

République Algérienne Démocratique et Populaire

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

Université M'hamed Bougara de BOUMERDES

**Faculté des Sciences
Département des Mathématiques**



Mémoire de fin d'études

**Pour L'obtention du Diplôme de Master en
Mathématiques Financières**

Thème

***Modélisation du coût des sinistres en
assurance automobile***

Réalisé par :

M^{elle} M'ahammedi Elkhansa

M^{elle} Tellache Samira

Soutenue le 17/11/2020 , devant le Jury composé de :

Mr.	MISSOUM.R	ENA	Président
Mr.	ZITOUNI.M	UMBB	Encadreur
Mme.	AIT HOCINE .F	UMBB	Examinatrice

Année Universitaire 2019/2020

DÉDICACE

Je dédie ce modeste travail à :

À mes chers parents : Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être.

À mes chères et adorable sœurs : TASNIME la prunelle de mes yeux,
NOUR, SARAH et AMINA .

À mon très cher frère OUSSAMA.

À ma très chère sœur KHADIDJA, son mari ABDELKADER Et leurs
petites filles.

À Mon grand-père AHMED

À tous les membres de la grande famille M'ahammedi et la grande famille
Gherab petites et grands.

À Toutes mes amies , particulièrement : HALIMA , IBTISSAM , AMIRA
qui m'ont soutenus et fournis leur aide de près ou de loin .

À mon binôme SAMIRA.

À tous les membres de ma promotion.

À toutes personnes qui en contribue de près ou de loin.

El-khansa

DÉDICACE

Je dédie ce travail à :

À ceux qui m'ont donné la vie, je leurs témoigne mon amour et m'a
reconnaissance,

À ma mère qui m'a appris à être patiente pour surmonter les difficultés

À mon défunt père que dieu ait pitié de lui

À mes très chers sœurs ,A mon cher fiancé RABAH

À mon unique et adorable frère LAMINE.

À toute ma famille, A mes collègues, mes amies ,À mon binôme
EL-KHANSA.

À toutes personnes qui me connaît et m'aime de près ou de loin.

Samira

REMERCIEMENTS

« Louange à Allah qui nous a guidés à ceci. Nous n'aurions pas été guidés, si Allah ne nous avait pas guidés »

{Sourate 7. Al Araf verset 43 }

Toute notre gratitude, grâce et remerciement est à Allah le plus puissant qui nous a donnée la force, le courage et la volonté pour ce travail.

Tout d'abord, nous exprimons toutes nos gratitude et nos sincères remerciements à Dr Y.FIRANI Pour ses conseils, et encouragements. Vos qualités humaines et professionnelles sont pour nous un exemple à suivre.

Nous remercions également, le directeur de mémoire Mr. ZITOUNI pour notre encadrement, ses directives et sa disponibilité.

Nos remerciements s'adressent également aux membres du jury, c'est un grand honneur pour nous de vous accueillir d'examiner et discuter notre travail.

Je tiens à remercier également tous mes enseignants pour leurs bonnes orientations et pour leur aide précieuse.

Enfin, j'adresse mes plus sincères remerciements à tous mes proches et amies qui m'ont toujours soutenue et encouragée au cours de la réalisation de ce travail.
Merci à tous pour leurs aides et leurs encouragements.

Merci

Table des matières

LISTE DES ABRÉVIATIONS	7
1 Introduction générale	8
2 Généralités et concepts de base sur les assurances automobile	10
2.1 Concepts généraux d'assurance	11
2.1.1 Définition De L'assurance	11
2.1.2 Les différents types d'assurances	12
2.1.3 le rôle de l'assurance	13
2.1.4 le contrat d'assurance	14
2.2 le contrat d'assurance automobile	15
2.2.1 La définition de l'assurance automobile	15
2.2.2 La souscription du contrat d'assurance automobile	15
2.2.3 formation et durée du contrat	16
2.3 Les garanties d'un contrat d'assurance automobile	17
2.3.1 Les garanties obligatoire	17
2.3.2 les garanties facultatives « non obligatoires » :	17
2.4 Principe de la tarification	20
2.4.1 Définition de la tarification d'une assurance automobile	20
2.4.2 Types d'un tarification d'une assurance automobile	20
2.4.3 Critères de tarification d'une assurance automobile	21
2.4.4 Les facteurs de tarification en assurance automobile	25
3 Outils mathématiques utilisés pour modéliser le coût des sinistres	28
3.1 La prime d'assurance automobile	29
3.1.1 Le calcul de la prime	29
3.1.2 Les composantes d'une prime d'assurance automobile :	30
3.2 Les lois utilisées pour la modélisation des nombres et des coûts des sinistres	31
3.2.1 Les lois utilisées pour la modélisation des nombres des sinistres . .	31
3.2.2 Les lois utilisées pour la modélisation des coûts des sinistres	33
3.3 Modèle linéaire généralisé GLM	35

4	Modélisation du coût des sinistres en assurance automobile	46
4.1	Présentation de la base des données et des variables :	47
4.2	Application du GLM au portefeuille	48
4.2.1	Modélisation de la fréquence des sinistres :	48
4.2.2	Modélisation des montants (coût) des sinistres :	56
5	Conclusion générale	64
	Annexe A : Interface pour le calcul de la prime Total	70
	Annexe B : La base de données	73

Table des figures

3.1	Calcul la prime Totale	29
3.2	Schéma de construction d'un modèle linéaire généralisé	36
3.3	Allures possibles des QQ-plot et interprétation	45
4.1	Calcule le coefficient de pearson dans R.	48
4.2	Principe de la fonction « step » avec une méthode descendante	50
4.3	Eléments de la fonction « summary » d'un GLM	51
4.4	Résultats issus de la fonction « anova ».	52
4.5	Résultats issus de la fonction « drop1 ».	53
4.6	Représentations Graphiques d'analyse des résidus en fonction des valeurs prédites	53
4.7	Représentations Graphiques d'analyse des résidus à une loi normale.	54
4.8	Représentations Graphiques de la racine des résidus	54
4.9	Représentations Graphiques des distances de Cook.	55
4.10	Principe de la fonction « step » avec une méthode descendante	57
4.11	Eléments de la fonction « summary » d'un GLM	58
4.12	Résultats issus de la fonction « anova ».	59
4.13	Résultats issus de la fonction « drop1 »	60
4.14	Représentations graphiques d'analyse des résidus en fonction des valeurs prédites.	61
4.15	Représentations graphiques d'analyse des résidus à une loi normale.	61
4.16	Représentations graphiques de la racine des résidus	62
4.17	Représentations graphiques des distances de Cook	62
5.1	Interface pour le calcul de la prime Total (Source : Trust assurances) . . .	70
5.2	Interface pour le calcul de la prime Total (Source : Trust assurances) . . .	71
5.3	Interface pour le calcul de la prime Total (Source : Trust assurances) . . .	72
5.4	Les données d'un assurance Automobile (Source : https://www.cna.dz) .	73
5.5	Les données d'un assurance Automobile(Source : https://www.cna.dz) . .	74
5.6	Les données d'un assurance Automobile(Source : https://www.cna.dz) . .	75
5.7	Les données d'un assurance Automobile(Source : https://www.cna.dz) . .	76
5.8	Les données d'un assurance Automobile(Source : https://www.cna.dz) . .	77

5.9	Les données d'un assurance Automobile(Source : https://www.cna.dz)	. . . 78
-----	--	----------

Liste des tableaux

2.1	Le bonus accordé aux assurés n'ayant pas été responsable de sinistres en Algérie.	24
2.2	Les malus accordé aux assurés ayant été responsable sinistres en Algérie . .	24
3.1	Lois de la famille exponentielle classiquement utilisée en modélisation linéaire généralisée	38
3.2	Liens canoniques associés aux lois de probabilité usuelles de la famille exponentielle	40
4.1	Résumé du modèle de poisson	56

LISTE DES ABRÉVIATIONS :

GLM	modèles linéaires généralisés.
RC AUTO	responsabilité civile Automobile.
AIC	Critère d'Information d'AKAIKE.
BIC	bayesian information Criterion.
ZIP	Zero Inflated Poisson.
ZINB	Zero Inflated negative Binomial.
FGA	la contribution pour le fond de garantie Automobile.
TVA	Taxe sur la valeur Ajoutée.
DASC	Domage avec sans collision.
DC	Domage collision.

1

Introduction générale

La sinistralité automobile est un problème important pour les assureurs, elle se mesure en termes de fréquence des accidents et de montant de ces accidents. Dans ce marché fortement concurrentiel de l'assurance automobile, qui représente la branche la plus importante de l'assurance non-vie, l'assureur cherche à déterminer des facteurs qui contribuent à expliquer la sinistralité. Ces facteurs lui permettent, en construisant des classes de risque, de segmenter son portefeuille et de hiérarchiser ces classes à l'aide d'indicateurs de sinistralité, comme la prime pure. Cette démarche vise à obtenir une bonne adéquation entre la sinistralité et les primes payées par les assurés.

la tarification est un processus instantané d'évaluation des risques où la prime doit être proportionnelle au risque encouru par les assurés; en assurance automobile on distingue deux types de tarification, une tarification a priori permettant de fixer le prix à la souscription du contrat en fonction des caractéristiques de l'assuré. L'autre type de tarification est celle a posteriori se basant sur la connaissance ultérieure des antécédents en matière de sinistre permettant de corriger la prime a priori.

Nous proposons, dans cette étude, de formaliser un modèle optimal de tarification basé sur les caractéristiques des assurés (un modèle a priori) et sur le nombre d'accidents causés des individus (un modèle a posteriori), afin d'ajuster les primes individuelles selon le degré de risque intrinsèque à travers le temps, de sorte que chaque assuré paye une prime proportionnelle à sa fréquence d'accident et que l'assureur soit financièrement équilibré. Le modèle utilisé s'inspire de deux travaux à ce sujet De là ressort des questions importantes qui se formalise comme suit :

Quels sont les facteurs influents sur la fréquence des sinistres automobile en algerie ?

Quels sont les facteurs influents sur le montant des sinistres automobile en Algérie ?

Quels est la distribution utilisée pour modéliser le nombre des sinistres ?

Quels est la distribution utilisée pour modéliser le montant des sinistres ?

Notre travail se divise en trois chapitres, dans le chapitre 1 nous présenterons les généralités sur l'assurance automobile.

Le chapitre 2 nous allons présenté les lois de probabilité utilisées pour modéliser le cout des sinistres et la méthode GLM .

Dans Le chapitre 3 nous allons présenté les données dont on a appliquée le modèle GLM et enfin Nous allons terminé par une conclusion générale.

2

Généralités et concepts de base sur les assurances automobile

2.1 Concepts généraux d'assurance

2.1.1 Définition De L'assurance

Nous allons tout d'abord définir l'assurance de façon générale, technique, Juridique puis économiques.

a) Définition générale :

D'une manière générale, l'assurance peut être définie 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 »¹.

b) Définition technique :

L'assurance est une opération par laquelle un assureur organisé en mutualité une multitude d'assurés exposés à la réalisation de certains risques, indemnise ceux d'entre eux qui subissent un sinistre grâce à la masse commune des primes collectées, cette mutualisation permet aux assureurs d'équilibrer leur comptes et de prospérer².

c) Définition juridique :

L'article 2 de l'ordonnance n°95-07 du 25 janvier 1995 relative aux assurances définit l'assurance en référence à l'article 619 du code civil en Algérie comme suit :

« L'assurance est 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) Définitions économiques :

« L'assurance c'est la volonté des agents économiques (les individus, mais aussi les entreprises) de se protéger contre les aléas de l'existence, qu'il s'agisse de dommages aux

1. Francois couibault, constant eliasberg, « les grands principes de l'assurance », 2011, 10^e éditions, p57 Yeatman : « Manuel international de l'assurance », éd Economica ; 1998

2. Yeatman : « Manuel international de l'assurance », éd Economica ; 1998, P 17.

biens (automobiles, habitations) ou aux personnes (santé, invalidité, décès), que ceux-ci aient été causés involontairement à autrui (responsabilité civile) ou à soi-même »³

2.1.2 Les différents types d'assurances

Il existe deux grandes catégories d'assurances :

celles qui couvrent une personne physique et celles qui couvrent les biens. Ainsi et en fonction de leur objet et de leur portée, les opérations d'assurances peuvent être classées selon deux grandes familles⁴ :

a) Assurances de dommages :

L'assurance des dommages permet d'obtenir une indemnisation en cas de sinistre. Elle regroupe à la fois la protection de responsabilité (responsabilité civile, responsabilité civile familiale ou responsabilité professionnelle) et celle de biens (dommages causés au véhicule, protection des biens meubles ou immeubles). Elles englobent les deux catégories suivantes :

- **Assurances de biens :** elles garantissent l'indemnisation des préjudices subis par l'assuré suite à des dommages et pertes causés aux biens lui appartenant.
- **Assurances de responsabilité :** Elles garantissent les conséquences pécuniaires de la responsabilité civile que l'assuré peut encourir à raison de dommages corporels, matériels ou immatériels causés aux tiers.

b) Assurances de personnes :

Une assurance de personnes a pour objet de couvrir les risques relatifs aux individus comme les accidents corporels, la maladie, le décès ou encore l'invalidité. L'assurance de personnes peut être souscrite soit à titre individuel soit à titre collectif. Ces assurances peuvent être classées comme suit :

- **Assurances vie :** qui ont pour objet la couverture des risques dont la survenance dépend de la survie ou du décès de l'assuré ;
- **Assurances des accidents corporels et maladie/maternité :** dont l'objet est la couverture des risques portant atteinte à l'intégrité physique de la personne assurée, des risques liés à la maladie ou à la maternité ou des risques d'incapacité et d'invalidité.

3. Pierre PICARD, " ASSURANCE - Économie de l'assurance ", Encyclopædia Universalis : <http://www.universalis.fr/encyclopedie/assurance-economie-de-l-assurance/>

4. source : <https://www.acaps.ma/particuliers>

- **La capitalisation :** qui a pour objet de constituer une épargne avec un rendement minimum garanti.

2.1.3 le rôle de l'assurance

a) Le rôle économique de l'assurance

- L'assurance finance l'économie du pays grâce au fonds collectés par les sociétés d'assurance lui permette d'investir.

Dans l'attente de reverser ces sommes aux assurés sous formes d'indemnisation, les sociétés placent ces fonds sur les marchés financiers en achetant des titres émis par les entreprises en quête de financement. Ces titres sont représentatifs soit de capitaux propres (les actions), soit de dettes à long terme (les obligations). Ces placements contribuent donc au financement des entreprises.

- protection de l'outil productif national et renouvellement des capacités productive ayant subi les dommages.

b) Le rôle social de l'assurance :

- **Fonction réparatrice de l'assurance :** L'assurance permet d'indemniser les préjudices résultants de la réalisation des risques. Elle joue généralement ce rôle dans l'intérêt de l'assuré lui-même car elle lui permet de conserver l'équilibre de son patrimoine.

Mais l'assurance est aussi utilisée pour garantir au tiers la réparation du préjudice dont il est victime. C'est le but essentiel des assurances de responsabilité obligatoire.

- **Fonction d'épargne :** L'assurance vie peut aussi permettre à l'assuré de se constituer un capital ou une rente dans en bénéficiant des avantages de la fiscalité de l'assurance vie cumulés avec ceux liés à la transmission du patrimoine.

- **Fonction de prévention :** Le rôle de l'assureur est aussi d'aider l'assuré à éviter un sinistre en le sensibilisant au risque, en le dissuadant de prendre des risques inconsidérés et incite les assurés à observer un comportement prudent afin d'éviter les sinistres.

- **Rôle de protection sociale :** Les assurances privées jouent un rôle très important pour compléter le rôle de l'état dans le cadre de la protection sociale en matière de santé (complémentaire santé), de retraite (contrat retraite) et de dépendance (contrat dépendance)

2.1.4 le contrat d'assurance

a) définition du contrat d'assurance :

le contrat est une convention par laquelle une entreprise d'assurances s'engage, en cas de réalisation du risque ou au terme fixé au contrat, à fournir à l'assuré une prestation pécuniaire en contrepartie d'une rémunération appelée prime ou cotisation⁵ Définition du contrat d'assurance.

b) les acteurs d'un contrat d'assurance :

Cinq éléments découlent d'une opération d'assurance :⁶.

- **L'assuré** : Personne physique ou morale sur la tête ou sur les intérêts de qui pèse le risque assuré. L'assuré est la personne à laquelle s'appliquent les garanties du contrat d'assurance, sans qu'elle en soit nécessairement le souscripteur.

- **Le souscripteur** : Le souscripteur est la personne qui souscrit un contrat d'assurance, c'est à dire qui signe les différents documents du contrat d'assurance (devis ou proposition d'assurance, questionnaire, conditions particulières) et qui s'engage à payer les primes dues à l'assureur.

Le souscripteur n'est pas obligatoirement l'assuré : il peut souscrire un contrat d'assurance pour son propre compte, ou pour celui d'autres personnes indiquées aux conditions particulières.

- **Le bénéficiaire** : Est une personne physique ou morale qui recevra les prestations promises par l'assureur en cas de la réalisation du risque prévu au contrat d'assurance.

- **Le tiers** : En assurance, le terme « tiers ». est employé pour désigner deux catégories bien spécifiques de personnes :

En responsabilité civile, les tiers représentent l'auteur et la victime de l'accident, dès lors qu'aucun lien contractuel n'est établi entre eux. Un dommage est causé à un tiers par un autre tiers.

