **Âge.c** affiche une probabilité largement supérieure par rapport aux autres variables. De ce fait, nous attendons un degré d'influence faible de la part de la variable **Âge.c** sur la fréquence et le coût moyen.

Nous choisissons donc de toutes les garder aussi bien dans le modèle de coût que le modèle de fréquence.

### Conclusion

- L'analyse univariée montre que la variable :

**Genre** a une influence significative sur la sinistralité. Nous avons décelé que les femmes sont caractérisées par une fréquence moyenne élevée et un coût moyen faible relativement aux hommes.

**Age du véhicule** a également une influence sur la sinistralité. Nous avons constaté que plus le véhicule vieillit moins il est sinistré.

**Zone** a été éliminée du fait que la région Nord contient à elle seule 97% des observations de la variable **zone**.

**Valeur du véhicule (VV)** a une influence positive sur la sinistralité

- L'analyse bivariée montre que :

Les variables **valeur du véhicule (VV)** et sa **puissance fiscale (nb.CH)** sont parfaitement positivement associées ( $V=1$ ). De ce fait, la variable nombre de chevaux a été éliminée.

Toutes les variables ont une influence significative sur la fréquence et le coût de sinistres. Cependant la variable âge du client (**Age.C**) affiche une influence moins significative relativement aux autres variables.

# **CHAPITRE IV**

## **MODÉLISATION DE LA FRÉQUENCE ET LE COÛT DE SINISTRES**

Dans le chapitre précédent, nous avons retenu 5 variables pour la segmentation. Nous avons conclu qu'il n'était pas judicieux d'utiliser la variable **zone** (zone géographique) dans notre modèle de tarification juste parce que la modalité zone1 représentant les wilayas de la région Nord contient à elle seule 97% des observations de la variable **zone**. Par ailleurs, après l'étude de la corrélation entre variables tarifaires, nous avons envisagé la suppression de la variable nombre de chevaux (**nb.CH**) pour ne pas paramétrer notre modèle de tarification avec des variables qui ont la même influence sur la sinistralité.

Dans ce chapitre, nous allons nous intéresser à la modélisation de la fréquence et le coût de sinistres. La modélisation de la charge de sinistres est fondée sur l'hypothèse d'un modèle collectif, dont l'une des propriétés est de permettre l'estimation de la charge moyenne des sinistres en séparant l'étude de cette dernière en l'étude de la fréquence moyenne et du coût moyen de sinistre selon la formule :

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

Avec :

S la charge de sinistre

N la fréquence (le nombre) de sinistres

C le coût de sinistre.

Notre démarche est la suivante :

Nous commencerons par la construction du modèle de fréquence, puis le modèle de coût. Dans chaque partie, nous allons respecter les étapes suivantes qui sont nécessaires à la validation des modèles.

1. Choix de la distribution théorique ajustée
2. Estimation des paramètres du modèle
3. Sélection des variables
4. Validation du modèle

## Chapitre 4 : modélisation de la fréquence et coût de sinistres

### Section 1 : Modèle pour la fréquence

Dans une première partie, notre variable d'intérêt est la fréquence de sinistres. Notons que la fréquence est égale au nombre de sinistres déclarés divisés par l'exposition, car nous cherchons à une fréquence exprimée à une échelle d'année.

#### 1.1. Choix de la distribution théorique

En actuariat<sup>44</sup>, les principales lois pour la variable aléatoire de fréquence sont les lois de Poisson et binomiale négative. Dans ce cas, il nous revient de faire le choix entre la loi de poisson et la loi binomiale négative en prenant la plus adaptée à nos données.

Une technique à l'aide de logiciel **easyfitt** consiste à analyser la distribution empirique et la rapprocher à la loi théorique la plus adaptée en procédant par classement sur la base du teste de **Anderson-Darling** et **Kolmogorov-Smirnov**. Les résultats de l'ajustement sont illustrés ci-après :

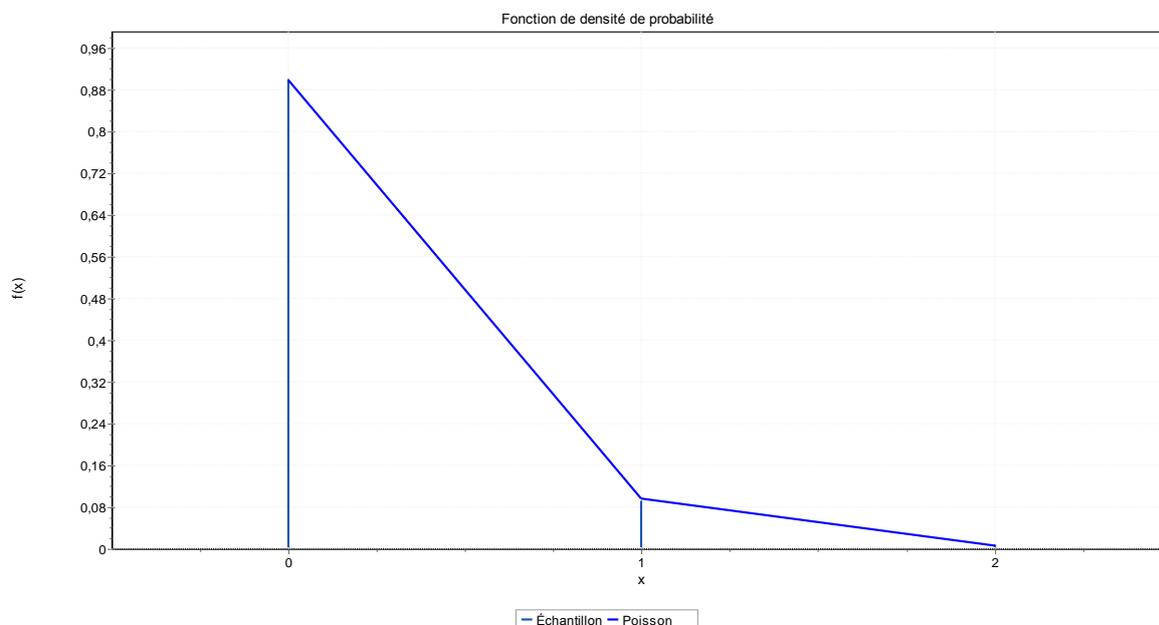


Figure 19: Ajustement de la distribution théorique à une distribution de la loi de poisson

