

**Mémoire d'actuariat présenté devant
le Jury de l'Institut des Actuaires
pour l'obtention du diplôme d'actuaire
du Centre d'Etudes Actuarielles
et l'admission à l'Institut des Actuaires
le : _____**

Par : Angelo CARIA

Titre : « CONTRAT DE FIDELITE BANCASSURANCE »
MODELISATION DE LA VALEUR D'UN CLIENT BANCASSURANCE A L'AIDE D'UN SCORE
ET PROPOSITION D'UN TARIF ADAPTE SUR LE PRODUIT D'APPEL QU'EST
L'AUTOMOBILE EN CAS DE SOUSCRIPTION D'UN CONTRAT DE FIDELITE

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

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

Membre présent du jury de l'Institut
des Actuaires :

signature

Entreprise :

Nom : _____

Signature : _____

Membres présents du jury du Centre
d'Etudes Actuarielles :

Directeur de mémoire en entreprise :

Thomas BEHAR

Nom : _____

Vincent DAMAS

Signature : _____

Gérard CROSET

Invité :

Arnaud COHEN

Nom : _____

Jean-Pierre DIAZ

Signature : _____

Brigitte DUBUS

Paul ESMEIN

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

(après expiration de l'éventuel délai de
confidentialité)

Michel FROMENTEAU

Stéphane MENART

Christophe IZART

Signature du responsable entreprise

Pierre PETAUTON

Florence PICARD

Olivier LOPEZ

Signature du candidat

Secrétariat :

Bibliothèque :

PROMOTION : 2004/9

Mémoire d'actuariat présenté devant

le jury du CEA

Pour l'obtention du

**Diplôme d'Actuaire du CENTRE D'ETUDES
ACTUARIELLES**

Par Mr. Angelo CARIA

Sur le sujet :

« CONTRAT DE FIDELITE BANCASSURANCE »

**MODELISATION DE LA VALEUR D'UN CLIENT BANCASSURANCE
A L'AIDE D'UN SCORE ET PROPOSITION D'UN TARIF ADAPTE
SUR LE PRODUIT D'APPEL QU'EST L'AUTOMOBILE EN CAS DE
SOUSCRIPTION D'UN CONTRAT DE FIDELITE**

Devant un jury composé de :

Liste du jury :

Thomas BEHAR
Vincent DAMAS
Gérard CROSET
Arnaud COHEN
Jean-Pierre DIAZ
Brigitte DUBUS
Paul ESMEIN
Michel FROMENTEAU
Stéphane MENART
Christophe IZART
Pierre PETAUTON
Florence PICARD
Olivier LOPEZ

Directeur du mémoire :

Arthur CHARPENTIER

Invité :

La reproduction ou la diffusion – totale ou partielle – de ce document, par quelque procédé que ce soit, électronique ou mécanique, y compris la photocopie et l'enregistrement, ou par tout autre moyen sont interdits sans l'autorisation écrite de Angelo Caria. Ce document est destiné à l'usage exclusif de la personne, des dirigeants ou du personnel appartenant à l'organisme auquel il a été initialement confié. La diffusion de ce document en est contrôlée. Des exemplaires supplémentaires à usage interne sont disponibles sur demande, mais ce document ne peut être mis à la disposition de tiers sans l'accord exclusif de Angelo Caria.

RESUME:

« CONTRAT DE FIDELITE BANCASSURANCE » MODELISATION DE LA VALEUR D'UN CLIENT BANCASSURANCE A L'AIDE D'UN SCORE ET PROPOSITION D'UN TARIF ADAPTE SUR LE PRODUIT D'APPEL QU'EST L'AUTOMOBILE EN CAS DE SOUSCRIPTION D'UN CONTRAT DE FIDELITE.

Mots clés : contrat de fidélité, multi équipement, PNB, valeur client, prime pure, S/P, CNAT, attrition, rétention, Bâle II.

Les sociétés d'assurance et les mutuelles ont profité depuis près d'une décennie de la bonne tenue du marché non vie aussi bien en termes d'évolution du chiffre d'affaires qu'en termes de résultat technique. Nombreux sont ceux qui ont essayé de se différencier au niveau produit en segmentant leur tarif, le voulant compétitif sur des garanties quasi identiques.