En assurance, les tiers représentent les personnes pouvant prétendre aux bénéfices d'un contrat d'assurance sans y être explicitement mentionnées : bénéficiaires d'une garantie décès ou d'une assurance pour compte, victimes, personnes possédant un contrat de capitalisation au porteur frappé d'opposition.

- **L'assureur** : Un assureur est une personne physique ou morale qui s'engage, par un contrat d'assurance, à indemniser un assuré en cas de réalisation d'un risque

5. www.cga.gov.tn

6. Source : <https://www.index-assurance.fr/dictionnaire/assure>

2.2 le contrat d'assurance automobile

2.2.1 La définition de l'assurance automobile

L'assurance automobile englobe deux grandes catégories des garanties.

- La garantie obligatoire : La garantie obligatoire qui est la responsabilité civile du conducteur.
- Les garanties facultatives : Les garanties facultatives qui concernent les dommages du véhicule assuré et d'autres garanties comme les dommages au conducteur, la protection juridique, et l'assistance.

L'assurance auto est obligatoire pour tous les véhicules motorisés. En échange d'une cotisation annuelle payée par l'assurée, appelée aussi prime, l'assurance auto couvrira la responsabilité civile du conducteur, c'est à dire prendra en charge les dommages causés par le véhicule lors d'un accident quel que soit l'objet (piétons, passagers, autre véhicule, éléments de voirie) à l'exception du conducteur et de la voiture elle-même.

2.2.2 La souscription du contrat d'assurance automobile

Il existe deux types de contrats d'assurance automobile :

Les contrats des particuliers et Les contrats « flottes »

a) le contrat particulier : Le contrat pour particulier est destiné pour couvrir un seul véhicule, la distinction de particulier est un peu différent du terme employé en général car dans ce contexte le terme particulier ne veut pas dire uniquement l'usage privé mais aussi que le contrat prend en charge un seul véhicule car il peut exister des contrats flotte pour particuliers (un particulier qui possède plusieurs véhicules).

Contrairement aux contrats flotte, les contrats pour particuliers n'ont pas d'avantage de réduction, d'absence de franchise ou d'absence de bonus-malus⁷.

b) contrat flotte : La flotte est l'ensemble de véhicules à moteur couverts au sein d'une même police automobile, elle est divisée en deux catégories (c.jean.p., 2016)⁸ :

- **Les flottes naturelles :** Sont constituées d'un ensemble de véhicules appartenant ou exploités par un même propriétaire ou entité juridique ayant souscrit un contrat collectif pour la couverture globale de son parc. Ainsi tous les véhicules sont soumis aux mêmes règles tarifaires, les primes de chacun d'eux sont recouvrées en une seule fois et les conditions du contrat sont applicables indistinctement à tous.

7. Source : support de cours de droit des assurances Université de Djilali B, Khemis Miliana.

8. SylvieC.jean.p. "manuel des l'assurance automobile ",5eme éditions ,l'agrur ,2016,paris ,p40

- **Les flottes artificielles :** Correspondent au regroupement « mutualisé » de contrats automobiles couvrant des clients distincts d'un prescripteur ayant les mêmes besoins en termes d'assurance, chacun acquittant la prime relative à son véhicule.

2.2.3 formation et durée du contrat

a) Effet et durée du contrat :

Le présent contrat est parfait dès l'accord des parties et notamment dès sa signature par l'assuré et l'assureur.

Il prend effet le lendemain à midi de paiement de la première cotisation et au plus tôt aux dates et heures fixées aux dispositions particulières.

Les mêmes conditions s'appliquent à tout avenant intervenant au contrat⁹.

Le contrat est conclu pour une durée qui ne saurait être inférieure à une année .

Lorsque le contrat contient une clause de tacite reconduction, il est, à son expiration, reconduit d'année en année, sauf dénonciation par l'une ou l'autre des parties, un mois (01) au moins avant l'échéance annuelle de la prime.

b) Résiliation du contrat :

Le contrat peut être résilié avant sa date d'expiration normale dans les cas suivants¹⁰

- Par le souscripteur de à chaque échéance annuelle de la prime, moyennant, sous réserve que le contrat ait au moins d'une année d'existence, En cas de transfert de propriété du risque assuré, En cas de changement ou de cessation définitive d'activité professionnelle de l'assuré.
- Par l'assureur En cas de non-paiement des primes ou en cas d'aggravation des risques lorsque l'assuré refuse de s'acquitter de la différence de prime réclamée par l'assureur et En cas d'omission ou d'inexactitude dans la déclaration du risque à la souscription ou en cours de contrat lorsque l'assuré refuse de s'acquitter de l'augmentation de prime proposée par l'assureur¹¹.
- Par le souscripteur Si des circonstances aggravantes mentionnées dans le contrat disparaissent, et si l'assureur refuse de réduire la prime en conséquence et si l'assureur concerné par le sinistre résilié un autre contrat du souscripteur après sinistre.
- De plein droit En cas de retrait total de l'agrément de l'assureur et En cas de disparition totale du risque assuré, résultant d'un événement non garanti.

9. Article 17 de l'ordonnance n°95 – 07 du 25 janvier 1995 relative à l'assurance

10. Police d'assurance CAAR "assurance automobile" code 3.1.111 p20 .

11. Article 33 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.

2.3 Les garanties d'un contrat d'assurance automobile

2.3.1 Les garanties obligatoire

a) Assurances responsabilité civile automobiles RC :

Elle concerne le responsabilité civile du conducteur¹². Tout propriétaire d'un véhicule doit, avant de le mettre en circulation, souscrire une assurance couvrant les dommages causés aux tiers par ce véhicule.

Le mot véhicule désigne dans le présent texte, tout véhicule terrestre à moteur ainsi que ses remorques ou semi-remorques et leur chargement. Par remorques et semi- remorques, il faut entendre :

- Les véhicules terrestres construits en vue d'être attelés à un véhicule terrestre à moteur et destinés au transport de personnes ou de choses.
- Tout appareil terrestre attelé à un véhicule terrestre à moteur.
- Tout autre engin pouvant être assimilé, par voie de décret, aux remorques ou semi remorques.

La garantie RC s'exerce en circulation comme à l'arrêt, dans quelque lieu que ce soit ; En cas d'utilisation du véhicule à l'insu de l'assuré (vol, violence), la garantie RC reste acquise à ce dernier et ce, quelques soient les conditions dans lesquelles s'est déroulé l'accident.

b) Autres cas d'assurance et extensions de la responsabilité civile (RC) :

les véhicules qui viennent de l'étranger sont soumis aussi à l'obligation d'assurance en Algérie, cette dernière est dénommée L'assurance frontière. De même pour les véhicules immatriculés en Algérie désirant se rendre dans un pays arabe doivent avoir une assurance de responsabilité civile couvrant leur responsabilité en cas de dommages causés à autrui, cette dernière (carte orange) est délivrée sous certaines conditions.

2.3.2 les garanties facultatives « non obligatoires » :

a) Dommage collision D.C :

La garantie « tierce collision » ou « dommage collision » ne joue que s'il y a heurt avec un tiers identifié. En règle générale, les événements garantis sont :

- le choc avec un véhicule en mouvement ou à l'arrêt appartenant à une personne identifiée.
- le choc avec un piéton identifié.

12. L'article 1er de l'ordonnance 74-15 du 30 janvier 1974 modifiée et complétée par la loi n88 – 31 du 19 juillet 1988 relative à l'obligation d'assurance des véhicules automobiles

- le choc avec un animal dont le propriétaire est identifié.

A contrario, ne sont généralement pas garantis :

- le choc avec un véhicule en stationnement.
- le choc avec un véhicule ou un animal appartenant à un membre de la famille de l'assuré ou de la famille du conducteur.
- les dommages consécutifs à un événement prévu dans le cadre de la garantie contre le vol.

b) Dommages avec ou sans collision « D.A.S.C. »

La garantie « tous risques » couvre les événements de la garantie tierce collision complétés par les événements suivants ¹³

- Le choc avec un corps fixe ou mobile extérieur au véhicule (arbre, pierre, automobile, bicyclette, piéton, animal ...).
- Le renversement du véhicule assuré.
- La chute accidentelle du véhicule assuré (dans un cours d'eau, un étang ...).
- L'inondation imprévisible du véhicule assuré en stationnement résultant de la brusque montée des eaux.
- Les actes de vandalisme ou de malveillance. La mise en jeu de cette garantie est subordonnée à la remise par l'assuré d'un certificat de dépôt de plainte.
- Le terme « tous risques » peut être source de confusion. En effet, cette garantie comporte toujours des exclusions de risques et quasiment toujours des franchises.

c) Les dommages corporels du conducteur « personnes transportés » :

Après un accident garanti, la garantie des dommages corporels du conducteur a pour objet de permettre au conducteur assuré ou à ses proches (en cas de blessures ou de décès consécutifs à cet accident) de percevoir une aide financière immédiate, des indemnités s'il est responsable ou des avances de fonds s'il ne l'est pas.

Variables selon les contrats, les garanties qui peuvent être accordées au titre du contrat ou en option sont, en cas de blessures du conducteur, un capital invalide au montant variable selon le taux d'invalidité permanente subsistant après consolidation et, en cas de décès du conducteur, un capital versé au conjoint de l'assuré.

Un capital est parfois accordé pour les enfants mineurs de l'assuré.

d) L'incendie-explosion :

La garantie incendie-explosion couvre les détériorations du véhicule consécutives à un incendie, c'est-à-dire à une combustion spontanée vive, à la chute de la foudre sur le véhicule ou une explosion autre que celle causée par des explosifs transportés.

13. Source : <http://www.cna.dz>. Le portail s'assurance en Algérie ; Assurance automobile, consulté le 17 aout 2020

e) Le bris de glace :

Cette garantie couvre non seulement le pare-brise, mais aussi la lunette arrière et les glaces latérales du véhicule assuré, par projection de cailloux, de gravillons ou autres corps. Sur option, la garantie peut être étendue au bris du toit ouvrant et des optiques de phares.

La garantie est acquise que le véhicule soit en mouvement ou à l'arrêt.

f) Les dommages naturels :

Les événements garantis couvrent les hautes eaux, inondations, grêle, pour autant qu'ils aient été imprévisibles ainsi que les chutes de pierres, éboulements de rochers, glissements de terrain. Ces garanties sont accordées à l'exclusion de tout autre cataclysme, notamment les tremblements de terre.

g) Le vol :

En cas de vol ou tentative de vol du véhicule assuré, sont couverts au titre de la garantie vol les dommages résultant de sa disparition ou de sa détérioration à l'exclusion des dommages indirects et les frais engagés par l'assuré, légitimement ou avec l'accord de l'assureur, pour sa récupération.

En outre, les garanties complémentaires souscrites par l'assuré peuvent produire leurs effets. sont couverts les pneumatiques, accessoires et pièces de rechange fournies avec le véhicule.

h) La protection juridique :

La garantie de protection juridique propre à l'usage d'un véhicule à moteur est également appelée « défense-recours ». Cette garantie a, le plus souvent, pour objectif de défendre l'assuré ainsi que les personnes ayant la qualité d'assuré, à l'amiable ou devant les tribunaux, suite à un accident susceptible de mettre en jeu la garantie « responsabilité civile »

Cette garantie est automatiquement associée à la garantie « responsabilité civile ».

L'assureur prend en charge les frais provoqués par la défense de l'assuré devant toute juridiction y compris devant les juridictions pénales devant laquelle il serait poursuivi pour des faits liés au véhicule assuré (garantie « défense »)

. En outre, l'assureur s'engage à réclamer, à ses frais, à l'amiable ou par voie judiciaire la réparation des dommages, corporels ou matériels, subis par les personnes assurées à la suite d'un accident de la circulation impliquant le véhicule assuré et causé par un tiers responsable (garantie « recours »).

Les personnes bénéficiaires de la garantie sont :

- L'assuré, le propriétaire du véhicule assuré, le conducteur (ou le gardien) autorisé du véhicule assuré.

- Leurs conjoints, ascendants et descendants.

Toutes les personnes dont l'assureur garantit les responsabilités bénéficient, dans les mêmes conditions, de la garantie "défense" à l'exception des personnes poursuivies pour la conduite en état d'ivresse ou délit de fuite et des personnes utilisant le véhicule de l'assuré sans son accord ou contre son gré.

Le remboursement des amendes, qui constituent une peine que la loi interdit d'assurer, n'est jamais couvert.

i) Assistance automobile :

L'assuré peut bénéficier, au sein de contrat d'assurance automobile, de prestations d'assistance.

Son rôle est d'aider l'automobiliste à résoudre les difficultés rencontrées au cours de ses déplacements avec le véhicule assuré.

2.4 Principe de la tarification

2.4.1 Définition de la tarification d'une assurance automobile

La tarification de l'assurance automobile est une opération très importante dans la mesure où elle aboutit à la détermination de la prime à payer. Les tarifs d'assurance automobile sont libres. Ils sont établis à partir de statistiques qui portent sur le nombre et le coût des accidents. Chaque société d'assurances étudie les caractéristiques de ses propres assurés et procède à des études de marché. Les tarifs varient donc d'un assureur à l'autre. La cotisation d'assurance n'est donc pas la même pour tous les véhicules ni pour tous les assurés. L'assureur calcule la prime qui doit être payée par l'assuré pour se voir garantir un risque.

En Algérie la tarification de la prime d'assurance automobile pour la garantie responsabilité civile est imposée par les pouvoirs publics par contre celle des garanties facultatifs est calculée par la compagnie d'assurance¹⁴.

2.4.2 Types d'une tarification d'une assurance automobile

a) La tarification a priori

La tarification a priori consiste à déterminer une prime de risque et s'appuie sur l'observation de certaines variables relatives au véhicule et au conducteur influençant réellement le risque automobile. Le Ministère de Finance impose la classification suivante :

14. Conditions générales, assurance auto, société nationale d'assurance n01/MF/DGT/DASS/ du 15/03/2010, p 33

- Les critères liés au véhicule : Le genre, la zone de circulation, l'usage, la puissance fiscale.
- Les critères liés à l'assuré lui-même : La Catégorie socio Professionnelle, l'âge de l'assuré, l'ancienneté du permis de conduire, ainsi que le sexe.

b) La tarification a posteriori

Chaque assuré doit payer une prime qui permet de couvrir exactement le risque encouru selon le nombre de sinistres dont il est responsable. Il s'agit du système Bonus- Malus qui peut être défini comme étant « une note » personnelle qui reflète l'historique de conducteur. De ce fait, la préoccupation des compagnies d'assurance est d'avoir une base de données exploitable permettant d'approcher une modélisation satisfaisante de la survenance d'accidents en estimant l'évolution du portefeuille automobile ; de façon à déterminer une prime équitable permettant de couvrir le risque assumé¹⁵

2.4.3 Critères de tarification d'une assurance automobile

a) Définition des critères de tarification d'une assurance automobile

Tarifier un risque automobile, c'est apprécier la probabilité que le véhicule à qui est accordé une garantie soit impliqué dans un accident de la circulation. Pour déterminer cette probabilité, l'assureur se base sur :

- Des études statistiques
- La fréquence et le coût des accidents.

C'est le travail des actuaires qui vont déterminer les différents paramètres à prendre en considération pour tarifier un risque donné, en partant de l'analyse globale de l'ensemble des données statistiques relatives à la circulation automobile.

Ce travail fondamental permet à chaque assureur de déterminer un tarif indispensable à l'équilibre de ses résultats dans la branche « automobile ».

b) les critères de la tarification automobile

Les critères pris en compte pour calculer la prime d'assurance automobile et les éventuelles majorations tarifaires sont librement fixés par l'assureur. En fonction de trois éléments :

- **cotisation de référence** : Dépend de plusieurs critères liés au véhicule tel que sa puissance, son utilisation et le mode de garage. Et d'autre liée au conducteur, le profit

15. Thèse doctorant à l'ENSSEA IAD Meriem .Maitre assistante à Université de Tipaza .Chercheur Laboratoire "LASAPMODELE DE TARIFICATION OPTIMAL EN ASSURANCE AUTOMOBILE DANS LE CADRE D'UN MARCHE REGLEMENTE CAS DE L'ALGERIE" p 215.

de ce dernier est le critère central de la tarification d'une assurance auto. Ses critères sont généralement pris en compte dans les statistiques des assureurs tel que :

- **L'âge :** les jeunes conducteurs dessous de 25 ans causent 2 à 3 fois plus d'accidents que les personnes de plus de 25 ans. Les jeunes devront donc faire face à des prix d'assurances plus élevés que les personnes plus âgées. Les personnes plus âgées (seniors) causent elles aussi plus d'accidents. Le prix de l'assurance automobile sera, pour les personnes âgées aussi plus élevé.

- **Le lieu de résidence :** les compagnies d'assurance tiennent compte de ces facteurs pour définir le prix de l'assurance. Étant donné que plus d'accidents ont lieu en ville, le tarif d'assurance sera plus élevé pour un habitant de la ville que pour un habitant de la campagne.

- **La situation professionnelle et l'expérience du conducteur :** si le véhicule est utilisé à des fins professionnelles, les accidents en tort ou accidents en droit vont avoir un impact sur le tarif assurance. Chaque assureur auto est tenu de délivrer une attestation mentionnant le nombre d'accidents dans lesquels le conducteur a été impliqué les 5 dernières années.