Résultats d'ajustage Qualité d'ajustement – Synthèse

#	Distribution	Paramètres	#	<u>Distribution</u>	<u>Kolmogorov Smirnov</u>		<u>Anderson Darling</u>	
					Statistique	Rang	Statistique	Rang
1	Geometric	p=0,90264	1	<u>Geometric</u>	0,90264	2	7129,3	2
2	Neg. Binomial	n=2 p=0,96	2	<u>Neg. Binomial</u>	0,92159	3	8140,0	3
3	Poisson	$\lambda=0,10786$	3	<u>Poisson</u>	0,89775	1	6934,9	1
4	Bernoulli	Pas d'ajustage	4	Bernoulli	Pas d'ajustage (max > 1)			
5	Binomial	Pas d'ajustage	5	Binomial	Pas d'ajustage			
6	D. Uniform	Pas d'ajustage	6	D. Uniform	Pas d'ajustage			
7	Hypergeometric	Pas d'ajustage	7	Hypergeometric	Pas d'ajustage			
8	Logarithmic	Pas d'ajustage	8	Logarithmic	Pas d'ajustage (min < 1)			

Tableau 08: résultats et Qualité d'ajustement - Synthèse

D'après ces deux tests, notre distribution se rapproche le mieux à une distribution poissonnienne avec un paramètre  $\lambda=0.1$ . Une fois la distribution théorique y est nous passons à l'estimation des coefficients du modèle.

### 1.2. Estimation des coefficients de la régression

L'estimation des coefficients du modèle est réalisée à l'aide de la fonction **glm()** sous R. les variables intégrées (avant sélection) au modèle initial ainsi que leur signification sont les suivantes :

Codification	Signification	Nombre de modalités
Age.C	Âge du client	5
Age.V	Âge du véhicule	4
VV	Valeur du véhicule	7
Garantie	La garantie souscrite	3
Genre	Le genre du client	2
Type.client	Type client	2

Tableau 09: variables intégrées au modèle initial de la fréquence de sinistres.

La variable exposition est intégrée comme une variable **Offset**.

La fonction de lien appliquée sur la variable à expliquer est la fonction log, ce qui donne la relation à estimer suivante :

$$\log(E[F |]) = \beta_0 + \beta_1x_1 + \dots + \beta_px_p$$

- Coefficients estimés :**

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0539  -0.3020  -0.1972  -0.1130   4.6987

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.870965   0.063559  -45.170 < 2e-16 ***
VVVV2         0.021277   0.045498   0.468  0.6400
VVVV3        -0.043982   0.048850  -0.900  0.3679
VVVV4         0.000942   0.051540   0.018  0.9854
VVVV5        -0.025959   0.055306  -0.469  0.6388
VVVV6        -0.311729   0.067334  -4.630 3.66e-06 ***
VVVV7        -0.554265   0.066070  -8.389 < 2e-16 ***
garantieDASC  2.236835   0.034929  64.039 < 2e-16 ***
garantievol-incendie
-1.792740   0.083706 -21.417 < 2e-16 ***
genreM        -0.056617   0.040003  -1.415  0.1570
genreND       -0.156548   0.038689  -4.046 5.20e-05 ***
Age.VAV2     -0.060808   0.035710  -1.703  0.0886 .
Age.VAV3     -0.161385   0.031805  -5.074 3.89e-07 ***
Age.VAV4     -0.230465   0.058217  -3.959 7.54e-05 ***
type.clientSociété
-0.436404   0.041429 -10.534 < 2e-16 ***
Age.CAC2      0.024659   0.035442   0.696  0.4866
Age.CAC3     -0.013872   0.036402  -0.381  0.7032
Age.CAC4     -0.025873   0.039115  -0.661  0.5083
Age.CAC5      0.003849   0.041612   0.093  0.9263
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 32181  on 71666  degrees of freedom
Residual deviance: 20527  on 71648  degrees of freedom
AIC: 32978