Toutefois, l'approche différenciée au niveau produit ne garantit pas d'une manière générale la fidélité du client, ni ne permet l'équipement systématique d'un contrat d'assurance supplémentaire.

Aussi est-il plus judicieux d'attacher un intérêt particulier pour les clients qui contribuent de manière significative à la mutualisation du risque et garantissent ainsi notre activité Assurance. C'est pourquoi l'objet de mes travaux portera principalement sur l'identification des clients à forte valeur ajoutée et sur le produit d'appel « automobile » qui leur sera proposé à tarif avantageux.

Au travers de projections des différents taux d'équipement et de pénétration Auto sur des cartes de la région Nord-Pas-de-Calais et d'un schéma de rétention¹ (en fonction de la détention d'un contrat IARD), nous démontrerons l'intérêt du « contrat Fidelio ».

L'objectif étant de détecter en amont les clients susceptibles de quitter l'assurance et de leur proposer le « contrat Fidelio », nous tenterons également, par une modélisation de la durée de vie d'un client bancassurance, de donner une première réponse à la problématique de l'attrition¹.

Dans le cadre de la souscription d'un contrat de fidélité ou d'une problématique de multi équipement, la première difficulté sera d'approcher le client dans sa globalité en déterminant sa valeur au niveau bancassurance. La seconde difficulté sera de définir un avantage tarifaire qui suscitera chez ce dernier l'intérêt de souscrire un nouveau contrat ou de continuer à rester fidèle à sa société d'assurance.

Concernant la qualification de nos clients bancaires non assurés, nous proposons de déterminer un indice de qualité, par le biais d'une modélisation d'un score basés sur nos clients bancaires assurés en Iard et ayant un ratio S/P² nul. S'agissant de la modélisation en fréquence et en coût moyen des contrats automobile, nous intégrerons les différents indicateurs permettant de définir la qualité d'un client bancassurance. Ainsi, nous pourrions déterminer une prime pure³ chargée modélisée que nous comparerons avec la prime pure chargée empirique suivant ces critères de tarification. Nous serons alors en mesure d'affirmer ce que nous constatons depuis des années au niveau bancassurance, à savoir qu'« un bon risque bancaire est un bon risque en assurance », notamment à travers la note Bâle II⁴ du client qui n'est autre qu'une vision de la qualité du client au niveau risque bancaire. Nous

terminerons par une estimation des marges futures sur 5 ans, aussi bien au niveau bancaire qu'au niveau assurance, dans le cadre, de la mise en marché du « contrat Fidelio ».

¹ Attrition – L'attrition est la perte de client. C'est l'inverse de la rétention.

² S/P – ratio entre la charge sinistre et les primes acquises depuis la création du client.

³ 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.

⁴ Bâle II – Les accords de Bâle II ont été signés en 2004 par les gouverneurs des banques centrales du comité de Bâle pour fixer la réglementation bancaire. Ils proposent un certain nombre de normes prudentielles pour limiter les risques bancaires comme le risque de crédit ou de contrepartie, en jouant sur les exigences de capitaux propres. En particulier, ils proposent la mise en place du ratio McDonough, qui limite le montant des prêts accordés en fonction du niveau des capitaux propres et du risque des prêts. Ces normes prennent la suite des normes issues des accords de Bâle I, dont le ratio Cooke, qui ne prenait pas en compte finement le risque. Elles sont entrées pleinement en vigueur au 1er janvier 2008.

ABSTRACT :

« STEADINESS CONTRACT FOR BANKING AND INSURANCE » THE VALUE OF A CLIENT THANKS TO A SCORING. PROPOSAL FOR A SUITABLE PRICE SCALE FIXING FOR INSURANCE CAR (PRODUCT OF APPEALS) IN CASE OF SUBSCRIPTION OF A STEADINESS CONTRACT.

Key words: steadiness contract, multi equipment, net banking income, customer value, pure premium, S/P, CNAT, attrition, retention.

Insurance Companies and Mutual Health Companies have been making use for nearly a decade of the good performance of the market “non life”, both in terms of evolution of total revenue, than in technical results.

Many Companies have been trying to differentiate their products by segmenting their price list in order to be more competitive with almost identical guarantees.

However, working on the product with a differentiated approach gives no guarantee concerning the steadiness of customers, nor allows systematically the sale of an additional contract of insurance. Regarding to these observations, we may think that the most judicious thing to do is to carry a particular interest on one kind of customers.