Si certains profils considérés comme à risques statistiquement se voient infliger des surprimes parfois très élevées (conducteur est jeune, que son permis de conduire est récent, qu'il est célibataire de sexe masculin et qu'il a des antécédents), d'une autre sont plutôt valorisés par les assureurs. C'est notamment le cas des femmes ou des pères de familles, tous considérés comme plus prudents et donc moins sujets à des accidents graves aux indemnisations élevées. Beaucoup d'éléments entrent en considération pour le calcul de la prime d'assurance auto, de même que chaque compagnie d'assurance auto établit son propre mode de répartition des risques. Dès lors, il conviendra de comparer les prix qui vous seront proposés.

Liées au profit de l'assuré et à son comportement passé, les personnes qui ont passés leurs permis en apprentissage. Anticipé bénéficient d'une réduction de 50% de la surprime des conducteurs nocifs dès La première année de cotisation.

La surprime est ensuite réduite de moitié la seconde année si le conducteur n'a pas été responsable d'un accident, certains conducteurs peuvent subir une majoration de cotisations de référence. les assureurs sont libres de fixer cette majoration dans la limite des pourcentages suivants pour Assuré responsable d'un accident alors qu'il conduisait sous l'emprise d'un état alcoolique 150%.

il est pour l'assuré responsable d'un accident, ou coupable d'une infraction, ayant entraîné la suspension ou l'annulation du permis de conducteur (suspension de deux à six mois 50% , suspension de plus de six mois 50% et annulation ou plusieurs suspensions de plus de deux mois au cours de la même période annuelle de référence 200%) .

Le pourcentage de délit de fuite après accident est de 100%.

Hors dans le cas de Non déclaration des accidents ou des circonstances aggravantes précitées ou des accidents dont ils ont été responsables au cours des trois années précédentes 100% .

Dans le cas de Fréquence d'accidents anormale par rapport à la fréquence moyenne (il s'agit de trois accidents et plus au cours d'une période d'un an précédant de deux mois l'échéance annuelle du contrat) 50% .

Ces différentes majorations se cumulent sans toutefois pouvoir dépasser 400% . après deux années, les majorations sont supprimées.

• **Le bonus-malus Algérien :** Une partie intégrante du tarif obligatoire au titre de l'assurance. Ce système consiste à accorder des réductions sur la prime responsabilité civile pour les assurés n'ayant pas fait l'objet d'accident ou non responsable d'accidents au cours de la période d'observation , et majorer la prime responsabilité civile pour les assurés dont cette responsabilité est engagée ,totalement ou partiellement dans la survenance d'un sinistre, au cours de la période d'observation.

La période d'observation correspond à deux années précédant la date du renouvellement du contrat. La prime retenue pour l'application du bonus-malus est la prime fixée au tarif de référence et afférente à la garantie responsabilité civile en matière d'assurance automobile.

• **Taux du bonus :** Le bonus est accordé aux assurés n'ayant pas été responsable de sinistres durant la période d'observation¹⁶ comme illustré dans ce tableau :

16. Source : document interne de la CAAR.

Durée cumulée d'assurance durant la période d'observation	Taux du bonus
Durée inférieure à 12 mois	0%
Durée égale ou supérieure à 12 mois et inférieure à 24 mois	25%
Durée égale ou supérieure à 24 mois	35%

Table 2.1 : Le bonus accordé aux assurés n'ayant pas été responsable de sinistres en Algérie.

Les taux du bonus sont déterminés de la manière suivante : Dans une durée inférieure à 12 mois le taux appliqué est de 0%, entre 12 et 24 mois le taux appliqué est de 25% et à partir de 24 mois le taux est de 35%.

• **Taux du malus :** Le malus est accordé aux assurés ayant été responsable de sinistre durant la période d'observation¹⁷ comme illustré dans ce tableau :

Nombre de sinistres survenus	Taux du malus
Assuré n'ayant pas de bonus au titre du contrat précédent	
01 sinistre	50%
02 sinistre	100%
03 sinistre	200%
Assuré ayant un bonus au titre du contrat précédent	
01 sinistre	0%
02 sinistre	50%
03 sinistre	100%
04 sinistre	200%

Table 2.2 : Les malus accordé aux assurés ayant été responsable sinistres en Algérie

Les taux du malus sont déterminés en fonction de deux situations, la première situation est celle des assurés n'ayant pas de bonus au titre du contrat précédent. Dans ce cas le taux du malus appliqué est de 50% pour un sinistre, de 100% pour deux sinistres et de 200% pour trois sinistres et plus pendant la période d'observation. La deuxième situation concerne les assurés ayant un bonus au titre du contrat précédent. Dans ce cas le taux du malus appliqué est de 0% pour un seul sinistre, 50% pour deux sinistres, 100% pour trois sinistres et 200% pour quarts sinistres et plus pendant la période d'observation.

17. document interne de la CAAR 2016

2.4.4 Les facteurs de tarification en assurance automobile

A) La tarification de la RC automobile

La tarification auto dépend de plusieurs critères dont certains sont propres au véhicule et d'autres au comportement du conducteur au volant. La zone géographique dans laquelle circule le véhicule peut également avoir une influence sur la tarification.

A.1) Les critères propres au véhicule Pour calculer la prime, un assureur a besoin d'un ensemble d'informations concernant le véhicule : catégorie, usage, puissance,...

• **L'usage du véhicule** L'idée est de créer des classes homogènes de risques, basées sur le kilométrage parcouru, car plus un véhicule circule, plus il risque d'être impliqué dans un accident. Traditionnellement les assureurs distinguent entre les usages privés (tourisme promenade), semi privés (promenade et trajet domicile travail), et professionnels (représentants du commerce) Les principaux usages retenus sont :

- . Affaires
- . Tourisme
- . Commerce
- . Transport public de marchandises
- . Transport public de voyageurs
- . Auto-écoles
- . Ambulances
- . Taxis
- . Transport du personnel

• **Les autres caractéristiques du véhicule** - La puissance du véhicule :

Généralement exprimée en nombre de chevaux-vapeur.

- La source d'énergie : moteur à essence ou de type diesel.

- Le nombre de places autorisées : chiffre prélevé sur la carte grise du véhicule.

- La valeur du véhicule : pour la garantie dommages facultative.

A.2) La zone géographique La tarification peut être différente selon la zone dans laquelle le véhicule est amené à circuler. L'intensité du trafic routier influe sur la fréquence des accidents. C'est le cas des provinces sahariennes qui bénéficient d'un abattement de 50%.

A.3) Le bonus-malus Le système de bonus-malus a pour but de favoriser la prévention des accidents. A chaque renouvellement, un taux de réduction ou de majoration sera appliqué sur la prime RC, en fonction de la survenance ou non d'un sinistre dont l'assuré est responsable au cours d'une certaine période.

B) La tarification des garanties facultatives pour les mono-véhicules

De façon générale, les principaux critères de tarification de garantie facultative sont :

- L'usage du véhicule
- Les valeurs assurées
- Les franchises

B.1) Les valeurs d'assurance En assurance dommages, la valeur assurée est mentionnée dans le contrat. Selon le principe indemnitaire, le montant de la réparation ne peut pas être supérieur au montant de la valeur du véhicule telle que définie lors de la souscription. Les assureurs auto retiennent l'une des valeurs suivantes :

- La valeur vénale : elle correspond à la valeur de revente du véhicule compte tenu de son prix neuf, de son âge, du kilométrage parcouru et de son état.
- La valeur à neuf : correspond à la valeur du véhicule non pas en fonction de son prix d'achat mais en tenant compte du coût du véhicule équivalent au moment du sinistre sans dépréciation. Cette garantie est, soit accordée en option, soit incluse dans la garantie « tous risques ». Cette option joue pendant une durée limitée : dans la plupart des cas, les six ou douze premiers mois de mise en circulation du véhicule assuré et n'intervient qu'en cas de perte totale.
- La valeur à dire d'expert :
correspond à la valeur de remplacement arrêtée par l'expert.
- La valeur conventionnelle : correspond à la valeur fixée de manière forfaitaire dans le contrat d'assurance.

B.2) Les franchises Une franchise se définit comme la part que supporte l'assuré dans l'indemnisation de son dommage. L'application de la franchise permet de moraliser le risque et de réduire la prime payée par l'assuré. La franchise peut être exprimée en montant (une somme fixe) ou en pourcentage du sinistre.

C) La tarification des flottes automobiles

La tarification est établie à partir des informations fournies par l'assuré. Pour le risque RC obligatoire, la prime est à lecture directe (tableau où les primes sont affichées en fonction de la catégorie du véhicule, de son usage, de sa puissance et de la source d'énergie). Pour les risques facultatifs dommages (vol, incendie, bris de glace,...), la prime des contrats flottes de véhicules tient également compte des statistiques et des caractéristiques de la flotte assurée. Des réductions de primes peuvent être accordées en fonction du nombre de véhicules assurés. Informations usuelles :

- Activité de l'entreprise
- Nature du parc à garantir :

- liste de véhicules en précisant leur genre, marque, puissance, année de première mise en circulation, tonnage, valeur à neuf
- Usage des véhicules et, éventuellement, la nature des marchandises transportées
- Protection contre le vol installée sur les véhicules, s'il y a lieu
- Garanties demandées
- Franchises souhaitées
- Statistiques sur les trois à cinq dernières années, avec, en particulier :
 - L'effectif des véhicules par année
 - Un état des sinistres par année
 - Les garanties et franchises en cours, pendant la période d'observation statistique
 - Les événements ayant influencé la sinistralité pendant cette période

3

Outils mathématiques utilisés pour modéliser le coût des sinistres

3.1 La prime d'assurance automobile

3.1.1 Le calcul de la prime

Le paiement des primes est une obligation pour l'assuré¹.

La prime d'assurance automobile correspond au montant réglé chaque année par le conducteur pour bénéficier des garanties de son assureur, en cas de sinistre.

$$\begin{array}{c} \text{PRIME TOTALE} \\ = \\ \text{PRIME PURE} \\ + \\ \text{COÛT DE POLICE} \\ + \\ \text{TVA} \\ + \\ \text{DROIT DE TIMBRE DE DIMENSION} \\ + \\ \text{DROIT DE TIMBRE GRADUE} \\ + \\ \text{FGA} \end{array}$$

Figure 3.1 : Calcul la prime Totale

a) La prime totale :

La prime réellement payée par l'assuré (ou le souscripteur) est la prime totale, elle est égale à la prime commerciale augmentée des frais accessoires et des taxes.

Les frais accessoires sont une petite somme forfaitaire qui représente la participation de l'assuré dans le coût matériel de l'établissement du contrat (papier, rédaction, tirages informatiques...).

b) La prime pure :

Elle correspond à la part des sinistres de l'assuré gérée au sein de la mutualité. En assurance dommage, la prime pure se calcule en multipliant la fréquence des sinistres par le coût moyen des sinistres.

1. Landel J. et pechinot J. "les assurances automobiles", 2ème édition, l'argus, 1996, paris, p 83.

$$\text{Prime Pure} = \text{Fréquence des sinistres} \times \text{le coût moyen des sinistres}$$

c) **Le coût de police** : Il correspond aux frais engagés pour la réalisation du contrat, c'est un chiffre d'affaires pour la société. Il est de l'ordre de 200 DA (50DA pour le coût d'avenant)

d) **La TVA** :

Taxe sur la valeur ajoutée, elle est de 17% du montant (prime nette + coût de police) et est reversée.

e) **Droit de Timbre de dimension** :

Il est à hauteur de $40DA \times \text{nombre de pages}$ et est reversé.

f) **Droit de Timbre gradué** :

Dédié uniquement à l'assurance automobile et matériel agricole et il est reversé.

g) **FGA** :

C'est la contribution pour le fond de garantie automobile afin d'alimenter ce dernier pour l'accomplissement de sa mission. Le taux est de l'ordre de 3% de (prime nette RC + coût de police) et est reversé au FGA.

3.1.2 Les composantes d'une prime d'assurance automobile :

La prime d'assurance payée par l'assuré se compose de plusieurs parties :

- **La prime pure** : c'est le montant du sinistre moyen auquel devra faire face l'assureur pour le risque. Mathématiquement, la prime pure est égale à l'espérance des pertes.

- **Le chargement de sécurité** : ce montant vient s'ajouter à la prime pure. Il autorise l'assureur de pouvoir résister à la volatilité naturelle des sinistres.

- **Le chargement pour frais de gestion** : ces frais comportent autant les frais de gestion des sinistres que la rémunération des apporteurs (agents généraux ou courtier).

- **Les taxes** : La prime ainsi définie est une prime entièrement technique. Cette prime est modifiée selon la politique commerciale de la compagnie d'assurance.

3.2 Les lois utilisées pour la modélisation des nombres et des coûts des sinistres

3.2.1 Les lois utilisées pour la modélisation des nombres des sinistres

Dans ce paragraphe, nous définirons les lois que nous avons utilisées pour modéliser le nombre de sinistres. Pour chaque loi, nous définirons la loi de probabilité, l'espérance ainsi que la variance. Dans la modélisation des processus de comptage, ici de la fréquence des sinistres, deux sortes de modèle sont couramment mis en ; le modèle de poisson et le modèle binomial négatif. On trouve une littérature abondante sur l'utilisation de ces modèles : GREENE (1996), WOOLDRIDGE (1997), CAMERON et TRIVEDI (1998), WINKILMANN (2000), YAU et al. (2003), YANG et al. (2007). Rappelons les définitions et les propriétés de ces modèles pour bien comprendre par la suite l'emploi des modèles ZIP et ZINB.

A) La loi de Poisson :

La loi de Poisson est une des lois de probabilités classiques discrètes qui s'applique aux événements rares. Cette loi est adaptée à la description du nombre des événements se produisant dans une période de temps fixé et indépendamment du temps passé depuis l'événement précédent. Cette loi est caractérisée par le paramètre λ strictement positif qui est le nombre moyen d'occurrences d'un événement donné.

La loi de Poisson n'est pas limitée aux intervalles temporels et peut cependant être appliquée pour d'autres types d'intervalles comme les intervalles spatiaux.

Si X est une variable aléatoire comptant le nombre d'apparitions des événements rares et qui suit la loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$ qui est le nombre moyen d'occurrences dans un intervalle donné, alors la probabilité d'avoir k occurrences dans cet intervalle est donné par :

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \forall k \in \mathbb{N}.$$

Une loi Poisson de paramètre λ sera notée par la suite $L(X) = P(\lambda)$.

L'espérance d'une variable aléatoire X suivant une loi Poisson $P(\lambda)$ est notée :

$$E[X] = \lambda.$$

La variance d'une variable aléatoire X suivant une loi Poisson $P(\lambda)$ est notée :

$$Var[X] = \lambda.$$

B) La loi binomiale négative :

Une loi alternative à la loi de Poisson est la loi Binomiale Négative connue aussi sous le nom de loi de Polya, Cette loi apparait dans des études des événements qui peuvent se réaliser dans un même laps de temps mais qui n'ont pas la même probabilité de survenance. Cet aspect rend cette loi très utile dans le domaine d'assurance lorsque nous voulons étudier le nombre de sinistres dans un portefeuille avec des risques hétérogènes. Une variable aléatoire X suit une loi Binomiale Négative de paramètres $\nu \in \mathbb{R}_+^*$ et $p \in]0; 1[$ si la loi de probabilité, prenant ses valeurs dans $\mathbb{N}$ est définie par :

$$P(X = k) = \frac{\Gamma(\nu + k)}{\Gamma(\nu) k!} p^\nu (1 - p)^k \quad \text{avec } k \in \mathbb{N} (*)$$

Une loi Binomiale négative de paramètres $(\nu; p)$ sera notée par la suite $L(X) = BN(\nu; p)$. L'espérance d'une variable aléatoire X suivant une loi Binomiale négative $BN(\nu; p)$ est notée :

$$E[X] = \frac{\nu(1 - p)}{p}$$

La variance d'une variable aléatoire X suivant une loi Binomiale négative $BN(\nu; p)$ est notée :

$$Var[X] = \frac{\nu(1 - p)}{p^2}$$

Nous attirons l'attention sur le fait que $V[X] > E[X]$ ce qui veut dire que nous calibrons la loi Binomiale négative pour un échantillon d'observations dont la variance est supérieure à la fréquence moyenne.

C) Les modèles ZIP et ZINB :

Les modèles ZIP (Zero Inflated Poisson) et ZINB (Zero Inflated negative Binomial). Le nombre de sinistres observé est décomposé en produit de deux variables :

$$Y = BY^*$$

Y elle désigne le nombre des sinistres

B est une indicatrice égale à 1 si le sinistre est déclaré et 0 sinon (elle n'est donc pas observable).

Y^* est supposé suivre une loi de Poisson (modèle ZIP) ou binomiale négative (ZINB).

On a donc typiquement des équations du type :

$$\boxed{\begin{aligned} P(Y = 0/Xi) &= qi + (1 - qi)e^{-\lambda} \\ \text{Avec } qi &= \frac{\exp(X'iB)}{1 + \exp(X'iB)} \end{aligned}}$$

Et pour y_i non nulle, on a :

$$\boxed{P(Y = y_i/Xi) = (1 - q_i)e^{-\lambda_i} \frac{\lambda_i^{y_i}}{y_i!}}$$

La probabilité du nombre de sinistres conditionnellement à $bi=1$ est égale à la probabilité, non conditionnelle, de la variable inobservée Y_i^* .