Number of Fisher Scoring iterations: 7
    
```

Figure 20: estimation des coefficients du modèle initial de fréquence

L'estimation du paramètre  $\beta$  associé à chaque modalité est représentée par la colonne [estimate]. La colonne [std.Error] correspond à l'écart-type ( $\sigma_B$ ) du paramètre estimé.

Nous remarquons aussi l'absence de modalité pour chaque variable. Cette dernière est appelée modalité de référence, car son coefficient vaut 0 par construction et sa valeur est fixe. De ce fait, un individu se caractérisant par la modalité de référence pour chaque variable est appelé individu de référence.

### 1.3.Sélection des variables

Après avoir estimé les coefficients associés à chaque variable, il convient par la suite de procéder à la sélection des variables.

Nous distinguons parmi les méthodes de sélection conjointe des variables non significatives, les méthodes de sélection ascendante, descendante et pas à pas mixte. Dans notre cas, la sélection sera faite à l'aide de la méthode de sélection descendante. Le processus de sélection est réalisé par la fonction step() sous R.

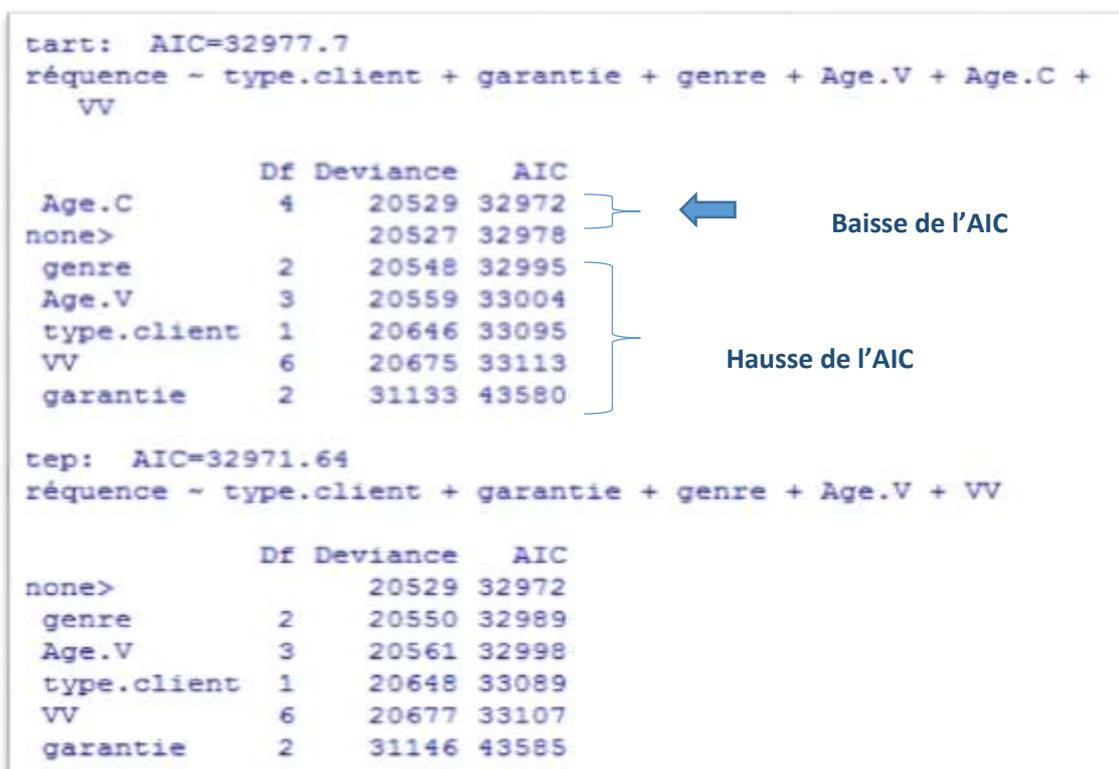


Figure 21: processus de sélection des variables

La variable Age.C a été sélectionnée comme une variable non significative puisque son retrait a baissé le critère d'AIC d'une manière significative. Cela, en fait, coïncide avec l'étude de l'influence des variables explicatives sur la sinistralité où la p-value associée à cette variable était relativement grande par rapport à celles des autres variables.

Le modèle final après sélection des variables non significatives figure ci-dessous :

```

Deviance Residuals:
  Min       1Q   Median       3Q      Max
-1.0402  -0.3019  -0.1972  -0.1129   4.6934

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.873314   0.060003  -47.886 < 2e-16 ***
VVVV2         0.021631   0.045492   0.475  0.6344
VVVV3        -0.043257   0.048842  -0.886  0.3758
VVVV4         0.001417   0.051536   0.027  0.9781
VVVV5        -0.025566   0.055301  -0.462  0.6439
VVVV6        -0.311374   0.067329  -4.625 3.75e-06 ***
VVVV7        -0.554191   0.066065  -8.389 < 2e-16 ***
garantieDASC  2.237303   0.034928  64.054 < 2e-16 ***
garantievol-incendie -1.793036  0.083705 -21.421 < 2e-16 ***
genreM       -0.055999   0.039993  -1.400  0.1614
genreND      -0.156493   0.038673  -4.047 5.20e-05 ***
Age.VAV2     -0.060128   0.035702  -1.684  0.0922 .
Age.VAV3     -0.161318   0.031807  -5.072 3.94e-07 ***
Age.VAV4     -0.229834   0.058209  -3.948 7.87e-05 ***
type.clientSociété -0.435396  0.041414 -10.513 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 32181  on 71666  degrees of freedom
Residual deviance: 20529  on 71652  degrees of freedom
AIC: 32972

