



Tarification à posteriori en assurance automobile : Évaluation des probabilités de risque par la méthode de l'apprentissage statistique

Posteriori pricing in automobile insurance: Evaluation of risk probabilities by the statistical learning method

Riad Meriem*, Centre Universitaire Morsli Abdellah- Tipaza- (Algérie),
r-meriem@hotmail.fr

Réception : 22/02/2022

Acceptation: 20/05/2022

Édition : 15/06/2022

Résumé :

Cet article a pour but d'essayer de proposer un outil d'aide à la tarification à posteriori en assurance automobile qui tiendra compte de la portée pécuniaire qu'aura un sinistre une fois survenu c'est ce que l'on nommera « Gravité du sinistre ». On cherche à mesurer l'impact de la prise en considération de cette variable dans la détermination de la prime à posteriori. On a fait recours aux techniques du Data Mining pour le traitement de nos données ; en adoptant l'approche Wait and See. De ce fait des classes de risques sont constituées à partir de caractéristiques de l'assuré et du véhicule. Cette segmentation est stratégique pour que le principe de mutualisation soit fonctionnel dans un environnement concurrentiel.

Mots-clés : Assurance ; Tarification à posteriori ; Gravité de l'accident ; Data Mining ; segmentation.

Abstract:

The purpose of this article is to try to propose a tool to assist with a posteriori pricing in automobile insurance which will take into account the financial impact of a claim once it has occurred, this is what we will call "Loss severity". We seek to measure the impact of taking this variable into account in determining the premium after the event. We have used data mining techniques to process our data, adopting a wait and see approach. As a result, risk classes are constituted on the basis of the characteristics of the insured and the vehicle. This segmentation is strategic so that the principle of mutualization is functional in a competitive environment.

Keywords : Insurance ; A posteriori pricing ; Accident severity ; Data Mining ; segmentation.

* Auteur correspondant : Riad Meriem

1. INTRODUCTION :

L'assurance automobile est une catégorie d'assurance omniprésente et diverse qui représente un marché très important. Elle constitue la principale branche du marché des assurances avec plus de 52% (CNA, 2019) de parts du marché et a enregistré une progression régulière de 12% durant ces dernières années en raison de l'accroissement du parc automobile national.

Chaque année, les assureurs enregistrent plus d'un million de déclarations de sinistres en raison de l'augmentation continue des accidents de la route avec 25.000 à 30.000 accidents corporels par an (DGSN, 2022).. Le risque automobile est la probabilité qu'un véhicule pour lequel une garantie est demandée, soit impliqué dans un accident de circulation. Cette probabilité est déterminée sur la base d'étude statistique par rapport à la fréquence des accidents.

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* selon (François Couilbault, 2003).

1.1. Problématique principale de l'étude :

Le principe le plus répandu en tarification *a posteriori* est le *Coefficient de Réduction Majoration* ; on parlera plutôt d'une segmentation quantitative (C'est à dire l'historique sinistre de l'assuré durant l'année « n » est pris en compte, au lieu de vérifier l'impact financier qu'a eu un accident sur le bilan de la compagnie).

Mais on doit noter que la branche reste déficitaire du fait que les primes encaissées ne couvrent pas les sinistres engendrés. D'où la nécessité de prendre une autre variable en compte, à savoir la « **Gravité de l'accident** » ; qui peut avoir plusieurs classifications selon la nature du sinistre. On peut l'expliquer comme étant une échelle qui nous permet de classer les accidents d'une année donnée par catégorie. De ce fait ; nous tenterons de répondre à cette interrogation :

Comment peut- on évaluer les probabilités de risque automobile pour la tarification à postérieure par le biais des arbres de décisions ?

1.2. Questions de recherche :

On doit noter que le système de tarification automobile utilisé en Algérie, se base essentiellement sur la puissance et l'usage du véhicule, et d'un système bonus-malus. Il

est réservé à l'usage privé et appliqué selon la même règle à tous les assurés par toutes les compagnies d'assurances. De ce fait on pourra s'interroger sur le fait de l'optimalité et l'efficacité du système de tarification en Algérie mais surtout le système Bonus – Malus et son applicabilité.

1.3. Hypothèse/ hypothèses :

Pour bien mener notre recherche des hypothèses doivent être émises. On se base sur l'hypothèse de la non optimalité du système de tarification instauré par le Ministère des Finances dont les primes du secteur ne pourront jamais faire face aux indemnités onéreuses. Ainsi ; sur l'insuffisance des variables tarifaires (Sexe, âge, usage du véhicule...Etc.) pour la fixation de la prime Responsabilité Civile. Cette hypothèse va être vérifiée au fur et à mesure que nous allons avancer dans notre article.

1.4. Importance de l'étude :

L'importance de notre article réside dans le fait de chercher à proposer un outil d'aide à la tarification à posteriori en assurance automobile qui tiendra compte de la portée pécuniaire qu'aura un sinistre une fois survenu c'est ce que l'on nommera « Gravité du sinistre ».