Pour un modèle ZINB, la probabilité est alors donnée, en comparaison avec (*), par :

$$\boxed{\begin{aligned} P(Y = y_i/X_i) &= q_i(1 - \min\{y_i, 1\}) + (1 - q_i) \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1)\Gamma(\nu)} \left(\frac{\nu}{\nu + \lambda_i}\right)^\nu \left(\frac{\lambda_i}{\nu + \lambda_i}\right)^{y_i} \\ \text{Avec } i &= 1, 2, \dots \end{aligned}}$$

3.2.2 Les lois utilisées pour la modélisation des coûts des sinistres

Dans cette partie, nous définissons les lois pour modéliser la sévérité de nos sinistres. Pour chaque loi de probabilité, nous définissons la densité, l'espérance ainsi que la variance.

a) La loi log normale :

Une variable aléatoire continue X suit une loi Log-Normale, de paramètres $\mu \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$

si $L(\ln(X)) = N(\mu, \sigma^2)$ avec $N(\mu, \sigma^2)$ la loi normale d'espérance μ et de variance σ^2 .

Elle admet alors pour densité de probabilité la fonction :

$$\boxed{\begin{aligned} f_X(t) &= \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(t) - \mu)^2\right) \mathbb{I}_{]0;+\infty[}(t) \\ \text{Avec } t &\in \mathbb{R} \quad (1) \end{aligned}}$$

Une loi Log-Normale de paramètres (μ, σ) sera notée par la suite $L(X) = LN(\mu; \sigma)$. L'espérance d'une variable aléatoire X suivant une loi Log-Normale $LN(\mu; \sigma)$ est notée :

$$\boxed{E[x] = e^{\mu + \frac{\sigma^2}{2}} \quad \forall \mu \in \mathbb{R} \text{ et } \forall \sigma > 0}$$

La variance d'une variable aléatoire X suivant une loi Log-Normale $LN(\mu; \sigma)$ est notée :

$$Var[X] = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1) \quad \forall \mu \in \mathbb{R} \text{ et } \forall \sigma > 0$$

b) La loi Gamma :

Une variable aléatoire X suit une loi Gamma de paramètres $\alpha, \lambda \in \mathbb{R}_+^*$ si elle admet pour densité de probabilité la fonction :

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

$$\text{Avec } \Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx, \forall \alpha \in \mathbb{R}_+^*$$

Une loi Gamma de paramètres (α, λ) sera notée par la suite $L(X) = GA(\alpha; \lambda)$.

L'espérance d'une variable aléatoire X suivant une loi Gamma $GA(\alpha; \lambda)$ est notée :

$$E(X) = \frac{\alpha}{\lambda}, \forall \alpha, \lambda \in \mathbb{R}_+^*$$

La variance d'une variable aléatoire X suivant une loi Gamma $GA(\alpha; \lambda)$ est notée :

$$Var(X) = \frac{\alpha}{\lambda^2}, \forall \alpha, \lambda \in \mathbb{R}_+^*$$

c) La loi exponentielle :

Une variable aléatoire continue X suit une loi Exponentielle, de paramètre $\lambda \in \mathbb{R}_+^*$ si elle admet pour densité de probabilité la fonction :

$$f(X) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Une loi Exponentielle de paramètre λ sera notée par la suite $L(X) = E(\lambda)$.

L'espérance d'une variable aléatoire X suivant une loi Exponentielle $E(\lambda)$ est notée :

$$E(X) = \frac{1}{\lambda} \quad \forall \lambda > 0$$

La variance d'une variable aléatoire X suivant une loi Exponentielle $E(\lambda)$ est notée :

$$Var(X) = \frac{1}{\lambda^2} \quad \forall \lambda > 0$$

Remarque : Nous avons $GA(1; \lambda) = E(\lambda)$ comme relation entre la loi Gamma et la loi Exponentielle.

3.3 Modèle linéaire généralisé GLM

Dans les modèles linéaires simples, la variable dépendante y est exprimée par une variable interprétée x , tandis que le multiple est exprimé par plusieurs variables interprétées x_i , alors que les modèles linéaires généralisés sont exprimés par la fonction de lien, pour des raisons d'explication Y nous créons une fonction à relier entre Y et X nous détaillons également cela ci-dessous.

La définition du modèle linéaire généralisé GLM :

Le modèle linéaire généralisé est une extension du modèle linéaire. Il a été introduit par Robert Wedderburn et John Nelder² en 1972.

Un modèle linéaire généralisé a pour but de relier des variables explicatives $X = (X_1, \dots, X_p)$ à une variable à expliquer Y . c'est grâce à le modèle linéaire généralisé que les assureurs déterminent la Prime pure pour chacun des assurés .

2. Hélène Compain, "Analyse du risque de provisionnement non-vie dans le cadre de la réforme Solvabilité II", Mémoire d'Actuariat présenté le 18 novembre 2010 devant l'Université Paris Dauphine et l'Institut des Actuaire, p.27.

La logique sous-jacente à un tel modèle peut alors être résumée à travers le schéma suivant :

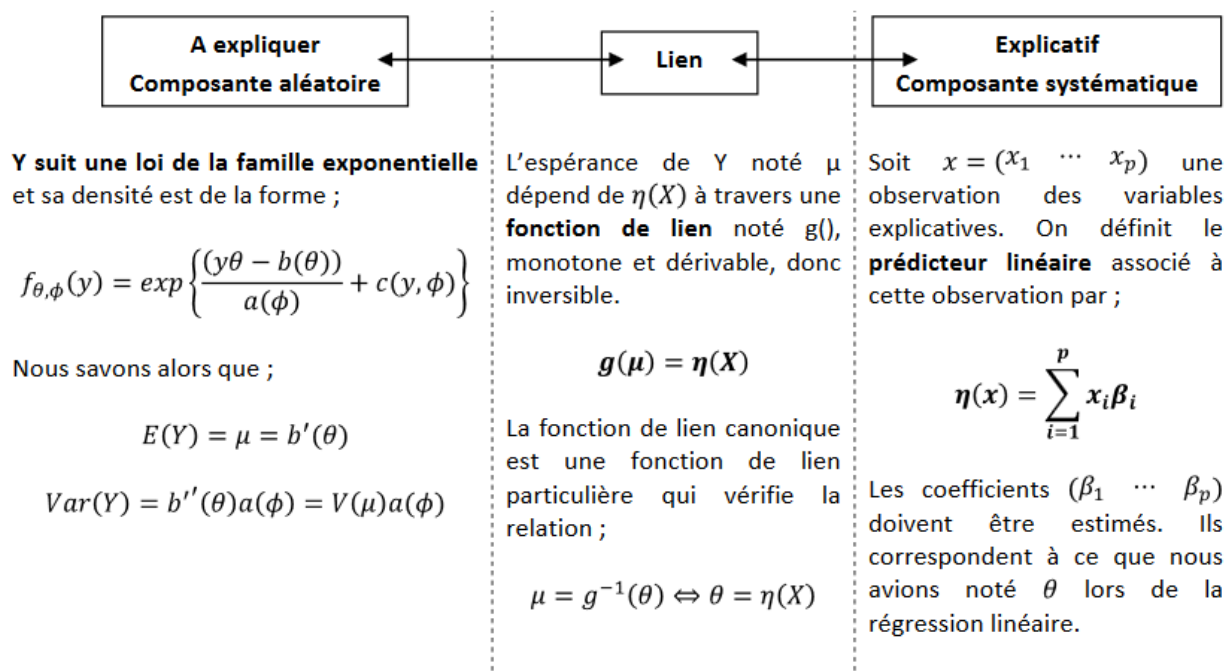


Figure 3.2 : Schéma de construction d'un modèle linéaire généralisé

Pour construire un modèle linéaire généralisé cinq étapes doivent être suivies :

- le choix de la loi de la distribution
- le choix de la fonction de lien
- l'estimation des coefficients de la régression
- la procédure de sélection des variables
- la vérification de la validité du modèle.

Etape 01 : le choix de la loi de la distribution.

il faut donc commencer par choisir la loi de Y dans la famille exponentielle, ce qui fixe les fonctions $a()$, $b()$ et $c()$. Ce choix peut être orienté à partir du tableau de la page 40 (table 3.1)

La loi de la variable aléatoire Y appartient à la famille exponentielle, si sa densité peut s'écrire sous la forme :

$$f(y; \Phi, \theta) = \exp \frac{y \cdot \theta - b(\theta)}{a(\Phi)} + c(y, \Phi)$$

Où :

- $a(\cdot), b(\cdot), c(\cdot)$ sont des fonctions connues et dérivables
- Φ est le paramètre de dispersion
- θ est le paramètre naturel. Il est lié aux deux premiers moments de la loi.

Et :

$$\begin{aligned} E(Y) &= \mu = b'(\theta) \\ Var(Y) &= \sigma^2 = b''(\theta) * a(\Phi) \end{aligned}$$

Exemple : Cas d'une distribution Gamma soit $Y = \Gamma(\alpha, \beta)$ Y admet donc pour densité :

$$f_{\alpha, \beta}(y) = \frac{1}{\alpha^\beta \times \Gamma(\beta)} \cdot y^{\beta-1} e^{-\frac{y}{\alpha}}$$

Que nous pouvons également écrire sous la forme :

$$f_{\alpha, \beta}(y) = \exp[(\beta - 1)y - \frac{y}{\alpha} - \beta \ln \alpha - \ln(\Gamma(\beta))]$$

Nous retrouvons alors les éléments de la famille exponentielle :

$$\begin{aligned} \alpha &: \text{identité} \\ \Phi &= \frac{1}{\beta} \\ \theta &= -\frac{1}{\alpha\beta} \\ b(\theta) &= -\ln(-\theta) \\ c(y, \Phi) &= -\ln(\Gamma(\beta)) + \beta \ln(y\beta) - \ln y \end{aligned}$$

Distribution	θ_i	$b(\theta_i)$	$a(\Phi)$	μ_i
Bernoulli $(1, \pi_i)$	$\ln(\frac{\pi_i}{1 - \pi_i})$	$\ln(1 + \exp(\theta_i))$	1	$\pi_i = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)}$
Binomiale négative $(r, p)(r \text{ cst})$	$\ln(p)$	$r * \ln(1 + \exp(\theta_i))$	1	$\exp(\theta_i) = \frac{r(1 - p)}{p}$
Gamma $\Gamma(V, \frac{v}{\mu_i})$	$-\frac{1}{\mu_i}$	$-\ln(-\theta_i)$	$\frac{1}{v}$	$-\frac{1}{\theta_i} = \mu_i$
Poisson (λ_i)	$\ln(\lambda_i)$	$\exp(\theta_i)$	1	$\exp(\theta_i) = \lambda_i$
Normale $N(\mu_i, \sigma^2)$	μ_i	$\frac{\theta_i^2}{2}$	σ^2	$\theta_i = \mu_i$

Table 3.1 : Lois de la famille exponentielle classiquement utilisée en modélisation linéaire généralisée

Etape 2 : le choix de la fonction de lien

Pour le choix de la distribution, nous utilisons le critère de AIC (Akaike Information Criterion) et BIC (Bayesian information Criterion) qui stipule de choisir la distribution donc le modèle linéaire généralisé a le plus petit AIC et BIC. Ces critères permettent de choisir dans un premier temps le meilleur modèle en termes d'ajustement aux données. Ils permettent également de choisir le meilleur modèle en termes de variables à retenir. Ils font tous deux intervenir dans leur calcul le nombre de paramètres du modèle ainsi que la Log Vraisemblance LL maximisée de ce dernier donnée par :

$$\max_{\beta} LL(y\theta(\beta), \Phi) = \max_{\beta} \ln(\prod_{i=1}^n f(Y_i|\theta, \Phi))$$

Critère AIC : Une méthode de comparaison de modèles s'appelle le critère AIC (Critère d'Information d'AKAIKE) qui s'applique aux modèles estimés par Maximum de Vraisemblance. Le principe consiste à calculer pour chacun des modèles le critère :

$$AIC = -2LL + 2(p + 1)$$

Où :

- LL est la Log Vraisemblance maximisée du modèle ;
 - $p + 1$, correspond au nombre de paramètres du modèle.
- le meilleur modèle est celui qui présente le critère minimal.

Critère BIC Une deuxième méthode de comparaison de modèle s'appelle le critère BIC (Bayesian Information Criterion)

$$BIC = -2LL + (p + 1)\log(n)$$

Où :

- LL est la log vraisemblance des paramètres associée aux données .
- $p+1$ correspond au nombre de paramètres du modèle .
- n est le nombre d'individus composant l'échantillon.

Ce critère pénalise de manière plus importante le nombre de variables présentes dans le modèle que le critère AIC. Le meilleur modèle est celui qui présente le critère BIC minimal.

Le modèle linéaire généralisé se distingue en trois composantes qui sont : la composante aléatoire, la composante déterministe et la fonction de lien :

Composante aléatoire :

La composante aléatoire dans notre cas est la variable à expliquer dont les densités Appartiennent à la famille exponentielle. Dans notre cas les variables à expliquer seront le nombre de sinistres d'un véhicule et le coût d'un sinistre.

Composante déterministe :

C'est l'ensemble des variables explicatives qui nous permettrons de déterminer les variables expliquées (nombre de sinistre et coût d'un sinistre).

La fonction de lien :

C'est une fonction de la forme :

$$g(E[Y_i/X]) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n .$$

la fonction de lien notée g est la liaison entre la composante aléatoire Y et une combinaison linéaire des variable explicatives.

g est une fonction bijective et différentiable.

$$\eta = g(E(Y)) = \sum_{j=1}^p (X_i \beta_j)$$

La fonction de lien canonique définie par $g(.) = (b')^{-1}$ simplifie les calculs théoriques puisque dans ce cas :

$$g(E(Y)) = ((b')^{-1}(b'(\theta))) = \theta$$

Chacune des lois de probabilités de la famille exponentielle possède une fonction de lien spécifique, dite tous « canonique ».

Le tableau(table 3.2 page 42) présent des liens classiques,dans le contexte de notre base de données automobile, nous allons mettre en œuvre la régression de Poisson et nous choisissons la fonction de lien canonique.

Loi de probabilité	Nom de la Fonction	Fonction de lien canonique
Normale	Fonction identité	$g(\mu) = \mu$
Poisson	Fonction logarithme	$g(\mu) = \ln \mu$
Gamma	Fonction inverse	$g(\mu) = \frac{1}{\mu}$
Binomiale	Fonction logit	$g(\mu) = \ln \mu - \ln(1 - \mu)$

Table 3.2 : Liens canoniques associés aux lois de probabilité usuelles de la famille exponentielle

Etape 03 : l'estimation des coefficients de la régression

Les coefficients de la régression $\beta_0, \beta_1, \dots, \beta_p$ ainsi que le paramètre de dispersion Φ sont inconnus et doivent être estimés. On se focalisera uniquement sur l'estimation des coefficients de la régression grâce à la méthode du Maximum de Vraisemblance : il s'agit donc de maximiser la Log-Vraisemblance.

Vraisemblance d'un échantillon La vraisemblance d'un échantillon est la probabilité d'observer cet échantillon. L'échantillon doit être composé d'un groupe ordonné de réalisations y_i de n variables Y_i .

Supposons n variables indépendantes, la probabilité d'observer cet échantillon est le produit des probabilités d'observer chacune des réalisations. La vraisemblance de l'échantillon, notée L , est alors :

$$L(y, \theta, \Phi) = \prod_{i=1}^n f(y_i | \theta, \Phi)$$

Maximisation de la Log Vraisemblance : La Log Vraisemblance LL correspond au log de la fonction de vraisemblance :

$$\begin{aligned} LL(y, \theta(\beta), \Phi) &= \ln \left(\prod_{i=1}^n f(y_i | \theta, \Phi) \right) = \sum_{i=1}^n \ln(f(y_i | \theta, \Phi)) \\ &= \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a(\Phi)} + c(y_i, \Phi) \end{aligned}$$

La recherche des estimateurs de vraisemblance revient à rechercher les $\beta_0, \beta_1, \dots, \beta_p$ qui vérifient les équations :

$$A_j = 0 \text{ pour } j \in \{0, 1, \dots, p\}$$

Où :

$$\begin{aligned} A_j &= \frac{\partial LL(y, \theta(\beta), \Phi)}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{\partial \ln(f(y_i | \theta, \Phi))}{\partial \beta_j} \end{aligned}$$

$$= \sum_{i=1}^n \frac{\partial}{\partial \beta_j} \left(\frac{y_i \theta_i - b(\theta_i)}{a(\Phi)} \right) + c(y_i, \Phi)$$

A_j est déterminé à partir de la formule suivante :

$$\frac{\partial \ln(f(y_i|\theta, \Phi))}{\partial \beta_j} = \frac{\partial \ln(f(y_i|\theta, \Phi))}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}$$

Comme $\mu_i = b(\theta_i)$, il en découle :

$$\frac{\partial \ln(f(y_i|\theta, \Phi))}{\partial \theta_j} = \frac{y_i - b'(\theta_i)}{a(\Phi)} = \frac{y_i - \mu_i}{a(\Phi)}$$

Et

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i)$$

De plus,

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = x_i^j \frac{\partial \mu_i}{\partial \eta_i}$$

Par conséquent,

$$\begin{aligned} \frac{\partial \ln(f(y_i|\theta, \Phi))}{\partial \beta_j} &= \frac{\frac{\partial \ln(f(y_i|\theta, \Phi))}{\partial \theta_i} \frac{\partial \mu_i}{\partial \beta_j}}{\frac{\partial \mu_i}{\partial \theta_i}} \\ &= \frac{(y_i - \mu_i) x_i^j \frac{\partial \mu_i}{\partial \eta_i}}{a(\Phi) b''(\theta_i)} = \frac{(y_i - \mu_i) x_i^j}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \end{aligned}$$

Et finalement :

$$A_j = \sum_{i=1}^n \frac{(y_i - \mu_i) x_i^j \frac{\partial \mu_i}{\partial \eta_i}}{a(\Phi) b''(\theta_i)} = \frac{(y_i - \mu_i) x_i^j}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i}$$

Le paramètre Φ n'apparaît plus dans l'expression de $(A_j)_{0 \leq j \leq p}$. Les équations de vraisemblance peuvent être résolues sans s'en préoccuper.