Number of Fisher Scoring iterations: 7
    
```

Figure 22: estimation des coefficients du modèle de fréquence après sélection des variables

Ce modèle nous permet de calculer la fréquence moyenne de chaque classe tarifaire. Prenons l'exemple de la variable « **genre** ». La modalité de référence est le genre féminin. Avec un lien logarithmique. Les femmes ont une fréquence moyenne d'avoir un sinistre plus élevée que celle des hommes. Par ailleurs, l'intercepte correspond la fréquence moyenne d'un

assuré ayant comme caractéristiques toutes les modalités choisies comme référence dans la construction des modèles.

### 1.4. Validation du modèle

La validation du modèle doit être guidée par la théorie économique. Cela revient à vérifier le signe des coefficients estimés ainsi que leurs ordres de grandeur en examinant les écarts entre les valeurs théoriques (obtenues avec le modèle) et les valeurs observées. Puis, il convient de contrôler la légitimité de ce modèle par sa déviance.

#### 1.4.1. Signe des coefficients :

Pour contrôler les signes des coefficients du modèle, nous allons vérifier la position des coefficients des modalités de la variable  $j$  par rapport au coefficient de sa modalité de référence. En effet, comme nous avons vu, à l'étape de l'estimation, le coefficient de la modalité de référence est fixé à 0 par construction. Avec un lien logarithmique, nous pouvons établir une relation entre le coefficient de la modalité de la variable  $j$  et le coefficient de sa modalité de référence.

$$\left\{ \begin{array}{l} \text{si } \beta > 0 \Rightarrow \exp(\beta) > \exp(0) \\ \text{Sinon} \Rightarrow \exp(\beta) < \exp(0) \end{array} \right.$$

$\beta$  : coefficient de modalité

En référence de ce qui précède et d'après la positivité de la variable à expliquer, nous obtenons la règle suivante pour contrôler le signe des coefficients de notre modèle de fréquence :

Si le coefficient  $\beta_j$  associé à la modalité  $j$  est positif, la fréquence moyenne de cette modalité doit être supérieure à la fréquence moyenne de la modalité de référence correspondante. De même, si ce coefficient est négatif, cela témoigne que la fréquence moyenne de la modalité  $y$  afférente est inférieure à la fréquence moyenne de la modalité de référence correspondante.

Suivant cette règle, les hommes ont une fréquence moyenne inférieure à celle des femmes (modalité de référence) du fait que le coefficient associé est négatif.

En effet, toutes les modalités ayant un coefficient positif dans notre modèle est caractérisée par une fréquence moyenne supérieure à la fréquence moyenne de la modalité de référence à l'exception de la variable valeur du véhicule « **VV** » qui déroge à cette règle.

Pour ne pas fausser la modélisation, nous avons jugé utile de l'éliminer de notre modèle de fréquence. Nous affichons dans le tableau ci-après la fréquence moyenne de chaque modalité en comparaison avec la modalité de référence.

Modalités	Coefficients	Fréquence moyenne observée	Fréquence moyenne de référence
<b>M</b>	-0.08329	11%	14%
<b>ND</b>	-0.19613	8%	
<b>DASC</b>	2.22831	35%	4%
<b>vol-incendie</b>	-1.79362	1%	
<b>AV2</b>	-0.01324	10%	11%
<b>AV3</b>	-0.09059	9%	
<b>AV4</b>	-0.13283	3%	
<b>Société</b>	-0.53585	7%	9%

**Tableau 10:** la fréquence moyenne par modalité en comparaison avec la fréquence moyenne de référence.

En ce qui concerne la progression des coefficients, Une comparaison de la fréquence moyenne observée avec la progression des coefficients estimés laisse apparaitre une adéquation quasi-parfaite entre ces derniers. En effet, plus l'âge du véhicule augmente plus la fréquence diminue.

### 1.4.2. Ordre de grandeur des coefficients :

Il s'agit de voir si la moyenne des valeurs théoriques est proche de la moyenne des valeurs observées dans un premier temps, ensuite de relever l'écart entre les valeurs du modèle et les valeurs observées en rapportant la somme des valeurs observées à la somme des valeurs du modèle.

```
summary(predictf)
  Min.   1st Qu.   Median     Mean   3rd Qu.   Max.
.0000234 0.0072749 0.0244772 0.0915624 0.0489544 0.4953335
summary(donnees$fréquence2)
  Min. 1st Qu.  Median     Mean 3rd Qu.   Max.
0.0000 0.0000 0.0000 0.1006 0.0000 7.3000
```

La moyenne des valeurs prédites (0,09) est très proche de la moyenne des valeurs observées du portefeuille (0,10).

### 1.4.3. Déviance du modèle

Si le modèle est en bonne adéquation avec les données, la déviance standardisée doit être proche de la valeur  $n-p$  ( $n$  étant le nombre d'observations et  $p$  variables explicatives).

Le modèle estimé est comparé avec le modèle dit saturé, c'est-à-dire le modèle possédant autant de paramètres que d'observations et estimant donc exactement les données. Cette comparaison est basée sur l'expression de la déviance  $D$  des log-vraisemblances  $l$  et  $l_{\text{sat}}$  :

$$D = -2(l - l_{\text{sat}})$$

Qui est le logarithme du carré du rapport des vraisemblances. Ce rapport remplace ou "généralise" l'usage des sommes de carrées propres au cas gaussien et donc à l'estimation par moindres carrées. On montre qu'asymptotiquement,  $D$  suit une loi du  $\chi^2$  à  $n-p$  degrés de liberté ce qui permet de construire un test de rejet ou d'acceptation du modèle selon que la déviance est jugée significativement ou non importante.

L'approximation<sup>45</sup> de la loi du  $\chi^2$  Peut-être douteuse. En pratique, sachant que l'espérance d'une loi du  $\chi^2$  est son nombre de degrés de liberté et, connaissant les aspects approximatifs

<sup>45</sup> PHILIPPE BESSE, Data mining II. Modélisation Statistique & Apprentissage, Laboratoire de Statistique et Probabilités — UMR CNRS C5583 Université Paul Sabatier — 31062 – Toulouse cedex, page 110.

des tests construits, l'usage est souvent de comparer les statistiques avec le nombre de degrés de liberté. Le modèle peut être jugé satisfaisant pour un rapport  $D/ddl$  plus petit que 1.

Dans notre cas de figure, la déviance standardisée (20677) est toujours inférieure au nombre de degrés de liberté (71658). Ainsi donc, notre modèle est pertinent.

## Section 2 : Modèle pour le coût

### 2.1. Choix du modèle

Dans cette étape, nous faisons recours au critère d'information AIC pour choisir la meilleure distribution théorique en termes d'ajustement aux données. Ce processus sera automatisé à l'aide de la fonction `fitDsitR()` sur R.

	Distribution	AIC	
16	Johnson SU	-5384.036	
3	Generalized normal	-5359.410	←
11	Generalized Trapezoidal	-4949.217	
12	Gamma	-4905.117	←
15	Gumbel	-4873.246	
2	Skewed-normal	-4804.970	
14	Laplace	-4795.274	
5	Scaled/shifted t-	-4760.269	
4	Log-normal	-4759.169	
6	Logistic	-4667.943	
9	Trapezoidal	-4652.428	
13	Cauchy	-4607.652	
1	Normal	-4578.591	
20	Arcsine	-4119.566	
8	Triangular	-3816.689	
10	Curvilinear Trapezoidal	-3809.456	
7	Uniform	-3801.696	
17	Johnson SB	-3797.032	
21	von Mises	-3793.640	
18	3P Weibull	-3791.640	
19	4P Beta	-3789.640	