Indeed, better is to focus on customers who really contribute to the pooling of risk, since they allow us to preserve our profession as insurers.

That is why the purpose of my work will mainly focus on the way we can identify customers with high added value products and on the leader contract “car insurance”, that we will propose to them but with advantageous rates.

Through out the maps of the Nord Pas de Calais area, we will demonstrate the interest of “Fidelio contract”.

To realize this demonstration, we will project the different equipment rates but also the penetration rates concerning “car and house” insurance products. For that, we will present a scheme of retention based on the possession of a general insurance contract.

The aim is to detect “upstream” the customers likely to leave the insurance Company in order to offer them “Fidelio contract”.

We will also try to give an initial response to the problem of attrition by modelling the life of a client in a banking and insurance group.

Within the subscription of a loyalty contract or of a case of additional sales, we will encounter two challenges.

Firstly, we will have to approach the client as a whole, determining his value regarding banking and insurance group.

Secondly, we will have to set up a tariff advantage, so that the client will subscribe a new contract or will just remain faithful to his insurance company.

About the scoring realised on clients in bank who have no insurance products, we suggest to determine a quality index.

In order to be able to do that, we need to do a modelisation and a scoring based on our clients in bank who also have insurance products and a S/P ratio propitious to the company.

Regarding the “modelling frequency” and average cost of insurance cars contracts, we will integrate the different indicators used to define the quality of a customer in banking and insurance group.

Thus, we can determine a pure premium “charged modeled” we compare with the pure premium “charged empirical” following the rating criteria. We will then be able to show what we have been noticing for years in banking and insurance group, that is to say that “a good -credit risk-” is “a good -insurance risk-”, particularly through the “Bâle II rating” of a customer. This rate is nothing but a vision of quality in the customer banking risk.

We will conclude with an estimate of future margins over 5 years (bank and insurance) with the marketing of the “Fidelio contract”.

REMERCIEMENTS

Je tiens tout d'abord à remercier A. CHARPENTIER, Actuaire ENSAE et professeur à l'IA, pour son expertise et ses conseils. Par la mise à disposition de certains de ses travaux ainsi que la communication de multiples références bibliographiques en matière de tarification auto, de réassurance et d'économétrie, il a grandement contribué à la réalisation de ce mémoire.

Je voudrais aussi remercier O. LOPEZ pour sa disponibilité, son soutien et ses précieux conseils au cours de la rédaction de mon mémoire.

Je remercie particulièrement mes supérieurs hiérarchiques directs au sein de la MRACA : JL. CATHERINE Directeur Général et P. METAIS Directeur Financier et Responsable des Etudes Techniques, sans oublier mes collègues de PACIFICA, notamment, P. DEGIOVANNI Directeur Général Adjoint et L. GARRIGA Responsable de l'Actuariat, pour leur soutien, leur relecture attentive ainsi que leurs conseils.

Je remercie également les collègues de la Caisse Régionale Nord De France, et plus particulièrement, F. SIMON pour la traduction du résumé et de la synthèse en anglais.

Enfin, mes dernières pensées vont à mes proches pour leur soutien et leur patience durant toutes les étapes de la réalisation de ce mémoire.