1.5. Objectifs de l'étude :

Le but étant de vérifier l'impact de la prise en considération de la variable gravité du sinistre sur la détermination de la prime *Responsabilité Civile* (RC) à posteriori ; nous allons d'ores et déjà analyser les statistiques relatives à d'autres variables qui ne sont pas prises par la réglementation actuelle lors de l'élaboration de la prime RC.

1.6. Méthodologie de l'étude :

Évaluer la prime pure permet d'augmenter la performance du modèle global et d'intégrer les dépendances éventuelles entre les risques souscrits. C'est pourquoi, tant d'un point de vue technique que commercial, nous modélisons le sinistre total du véhicule plutôt que de construire un modèle par risque. À cet effet ; nous allons nous focaliser sur les données du Conseil National de Prévention et de Sécurité Routière afin d'essayer de construire des classes de risques auxquelles l'assuré sinistré sera affecté. Selon la gravité de son sinistre encouru.

1.7. Recherches antérieures :

Depuis le recours à l'outil statistique en assurance dommages, le nombre de modèles de tarification proposés par les chercheurs est en perpétuelle hausse. Vu la sinistralité conséquente de la branche automobile mais surtout le déficit enregistré, beaucoup de travaux ont essayé de comprendre le déséquilibre entre une prime et un remboursement considérable.

Dans ce qui va suivre nous allons étaler certains travaux phares sur la tarification en assurance automobile ; nous avons choisi plutôt ceux qui rejoignent l'idée de cet article.

- ✓ Maki DAHCHOUR a proposé de faire l'étude sur des données panel dans le cadre de l'instauration du permis à point pour une incitation à la sécurité routière. La critique qu'on puisse apporter sur ce travail est qu'on se positionne dans un environnement où un assuré a déjà payé une prime tandis que le sinistre s'est réalisé. D'autant plus ce n'est qu'un échantillon d'une population diversifiée en termes de comportement de conduite (DAHCHOUR, 2002).
- ✓ Guillaume GONNET qui a développé l'idée du Pay As You Drive. Il a proposé l'introduction de nouvelles variables par le biais de la géolocalisation. Il a procédé à un épurement de données suivi d'une réalisation d'une cartographie de risques. Il a conclu que d'autres variables sont explicatives du sinistre. Son idée était propre pour le marché français à condition que tous les véhicules assurés soient dotés d'un GPS (GONNET, 2003).

1.8. Organisation de l'étude :

Pour répondre à cette problématique l'article va être scindé en trois parties la première pour l'explication du principe de tarification en assurance automobile, la seconde va être une étude descriptive de la base de données suivie par l'application de l'une des techniques du Data Mining pour le traitement de nos données.

2.Principes de calcul de prime en assurance automobile :

La tarification est un processus instantané d'évaluation des risques où la prime doit être proportionnelle au risque encouru par les assurés, de ce fait les assureurs sont à la recherche d'un modèle de tarification adéquat et optimal. Comme pour les autres assurances dommages, l'assurance automobile repose essentiellement sur le rapport de sinistralité. La clé de voûte de toute analyse économique de l'assurance est la relation entre ce que l'assureur perçoit comme primes et ce qu'il rembourse à ses assurés ou aux victimes de ses assurés sous forme de règlement de sinistre (Trivedi, 2005).

2.1.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. Celles qui sont prises par le par le Ministère de Finance et suivie par toutes les compagnies d'assurance en Algérie sont :

- ✓ *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.

2.2. La tarification a posteriori :

Elle est considérée comme étant la complémentarité de la tarification a priori, ou 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 (Michel Denuit, 2007).

De ce fait, la préoccupation des compagnies d'assurances 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 assumé. C'est d'ailleurs ce que nous tenterons de proposer à travers la partie pratique de cet article.

3. Présentation des données :

Nous avons essayé de constituer une base de données assez riche en informations dont nous aurons besoin pour l'élaboration de notre modèle (CNPSR, 2021). Les informations sont relatives à l'accident, son origine, sa survenance mais aussi son impact. Nous allons dans ce qui va suivre exposer les différentes rubriques retenues pour notre travail.

Notre base de données est relative aux accidents de la circulation pour la période entre le mois de *Décembre 2021* et *Janvier 2022* qui inclue les différentes caractéristiques liées aux sinistres automobiles. On doit juste préciser que notre étude se porte sur les accidents corporels, car la variable « Gravité » qu'on est censé quantifier et mesurer n'est pas déterminée pour les sinistres matériels (Direction Générale de la Sécurité Nationale, 2022).

Nous avons juger opportun de procéder à une explication de la notion d'accident corporel avant de passer à la spécification de notre base de données.

3.1. Définition d'un accident corporel :

Est défini comme un accident corporel (mortel et non mortel) de la circulation routière, tout accident qui (James Landel, 2008) : Implique au moins une victime, survient sur une voie publique ou privée, ouverte à la circulation publique, implique au moins un véhicule.