```

$fit
$fit$Normal
    
```

Figure 23: AIC des modèles pour les lois envisagées

Les deux modèles les plus classiques permettant de modéliser les coûts individuels de sinistre sont :

- Le modèle Gamma sur les coûts individuels  $Y_i$ ,
- le modèle log-normal sur les coûts individuels  $Y_i$ .

Vu que ces deux lois sont fréquemment utilisées dans la modélisation des coûts, nous avons jugé pertinent de nous servir de la loi Gamma pour modéliser nos coûts.

Comme lien, nous avons choisi une fonction de lien logarithmique. Notre choix a été motivé par le fait que ce lien va nous permettre par la suite de construire un modèle multiplicatif et mesurer l'influence de chaque modalité.

## 2.2. Estimation des coefficients de la régression

Une fois le modèle et la fonction de lien choisis, il convient d'estimer les paramètres de ce modèle. Pour rappel, les variables intégrées au modèle initial sont :

Codification	Signification	Nombre de modalités
Age.C	Age du client	5
Age.V	Age du véhicule	4
VV	Valeur du véhicule	7
Garantie	La garantie souscrite	3
Genre	Le genre du client	2
Type.client	Type client	2

**Tableau 11:** variables intégrées au modèle initial de la fréquence de sinistres.

- **Coefficients estimés :**

Le modèle à estimer s'écrit :

$$\log(E[C | X]) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5805  -1.1508  -0.6034   0.0090   6.9030

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.28346    0.12938  71.752 < 2e-16 ***
type.client[T.Société] -0.01306    0.08267  -0.158 0.874518
garantie[T.DASC]    0.81979    0.06900  11.880 < 2e-16 ***
garantie[T.vol-incendie] 0.88477    0.17350   5.099 3.51e-07 ***
VV[T.VV2]         0.30886    0.09078   3.402 0.000673 ***
VV[T.VV3]         0.42906    0.09767   4.393 1.14e-05 ***
VV[T.VV4]         0.64035    0.10305   6.214 5.52e-10 ***
VV[T.VV5]         0.65879    0.11007   5.985 2.29e-09 ***
VV[T.VV6]         0.58291    0.13383   4.356 1.35e-05 ***
VV[T.VV7]         0.93058    0.13041   7.136 1.07e-12 ***
genre[T.M]        0.24066    0.08100   2.971 0.002978 **
genre[T.ND]       0.25985    0.07885   3.296 0.000988 ***
Age.V[T.AV2]      0.11945    0.07132   1.675 0.093989 .
Age.V[T.AV3]     -0.03880    0.06324  -0.614 0.539548
Age.V[T.AV4]     -0.07315    0.11558  -0.633 0.526856
Age.C[T.AC2]     -0.06107    0.07098  -0.861 0.389542
Age.C[T.AC3]     -0.09245    0.07302  -1.266 0.205570
Age.C[T.AC4]     -0.13561    0.07866  -1.724 0.084784 .
Age.C[T.AC5]     -0.03060    0.08368  -0.366 0.714601
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 3.698458)

Null deviance: 9188.5 on 6046 degrees of freedom
Residual deviance: 8166.8 on 6028 degrees of freedom
AIC: 140093

Number of Fisher Scoring iterations: 7
    
```

Figure 24: estimation du modèle initial

### 2.3. Sélection des variables

À présent, nous essaierons de réduire le critère d'information en éliminant les variables moins significatives dans le modèle. Pour ce faire, nous utiliserons la même méthode adoptée pour le modèle de fréquence, à savoir, la sélection ascendante.

```

> set=step(modelcc, dir="backward")
start:  AIC=140093.4
sou.u ~ type.client + garantie + VV + genre + Age.V + Age.C

      Df Deviance   AIC
- Age.C      4  8179.7 140089
- type.client 1  8166.9 140091
- Age.V      3  8185.4 140092
:none>           8166.8 140093
- genre      2  8207.1 140100
- VV         6  8433.3 140153
- garantie   2  8611.2 140210

step:  AIC=140096.8
sou.u ~ type.client + garantie + VV + genre + Age.V
    
```

Figure 25: processus de sélection des variables

Nous remarquons que le retrait des trois premières variables permet la diminution du critère d'information. La fonction **step** procède à la suppression de la variable dont l'AIC est le plus bas. Ce processus est répété jusqu'à ce que le retrait d'une variable supplémentaire entraîne une augmentation significative de l'AIC.

Le processus de sélection a permis la suppression de la variable âge du client «**Age.C**». Tout retrait d'une variable supplémentaire augmente l'AIC.