SOMMAIRE

| | |
|--|--------|
| RESUME: | - 5 - |
| ABSTRACT : | - 7 - |
| REMERCIEMENTS | - 9 - |
| SOMMAIRE | - 10 - |
| CONTEXTE DE L'ETUDE | - 15 - |
| 1. Une nouvelle approche du client | - 15 - |
| 2. Les outils utilisés..... | - 15 - |
| INTRODUCTION | - 17 - |
| Chapitre 1 De la réflexion à la conception du « contrat Fidelio »..... | - 19 - |
| 1. Introduction | - 19 - |
| 2. Problématique commerciale et intérêts | - 19 - |
| 3. Présentation du schéma de rétention | - 22 - |
| 4. Conclusion de la première partie..... | - 23 - |
| Chapitre 2 Modélisation de la durée de vie d'un client bancassurance | - 24 - |
| 1. Objectif..... | - 24 - |
| 2. Principes..... | - 24 - |
| 2.1. Notions fondamentales..... | - 24 - |
| 2.2. Relation entre les différentes définitions | - 24 - |
| 2.3. Interprétation Marketing en durée de vie d'un client : fonction de répartition..... | - 25 - |
| 2.4. Taux de risque | - 26 - |
| 2.5. Spécificité des modèles de durée..... | - 26 - |
| 2.6. Différence entre troncature et censure..... | - 26 - |
| 2.7. Effet d'une censure..... | - 27 - |
| 3. Préparation des données | - 27 - |
| 3.1. Etape très importante : la mise en forme de la base de données..... | - 27 - |
| 3.2. La mise en forme de la base de données | - 27 - |
| 3.3. Remarques importantes | - 27 - |
| 4. Les estimateurs des modèles de durée..... | - 28 - |
| 4.1. L'estimation fonctionnelle (ou non paramétrique), | - 28 - |
| 4.2. L'estimation paramétrique, | - 28 - |
| 4.3. L'estimation semi-paramétrique,..... | - 28 - |
| 5. L'estimation non paramétrique | - 28 - |
| 5.1. Estimateur de Kaplan-Meier | - 28 - |
| 5.2. Tests d'égalité | - 30 - |
| 5.2.1 Test du log_rank (de Nathan Mantel)..... | - 31 - |
| 5.2.2 Test de Gehan- Wilcoxon..... | - 31 - |
| 5.3. APPROCHE PARAMETRIQUE sans variables explicatives..... | - 33 - |
| 5.3.1. Lois classiques en modèle de durées | - 34 - |
| a. Loi exponentielle, un paramètre | - 34 - |
| b. Loi de Weibull à deux paramètres..... | - 35 - |
| c. Loi lognormal, deux paramètres..... | - 35 - |
| d. Autres lois possibles..... | - 36 - |
| 5.3.2. Comment choisir une loi ?..... | - 37 - |
| a. Par la méthode du maximum de vraisemblance | - 37 - |
| i. Calcul du maximum de vraisemblance..... | - 38 - |
| ii. Interprétation de la vraisemblance..... | - 38 - |
| b. Par la méthode graphique : « le hazard plotting » | - 40 - |
| i. Calcul des paramètres par le hazard plotting..... | - 40 - |
| ii. Résultats du hazard plotting | - 41 - |
| c. Représentation en un même graphique de l'estimation paramétrique de KM et celle des différents modèles paramétriques testés. | - 42 - |
| d. Tracé des fonctions de risque issues des estimations non paramétrique et paramétrique..... | - 42 - |
| 5.4. APPROCHE PARAMETRIQUE avec variables explicatives | - 43 - |
| 5.4.1. Intégration de variables explicatives | - 43 - |
| 6. Conclusion de la deuxième partie | - 47 - |