Un accident corporel implique un certain *nombre de personne* ; parmi celles-ci on distingue : *Les personnes indemnes* (*Impliquées non décédées et dont l'état ne nécessite*

aucun soin médical du fait de l'accident) ; **Les victimes** (Impliquées non indemnes). Parmi *les victimes* on distingue : **Les personnes tuées** (Personnes qui décèdent du fait de l'accident, sur le coup ou dans les trente jours qui suivent l'accident) ; **Les personnes blessées victimes non tuées**.

Parmi les *personnes blessées*, il convient de différencier entre :

- ✓ Les blessés dit « hospitalisés » : victimes hospitalisées plus de 24 heures ;
- ✓ Les blessés légers : victimes ayant fait l'objet de soins médicaux mais n'ayant pas été admises
- ✓ Comme patients à l'hôpital plus de 24 heures.

Passons maintenant à l'explication des caractéristiques de la notre base de données.

3.2. Spécification de la base de données :

La base de données des accidents de la circulation est répartie en quatre rubriques reprises comme suit :

- ✓ La rubrique « **Caractéristiques** » qui décrit les circonstances générales de l'accident ;
- ✓ La rubrique « **Lieux** » qui décrit le lieu principal de l'accident. Est défini comme étant lieu d'accident : « *Situation spatiale de l'accident qui permet de le localiser afin de déterminer sa trajectoire ainsi que l'espace dans lequel est survenu ce dernier* » selon (Revue, 2003). Un accident peut être géo-localisé de plusieurs façons : Adresse, Catégorie de la route, Numéro de la route.
- ✓ La rubrique « **Véhicules** » impliqués ;
- ✓ La rubrique « **Usagers** » impliqués.

Il est à noter que chaque accident porte un numéro d'identifiant ainsi que la date de survenance avec l'heure exacte.

Après avoir présenté les informations nécessaires pour mieux analyser et lire notre base de données, qui englobent le détail des accidents de la circulation ; nous allons maintenant expliquer et codifier la variable cible de notre étude.

3.3. Description de la variable gravité de l'accident :

Il s'agit de « **La gravité de l'accident** ». Cette dernière est déterminante lors de l'établissement d'une prime à posteriori ; du moment où elle nous permet de quantifier les dommages survenus et ainsi connaître le montant des indemnités que les assureurs verseront à leur assurés et victimes. Elle permettra d'affecter à chaque assuré le score qui lui est adapté (Selon le poids qu'aura l'accident sur les remboursements et donc les provisions techniques de l'assureur). Le tableau suivant englobe les diverses codifications de cette variable.

Tableau N°01 : Informations relatives à la variable « Gravité » du sinistre

Variable	Codification
Grav : La gravité de l'accident en termes de dommages corporels.	1- Indemne ; 2- Tué ; 3- Blessé hospitalisé ; 4- Blessé léger

Source : Sur la base des informations regroupées de la base de données (Direction Générale de la Sécurité Nationale, 2022)

Dans ce qui va suivre nous allons présenter les informations descriptives de notre variable Gravité de l'accident.

4.Statistiques de la variable « Gravité » :

On peut déduire que la connaissance de l'implication de l'assuré dans l'accident (à travers la détermination des points de choc) ; soit une information supplémentaire qui permettra de mieux ajuster la prime à posteriori (Charpentier, 2004). Sachant que le CRM la néglige c'est la gravité de l'accident qu'on propose d'étudier. Cette dernière est spécifique aux conséquences physiques qu'aura un accident soit sur l'usager du véhicule avec les passagers soit sur l'adversaire (Piéton ...Etc).

Notre variable comporte *quatre* modalités : Indemne, Tué, Blessé hospitalisé, Blessé léger. Le tableau ci-après nous fournit plus d'explicitation sur la variable de notre étude.

Tableau N° 2 : Statistiques de la variable « Grav »

Statistique	Grav
Nb. D'observation	982
Modalité 1 : Indemne	383
Modalité 2 : Tué	44
Modalité 3 : Blessé hospitalisé	247
Modalité 4 : Blessé léger	308

Source : Sur la base des informations regroupées de la base de données (Direction Générale de la Sécurité Nationale, 2022)

Une première lecture de ce tableau nous indique que **383** des victimes d'accidents sont ressorties indemnes. Tandis que **555** présentent diverses blessures ; celles nécessitant l'hospitalisation de la victime sont de l'ordre de **247** blessés et le reste n'ont que des blessures légères.

Malheureusement ; notre base de données présente des accidents mortels ayant causés le décès de **44** personnes.

Cette analyse est primordiale pour les compagnies d'assurances qui voudront adopter une tarification individuelle des assurés. Du moment où la gestion d'un dossier sinistre diffère selon le type : *Corporel / Matériel*.

La disparité réside en premier lieu en provisionnement (Méthodes de provisionnement) mais aussi en indemnisation.

Un dossier sinistre est classé d'ordre corporel quand il engendre au moins un blessé nécessitant les premiers soins allant jusqu'à l'hospitalisation. Car l'assureur doit rembourser les différentes indemnités dues à la victime (Assuré ou adversaire) et aux ayants droit en cas de décès (Petauton, 2000).