Le modèle final après sélection des variables non significatives est affiché ci-après :

```

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    9.213971   0.122958  74.936 < 2e-16 ***
type.clientSociété             -0.008411  0.083058  -0.101  0.91934
garantieDASC                   0.822334   0.069379  11.853 < 2e-16 ***
garantievol-incendie          0.882791   0.173966   5.074 4.00e-07 ***
VVVV2                          0.321401   0.091231   3.523 0.00043 ***
VVVV3                          0.440133   0.098179   4.483 7.50e-06 ***
VVVV4                          0.648156   0.103575   6.258 4.17e-10 ***
VVVV5                          0.671297   0.110641   6.067 1.38e-09 ***
VVVV6                          0.598090   0.134458   4.448 8.82e-06 ***
VVVV7                          0.931332   0.131119   7.103 1.36e-12 ***
genreM                          0.246230   0.081418   3.024 0.00250 **
genreND                        0.255395   0.079213   3.224 0.00127 **
Age.VAV2                       0.117016   0.071702   1.632 0.10274
Age.VAV3                       -0.040201  0.063573  -0.632 0.52718
Age.VAV4                       -0.076456  0.116179  -0.658 0.51050
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 3.739255)

Null deviance: 9188.5 on 6046 degrees of freedom
Residual deviance: 8179.7 on 6032 degrees of freedom
AIC: 140097

Number of Fisher Scoring iterations: 7
    
```

**Figure 26:** estimation du modèle après sélection.

Si nous interprétons les paramètres de la régression, nous pouvons dire, en prenant comme exemple l'âge du véhicule, que le coût moyen pour les véhicules âgés entre 0 à 2 ans est caractérisé par une tendance haussière bien qu'il soit en diminution à partir de la 3<sup>ème</sup> année de circulation. Ce constat ayant été relevé dans la partie analyse univariée s'expliquerait soit par le fait que les assurés négligeaient la déclaration des petits sinistres, soit par l'acquisition de l'expérience au fur et à mesure avec le vieillissement du véhicule assuré et par conséquent la diminution de la charge de sinistre supportée par l'assureur. De plus, les hommes ont un coût moyen supérieur à celui des femmes, ce qui contredit nos constatations dans le modèle

de la fréquence. En effet, l'apport des modèles linéaires généralisés (GLM) réside dans la séparation entre l'étude de coût et l'étude de la fréquence.

## 2.4. Validation du modèle

Comme précédemment, nous commençant par vérifier le signe des coefficients estimés ainsi que leurs ordres de grandeur en examinant les écarts entre les valeurs théoriques (obtenues avec le modèle) et les valeurs observées. Puis, il convient de contrôler la légitimité de ce modèle par sa déviance.

### 2.4.1. Signe des coefficients estimés :

Avec un lien logarithmique, nous pouvons établir une relation entre le coefficient de la modalité de la variable  $j$  et le coefficient de sa modalité de référence.

$$\left\{ \begin{array}{l} \text{si } \beta > 0 \Rightarrow \exp(\beta) > \exp(0) \\ \text{Sinon} \Rightarrow \exp(\beta) < \exp(0) \end{array} \right.$$

$\beta$  : coefficient de modalité

D'après la positivité de la variable à expliquer, nous obtenons la règle suivante pour contrôler le signe des coefficients de notre modèle de coût :

Si le coefficient  $\beta_j$  associé à la modalité  $j$  est positif, le coût moyen de cette modalité doit être supérieur au coût moyen de la modalité de référence correspondante. De même, si ce coefficient est négatif, cela témoigne que le coût moyen de la modalité  $y$  afférente est inférieur au coût moyen de la modalité de référence correspondante.

Modalités	Coefficients	Coût moyen observé	Coût moyen de la modalité de référence
M	0.246230	35%	27%
ND	0.255395	37%	
VV2	0.321401	11%	7%
VV3	0.440133	13%	
VV4	0.648156	16%	
VV5	0.671297	16%	
VV6	0.598090	15%	
VV7	0.931332	21%	
DASC	0.822334	43%	17%
vol-incendie	0.882791	40%	
Société	-0.008411	46%	54%
AV2	0.117016	31%	30%
AV3	-0.040201	23%	
AV4	-0.076456	15%	

**Tableau 012:** Le coût moyen par modalité en comparaison avec le coût moyen de référence

La règle s'impose sur toutes les modalités à l'exception de la variable type du client. De ce fait, cette variable sera éliminée de la modélisation. Par ailleurs, nous constatons une adéquation quasi-parfaite entre la progression des coefficients et la progression des coûts moyens observés. En effet plus le coût augmente, plus le coefficient associé à la modalité est grand.

Le tableau ci-après met en évidence l'harmonisation entre la progression des coefficients et la progression des coûts moyens observés.

VV1	0	7%
VV2	0.321401	11%
VV3	0.440133	13%
VV6	0.598090	15%
VV4	0.648156	16%
VV5	0.671297	16%
VV7	0.931332	21%

**Tableau 13:** comparaison entre progression des coefficients du modèle et progression des coûts moyens observés

### 2.4.2. Ordre de grandeur des coefficients :

Il y a lieu de procéder à la vérification des grandeurs des coefficients par la comparaison de la moyenne des valeurs théoriques avec la moyenne des valeurs observées dans un premier temps. Par la suite, nous relèverons l'écart entre les valeurs du modèle et les valeurs observées en rapportant la somme des valeurs observées à la somme des valeurs du modèle.

```
summary(c$cou.u)
  Min.   1st Qu.   Median     Mean   3rd Qu.   Max.
  23.5   9787.6  18180.0  42763.0  35515.9 1725250.0
summary(predict)
  Min. 1st Qu. Median   Mean 3rd Qu.   Max.
 9298 30258 43986 42729 54961 84124
```

Nous observons que la moyenne des valeurs prédites (0,09) est très proche de la moyenne des valeurs observées du portefeuille (0,10).

```
> sum(predict)/sum(c$cou.u)
[1] 0.9992033
```

De plus, l'écart entre les valeurs du modèle et les valeurs observées est minime.