Les équations obtenues ne sont pas linéaires en β et leur résolution passe par l'utilisation de méthodes itératives. La détermination des coefficients de la régression se fera à l'aide du logiciel R.

Etape 04 : la procédure de sélection des variables trois méthodes sont habituellement utilisées :

• **La procédure Ascendante (Forward Selection)** : Elle consiste à utiliser un modèle linéaire avec seulement une variable explicative et à ajouter au fur et à mesure la variable explicative qui permet de faire diminuer l'AIC du modèle de manière la plus significative. Le processus est alors arrêté, lorsque l'ajout d'une variable supplémentaire entraîne une hausse de l'AIC.

• **La procédure descendante : (Backward Selection)** : Elle consiste, à l'opposé de la procédure Forward, à calculer l'AIC du modèle complet (avec toutes les variables explicatives disponibles) ainsi que l'impact du retrait de chaque variable sur l'AIC. Le processus supprime ainsi du modèle la variable dont le retrait permet une baisse significative de l'AIC.

• **La procédure pas à pas mixte (STEPWISE Selection)** : Cette méthode est une combinaison des deux méthodes précédentes. Les variables jugées les plus significatives pour le modèle sont sélectionnées. A chaque étape, la variable la plus significative restante est ajoutée, puis les variables qui ne sont plus significatives sont retirées. En effet, en raison des corrélations, l'ajout d'une variable explicative peut diminuer la significativité d'une autre variable explicative déjà présente. La procédure Stepwise consiste donc à effectuer une procédure ascendante et descendante à chaque étape.

Etape 05 : validité du modèle (Adéquation du modèle et tests de significativité)
la difficulté dans l'utilisation des modèles linéaires généralisés est de choisir le modèle le plus adéquat. Pour cela, il existe des critères permettant de comparer les modèles entre eux.

• **Déviance :**

Lorsque nous sommes intéressés à la régression linéaire, nous avons effectué des tests d'adéquation du modèle ou de significativité des variables à partir de la somme des carrés résiduels que nous avons noté SCR. En ce qui concerne les modèles linéaires généralisés, nous allons nous baser sur les résultats théoriques issus des tests de rapport de vraisemblances et de Pearson. Pour ce faire, nous allons définir ce que l'on appelle la déviance d'un modèle, ainsi que la statistique de Pearson.

Nous avons précédemment estimé le paramètre β par $\hat{\beta}$ telle que : les $\hat{\beta}$ sont des estimateurs du β .

Ceci nous permet d'exprimer la log-vraisemblance maximisée pour chaque observation, soit en fonction de $\widehat{\theta}$, soit en fonction de $\widehat{\mu}_i$

$$\begin{aligned}\Phi * l(y_i, \widehat{\beta}, \Phi) &= y_i \widehat{\theta} - b(\widehat{\theta}) + cte \\ \Phi * l(y_i, \widehat{\beta}, \Phi) &= y_i (b')^{-1}(\widehat{\mu}_i) - b(b'^{-1}(\widehat{\mu}_i)) + cte\end{aligned}$$

Or nous savons que si le modèle était parfait, la prévision $\widehat{\mu}_i$ par le modèle correspondrait à Y_i (ou à la moyenne des Y_i dans l'hypothèse où l'on possède plusieurs observations telles que $X = X_i$). On parle alors de modèle saturé (autant de paramètres que d'observations distinctes), et nous pouvons en calculer la log-vraisemblance maximisée comme suit (dans le cas où l'on possède une seule observation $X = (X_i)$)

$$\Phi * l_{\text{satur}}(Y_i) = Y_i (b')^{-1}(Y_i) - b((b')^{-1}(Y_i)) + cte$$

Nous définissons alors ce que l'on appelle la déviance, qui mesure l'écart entre la vraisemblance du modèle et celle du modèle saturé correspondant (ayant la même loi) :

$$D = \Phi \sum_{i=1}^n (l_{\text{satur}}(Y_i) - l(y_i, \widehat{\beta}, \Phi)) \geq 0$$

Nous définissons la Déviance standardisé D^* comme suit $D^* = D/\Phi$ Lorsqu'un modèle est adéquat, il paraît logique que ce dernier soit proche du modèle saturé, et donc que la déviance soit proche de 0. Inversement, plus la déviance d'un modèle sera importante, moins le modèle en question sera pertinent. Sous l'hypothèse H_0 la déviance standardisée D^* suit asymptotiquement une loi du Khi^2 à $n - p$ degrés de liberté (avec n le nombre d'observations distinctes et p le nombre de variables du modèle).

Nous pouvons ainsi mettre en place un test de significativité du modèle de niveau α en comparant la déviance standardisée observée avec le quantile d'ordre $1 - \alpha$ d'une loi du Khi^2 à $n - p$ degrés de liberté. Nous considérerons ainsi que le modèle est significatif lorsque la déviance standardisée observée est inférieure à ce quantile.

De plus, lorsque les données sont binaires, la déviance standardisée ne suit pas une loi du Khi^2 , et le test de déviance n'est alors plus valide. Dans cette hypothèse, nous pouvons utiliser le test d'Hosmer-Lemeshow. Ce dernier consiste à ordonner par ordre croissant les $\widehat{\mu}_i$, puis à les scinder en « g » groupes (en général $g = 10$). Il reste alors à mener un test en tout point semblable à celui de la déviance, mais portant sur une statistique suivant approximativement une loi du Khi^2 à g degrés de liberté (validée par simulation)

$$C^2 = \sum_{k=1}^g \left(\frac{(\sum_{i=1}^C k(Y_i) - m_k^* \overline{\mu_k})^2}{m_k^* \overline{\mu_k} (1 - \overline{\mu_k})} \right)$$

Avec m_k^* le nombre d'observations (distinctes) dans le k^{ime} groupe,
 $\overline{\mu_k} = \sum_{i=1}^C k \left(\frac{m_i}{m_k^*} \widehat{\mu}_i \right)$ avec m_i le nombre d'observations
telles que $X = X_i$ dans le k^{ime} groupe.

Nous définissons de plus la statistique de Pearson, parfois nommée le Khi^2 de Pearson généralisé :

$$X^2 = \sum_{i=1}^n \left(\frac{n_i(Y_i - \hat{\mu}_i)^2}{\hat{Y}_i} \right)$$

Pour comparer deux modèles emboîtés, il suffit de comparer leurs déviations. En notant D_1 et D_2 les déviations respectives d'un modèle (1) et d'un sous modèle (2), nous pouvons nous apercevoir que correspond en réalité à un rapport de vraisemblance, la différence de déviance suit asymptotiquement une loi du Khi^2 à degrés de liberté avec le nombre de paramètres du sous modèle et le nombre de paramètres du modèle (Tout se passe comme si le modèle D_2 était le modèle saturé et le modèle D_1 le modèle à tester). Si la différence observée est inférieure au quantile du Khi^2 associé, alors on retiendra le modèle initial. En revanche si la différence observée est supérieure au quantile, alors on préférera le sous modèle au modèle complet. Nous représentons ci-dessous l'expression de ces deux indicateurs :

$$\begin{aligned} \text{AIC} &= -2L + 2P \\ \text{BIC} &= -2L + 2p \log(n) \text{ avec } n \text{ le nombre d'observation} \end{aligned}$$

L : log vraisemblance du modèle p : nombre de paramètres

Utilisation : Le modèle linéaire généralisé est un outil courant en assurance dommage et de nombreux logiciels ont développé des outils facilitant son utilisation (SAS, R, ...).

- **Vérifications graphique :** Nous allons maintenant décrire un ensemble d'outils visuels qui permettent de vérifier que les résidus obtenus possèdent bien toutes les propriétés auxquelles on s'attend.

La nullité de l'espérance des résidus et leur caractère homoscedastique Ces propriétés peuvent être observées sur le tracé de la racine des valeurs absolues des résidus en fonction des prédictions du modèle (« residuals versus fitted »). Le nuage de points doit être réparti aléatoirement. L'absence de tendance et la constance de la variabilité de l'erreur viennent confirmer ces deux hypothèses.

L'adéquation des résidus à une loi Normale : L'adéquation des résidus à une loi normale peut être vérifiée à l'aide d'un QQ-plot (Quantile to Quantile Plot). Ce dernier correspond à la représentation graphique des quantiles des résultats théoriques (en abscisse) en fonction des quantiles des données observées (en ordonnée).

Si l'hypothèse de normalité des résidus est vérifiée, alors les points de la représentation graphique sont approximativement alignés autour de la première bissectrice.

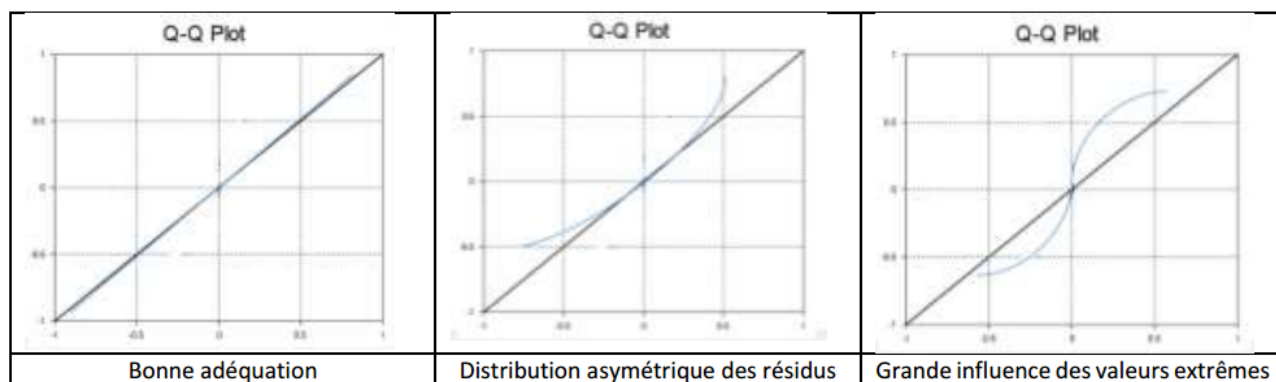


Figure 3.3 : Allures possibles des QQ-plot et interprétation

Mesurer l'influence d'une observation sur la modélisation : distance de Cook

L'un des critères les plus utilisés est la distance de Cook de la i^{me} observation (DC_i) qui peut être exprimée en fonction des résidus Standardisés où p représente le nombre de variables explicatives :

$$DC_i = \frac{1}{p} \frac{h_i}{(1 - h_i)} t_i^2$$

Le seuil de tolérance associé à cette distance de Cook le plus souvent employé est 1. En règle générale, toute observation pour laquelle la distance de Cook est élevée sera retirée de l'étude, ou son influence sur les coefficients surveillée de près.

4

Modélisation du coût des sinistres en assurance automobile

4.1 Présentation de la base des données et des variables :

Notre base de donnée [claimdata](#) port sur le portefeuille d'assurance automobile de l'année 2013, elle contient 119 lignes (les contrats d'assurances) Pour chaque contrat, nous disposons de variables qui caractérisent l'assuré, son contrat ainsi que son véhicule, savoir :

► Puisveh : la puissance de véhicule

► Sex : le genre de l'assuré (M : homme, F : femme)

► Ageveh : l'âge du véhicule sous forme de classes allant de 1 à 4 avec pour correspondance les intitulés suivants :

1. Véhicule très récent
2. Véhicule récent
3. Véhicule ancien
4. Véhicule très ancien

► Agecond : l'âge de l'assuré sous forme de classes allant de 1 à 6 avec pour correspondance les intitulés suivant :

1. Très-jeunes
2. Jeunes
3. Jeunes actifs
4. Actifs
5. Personnes âgées
6. Personnes très âgées

► garantie : garantie enregistré pour chaque assuré :

Garan1 : La sécurité obligatoire comprend toute responsabilité civile en plus de la garantie de défense et de suivi.

Garan2 : Comprend Garan1 en plus d'une garantie contre les dommages par collision.

Garan3 : Comprend toutes les autres possibilités.

► carbrantutil : le carburant utilisé.

1 : essence

2 : Diesel

► USAGEE : le type d'utilisation du véhicule :

► Zone : la zone géographique de résidence de l'assuré (codée de A à F)

Nous disposons également d'informations sur l'historique de sinistralité de l'assuré durant la période d'observation à travers les variables :

Bonus-malus : coefficient de réduction majoration(CRM) qui est entre 0.65 et 2

nb : les nombres de sinistres par police.

sinistre : les montants de sinistres pour chaque police.

Remarque : On considère la durée d'un contrat d'assurance dans une année.

4.2 Application du GLM au portefeuille

Avec les lois et les variables sélectionnées, nous appliquons le GLM sur les données en utilisant le logiciel R avec sa fonction GLM.

La prime pure est ensuite calculée en multipliant le coût moyen par la fréquence moyen.

La prime pure = $E(N) E(X)$ Telle que :

N : nombre de sinistre et X : le cout de sinistre

$E(N)$ c'est la fréquence moyen et $E(X)$ c'est le cout moyen.

Calcule le coefficient de Pearson dans R :

Nous allons cette tableau pour vérifier l'indépendance entre le nombre des sinistres et le cout des sinistres

```
> cor.test(claimdata$sinistre,claimdata$nb)

Pearson's product-moment correlation

data:  claimdata$sinistre and claimdata$nb
t = -0.32545, df = 117, p-value = 0.7454
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2089389  0.1507371
sample estimates:
          cor
-0.03007439
```

Figure 4.1 : Calcule le coefficient de pearson dans R.

la fonction **Cor.test** sur R nous donne Le coefficient de corrélation qui égale (-0.03007439) proche de 0 donc les variables **Sinistre et nb** sont non corrélées (indépendantes) et pour calculer la prime pure , nous allons modélisé le nombre et le montant des sinistres .

4.2.1 Modélisation de la fréquence des sinistres :

La variable que l'on cherche à expliquer est le nombre de sinistres (fréquence), La fonction de lien logarithme étant choisi pour notre modélisation, la formule de la fréquence moyen s'écrit comme :

$$\ln(E(N)) = \beta_0 + \beta_1 \times x_{i,1} + \dots + \beta_p \times x_{i,p}$$

$E(N)$: la fréquence moyen (le nombre moyen de sinistre).

Où :

P : désigne le nombre de variables explicatives sélectionnées pour la fréquence.

β : représente le vecteur des estimations des paramètres du GLM pour chacune des variables sélectionnées pour la fréquence.

$x_{i,p}$: représente les modalités des variables tarifaires pour l'individu i , prenant les valeurs 1 si l'individu présente la caractéristique, 0 si non.

Pour le choix du modèle : Le modèle du nombre de sinistre dans ce mémoire est le modèle de Poisson et la fonction de lien log, de sorte que le modèle binomial négatif et le modèle ZIP et modèle ZINB n'étaient pas efficaces avec les données collectées. Une fois que les données ont été saisies et que les commandes ont été utilisées dans un programme R, une erreur a été signalée et les commandes utilisées étaient les suivantes :

► Pour un modèle binomial négatif :

```
glm.nb(formula=nb~., claimdata=x)
```

► Pour un modèle ZIP :

```
zeroinf(nb~., claimdata=x, na.action=na.omit, dist="poisson")
```

► Pour un modèle ZINB :

```
zeroinf(nb~., claimdata=x, na.action=na.omit, dist="negbin")
```

a). Nous réalisons une modélisation poisson avec toutes les variables explicatives :

```
> View(claimdata)  
> QQ=glm(formula=(claimdata$nb)~claimdata$puisveh+claimdata$sex+claimdata$bonusmalus  
+ claimdata$ageveh+claimdata$agecond+claimdata$garantie+claimdata$carbrantutil+  
claimdata$USAGEE+ claimdata$zone, family=poisson (link = "log"))
```

b). Nous ne gardons que les variables les plus pertinentes à partir de la fonction « step », décrite plus bas, et utilisant ici une méthode descendante (backward qui ne sélectionner que les variables significatives)

```
> QQQ=step(QQ,dir='backward')
```

```
Start:  AIC=383.64
(c.laimdata$nb) ~ c.laimdata$puisveh + c.laimdata$sex + c.laimdata$bonusmalus +
  c.laimdata$ageveh + c.laimdata$agecond + c.laimdata$garantie +
  c.laimdata$carbrantutil + c.laimdata$USAGEE + c.laimdata$zone

- c.laimdata$zone          Df Deviance    AIC
- c.laimdata$USAGEE       3    71.709 378.61
- c.laimdata$ageveh       1    70.802 381.70
- c.laimdata$agecond      1    71.245 382.14
- c.laimdata$garantie     1    71.344 382.24
- c.laimdata$carbrantutil 1    71.640 382.54
- c.laimdata$sex          1    72.392 383.29
<none>                    70.739 383.64
- c.laimdata$bonusmalus   1    74.003 384.90
- c.laimdata$puisveh      1    75.312 386.21

Step:  AIC=375.42
(c.laimdata$nb) ~ c.laimdata$puisveh + c.laimdata$sex + c.laimdata$bonusmalus +
  c.laimdata$ageveh + c.laimdata$agecond + c.laimdata$garantie +
  c.laimdata$carbrantutil + c.laimdata$USAGEE

- c.laimdata$USAGEE       3    73.539 370.44
- c.laimdata$ageveh       1    72.660 373.56
- c.laimdata$agecond      1    73.162 374.06
- c.laimdata$garantie     1    73.270 374.17
- c.laimdata$carbrantutil 1    73.605 374.50
<none>                    72.518 375.42
- c.laimdata$sex          1    74.556 375.45
- c.laimdata$bonusmalus   1    75.478 376.38
- c.laimdata$puisveh      1    76.902 377.80

      .
      .
      .

Step:  AIC=364.52
(c.laimdata$nb) ~ c.laimdata$puisveh + c.laimdata$sex + c.laimdata$bonusmalus

- c.laimdata$sex          1    77.545 364.44
<none>                    75.627 364.52
- c.laimdata$bonusmalus   1    77.864 364.76
- c.laimdata$puisveh      1    79.081 365.98

Step:  AIC=364.44
(c.laimdata$nb) ~ c.laimdata$puisveh + c.laimdata$bonusmalus

- c.laimdata$bonusmalus   1    80.103 365.00
- c.laimdata$puisveh      1    80.873 365.77
```

Figure 4.2 : Principe de la fonction « step » avec une méthode descendante

c). Contrôle de la pertinence du modèle et des variables explicatives, la fonction « summary » :

L'estimation du paramètre β associé à chaque modalité (colonne Estimate)

► L'estimation de l'écart-type du paramètre associé à chaque modalité (colonne Std.error)

► Intercept : le terme constant dans la modélisation de la moyenne de la fréquence.