Dans ce qui va suivre on va faire une analyse de la variable « Gravité » par rapport aux autres variables constituant notre base de données. On procédera à une analyse statistique comparative des diverses modalités en fonction des autres variables afin de faire ressortir des relations par rapport à la gravité d'accident.

Rappelons que les rubriques précitées comportent plusieurs variables qui sont déterminantes lors de l'étude de la variable principale gravité de l'accident. Voici un schéma récapitulatif.

Figure N° 1 : Composition des rubriques étudiées.

Rubrique caractéristique	Rubrique Usagers	Rubrique Lieux	Rubrique Véhicule
<ul style="list-style-type: none"> • Conditions d'éclairage; • Conditions atmosphériques; • Type de collisions. 	<ul style="list-style-type: none"> • Sexe; • Age. 	<ul style="list-style-type: none"> • En agglomération; • Hors agglomération. 	<ul style="list-style-type: none"> • Obstacle fixe; • Obstacle mobile; • Points de choc.

Source : Sur la base des informations regroupées de la base de données (Direction Générale de la Sécurité Nationale, 2022)

4.1. Analyse de la gravité par rapport aux conditions d'éclairage :

Notre variable comportait cinq modalités décrivant les conditions d'éclairage lors de la survenance des accidents. Le tableau ci-après expose les chiffres comparatifs de cette variable avec celle de la gravité de l'accident.

Tableau N° 3 : Statistiques de la variable « Grav » par rapport à l'éclairage

		Gravité d'Accident				Total
		Indem ne	Tué	Blessé Hospitali sé	Blessé	
Eclairage	Plein Jour	266	32	171	217	686
	Crépuscule ou Aube	17	2	13	10	42
	Nuit sans éclairage	25	4	16	22	67
	Nuit avec éclairage non allumé	5	1	2	4	12
	Nuit avec éclairage allumé	70	5	45	55	175
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

On arrive au fait que la plupart des accidents soient arrivés en plein jour avec un nombre de l'ordre de 686 accidents du total. Mais on peut interpréter autrement ce résultat ; par le fait que les gens circulent plus durant la journée ce qui engendre des flux de circulation assez denses et donc un nombre élevé d'accident (Embouteillage par exemple). Tandis que durant la nuit les gens circulent moins et les conducteurs ont tendance à être plus prudent au volant à cause de la luminosité.

4.2. Analyse de la gravité par rapport aux conditions atmosphériques :

Les conditions dans lesquelles est survenu un accident sont départagées en *cinq modalités* ; nous voulons à travers l'analyse qu'aura l'impact des conditions atmosphériques, sur le taux de la sinistralité routière, et par conséquent sur notre variable gravité. C'est à partir du tableau ci-après que nous allons procéder :

Tableau N° 4 : Statistiques de la variable « Grav » par rapport aux conditions atmosphériques.

		Gravité d'Accident				Total
		Indemne	Tué	Blessé Hospitalisé	Blessé	
Conditions Atmosphériques	Normal	312	38	209	253	812
	Pluie	51	5	22	40	118
	Neige et Grêle	2	0	1	2	5
	Vents	17	1	13	11	42
	Autres	1	0	2	2	5
	Total		383	44	247	308

Source : Effectué sur SPSS 24 sur notre base de données initiale.

Le tableau nous informe que 881 accidents sont survenus en plein jour avec un nombre de blessés qui est de l'ordre de 462. Tandis que durant les intempéries la mortalité sur les routes est en baisse en hiver. Les mauvaises conditions météorologiques dissuadent, en effet, les conducteurs de prendre la route. Durant les beaux jours au contraire, les conducteurs sont plus souvent amenés à conduire ; notamment en raison des longs week-ends et des vacances.

4.3. Analyse de la gravité par rapport à la localisation de l'accident :

La localisation est subdivisée en deux modalités ; soit en agglomération ou bien hors agglomération. Lors du chapitre précédent nous avons fait une étude par zone (Urbaine / Rurale) et nous sommes arrivés à la conclusion que la majorité des accidents survient en zone rurale et ceci par rapport à l'inadaptation des infrastructures routières. Le tableau suivant reprend le récapitulatif des informations.

Tableau N° 5 : Statistiques de la variable « Grav » par rapport à la localisation de l'accident.

		Gravité d'Accident				Total
		Indemn e	Tué	Blessé Hospitalisé	Blessé	
Localisation d'Accident	Hors Agglomération	131	20	87	99	337
	En Agglomération	252	24	160	209	645
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

D'après les chiffres exposés ; on voit que finalement la majorité des accidents sont survenus en agglomération ; ils sont de l'ordre de 645 que l'on pourra expliquer par le flux routier important que connaissent les routes nationales.

Dans ce qui va suivre ; nous allons étudier la relation entre la variable gravité et les conséquences de l'accident par rapport à l'objet assuré, à savoir le véhicule. Commençons d'abord par les types de collisions et les points de choc qui son déterminant lors de l'affectation des taux de responsabilité dans un accident. Ainsi un *CRM* par rapport à la sinistralité de l'assuré.