| | | |
|------------|--|-------|
| Chapitre 3 | Définition des indicateurs représentant la qualité d'un client en bancassurance et Modélisation de la qualité d'un client au niveau Assurance Iard | 48 - |
| A. | Définition des indicateurs représentant la qualité d'un client en bancassurance | 48 - |
| 1. | Objectif : | 48 - |
| 2. | Définition des données utiles à la construction des différentes variables de qualité au niveau du partenaire commercial : | 48 - |
| B. | Modélisation de la qualité d'un client au niveau Assurance Iard | 50 - |
| 1. | Problématiques de l'étude et Intérêts | 50 - |
| 2. | Monitoring | 50 - |
| 3. | Construction de la vae (variable à expliquer) | 50 - |
| 4. | Exploration des données : examen des distributions | 51 - |
| a. | Statistiques univariées | 51 - |
| 5. | Sélection des variables explicatives corrélées à la variable à expliquer | 53 - |
| a. | Sélection des variables : liaison avec la variable à expliquer des variables explicatives quantitatives | 53 - |
| b. | Sélection des variables : liaison avec la variable à expliquer des variables explicatives qualitatives ou discrétisées | 55 - |
| c. | Sélection des variables quantitatives par analyse discriminante | 57 - |
| d. | Sélection des variables : régressions logistiques sur échantillons bootstrap multiples | 58 - |
| e. | Détermination des corrélations existantes entre variables explicatives | 61 - |
| f. | Synthèses des étapes de sélections de variables | 61 - |
| g. | Sélection de variables après leur classification | 65 - |
| 6. | Modélisation du score | 66 - |
| a. | Construction des échantillons d'apprentissage et de validation | 66 - |
| b. | Calcul de l'aire sous la courbe ROC | 67 - |
| c. | Régression logistique pas à pas sur l'ensemble des variables | 67 - |
| d. | les odds-ratios..... | 69 - |
| e. | Estimation des paramètres..... | 69 - |
| f. | Déviante et mesure de qualité d'un modèle | 71 - |
| g. | Présentation des résultats..... | 72 - |
| i. | Présentation des résultats sur l'échantillon d'apprentissage..... | 72 - |
| ii. | La table de classification d'un modèle | 79 - |
| iii. | Présentation des Courbes ROC et des Courbes LIFT sur l'échantillon d'apprentissage, de validation et total..... | 82 - |
| iv. | Calcul d'une grille de score..... | 84 - |
| C. | Conclusion de la deuxième partie | 92 - |
| Chapitre 4 | Modélisation de la fréquence et du coût moyen | 93 - |
| 1. | Objectif..... | 93 - |
| 2. | Monitoring du portefeuille et intégrité des données..... | 93 - |
| a. | Monitoring du portefeuille | 93 - |
| b. | Intégrité des données | 94 - |
| 3. | Les variables décrivant la sinistralité | 95 - |
| a. | Le nombre de sinistre NIN | 95 - |
| b. | L'occurrence des sinistres IND | 97 - |
| c. | Le coût des sinistres CTOT | 97 - |
| d. | Le coût des sinistres CTOT de 1997 à 2007..... | 98 - |
| 4. | La mesure de l'exposition au risque DUR | 98 - |
| a. | La mesure de l'exposition au risque : la variable DUR..... | 98 - |
| b. | Répartition du nombre de polices et de l'exposition issue du portefeuille 2007 suivant de la durée de la couverture..... | 98 - |
| 5. | Définition des variables à expliquer | 99 - |
| a. | Variables calculées..... | 99 - |
| 6. | Caractéristiques du preneur d'assurance | 99 - |
| a. | Variable Sexe | 99 - |
| 7. | Caractéristiques du véhicule..... | 102 - |
| a. | Variable Type de véhicule..... | 102 - |
| 8. | Caractéristiques de la police..... | 102 - |
| a. | Variable présence d'un conducteur novice occasionnel | 102 - |
| 9. | Interaction entre variables tarifaires | 103 - |
| 10. | Premier tri parmi les variables tarifaires | 104 - |