> summary(QQQ)

```
Call:
glm(formula = (claimdata$nb) ~ claimdata$puisveh + claimdata$bonusmalus,
     family = poisson(link = "log"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0133  -0.6145  -0.3749   0.3237   2.1669

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.63933    0.45106   1.417   0.156
claimdata$puisveh  0.10230    0.05646   1.812   0.070 .
claimdata$bonusmalus -0.59338    0.38435  -1.544   0.123
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 82.912  on 118  degrees of freedom
Residual deviance: 77.545  on 116  degrees of freedom
AIC: 364.44

Number of Fisher Scoring iterations: 5
```

Figure 4.3 : Eléments de la fonction « summary » d'un GLM

les variables expliquées le nombre des sinistres dans le modèle sont : puisveh et bonusmalus

- Le signe négatif des coefficients des variables dans le modèle montre également la relation inverse entre le nombre de sinistre et ses variables
- En ce qui concerne le contrôle de la légitimité du modèle La légitimité du modèle peut être contrôlée par le biais de l'étude de la déviance Standardisée du modèle.

- La déviance de notre modèle (Résiduel déviance), vaut 77.545 et le nombre de degré de liberté 116. Pour obtenir la déviance Standardisée, il faut diviser cette déviance par le paramètre de dispersion, ici $\phi = 1$.

Notre déviance Standardisée (77.545) étant inférieure au nombre de degrés de liberté : on peut donc admettre que le modèle est pertinent.

d). Contrôle de la pertinence du modèle et des variables explicatives, la fonction « anova » :

```
> anova(QQQ, test='Chisq')
```

Analysis of Deviance Table

Model: poisson, link: log

Response: (claimdata\$nb)

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			118	82.912	
claimdata\$puisveh	1	2.8087	117	80.103	0.09376.
claimdata\$bonusmalus	1	2.5583	116	77.545	0.10971

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 4.4 : Résultats issus de la fonction « anova ».

e). Contrôle de la pertinence du modèle et des variables explicatives, la fonction « drop1 ».

```
> drop1(QQQ, test='Chisq')
```

Single term deletions

Model:

(claimdata\$nb) ~ claimdata\$puisveh + claimdata\$bonusmalus

Df Deviance AIC LRT Pr(>Chi)

<none> 77.545 364.44

claimdata\$puisveh 1 80.873 365.77 3.3283 0.0681 .

claimdata\$bonusmalus 1 80.103 365.00 2.5583 0.1097

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 4.5 : Résultats issus de la fonction « drop1 ».

Nous constatons que le coefficient de la variable puisveh est significativement différents de zéro au risque ($\alpha = 10\%$) dans notre régression puisque les probabilités critiques(p-value)[$Pr(> |z|)$] est inférieures à 0.1.

Bien que la variable bonus malus n'ait aucune signification dans ce modèle.

```
> par(mfrow=c(2,2))
> plot(QQQ,which = 1:4,ask=F)
```

L'utilisation de la fonction «plot» appliquée aux modèles linéaires généralisés nous permet d'effectuer une analyse des résidus, et plus particulièrement des résidus de déviance.

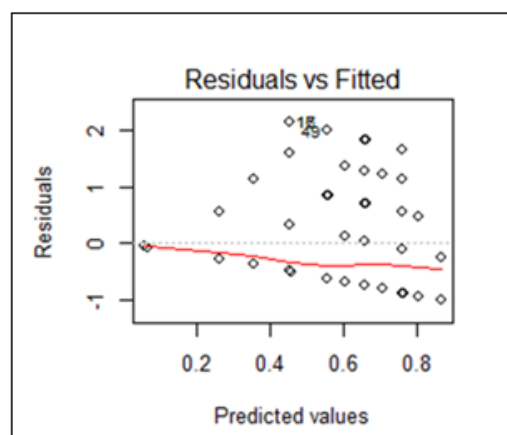


Figure 4.6 : Représentations Graphiques d'analyse des résidus en fonction des valeurs prédites

Le premier graphique (Figure 4.6) est une représentation des résidus en fonction des valeurs prédites, l'absence de tendance significative et l'équidispersion des points autour de l'ordonnée 0 (comme figure 4.6) indique une bonne adéquation du modèle au problème.

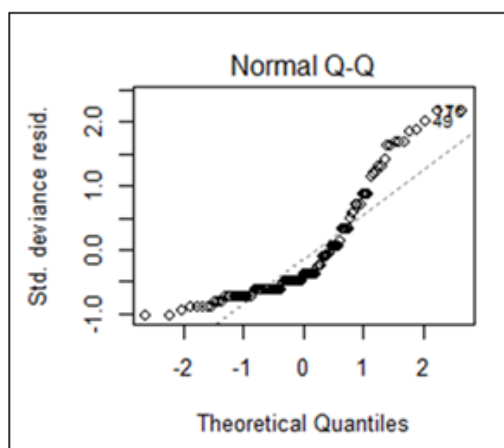


Figure 4.7 : Représentations Graphiques d'analyse des résidus à une loi normale.

Le second graphique(Figure 4.7) permet de contrôler l'adéquation des résidus à une loi normale.

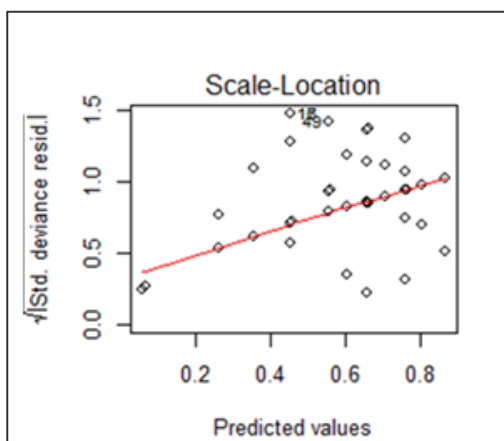


Figure 4.8 : Représentations Graphiques de la racine des résidus

Le troisième graphique(Figure 4.8) est une représentation de la racine des résidus (en valeurs absolues) en fonction des valeurs prédites. Comme pour le premier graphique, l'absence de tendance est la preuve d'une bonne adéquation.

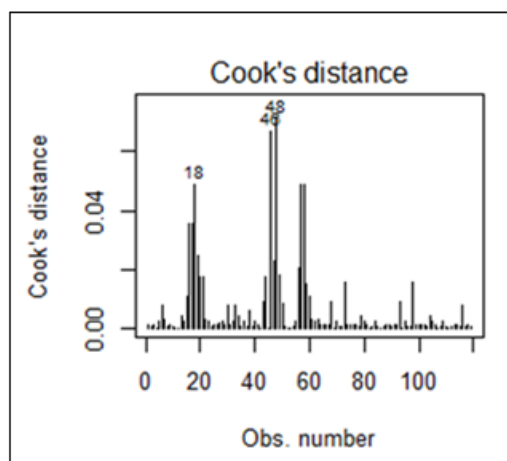


Figure 4.9 : Représentations Graphiques des distances de Cook.

Enfin le dernier graphique (Figure 4.9) est celui des distances de Cook. Comme nous l'avons vu précédemment, une distance supérieure à 1 .

Sera considérée comme anormale, nous remarquons que toutes les observations de l'étude sont convenables ici.

Les variables	Intercept	puisveh	sex	bonusmalus	ageveh
Coeff (β_i)	0.63933	0.10230	0	-0.59338	0

Les variables	agecond	garantie	carbrantutil	USAGEE	zone
Coeff (β_i)	0	0	0	0	0

Table 4.1 : Résumé du modèle de poisson

La formule du nombre moyen donc comme suit :

$$\text{Log}(\mathbf{E}(\mathbf{N})) = \beta_0 + \beta_1 \times [\text{puisveh}] + \beta_2 \times [\text{bonusmalus}]$$

D'où :

$$\mathbf{E}(\mathbf{N}) = \exp(0.63933 + 0.10230 \times [\text{puisveh}] + (-0.59338) \times [\text{bonusmalus}])$$

Avec $\mathbf{E}(\mathbf{N})$ le nombre moyen des sinistres .

4.2.2 Modélisation des montants (coût) des sinistres :

La variable que l'on cherche à expliquer est le cout de sinistres, La fonction de lien inverse étant choisi pour notre modélisation, la formule de le cout moyen s'écrit comme :

$$\mathbf{E}(\mathbf{X})^{-1} = \alpha_0 + \alpha_1 \times x_{i,1} + \dots + \alpha_q \times x_{i,q}$$

Où :

q : désigne le nombre de variables explicatives sélectionnées pour le coût.

α : représente le vecteur des estimations des paramètres du GLM pour chacune des variables sélectionnées pour le coût.

$X_{i,p}$: représente les modalités des variables tarifaires pour l'individu i , prenant les valeurs 1 si l'individu présente la caractéristique, 0 sinon.

En règle générale, les montants de sinistres sont modélisés à partir d'une loi Gamma :

Nous allons alors pouvoir commencer la modélisation. Suite à une confrontation de la fonction de lien log et la fonction de lien inverse pour le modèle complet (avec toutes les variables), nous décidons de choisir comme fonction de lien la fonction inverse correspondant à un AIC le plus petit (2740.808 contre 2745.211) et à une déviance plus faible (74.83382 contre 77.3875).

a). Nous réalisons une modélisation Gamma avec toutes les variables explicatives :

```
>QQw=glm(formula=(claimdata$sinistre)~claimdata$puisveh+claimdata$sex+claimdata$bonusmalus+claimdata$ageveh+claimdata$agecond+claimdata$garantie+claimdata$carbrantutil+claimdata$USAGEE+ claimdata$zone,family=Gamma(link="inverse"))
```

b). Nous ne gardons que les variables les plus pertinentes à partir de la fonction « step », décrite plus bas, et utilisant ici une méthode descendante.

```
> QQQw=step (QQw, dir='backward')
```

```
Start:  AIC=2740.81
(claimdata$sinistre) ~ claimdata$puisveh + claimdata$sex + claimdata$bonusmalus +
  claimdata$ageveh + claimdata$agecond + claimdata$garantie +
  claimdata$carbrantutil + claimdata$USAGEE + claimdata$zone
```

	Df	Deviance	AIC
- claimdata\$zone	5	81.234	2738.7
- claimdata\$garantie	1	74.835	2738.8
- claimdata\$ageveh	1	75.341	2739.4
- claimdata\$puisveh	1	75.352	2739.4
- claimdata\$agecond	1	75.357	2739.4
- claimdata\$sex	1	75.709	2739.9
- claimdata\$USAGEE	3	79.272	2740.2
<none>		74.834	2740.8
- claimdata\$bonusmalus	1	76.890	2741.3
- claimdata\$carbrantutil	1	84.845	2751.1

```
Step:  AIC=2741.6
(claimdata$sinistre) ~ claimdata$puisveh + claimdata$sex + claimdata$bonusmalus +
  claimdata$ageveh + claimdata$agecond + claimdata$garantie +
  claimdata$carbrantutil + claimdata$USAGEE
```

Figure 4.10 : Principe de la fonction « step » avec une méthode descendante .

La méthode backward permettant d'obtenir un modèle ne contenant que des variables significatives, c'est-à-dire que chaque variable retenue a un coefficient significatif sur la variable à expliquer , ici le cout. Le but de la sélection de variables est de choisir les variables les plus pertinentes.

c). Contrôle de la pertinence du modèle et des variables explicatives, la fonction « summary »

```
> summary(QQQw)
```

```
call:
glm(formula = (claimdata$sinistre) ~ claimdata$puisveh + claimdata$sex +
    claimdata$bonusmalus + claimdata$ageveh + claimdata$agecond +
    claimdata$garantie + claimdata$carbrantutil + claimdata$USAGEE,
    family = Gamma(link = "inverse"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4451  -0.8508  -0.3427   0.3684   2.2599

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.608e-05  2.452e-05   1.471  0.14407
claimdata$puisveh -1.198e-06  2.064e-06  -0.581  0.56278
claimdata$sex     -3.980e-06  6.855e-06  -0.581  0.56275
claimdata$bonusmalus -1.762e-05  5.997e-06  -2.939  0.00403 **
claimdata$ageveh   1.577e-06  2.659e-06   0.593  0.55431
claimdata$agecond  -2.025e-06  1.873e-06  -1.081  0.28195
claimdata$garantie -2.191e-07  3.198e-06  -0.069  0.94551
claimdata$carbrantutil 2.012e-05  6.785e-06   2.965  0.00372 **
claimdata$USAGEECommerce 7.080e-06  1.342e-05   0.528  0.59889
claimdata$USAGEEFonctionnaire 4.355e-05  4.096e-05   1.063  0.29001
claimdata$USAGEETaxi -6.729e-06  7.365e-06  -0.914  0.36294
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.8974108)

Null deviance: 105.106 on 118 degrees of freedom
Residual deviance: 81.234 on 108 degrees of freedom
AIC: 2741.6

Number of Fisher Scoring iterations: 7
```

Figure 4.11 : Eléments de la fonction « summary » d'un GLM

les variables expliquées le coût des sinistres dans le modèle sont : puisveh, et bonusmalus, le sex, ageveh, agecond, garantie, carbrantutil, et USAGEE.

La déviance de notre modèle (Residual deviance), vaut 81.234, et le nombre de degré de liberté 108. Pour obtenir la déviance Standardisée, il faut diviser cette déviance par le paramètre de dispersion, ici $\phi = 0.8974108$.

Notre déviance Standardisée (90.52) étant inférieure au nombre de degrés de liberté : on peut donc admettre que le modèle est pertinent.

d). Contrôle de la pertinence du modèle et des variables explicatives, la fonction « anova »

```
> anova(QQQw,test='F')
```

```
Analysis of Deviance Table
Model: Gamma, link: inverse
Response: (claimdata$sinistre)
Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL			118	105.106		
claimdata\$puisveh	1	0.9883	117	104.118	1.1013	0.296326
claimdata\$sex	1	0.1901	116	103.928	0.2118	0.646290
claimdata\$bonusmalus	1	7.9832	115	95.945	8.8958	0.003534 **
claimdata\$ageveh	1	0.1112	114	95.834	0.1239	0.725570
claimdata\$agecond	1	2.0423	113	93.791	2.2758	0.134328
claimdata\$garantie	1	0.1075	112	93.684	0.1198	0.729894
claimdata\$carbrantutil	1	9.6948	111	83.989	10.8031	0.001367 **
claimdata\$USAGEE	3	2.7549	108	81.234	1.0233	0.385357

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.12 : Résultats issus de la fonction « anova ».

La fonction « anova » permet de réaliser un test de significativité des variables explicatives. Selon la même logique que précédemment, une p-value (dernière colonne) faible est synonyme de significativité de la variable. Il est à noter que la fonction « anova » tient compte de l'ordre d'introduction des variables (« Terms added sequentially, first to last »). En effet, le premier test confronte le modèle NULL (aucune variable explicative) au modèle avec puisveh comme seule variable, le second confronte ce dernier au modèle avec puisveh et sex et ainsi de suite. Nous constatons ici que le rajout de la variable sex au modèle avec la variable puisveh n'est pas pertinent, ce qui ne signifie pas pour autant que la variable n'est pas significative dans le modèle complet. Dans le but de valider la significativité des variables d'un modèle, nous préférons utiliser un test de significativité indépendant de l'ordre d'introduction des variables, testant le modèle complet contre le

modèle sans une variable. Pour ce faire, nous allons utiliser la fonction « drop1 », plus pertinente ici que la fonction « anova ».

e). Contrôle de la pertinence du modèle et des variables explicatives, la fonction « drop1 ».