Comme nous l'avons expliqué précédemment le système Bonus-Malus Algérien n'est pas optimal ; du moment que les assurés sont tous regroupés de la même manière qui ne dépend que de leur historique de sinistre (Nombre d'accidents durant l'année « n ») ; tout en négligeant de mettre l'accent sur l'implication du conducteur dans l'accident.

Cet examen a pour but de prouver la nécessité de prendre en considération la responsabilité de l'assuré par rapport à la gravité de l'accident. Un conducteur qui a eu un accident matériel ne doit pas avoir le même *CRM* qu'un conducteur qui a fait un accident corporel.

4.4. Analyse de la gravité par rapport aux types de collision :

Les types de collisions sont répartis en *sept modalités* qui décrivent les circonstances d'accident avec le véhicule adverse. Le tableau suivant nous fournit les informations suivantes :

Tableau N° 6 : Statistiques de la variable « Grav » par rapport au type de collision

		Gravité d'Accident				Total
		Ind em ne	Tué	Blessé Hospitalis é	Blessé	
Type de Collision	Frontal	38	6	20	30	94
	Arrière	56	2	32	31	121
	Coté	114	12	74	82	282
	En Chaîne	11	0	5	12	28
	Multiple	14	1	5	8	28
	Autre	110	18	82	108	318
	Sans collision	40	5	29	37	111
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

À partir des chiffres énumérés au tableau on remarque que le nombre de blessé le plus élevé est quand la collision est affectée à la sixième modalité. Pour les victimes indemnes la collision est soit sur le côté (Donc partage de responsabilité) où à l'arrière donc c'est l'assuré qui est victime. On ne doit pas omettre qu'un accident avec un adversaire engage les deux garanties : Responsabilité civile et Défense et Recours ; qui représentent des dépenses supplémentaires pour la compagnie d'assurance.

4.5. Analyse de la gravité par rapport aux points de choc :

On adoptera le même raisonnement que pour la variable précédente. Car les points de choc seront aussi déterminants des taux de responsabilité dans un accident de la circulation. Les statistiques des points de choc en fonction de la gravité sont résumées à travers le tableau suivant :

Tableau N° 7 : Statistiques de la variable « Grav » par rapport aux points de choc.

		Gravité d'Accident				Total
		Indemn e	Tué	Blessé Hospitalis é	Blessé	
Point de Choc	Avant	243	34	169	218	664
	Arrière	51	2	28	31	112
	Coté	72	6	38	53	169
	Autre	17	2	12	6	37
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

On lit à travers le tableau que 664 accidents ont été provoqués par des chocs à l'avant ; d'après l'infra-code c'est le véhicule avec des chocs à l'avant qui est responsable. Le nombre le plus élevé des victimes en sortie indemnes d'un accident sont celles ayant subi un choc à l'avant, ceci s'explique par la présence des airs bag et l'obligation de la ceinture de sécurité.

Le même constat pour les blessés ; qu'on peut affectés à des passagers dans le véhicule.

Les variables « Choc » et « Col » sont les plus déterminantes du partage de responsabilité, donc les compagnies d'assurance devraient en tenir compte lorsque les assurés remplissent le constat amiable mais surtout demandaient aux experts de bien définir les conséquences matérielles.

On passe maintenant à la comparaison du nombre de victimes par rapport aux obstacles heurtés.

4.6. Analyse de la gravité par rapport aux obstacles fixes / mobiles heurtés :

Les obstacles fixes heurtés sont départagés en *quatre modalités* ; les résultats sont visibles à travers le tableau suivant :

Tableau N° 8. Statistiques de la variable « Grav » par rapport à l'obstacle fixe heurté.

		Gravité d'Accident				Total
		Indemne	Tué	Blessé Hospitalisé	Blessé	
Obstacle Fixe Heurté	Véhicule en Stationnement	92	13	52	82	239
	Arbre	25	3	17	22	67
	Glissière	73	7	44	60	184
	Autre	193	21	134	144	492
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

Alors on peut observer que la quatrième modalité a le plus grand effectif avec 492 victimes où le nombre de blessés est de l'ordre de 278 victimes. Cette modalité regroupe des obstacles tels que : Les bâtiments, les Poteaux ...Etc.

On doit spécifier que tout bien public détruit induit des poursuites judiciaires, donc des dépenses supplémentaires pour la compagnie d'assurance.

Pour l'autre type d'obstacles, c'est une variable à quatre modalités aussi (Piéton, Véhicule, Animal, Autre) ; où la quatrième modalité peut regrouper des obstacles tels que : Des animaux, des véhicules sur rail...Etc. L'effectif le plus élevé est celui d'un véhicule heurté ; avec 759 accidents survenus entre deux ou plusieurs véhicules. Les blessés représentent 437 de notre échantillon. Il peut s'agir de victimes des deux véhicules (Conducteurs, passagers).

On termine notre analyse par l'étude de l'impact qu'ont eu l'âge et le sexe des conducteurs sur la variable « Grav ». On ne doit pas omettre de noter que les deux variables sont déjà comprises dans le système de tarification de l'assurance automobile partout dans le monde.