### 2.4.3. Déviance du modèle

Si le modèle est en bonne adéquation avec les données, la déviance standardisée doit être proche de la valeur  $n-p$  ( $n$  étant le nombre d'observations et  $p$  variables explicatives).

Malheureusement, la précision du test est douteuse dès lors que l'échantillon utilisé pour réaliser le test est de petite taille. De ce fait, l'usage est souvent de comparer  $DS$  (déviance standardisée) avec le nombre de degrés de liberté ( $ddl$ ).

Un modèle est jugé suffisant si ce rapport reste inférieur à 1.

Dans notre cas de figure, la déviance standardisée (2192) est toujours inférieure au nombre de degrés de liberté (6033). Ainsi donc, notre modèle est pertinent.

## 2.5. Comparaison tarifaire Homme/Femme

Une fois la fréquence moyenne  $E(N)$  et le coût moyen  $E(C)$  sont estimés, il convient de déduire une estimation du taux de prime. En effet, la combinaison des résultats des deux modèles linéaires généralisés mis en place dans les sections précédentes nous permet de déduire la prime pure pour chaque segment tarifaire. Pour mettre en évidence les résultats du travail, nous verrons à travers le tableau suivant la prime pure de la garantie **bris de glace** (BDG) par classe de valeur du véhicule en croisement avec le genre du conducteur. Le reste des résultats ne sera pas présenté pour des raisons de confidentialité.

PRIME PURE en (DA)	BDG/H	BDG/F
AV1/VV1	630,4665	535,788771
AV1/VV2	869,2754	738,735544
AV1/VV3	979,2725	832,214317
AV1/VV4	1205,332	1024,32588
AV1/VV5	1232,981	1047,82304
AV1/VV6	1311,277	1114,36127
AV1/VV7	1598,003	1358,02946

**Tableau 14** : estimation du taux de prime par valeur du véhicule et genre du conducteur

Nous avons vu dans la modélisation de la fréquence et du coût des sinistres que les femmes sont caractérisées non seulement par une fréquence élevée, mais aussi par un coût moindre par rapport aux hommes. La combinaison de ces deux résultats témoigne que la prime payée par les femmes est inférieure à celle payée par les hommes.

### Conclusion

- les variables explicatives retenues pour notre modèle de coût sont
  - **Pour l'assuré** : son genre et son type
  - **Pour le véhicule** : sa classe d'âge.
- les variables explicatives retenues pour notre modèle de fréquence sont :
  - **Pour l'assuré** : son genre
  - **Pour le véhicule** : sa valeur et sa classe d'âge.

# CONCLUSION

Nous avons vu à travers ce mémoire que la détention du maximum d'informations pertinentes sur le risque à assurer est cruciale pour l'assureur pour pouvoir maîtriser la segmentation et la tarification en assurance automobile. En effet, l'introduction de nouveaux critères dans les tarifs permet de trouver de nouveaux segments rentables et de répondre au jeu de la concurrence.

Tout au long de ce mémoire, nous avons cherché à nous assurer que la tarification adoptée par la CASH assurances est toujours en adéquation avec la sinistralité constatée de ses assurés. Un traitement particulier a été apporté, dans un premier temps, à l'étude des données. Il est en effet important de s'attarder, dans le cadre de la mise en place de modèles statistiques, sur la qualité des données et les liens existant entre nos variables de tarification pour veiller à la qualité de notre modélisation. En effet, l'analyse descriptive, qui est une étape cruciale dans la modélisation du risque, nous a fourni un grand nombre d'intuitions concernant, entre autres, les variables qui ne sont pas prêtes à segmenter notre portefeuille comme la variable zone (zone géographique) du fait que la modalité zone1 représentant les wilayas de la région Nord contient à elle seule 97% des observations de la variable zone. De même, elle nous a permis également de faire état de la situation actuelle du portefeuille et de détecter d'éventuelles erreurs ayant préalablement échappé lors du traitement de la base de données.

Nous avons vu ensuite une méthodologie de la mise en place d'un modèle de tarification en assurance automobile par les modèles linéaires généralisés. D'après les résultats fournis, nous avons vu que l'apport des modèles linéaires généralisés (GLM) réside dans la séparation entre l'étude de coût et l'étude de la fréquence. Le premier modèle mis en place est une régression de la fréquence de sinistres observés sur un historique de sinistralité allant de 2010 à 2016. Dans ce cadre, nous tenons à préciser que la fréquence n'est pas l'observation du nombre d'événements réalisés rapportés au nombre d'observations, mais plutôt le nombre de sinistres divisé par la durée d'exposition qui permet de quantifier un risque donné sur une période donnée. En effet, un individu ayant un sinistre en quatre mois d'exposition présente le même risque qu'un individu ayant trois sinistres en an d'exposition. À l'aide de ce modèle de fréquence, nous avons décelé la qualité des assurés par l'estimation de la fréquence moyenne de chaque classe tarifaire. Le deuxième modèle mis en place est une régression du coût

moyen de sinistres. Nous avons ainsi utilisé plus particulièrement les lois Gamma et Poisson, classiquement employées pour la modélisation de coûts moyens et de fréquences.

L'application de l'approche fréquence-coût nous a permis de dégager certains points importants concernant la tarification du risque automobile et de répartir la charge de sinistre d'une manière plus équitable entre les différents segments relevés suivant leurs caractéristiques et en fonction de risque qui leur est associé. Cependant, nous tenons tout de même à relativiser nos conclusions. En effet, nous rappelons que le portefeuille d'étude comporte 71 669 contrats ayant souscrit au moins une garantie dommages et nous soulignons également que ce nombre d'observations demeure relativement faible et pourrait nous conduire à des résultats moins précis.