```
> drop1(QQQw,test='F')
```

```
Single term deletions

Model:
(claimdata$sinistre) ~ claimdata$puisveh + claimdata$sex + claimdata$bonusmalus +
  claimdata$ageveh + claimdata$agecond + claimdata$garantie +
  claimdata$carbrantutil + claimdata$USAGEE
              Df Deviance    AIC F value    Pr(>F)
<none>                81.234 2741.6
claimdata$puisveh      1    81.539 2739.9   0.4054 0.5256565
claimdata$sex          1    81.558 2740.0   0.4304 0.5131814
claimdata$bonusmalus   1    87.246 2746.3   7.9922 0.0055994 **
claimdata$ageveh       1    81.550 2739.9   0.4197 0.5184587
claimdata$agecond      1    82.293 2740.8   1.4080 0.2379884
claimdata$garantie     1    81.238 2739.6   0.0056 0.9404778
claimdata$carbrantutil  1    91.030 2750.5  13.0232 0.0004678 ***
claimdata$USAGEE       3    83.989 2738.7   1.2209 0.3057235
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.13 : Résultats issus de la fonction « drop1 » .

Nous constatons que le coefficient de la variable bonusmalus et le coefficient de la variable carbrantutil sont significativement différents de zéro au risque ($\alpha = 5\%$) et ($\alpha = 1\%$) dans notre régression puisque les probabilités critiques (p-value) [$Pr(> |z|)$] est inférieures à 0.01.

Bien que les autres variables n'ait aucune signification dans ce modèle.

```
> par(mfrow=c(2,2))
> plot(QQQw,which = 1:4,ask=F)
```

l'utilisation de la fonction « plot » appliquée aux modèles linéaires généralisés nous permet d'effectuer une analyse des résidus, et plus particulièrement des résidus de déviance.

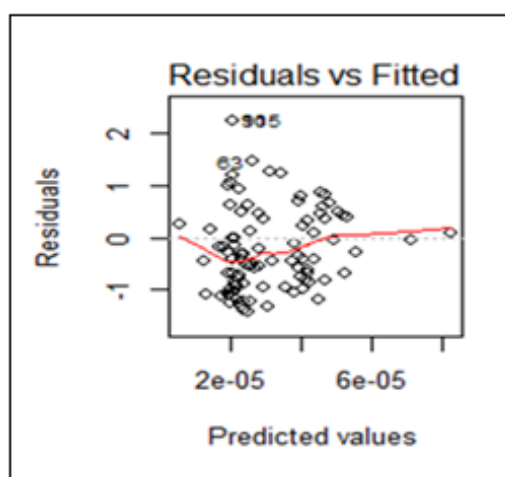


Figure 4.14 : Représentations graphiques d'analyse des résidus en fonction des valeurs prédites.

Le premier graphique (Figure 4.14) est une représentation des résidus en fonction des valeurs prédites, l'absence de tendance significative et l'équidispersion des points autour de l'ordonnée 0 (comme c'est le cas ici) indique une bonne adéquation du modèle au problème.

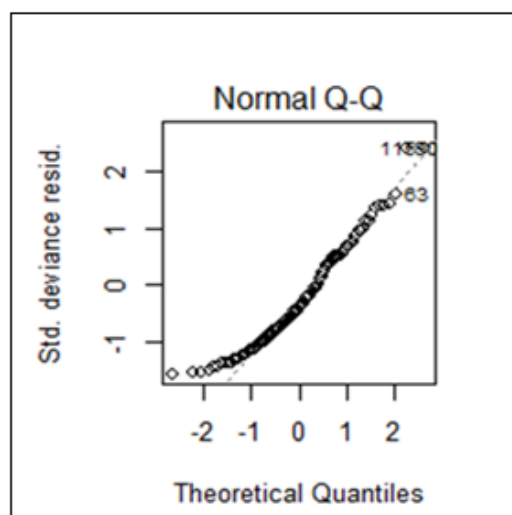


Figure 4.15 : Représentations graphiques d'analyse des résidus à une loi normale.

Le second graphique (Figure 4.15) permet de contrôler l'adéquation des résidus à une loi normale.

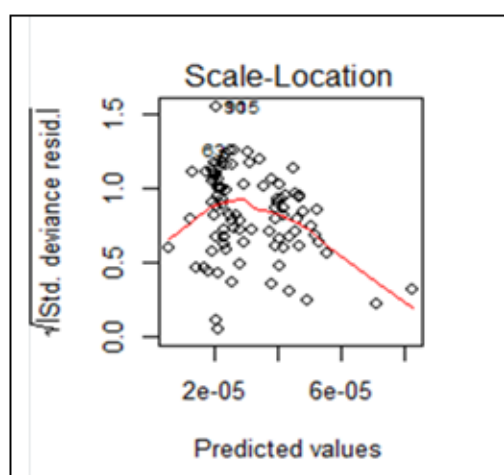


Figure 4.16 : Représentations graphiques de la racine des résidus .

Le troisième graphique (Figure 4.16) est une représentation de la racine des résidus (en valeurs absolues) en fonction des valeurs prédites. Comme pour le premier graphique, l'absence de tendance est la preuve d'une bonne adéquation.

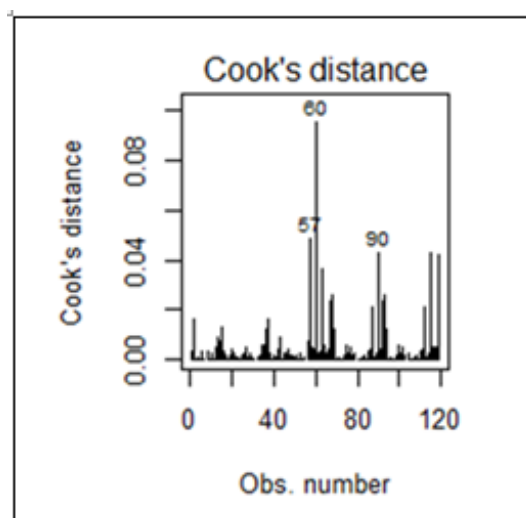


Figure 4.17 : Représentations graphiques des distances de Cook .

Enfin le dernier graphique (Figure 4.17) est celui des distances de Cook. Comme nous l'avons vu précédemment, une distance supérieure à 1 sera considérée comme anormale, nous remarquons que toutes les observations de l'étude sont convenables ici.

La formule du coût moyen donc comme suit :

$$E(X)^{-1} = \alpha_0 + \alpha_1 \times [\text{puisveh}] + \alpha_2 \times [\text{sex}] + \alpha_3 \times [\text{bonusmalus}] \\ + \alpha_4 \times [\text{ageveh}] + \alpha_5 \times [\text{agecond}] + \alpha_6 \times [\text{garantie}] + \alpha_7 \times [\text{carbrantutil}] + \alpha_8 \times [\text{USAGEE}]$$

$E(X)^{-1} = 3.608e^{-5} - 1.198e^{-6} \times [\text{puisveh}] - 3.98e^{-6} \times [\text{sex}] \\ - 1.762e^{-5} \times [\text{bonusmalus}] + 1.577e^{-6} \times [\text{ageveh}] - 2.02598e^{-6} \times [\text{agecond}] \\ - 2.194e^{-7} \times [\text{garantie}] + 2.012e^{-5} \times [\text{carbrantutil}] + 7.08e^{-6} \times [\text{USAGEE}]$
--

Avec **E(X)** le **coût moyen** des sinistres.

5

Conclusion générale

Dans ce mémoire nous avons appliqué la méthode **GLM** (le modèle linéaire généralisé) sur les données d'assurance automobile d'où nous avons établi un modèle pour la fréquence et l'autre pour le coût moyen .

Pour l'établissement d'un modèle, pour le coût moyen nous avons utilisé la loi Gamma et la fonction de lien « **inverse** » et pour le modèle de la fréquence des sinistres Nous avons utilisé la loi de Poisson et la fonction de lien « **Log** » .

Nous avons trouvé que les variables **puisveh**, **bonusmalus**, **le sex**, **ageveh**, **agecond**, **garantie**, **carbrantutil**, et **USAGEE** sont des critères qui influent sur le coût des sinistres (des variables significatives) et les variables **puisveh**, et **bonusmalus** sont des critères influents sur la fréquence des sinistres.

Remarque : Nous n'avons pas abordé la prime du risque à cause de manque des données, elle sera peut être traitée dans une étude ultérieure .

Bibliographie

- [1] **Allain E., Brenac T. (2001)** , « *Modèles linéaires généralisés appliqués à l'étude des nombres d'accidents sur des sites routiers : le modèle de Poisson et ses extensions* » , Recherche Transports Sécurité N°72,p 3-18.
- [2] **Article 17 de l'ordonnance n°95-07** « *de 25 janvier 1995 relative au assurance* »
- [3] **Article 1er de l'ordonnance 74-15** « *du 30 janvier 1974 modifiée et complétée par la loi n°88-31 du 19 juillet 1988 relative à l'obligation d'assurance des véhicules automobiles* »
- [4] **Article 251-2** « *de code des assurances*»
- [5] **Article 33** « *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
- [6] Articles 991 et suivants du Code général des impôts
- [7] **Assurance automobile** (cycle I MST-A) par Agbodo Koffi Joseph (Auteur)
- [8] **BESSON J.L, PARTRAT C. (2004)**, Assurance non-vie. Modélisation, simulation. Economica.
- [9] **BOUCHER, J.Ph., DENUIT, M. and GUILLEN, M. (2007)** Risk classification for claims counts : A comparative analysis of various zero-inflated mixed Poisson and hurdle models. North American Actuarial Journal, 11, 110-131
- [10] **BOUCHER, J.Ph., DENUIT, M. and GUILLEN, M. (2008)** Models of insurance claim counts with time dependence based on generalization of Poisson and Negative Binomial distribution. Variance 2, 135-162.
- [11] **Conditions générales, assurance auto**, société nationale d'assurance n°01/MF/DGT/DASS/ du 15/03/2010, p33.
- [12] **Foukroun.j** garantie responsabilité civile en assurance automobile. Mémoire de post graduation spécialité en actuariat .USTHB ,1999.
- [13] **Francois couibault, constant eliasberg**, « les grands principes de l'assurance »,2011 , 10ème éditions.
- [14] **GUILLOU C. (2011)**, Les risques en IARD et les impacts de Solvabilité II. Mémoire EURIA.

- [15] **J. Yeatman** :” Manuel international de l’assurance” éd Economica, 1998, P 11.
- [16] **JASSON S, TAILLEU F. (2013)**, La réforme du FGAO, quels impacts pour l’assurance automobile
- [17] **Julien Labreuche**, « Les différents types de variables, leurs représentations graphiques et paramètres descriptifs » ;,2011.
- [18] **La loi n° 06-04** du 20 février 2006 modifiée l’ordonnance n° 95-07 du 25 janvier 1995 relative au assurance
- [19] **Lambert D. (1992)**, «Zero-inflated Poisson regression, with an application to defects in manufacturing », Technometrics, vol. 34, p 1-14.
- [20] Landel J.etpechinot J. « les assurances automobiles », 2 ème édition, l’argus, 1996, paris, p 83.
- [21] **Les grands principes de l’assurance auteur** (François Couilbault, Constant Eliashberg) (Auteur)
- [22] **Luc grynbaum « assurances »**, éditions L’ARCUS de l’assurance 2011, p211
- [23] Mémoire de fin de cycle « nouvelle classification de tarification en assurance RC automobile » promotion recherche opérationnelle 2000.
- [24] Nouredine, B. (2009). MODELISATION DE LA FREQUENCE DES SINISTRES EN ASSURANCE AUTOMOBILE. Modélisation De La Fréquence Des Sinistres En Assurance Automobile
- [25] **Patrick rubise**, « l’assurance des risques techniques » 1999, 2 édition p 214.
- [26] Olga A. Vasechko ,Michel Grun-Rehomme et Nouredine Benlagha,”Modélisation de la fréquence des sinistres en assurance automobile”, Bulletin français d’actuariat, Vol 9,N°18, juillet-décembre 2009.
- [27] **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, École des Hautes Études Commerciales (HEC) MontrVal.
- [28] **Police d’assurance CAAR**, « assurance automobile » code 3.1.111.
- [29] **Support de cours de droit des assurances** Université de Djilali BOUNAAMA, Khemis Miliana
- [30] **Sylvie C.jean.p.** « manuel des l’assurance automobile »,5 eme éditions ,l’agrus ,2016,paris
- [31] **Thèse doctorant à l’ENSSEA IAD Meriem** , Maitre assistante à Université de Tipaza Chercheur Laboratoire « LASAPMODELE DE TARIFICATION OPTIMAL EN ASSURANCE AUTOMOBILE DANS LE CADRE D’UN MARCHE REGLEMENTE CAS DE L’ALGERIE ».
- [32] **Wedderburn R.W.M. (1974)**, « Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method », Biometrika 61, p 439-447.

- [33] **Yang Z., Hardin J.W., Addy C.L., Vuong Q.H. (2007)**, «Testing approaches for Overdispersion in Poisson Regression versus the Generalized Poisson Model », *Biometrika*, vol. 49, p 565-584.
- [34] **Yau K.K, Wang K., Lee A.H. (2003)**, «Zero-Inflated Negative Binomial Mixed Regression Modelling of Over-Dispersed Count Data with Extra Zeros », *Biometrika*, vol. 45, p 437-452.
- [35] **éd Economica ;1998**, « Manuel international de l'assurance »,

- [36] **Source** : <https://www.lesfurets.com> , site consulté le 22 mai 2020
- [37] **Source** : <http://www.cna.dz> , site consulté le 20 juin 2020
- [38] **Source** : [http://www.ffa.fr/sites/jcms/p183748/taxes-et-contributions-sur-les-cotisations-d'assurance \$cc = c_51663\$](http://www.ffa.fr/sites/jcms/p183748/taxes-et-contributions-sur-les-cotisations-d'assurance_cc=c51663) , site consulté le 15 Aout 2020
- [39] **Source** : <http://www.vbv.ch/lexikon/> [archive] , site consulté le 20 aout 2020
- [40] **Source** : <https://assurance-auto.ooreka.fr/astuce/voir/96164/les-criteres-d-attribution-de-valeur-pour-une-assura>, site consulté le 20 mai 2020
- [41] **Source** : <https://assurancesplus.blogspot.com> , site consulté le 5 juin 2020
- [42] **Source** : <https://www.assurland.com> , site consulté le 5 juin 2020.
- [43] **Source** : <https://www.compare-assurance.be> , site consulté le 17 juin 2020.
- [44] **Source** : <https://www.index-assurance.fr/dictionnaire/assure> , site consulté le 28 mai 2020.
- [45] **Source** : <https://www.wikifin.be/fr/thematiques/assurer/assurance-vehicules/votre-assurance-auto-en-bref/la-prime-dassurance> , site consulté le 7 juin 2020.
- [46] **Source** : <http://essai.academia.edu/DhaferMalouchedhafer.malouche@me.com> , site consulté le 26 aout 2020.
- [47] **Source** : <http://www.essai.rnu.tn> , site consulté le 26 aout 2020.
- [48] **Source** : <https://aermq.qc.ca/wp-content/uploads/2016/10/facteurs-tarifs-auto> , site consulté le 06 juin 2020 .
- [49] **Source** : <https://www.assurland.com> , site consulté le 20 Juillet 2020.

Annexe A :Interface pour le calcul de la prime Total

Calcul de la prime pure avec logiciel d'assurance automobile :
Pour un assuré ayant le scénario suivant :

The screenshot shows the 'PoliceAvenant - POLICY' v1.3 interface. The left sidebar contains buttons for 'Détails Police', 'Risque assuré', 'Couvertures', 'Réduction Flotte', 'Processus', 'Nouvel Avenant', 'Avis Aliment', 'Aven... Aliment', 'Renouvellement', 'Sinistre', 'Réassurances', 'Versement', 'Convert To Pol', 'Comm Sharing', 'Impression', 'Contr...', 'Quitta...', 'Attestation', 'Détails Flotte', 'Recherche', 'Voir Prime', 'LR Po...', 'Attach', and 'Change Pay.T...'. The main area is divided into two columns. The left column contains fields for 'N° Police' (A1644 20 1128 3379), 'Direction' (2), 'Agence' (A1644), 'Département' (1), 'Assurance' (1), 'Classe' (11), 'Branche' (1128), 'Document' (8), 'Transaction' (1), and 'An./N° Police' (2020 3379). The right column contains fields for 'Date Souscription' (26/10/2020 11:37), 'Org Ref Key', 'Payment' (2), 'Compte' (1), 'Assuré' (100144379), 'Bénéficiaire' (100144379), 'Code' (4), 'Annexe' (0), 'Durée Connue?' (Oui), 'Durée' (12), 'Mois', 'Total', 'Effet' (27/10/2020 11:40), 'Expiration' (26/10/2021 23:59), 'Observation', 'Remarque (NTV)', 'Remarque (FRN)', 'Sharing Commission', and 'Discount Agr Code'. The bottom status bar contains checkboxes for 'Prod. Emise', 'Comm. Posté', 'Traité', 'Facultatif', 'Sinistre', 'Annulé', and 'Tacite recon' (checked).