4.7. Analyse de la gravité par rapport au sexe du conducteur :

Cette variable comporte *deux* modalités, où les statistiques sont récapitulées dans le tableau ci-après.

Tableau N° 9 : Statistiques de la variable « Grav » par rapport au sexe des conducteurs.

		Gravité d'Accident				Total
		Indemn e	Tué	Blessé Hospitalisé	Blessé	
Sexe	Homme	296	35	172	202	705
	Femme	87	9	75	106	277
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

Les femmes sont moins dangereuses que les hommes au volant (CNPSR, 2021), ceci est appuyé par les chiffres ci-haut. Avec **71%** du total du nombre d'accident, les conducteurs hommes sont à l'origine de **374** blessés par rapport à **181** blessés provoqués par les conductrices femmes.

Une tarification personnalisée par rapport au sexe des conducteurs devrait faire l'objet d'étude par les actuaires des compagnies d'assurance.

4.8. Analyse de la gravité par rapport à l'âge du conducteur :

L'âge des assurés (Ou conducteur) sont pris en compte par la combine d'assurance lors de la détermination de la prime. La preuve une majoration d'âge est instaurée par quand l'âge du conducteur est moins de *25 ans* de l'ordre de 15% de la prime RC. On propose à travers le tableau suivant l'étude de cette variable qui comporte *cinq modalités* sous forme d'intervalle d'âge.

Tableau N° 10. Statistiques de la variable « Grav » par rapport aux tranches d'âges des conducteurs.

		Gravité d'Accident				Total
		Indem ne	Tué	Blessé Hospitalis é	Blessé	
Age en Classes	< 25 ans	71	9	65	108	253
	Entre 25 et 35	78	3	61	65	207
	Entre 35 et 45	80	3	29	32	144
	Entre 45 et 65	113	12	65	66	256
	> 65 ans	41	17	27	37	122
Total		383	44	247	308	982

Source : Effectué sur SPSS 24 sur notre base de données initiale.

D'après les chiffres du tableau on voit clairement que les tranches d'âge impliquées dans la survenance des accidents routiers sont celles de *moins de 25 ans* avec 253 accidents dont 173 ayant causé de blessés à différents degrés. Ce constat ne nous étonne pas du moment que c'est la tranche d'âge qui est majorée en prime RC par les compagnies d'assurances. L'explication plausible qui est donnée ; et que cette catégorie de conducteurs à moins d'expérience au volant (Nouveaux permis de conduire). Parmi les fautes dont la recrudescence est notoire chez les jeunes, on note particulièrement les excès de vitesse et les dépassements irréguliers. La fatigue et l'éblouissement sont également plus fréquents chez les jeunes.

Une autre lecture du tableau ; nous informe que 256 accidents routier sont provoqués par les conducteurs ayant un âge entre [45 ; 65] ans. On note que presque la moitié des victimes se sont en sorties indemnes de l'accident. Pour les personnes de cette catégorie, les fautes dont la fréquence augmente le plus vite semblent mettre en cause des erreurs d'appréciation et peut- être une certaine ignorance des dispositions règlementaire introduite récemment dans le code de la route.

5. Classification du risque automobile sur la base de la « gravité de l'accident » :

Dans cette partie on cherche à faire ressortir la relation entre notre variable d'étude « Gravité » de l'accident et les autres variables de l'échantillon. Le but étant de trouver des classes tarifaires de risque. La méthode choisie est celle de l'apprentissage par arbre de décision comme modèle prédictif mais aussi comme point de départ au processus de décision.

C'est une technique d'apprentissage supervisé : on utilise un ensemble de données pour lesquelles on connaît la variable- cible (Gravité de l'accident) afin de construire l'arbre, puis on extrapole les résultats à l'ensemble des données de test (Gasso, 2005).

5.1. Préliminaires à la théorie statistique d'apprentissage :

Les arbres de décisions sont des règles de classification qui basent leur décision sur une suite de tests associés aux attributs, les tests étant organisés d'une manière arborescente. Les nœuds de l'arbre sont appelés « **nœuds de décisions** » ; chaque nœud est étiqueté par un test qui peut être appliqué à toute description d'un individu de la population (Christian Partrat, 2004).

En général ; chaque test examine la valeur d'un unique attribut de l'espace des descriptions. Les réponses au test correspondent aux étiquettes des arcs issus de ce nœud. Les feuilles sont étiquetées par une classe appelée *classe par défaut*. Les arbres de décisions ont deux qualités appréciables : Les décisions sont aisément interprétables et la classification est très rapide.

5.2. Algorithme d'apprentissage :

La première étape consiste à construire un petit arbre consistant avec la plupart des données. C'est-à-dire on construit l'arbre de la racine vers les feuilles, gloutonnes et récursives. On résumera la suite des étapes à travers les points suivants :

- ✓ Décider si un nœud est terminal : Si le nœud doit être étiqueté comme une feuille u porter un test ;
- ✓ Si le nœud n'est pas terminal, on sélectionne le test à lui associer ;
- ✓ Si le nœud est terminal ; lui affecter une classe.