Ce n'est qu'une approche de tarification dans l'évolution constante des méthodes de tarification et de l'individualisation du risque automobile dans le souci d'adapter le tarif avec l'évolution du portefeuille et du risque assuré. Cette individualisation du risque vise en premier lieu à fidéliser les bons assurés, car chaque assuré paie pour son risque, et à faire face à la pression de la concurrence par la détection de nouveaux segments rentables en deuxième lieu.

# **BIBLIOGRAPHIE**

## Ouvrages

- ✓ PHILIPPE BESSE, Data mining II. Modélisation Statistique & Apprentissage, Laboratoire de Statistique et Probabilités — UMR CNRS C5583 Université Paul Sabatier — 31062 – Toulouse cedex.
- ✓ B. Barthélémy, Gestion des risques méthode d'optimisation globale, l'organisation novembre 2001.
- ✓ Murielle PAULIN, économie et organisation de l'assurance, Séfi édition, Québec, canada, 2007.
- ✓ François COUILBAULT, Constant ELIASHBERG, Les grands principes de l'assurance, l'argus, 9<sup>ème</sup> édition, 2009.
- ✓ Michel Latrasse, ComstantEliashberg et FrancoisCuilbault, Les grands principes de l'assurance, Edition l'Argus, 5<sup>ème</sup> édition, Paris 2002.
- ✓ Frédéric Malaval, Développement durable, assurance et environnement, Édition Economica, Paris 1999.
- ✓ Jerome Yeatman, Manuel international de l'assurance, Edition Economica, Paris 1998.
- ✓ François Ewald-Jean Hérné Lorenzi, Encyclopédie d'assurance, Economica, 1997.
- ✓ Théodore Corfias, Assurance vie : technique et produits, Édition l'Argus, Paris 2003.
- ✓ COUILBAULT François, Constant ELIASHBERG, LATRASSE Michel, Les grands principes de l'assurance, Edition l'Argus, 3<sup>ème</sup> Édition.
- ✓ YEATMAN Jérôme, manuel international de l'assurance, édition ECONOMICA, 1998.
- ✓ Frédéric Planchet, Pierre Therond et Julien Jacquenin, Modèles financiers en assurance, Edition Economica, 2005.
- ✓ Jean-pierredaniel préface de jacques vandier « l'assurance automobile aujourd'hui », 2003.
- ✓ P.J. Delaporte, Les Mathématiques de l'assurance Automobile, THE ASTIN BULLETIN, VOL. VI, PART (03), Editor H. G. Verbeek, Amsterdam, MAY 1972.
- ✓ Jean Luc Besson et Christian Partrat, Assurance non vie, modélisation et simulation, Édition Economica, Paris 2005.
- ✓ Francis Noël, Francis Noël, Économie et Organisation de l'Assurance, BTS Assurances.
- ✓ Jacques charbonnier « l'assurance du risque automobile » éditions larcier, 2012.

## **Articles et Thèses de recherche**

- ✓ RIAD Meriem Maitre assistante à Université de Tipaza; Doctorante à l'ENSSEA ; Chercheur Laboratoire LASAP «MODELE DE TARIFICATION OPTIMAL EN ASSURANCE AUTOMOBILE DANS LE CADRE D'UN MARCHE REGLEMENTE.
- ✓ AldjiaMezaie, *Tarifification a posteriori en assurance automobile par le principe de maximisation de l'utilité espéré*, mémoire de magister, option Finance et Actuariat, INPS 2006/2007.
- ✓ Olfa N. Ghali, *Un modèle de tarification optimal pour l'assurance automobile dans le cadre d'un marché réglementé : application à la Tunisie*, Cahier de recherche 01-09, Décembre 2001.
- ✓ MériemMaatig, *Modélisation multivariée des comportements à risque des conducteurs d'automobile*, publié dans "42<sup>èmes</sup> Journées de Statistique, 24 juin 2010.
- ✓ Malika CHIKHI, Michel CHAVANCE, Estimation du modèle linéaire généralisé et application, Sciences & Technologie, université Mentouri Algérie A – N°35, Juin 2012, 13-21.
- ✓ DOUANLA Frédéric Hermann, Master Recherche Option Statistique Appliquée, Analyse Statistique des Taux de Prime des Garanties Dommages en assurances automobile
- ✓ ATIA Rachel. Master actuariat de dauphine, Mise en place de modèles de tarification alternatifs face à la suppression réglementaire d'une variable tarifaire en automobile,

## **Rapports et revues**

- ✓ CAAT : Compagnie Algérienne d'Assurance Total.
- ✓ Conseil National des Assurances EDITIONS & DOCUMENTATION, Responsabilité Civile automobile, Tarif, Indice & Barème d'indemnisation, 2001.
- ✓ L'article un (01) de l'ordonnance N° 74-15 du 30 janvier 1974.
- ✓ Mohamed DKHILI, Cours assurance automobile, IFID.
- ✓ T. Ben Jemia, « droit des assurances », IFID.
- ✓ Le management des risques de l'entreprise, Cadre de Référence – Techniques d'application, COSO II, Editions d'Organisation, paru en 2005 est la traduction française de l' « Enterprise Risk Management – Integrated Framework » publié par le Committee of Sponsoring Organisations of the Treadway Commission.

## Sites

- ✓ <http://www.ressources-actuarielles.net/memoires>
- ✓ <http://www.cnpsr.org.dz/>

# ANNEXES