Figure 5.1 : Interface pour le calcul de la prime Total (Source : Trust assurances)

Risque Assuré - INTEREST

Clé Référence: A1644 20 1128 3379

Transaction: 1 Nouvelle Affaire N°: 1 Active: 1 Oui Inactive Du: Inactive Pour:

Période: 27/10/2020 Heure: 11:40 - 26/10/2021 Heure: 23:59

Wilaya: 35 BOUMERDES N° Permis: 45345

Lieu Date Obtent: 3501 BOUMERD 01/01/2004 Catég. Permis: B Veh < 3.5 Tonnes

Matricule: 045451-112-35

Matricule Final: N° Série: GHGHJGH

N° Type: KJULJ Marque: 1 TOYOTA

Model: 1005 LAND CRUISER Année M.C: 2012

Genre: 00 Véhicules particuliers Usage: 00 Affaire

Puissance CV: 02 5 à 6 Cv Cylindrée: P.T.C (Tonne): 0

C.Utile (Quintaux): 0 Couleur: Zone Circulation: 1 Zone 1

Code Cond.: 1001443795 Nouv... Nom Cond. (NTV): TELACH

Nom Cond. (FRN): TELACH Nb Places: 5

Cond. Couvert? ☒ Turbo? ☐ Alarme? ☐ N° Attestation: 198545

Liquid Inflamm? ☐ Qte: 0 Taux:

Monnaie/Taux: 19 Dinars Algerien 1

Valeur Vénale: 1,400,000.00 ML 1,400,000.00

Valeur Agravée: 1,400,000.00 ML 1,400,000.00

Objet Assuré: Valeur Objet: .00

Chiffre d'aff.: .00 Réduction: 50

Franchise 1: Franchise 2: Franchise 3:

Anc. Accidents: Accordé par:

DaL Accor. Reçu: Remarque:

v2.2

Détails Risque Prime Sinistre Ok Attach Vale De Risque

Figure 5.2 : Interface pour le calcul de la prime Total (Source : Trust assurances)

Calcul Primes Automobile - AUTOPREM

Prime RC Limite Franchise RC Veh RC Rem Prime Total RC

1,400,000.00 2,054.41 2,054.41

Calcul des Primes

	Limites	Franchises	Primes
☒ TERCE - Tous Risques	1,400,000.00	3,000.00	63,000.00
☐ D2			.00
☐ D2C	.00	.00	.00
☐ V - Vol - Incendie	1,400,000.00	.00	14,000.00
☐ D2	5,000.00	.00	.00
☒ PT - B	100,000.00	.00	550.00
☐ Car Glass	100,000.00		
ASS	10,000.00		
Classic	0	0	800.00

☒ Tax Pollution

Total Primes Net 78,350.00

Réductions / Majorations			Taxes et Timbres	
REDUCTION 50 %	50	-38,500.00	FGA	78.63
Malus			TVA	7,962.34
Bonus			Coût Police	500.00
			Timbre Dimensionné	40.00
			Timbre Graduë	1,632.13
			Taxe Pol	1,500.00
Total Rédu. / Maj.			Total Taxes	11,701.10
V4 Prime Total 41,904.41			Total à Payer	53,605.51

Figure 5.3 : Interface pour le calcul de la prime Total (Source : Trust assurances)

Sa prime Total est de valeur : 53,605.51 DA .

Annexe B : La base des données d'assurance Auto

numpolic	Puis veh	sex	Bonus malus	Age veh	nb	sinistre	Age cond	Valeur vehi	garantie	Carbran tutil	USAGEE	zone
1100005424	5	2	0,65	4	5	48 555,00	2	650 000	2	1	Affaire	B
1100014779	6	2	1,5	1	5	192 699,54	4	650 000	2	1	Affaire	A
1100012990	9	1	0,65	3	4	19 200,00	2	1 700 000	2	1	Commerce	E
1100007133	8	2	1	1	1	19 200,00	2	910 000	1	2	Affaire	D
1100006335	4	1	1	2	2	23 302,90	2	750 000	1	2	Transport Public de Marchandises (TPM)	C
1100010216		2	0,75	4	3	11 240,00	4	550 000	2	2	Affaire	C
1100007266	7	2	1	3	1	66 764,78	4	1 000 000	2	2	Affaire	A
1100018819	8	1	1	2	2	39 425,50	6	1 200 000	1	2	Vehicules Speciaux	B
1100000207	30	1	1	2	2	12 659,93	3	1 000 000	1	1	Vehicules Speciaux	A
1100017875	31	2	1	3	1	28 000,93	4	550 000	1	1	Auto-école	A
1100014298	7	2	1	4	3	26 515,00	4	1 422 000	1	1	Commerce C.Bis	B
1100006060	12	1	1	3	3	190 200,00	2	1 464 000	1	1	Commerce	C
1100007414	8	1	1	1	4	21 076,25	2	550 000	1	1	Location	F
1100000196	5	2	1	1	4	13 200,00	6	1 100 000	2	1	Vehicules Speciaux	F
1100014170	31	1	1	3	5	17 200,00	5	1 464 000	2	1	Affaire	E
1100005964	4	2	1	3	2	19 200,00	2	550 000	2	1	Affaire	E
1100018833	6	1	1	2	1	17 200,00	4	850 000	1	1	Vehicules Speciaux	F
1100006723	32	2	1,5	2	2	39 375,99	2	900 000	3	1	Affaire	B

FIGURE 5.4 : Les données d'un assurance Automobile (Source : <https://www.cna.dz>)

numpolic	Puis veh	sex	Bonus malus	Age veh	nb	sinistre	Age cond	Valeur vehi	garantie	Carbran tutil	USAGEE	zone
1100019306		2	1	3	3	21 200,00	4	450 000	1	1	Affaire	D
1100013833	4	1	1	4	1	47 286,75	3	1 000 000	1	1	Commerce	C
1100013128	8	1	1	1	4	19 200,00	1	550 000	2	2	Affaire	A
1100014365	6	1	1	1	3	19 200,00	5	1 400 000	2	2	Affaire	C
1100017979	5	2	1	2	1	43 624,50	3	450 000	2	2	Affaire	B
1100006246	4	2	1	4	1	13 200,00	1	3 000 000	2	2	Affaire	B
1100013279		1	0,65	2	1	36 970,39	2	1 200 000	1	2	Commerce	A
1100011869	8	1	1	2	1	19 200,00	2	1 100 000	1	1	Affaire	C
1100015074	7	2	1	4	1	7 250,00	5	850 000	1	2	Affaire	C
1100011574		1	1	3	1	40 104,75	2	1 500 000	1	1	Affaire	C
1100007914	7	1	0,65	1	1	9 200,00	2	800 000	2	2	Affaire	C
1100013086	7	2	1	2	1	19 200,00	2	1 100 000	2	2	Affaire	C
1100009134	4	2	1	4	1	19 200,00	4	2 200 000	3	2	Taxi	C
1100014563	7	1	1	2	1	5 500,00	4	800 000	3	2	Affaire	A
1100019019	8	2	1	2	1	23 950,00	2	500 000	3	2	Affaire	F
1100010458		2	1	1	1	21 005,00	2	550 000	2	1	Affaire	D
1100017210	5	2	1	1	1	26 344,00	2	1 300 000	2	1	Affaire	D
1100009071	6	2	1	1	4	47 700,00	4	1 300 000	2	1	Affaire	A
1100009790	7	2	1	1	4	19 000,00	2	800 000	2	1	Affaire	A

FIGURE 5.5 : Les données d'un assurance Automobile(Source : <https://www.cna.dz>)

numpolic	Puis veh	sex	Bonus malus	Age veh	nb	sinistre	Age cond	Valeur vehi	garantie	Carbran tutil	USAGEE	zone
1100016720	5	2	1	1	4	12 424,60	3	500 000	2	1	Affaire	E
1100008320		2	0,75	4	5	11 600,00	4	650 000	2	1	Commerce C.Bis	F
1100000947		2	1	3	2	33 243,38	1	1 150 000	2	1	Commerce	B
1100005496	5	2	1	2	1	46 058,40	2	600 000	2	1	Affaire	C
1100011406		2	1	2	1	30 812,32	4	600 000	2	1	Fonctionnaire	C
1100003095	5	1	1	3	1	57 091,69	4	1 100 000	2	1	Affaire	B
1100013439	7	1	1	2	1	17 500,00	2	1 200 000	2	2	Affaire	E
1100004959	5	1	0,65	1	1	16 898,18	3	700000	3	2	Fonctionnaire	F
1100004249	5	1	0,65	1	4	15 087,50	6	1 000 000	3	2	Affaire	A
1100032836	5	2	1	2	1	26 350,00	2	914 000	3	2	Commerce	F
1100003597	9	1	0,65	2	2	23 500,00	1	1 500 000	3	2	Affaire	A
1100019946		2	0,75	2	3	17 258,00	4	914 000	3	2	Commerce	B
1100003627	9	1	1	2	1	13 150,00	1	750 000	2	2	Affaire	C
1100004848	7	2	1	4	2	225 625,80	2	1 500 000	3	2	Affaire	A
1100004308		2	0,65	1	3	27,00	3	900 000	2	2	Affaire	A
1100007944	8	1	1	1	1	47 960,00	4	1 130 000	3	2	Affaire	E
1100008139	7	2	0,65	3	4	174 316,49	2	800 000	1	2	Commerce C.Bis	F
1100013520		2	1	2	3	48 895,47	3	1 500 000	1	2	Affaire	B
1100003151	6	2	1	4	4	14 750,00	4	500 000	1	2	Fonctionnaire	C
1100004660	5	1	1	1	5	39 335,29	5	1000000	2	1	Affaire	F

FIGURE 5.6 : Les données d'un assurance Automobile(Source : <https://www.cna.dz>)

numpolic	Puis veh	sex	Bonus malus	Age veh	nb	sinistre	Age cond	Valeur vehi	garantie	Carbran tutil	USAGEE	zone
1100003220	4	1	0,65	1	5	17 100,00	1	600 000	1	2	Affaire	E
1100012764		2	1	4	4	13 729,41	4	800000	1	1	Fonctionnaire	B
1100014957	4	2	1	4	4	33 100,00	2	700000	1	1	Commerce C.Bis	B
1100003685	11	2	1	3	3	34 041,18	4	1 000 000	3	1	Affaire	C
1100008541	4	2	1	3	3	29 100,00	2	2500000	2	1	Affaire	E
1100017970	7	2	1	2	3	5 000,00	1	900 000	1	1	Affaire	F
1100015465	7	1	1	1	1	81 100,00	2	1 000 000	1	1	Commerce	A
1100005324	9	2	1	2	1	4 000,00	1	800 000	2	1	Affaire	C
1100010195	7	2	0,65	4	1	36 394,12	2	1 000 000	1	1	Affaire	A
1100017440	6	2	1,5	3	4	11 500,00	4	900 000	3	1	Affaire	B
1100006365		1	1	1	4	19 600,00	1	1 100 000	1	1	Affaire	B
1100000056	7	2	1	2	4	34 900,00	5	900 000	2	1	Affaire	B
1100004278	7	2	0,65	4	4	41 100,00	3	800 000	2	1	Commerce	C
1100013269	9	1	1	2	1	10 676,47	6	1 520 000	2	1	Fonctionnaire	C
1100003493	6	2	1	1	1	42 276,47	5	1 200 000	2	1	Affaire	A
1100000237	4	2	1	1	1	39 100,00	2	1 300 000	1	1	Affaire	E
1100011705		2	1	4	1	390 200,00	3	900 000	1	2	Fonctionnaire	E
1100008449	6	2	1	1	1	33 340,00	6	900 000	3	2	Vehicules Speciaux	C
1100007909	37	2	1	2	1	115 360,72	1	500 000	1	2	Vehicules Speciaux	C
1100003542	19	2	0,75	3	1	37 980,07	3	100000	2	2	Affaire	A

FIGURE 5.7 : Les données d'un assurance Automobile(Source : <https://www.cna.dz>)

numpolic	Puis veh	sex	Bonus malus	Age veh	nb	sinistre	Age cond	Valeur vehi	garantie	Carbran tutil	USAGEE	zone
1100006279	7	2	1	2	1	36 052,40	4	400000	3	1	Affaire	B
1100006585	6	2	1	2	1	22 415,75	1	2510000	1	1	Commerce	A
1100001809	8	2	1	2	1	19 200,00	4	900 000	2	2	Fonctionnaire	C
1100011352	7	2	0,75	2	1	28 700,00	2	1 200 000	3	1	Affaire	C
1100018204	4	2	1	1	2	14 700,00	3	800000	1	2	Commerce	E
1100002003	8	2	1	2	1	16 314,29	2	1500000	2	1	Commerce C.Bis	B
1100002106	30	2	1	2	1	47 500,00	2	200000	2	1	Commerce	A
1100006889	8	2	1	3	2	49 666,50	2	2 200 000	2	1	Affaire	F
1100008793		2	1	3	2	31 906,79	2	3 000 000	3	1	Affaire	E
1100011785	7	2	1	3	2	9 200,00	2	1 300 000	2	1	Transport Public de Voyageurs (TPV)	E
1100003651	14	2	0,65	1	1	13 700,00	4	2500000	2	1	Fonctionnaire	C
1100008461	6	2	1	2	1	20 572,77	5	3 000 000	3	1	Commerce	F
1100008097	5	2	1	1	1	19 200,00	5	1 250 000	2	1	Fonctionnaire	B
1100010084	7	2	0,65	4	1	15 200,00	1	900 000	1	1	Affaire	A
1100013145	9	2	1	1	1	19 200,00	1	1 500 000	3	1	Affaire	B
1100032251	10	2	1	2	1	11 200,00	1	1 400 000	3	1	Vehicules Speciaux	C
1100014262	26	2	1	3	1	19 675,00	4	900 000	3	1	Affaire	B
1100006389		2	1	2	1	20 163,75	2	700 000	3	1	Affaire	C
1100000499	5	2	1	4	1	21 485,23	3	1 500 000	3	1	Commerce	B
1100014233	5	2	1	2	1	14 200,00	2	2 000 000	3	1	Commerce	A

FIGURE 5.8 : Les données d'un assurance Automobile(Source : <https://www.cna.dz>)

numpolice	Puis veh	sex	Bonus malus	Age veh	nb	sinistre	Age cond	Valeur vehi	garantie	Carbran tutil	USAGEE	zone
1100006809	5	2	1	3	1	9 200,00	2	1 200 000	1	1	Affaire	F
1100004382	7	2	1	1	1	23 411,25	3	1 400 000	2	2	Affaire	E
1100004612	8	2	1	4	1	9 200,00	6	1 580 000	2	2	Affaire	C
1100004815	5	2	1	1	1	7 500,00	1	1 300 000	2	2	Affaire	D
1100007912	6	2	1	2	1	8 073,19	3	800 000	2	2	Affaire	A
1100001879	7	2	1	3	1	19 200,00	4	1 000 000	2	1	Affaire	B
1100005099	4	2	1	1	1	10 200,00	5	800 000	2	1	Affaire	B
1100006504	6	2	1	4	1	20 150,00	4	950 000	1	2	Transport Public de Voyageurs (TPV)	B
1100014643	15	2	1	4	2	15 200,00	3	800 000	2	1	Affaire	C
1100011628	8	2	1,5	4	2	16 041,96	2	700 000	2	2	Affaire	F
1100019075	5	2	1	2	1	19 200,00	2	900 000	2	1	Commerce	D
1100005646	5	2	1	4	1	9 131,76	3	900 000	2	1	Affaire	A
1100011181	8	2	1	3	2	8 200,00	5	800 000	3	1	Affaire	A
1100003693	8	2	1	3	2	20 087,03	2	900 000	3	1	Affaire	A
1100007519	7	2	1	3	2	45 481,75	3	2 000 000	2	1	Transport Public de Voyageurs (TPV)	E
1100004436	26	2	1	3	1	13 200,00	4	5 000 000	2	1	Affaire	B
1100010589	9	2	1	3	1	14 741,67	4	1 900 000	2	1	Affaire	F
1100012213		2	1	3	2	61 303,89	4	700 000	2	1	Commerce	C
1100003985	8	2	1	4	3	40 267,06	2	1 250 000	2	1	Commerce	A
1100007089	9	2	1	4	3	21 005,00	5	2 000 000	2	1	Affaire	C
1100008016	9	1	1	4	1	11 700,00	3	1 200 000	2	1	Affaire	B
1100018723	7	1	1	2	1	15 200,00	5	300 000	2	1	Affaire	A

FIGURE 5.9 : Les données d'un assurance Automobile(Source : <https://www.cna.dz>)

Résumé : Ce mémoire a pour objet l'étude de la tarification en assurance automobile. Nous sommes intéressés à l'assurance auto en général avant d'entrevoir les différentes étapes à suivre pour la modélisation du risque automobile, donc les premières notions abordées ici sont l'extraction de données et une méthode statistique (GLM) et l'appliquent sur les données d'une compagnie d'assurance automobile pour modéliser le coût moyen et la fréquence pour proposer la prime pure et cherche à sélectionner des facteurs qui contribuent à expliquer la sinistralité.

Mots clés : Assurance automobile, Tarification, Modèles linéaires généralisés, Coût moyen, fréquence.

Abstract : The purpose of this thesis is to study auto insurance pricing. We have thus been interested in auto insurance in general before glancing at the different steps to follow for the modeling of automobile risk, so the first concepts discussed here are data extraction and a statistical method (GLM) and apply it. on the data of an automobile insurance company to model the average cost and the frequency to propose the pure premium and seeks to select factors that contribute to explain the loss experience .

Keywords : Auto insurance, pricing, generalized linear models, average cost, frequency.