La schématisation de l'algorithme sera reprise dans l'annexe.

Notre choix s'est porté sur la méthode *Chi-Squared Automatic Interaction (CHAID)* vu la facilité d'interprétation. Elle détecte l'interaction entre variables dans un jeu de données. En l'utilisant on peut établir des relations de dépendance entre variables. **CHAID** est souvent utilisé comme technique d'exploration et est une alternative aux multiples régressions, en particulier quand le jeu de connaissances n'est pas parfaitement adapté aux analyses par régression.

Le schéma reprend les différentes informations relatives et nécessaires à la construction de notre arbre de décision en appuyant sur la méthode **CHAID**.

Figure N° 2 : Spécification du modèle d'apprentissage

Spécification	Méthode de développememt	Exhaustive CHAID
	Variable dépendante	Grav
	Variables indépendantes	an, jour, hrnm, lum, agg, atm, col, obs, obsm, choc, sexe, Age.
	Profondeur maximum de l'arbre	3
	Nombre minimum d'observations d'un noeud parent	100
	Nombre minimum d'observations d'un noeud	50
	Enfant	Age, col
Résultats	Variables indépendantes	8
	Incluses	5
	Nombre de noeuds	2
	Nombre de noeuds Terminaux	
	Profondeur	

Source : Sur la base de l'algorithme d'apprentissage selon l'annexe 1

5.3. Interprétation des liaisons à partir de l'arbre de décision :

Dans notre idée nous avons proposé l'introduction de nouvelles variables de tarification afin de mieux quantifier le risque automobile. Comme nous l'avons spécifié précédemment la pensée nous est venue de l'approche statistique du Wait and See (Gérard Dreyfus, 2008) où une fois le phénomène survenu on observe les interrelations qu'aura ce dernier avec les variables pouvant l'influer mais surtout l'impact de la prise en considération du phénomène sur une quelconque décision.

Dans la tarification en assurance automobile nous allons proposer d'affecter à chaque assuré sinistré des caractéristiques relatives aux divers facteurs influant la gravité du sinistre survenu. Par rapport à notre base de données et nos variables définies ; nous

avons procéder à une segmentation où notre variable cible c'était la « **Gravité** » avec ses quatre modalités auxquels on a déroulé un algorithme conditionnel afin de faire sortir les nœuds susceptibles de nous aider à prendre une décision. Le nombre de nœud terminaux est de l'ordre de **5 ségments** ; où les seules variables retenues comme critères de classification (Affectation à des classes de risques) sont la variable « **Age** » et la variable « **Collision** ».

6.Résultats :

L'arbre de décision ayant dégagé *cinq ségments* qui permettront de faire sortir des classes de risque ; nous allons présenter dans ce qui va suivre les différents résultats :

- ✓ **Segment 1** : Issus du segment primaire où la modalité dominante de la gravité est celle de *victimes indemnes* et sous la prise en compte de la variable « **Age** » avec la condition d'une *catégorie d'âge inférieure à 25 ans* et un accident ayant induit *plusieurs collisions* aux véhicules accidentés on obtient une probabilité d'appartenance à cette classe de **0.340278** ;
- ✓ **Segment 2** : Issus du segment primaire où la modalité dominante de la gravité est celle de *blessés hospitalisés* ; avec une condition d'**Age** toujours inférieur à **25ans** où la variable *collision* prend *deux modalités* (6 et 4) ; on arrive à une probabilité de l'ordre de **0.55** ;
- ✓ **Segment 3** : Issus du segment primaire où la modalité dominante de la gravité est celle de *victimes indemnes*, on remarque la prise en compte que de la variable « **Age** » et ce pour deux modalités (*Inférieur à 25 ans* et *entre 25 et 35 ans*) ; dégage une probabilité d'appartenance à la classe de l'ordre de **0.383420** ;
- ✓ **Segment 4** : Issus du segment primaire où la modalité dominante de la gravité est celle de *victimes indemnes*, avec la prise en compte de la variable « **Age** » avec deux classes celles de **[25 ; 35 ans]** et **[35 ; 40 ans]** nous fournit une probabilité de l'ordre de **0.571429** ;
- ✓ **Segment 5** : Issus du segment primaire où la modalité dominante de la gravité est celle de *victimes indemnes*, avec la prise en compte de la variable « **Age** » avec deux classes celles de **[35 ; 40 ans]** et **[50 ; 60 ans]** nous dégage une probabilité d'affectation de l'ordre de **0.436000**.

On doit mettre l'accent sur le fait que l'arbre n'a pris que deux variables pour la segmentation celle de l'**Age** et celle de la **Collision**. Ceci est spécifique à notre base de données. Or l'étude statistique des accidents nous a révélé l'influence des autres variables sur la gravité de ces derniers.

7. CONCLUSION :

La tarification automobile est réglementée ce qui cause aux compagnies d'assurances de déficits reflétés sur leur bilan et sur leur positionnement dans un marché concurrentiel c'est ce qui appuie l'hypothèse de notre travail en confirmant que le système de tarification automobile est non optimal vu que la branche est déficitaire depuis des années. À cet effet nous avons essayé tout au long de cet article de mettre l'accent sur l'importance de la prise en compte des facteurs influant non pas le nombre d'accidents mais plutôt la gravité engendrée.

L'objectif principal de notre travail est d'appliquer une nouvelle approche de détermination pratique de la prime, c'est la tarification a posteriori, c'est-à-dire trouvé une autre variable sur laquelle s'appuie le système *Bonus – Malus*. À cet effet ; nous avons fait recours aux arbres de décision pour faire ressortir les segments d'une tarification a posteriori qui sera un outil d'aide aux compagnies d'assurances afin d'affecter l'assuré à sa classe de risque.

Par la suite ; nous nous sommes rabattus sur la méthode *CHAID* pour la construction de l'arbre qui a permis de faire ressortir cinq segments qui serviront d'outil d'aide pour les compagnies d'assurance lors de l'élaboration de la prime d'assurance. Les variables de segmentation retenues vers la fin, étaient *l'âge des conducteurs* par catégorie qui est déjà un critère important de tarification mais négligé par les compagnies d'assurances qui ne majorent que les conducteurs jeunes or l'étude statistique nous a révélé qu'une autre catégorie d'âge était susceptible de causer des accidents (de **plus de 40 ans**). L'autre variable est collision résultante de l'accident (Entre deux ou plusieurs véhicules). On doit noter que cette l'analyse de cette variable permet la détermination des taux de responsabilité dans l'accident. Cette dernière étant négligée par le régime de tarification actuel.

Même si nous ne prétendons pas être arrivés à des résultats probants, on peut exposer les recommandations ci-dessous :

- La construction des classes de risque est primordiale pour une compagnie d'assurance ;
- La constitution de classes de risques doit se faire à partir de caractéristiques de l'assuré et du véhicule ;
- Les classes de risques se doivent d'être homogènes en matière de sinistralité ;
- Lors de la souscription d'un contrat automobile, la caractérisation des assurés par leur classes de risques peut être un outil d'aide à la tarification à postérieure qui permettra dans le futur d'éviter le déficit de cette branche ;

Finalement ; pour que le principe de mutualisation soit fonctionnel, dans un environnement concurrentiel ; on pourra dire que cette segmentation est tout aussi stratégique.

8. Bibliographie :

8.1. Livres :

- Charpentier, A. (2004). *Mathématiques de l'assurance non-vie*. France: Economica.
- Christian Partrat, J.-L. B. (2004). *Assurance non-vie : Modélisation, Simulation*. France: Eyrolles.
- François Couilbault, C. E. (2003). *Les grands principes de l'assurance*. France: Eyrolles.
- Gasso, A. R.-G. (2005). Introduction au Data-Mining. *Introduction au Data-Mining*. France: INSA Rouen.
- Gérard Dreyfus, J.-M. M. (2008). *Apprentissage statistique*. France: Eyrolles.
- Michel Denuit, X. M.-F. (2007). *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*. England: Wiley.
- James Landel, L. N. (2008). *Manuel de l'assurance automobile*. France: L'argus de l'Assurance.
- Petauton, P. (2000). *Théorie de l'assurance dommages*. France: Dunod.
- Trivedi, A. C. (2005). *Microeconometrics: Methods and applications*. England: Cambridge.

8.2. Mémoires et Thèses:

- DAHCHOUR, M. (2002). Tarification de l'assurance automobile, utilisation du permis à point et incitation à la sécurité routière : Une analyse empirique. Paris: Paris Nanterre.
- GONNET, G. (2003). Etude de la tarification et de la segmentation en assurance automobile. France: Université de Lyon.

Articles :

- Revue, A. (2003, 03 20). Contribution des systèmes d'informations géographiques à la sécurité routière. France, Paris: HAL Open Science.

8.3. Sites Internet :

CNA. (2019). *Secteur Algérien des assurances*. Algeria: Conseil national des assurances. (Consulté le 19/01/2022 à 21h01).

CNPSR. (2021, 12 30). *Conseil National de Prévention et de sécurité Routière*. From <https://www.interieur.gov.dz/index.php/fr/le-ministere/le-minist%C3%A8re/communiqu%C3%A9s-d%C3%A9clarations/1356-la-cr%C3%A9ation-de-la-d%C3%A9claration-nationale-de-s%C3%A9curit%C3%A9-routi%C3%A8re-est-1%E2%80%99une-des-importantes-no> (Consulté le 20/01/2022 à 10h05).

DGSN (2022). From Direction Générale de la Sécurité Nationale: <https://www.dgsn.dz/?Statistiques-des-accidents-corporels-de-la-circulation-routiere-enregistres,17751> (Consulté le 26/01/2022 à 12h45).

DGSN. (2022). *État comparatif des accidents corporels entre 2019-2021*. Algeria: DGSN. (Consulté le 20/01/2022 à 11h05)

9. Annexe : algorithme de l'arbre de décisions