| | | |
|---|---|---------|
| a. | Test du khi-carré d'indépendance | - 104 - |
| b. | Table de contingence croisant IND et SEXE | - 104 - |
| c. | Dépendances vraie et apparente | - 105 - |
| 11. | Analyse des fréquences de sinistre | - 105 - |
| 11.1. | Modèle de régression de poisson et de negative binomiale pour le nombre des sinistres | - 105 - |
| 11.2. | Explication de l'effet des paramètres sur la prime | - 106 - |
| 11.3. | Présentation des différents modèles de comptage sur un échantillon représentant le nombre de sinistres de 1997 à 2007 vu à fin juillet 2008..... | - 107 - |
| a. | concernant la garantie RCM..... | - 107 - |
| 12. | Analyse des coûts des sinistres..... | - 108 - |
| 1. | Analyse par garantie..... | - 108 - |
| 2. | Présentation des deux lois les plus usitées pour la modélisation du coût individuel de sinistre | - 108 - |
| 3. | Utilisation des Q-Q plots pour déterminer la loi de distribution des montants de sinistre et présentation des statistiques permettant d'avoir une vision globale de la répartition des montants de sinistre par garantie. - | 109 - |
| a. | Q-Q plots..... | - 109 - |
| b. | Présentation de l'échantillon des montants de sinistres de 1997 à 2007 vu à fin juillet 2008 suivant les différentes garanties, telles que :..... | - 110 - |
| 4. | Analyse de la distribution du coût moyen d'un sinistre suivant la garantie : | - 112 - |
| a. | X(i) Montant des sinistres RCM (Responsabilité Civile Matériel) non exaequo et a la franchise égale à 0 - | 112 - |
| b. | X(i) Montant des sinistres RCM (Responsabilité Civile Matériel) exaequo et a la franchise égale à 0- | 113 - |
| c. | X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) non exaequo et a la franchise égale à 0 | - 114 - |
| d. | X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) exaequo et a la franchise égale à 0 - | 115 - |
| e. | X(i) Montant des sinistres DREC (Défense Recours) non exaequo et a la franchise égale à 0...- | 116 - |
| f. | X(i) Montant des sinistres DREC (Défense Recours) exaequo et a la franchise égale à 0..... | - 117 - |
| g. | X(i) Montant des sinistres BDG (Bris De Gglaces) non exaequo et a la franchise égale à 0 | - 118 - |
| h. | X(i) Montant des sinistres DOMM(Dommage tous accidents) non exaequo et a la franchise égale à 0 - | 119 - |
| i. | X(i) Montant des sinistres VOL/INC(Incendie) non exaequo et a la franchise égale à 0..... | - 120 - |
| 13. | Modélisation de la fréquence et du coût moyen d'un sinistre suivant la garantie | - 121 - |
| 13.1. | Analyse de la distribution du coût moyen d'un sinistre suivant la garantie : | - 121 - |
| a. | Méthodologie | - 121 - |
| b. | Présentation des résultats | - 121 - |
| i. | Modèle de régression de la Negative Binomiale pour le nombre de sinistres suivant la garantie Dommage hors forfait IDA | - 122 - |
| ii. | Quelques explications des différentes sorties issues de la Proc Genmod de sas | - 123 - |
| iii. | Quelques explications sur la table de résultats " Model Information" | - 123 - |
| 1. | Sur la table de résultats " Model Information" | - 123 - |
| 2. | Sur la table de résultats "Criteria For Assessing Goodness of Fit"..... | - 124 - |
| iv. | Quelques explications sur la table de résultats "LR Statistics For Type 1 Analysis" and "LR Statistics For Type 3 Analysis" | - 124 - |
| v. | Observation de la distribution des résidus en fonction des paramètres. | - 125 - |
| vi. | Modèle de régression de la loi Gamma pour le coût moyen des sinistres suivant la garantie Dommage hors forfait IDA | - 126 - |
| vii. | Les sorties SAS issues de la Proc Genmod | - 127 - |
| viii. | Observation de la distribution des résidus en fonction des paramètres. | - 128 - |
| ix. | Conclusions | - 129 - |
| 14. | Présentation des facteurs représentant la qualité du client en bancassurance..... | - 132 - |
| 15. | Conclusion de la troisième partie | - 133 - |
| Chapitre 5 Présentation de l'estimation des marges futures aussi bien en assurance qu'au niveau bancaire dans le cadre de la mise en marché du « contrat Fidelio » | | |
| 1. | Objectif..... | - 134 - |
| 2. | Hypothèses | - 134 - |
| a. | Au niveau Assurance..... | - 134 - |
| b. | Au niveau Bancaire | - 135 - |
| 3. | Incertitude sur l'estimation des marges futures | - 137 - |

| | | |
|----|---|---------|
| a. | Le taux d'inflation pris en compte dans le calcul des charges d'exploitation | - 137 - |
| | PREVISIONS INFLATION MOYENNE 2012 (réalisées fin 2011 par différents organismes) | - 137 - |
| | INFLATION PAR GRANDS CONTRIBUTEURS | - 138 - |
| b. | Le taux d'actualisation pris en compte dans le calcul des primes Hors Taxes | - 138 - |
| c. | Les différents taux d'affaires nouvelles et de résiliations, sur les monodétenteurs de contrat Auto ou Mrh, sont estimés à partir des données historisées suivantes : | - 139 - |
| d. | Le taux de résiliations Auto et Mrh sur affaires nouvelles sont estimés à partir des données historisées suivantes : | - 139 - |
| e. | Le taux des marges techniques des contrats Auto et Mrh : | - 140 - |
| 4. | Estimation des marges futures | - 140 - |
| a. | Taux d'affaire nouvelle des partenaires mono détenteurs à 5.3% en Auto et en Mrh | - 140 - |
| b. | Taux d'affaire nouvelle des partenaires mon détenteurs à 7.8% en Auto et en Mrh | - 140 - |
| 5. | Conclusions | - 141 - |
| | SYNTHESE DES RESULTATS | - 142 - |
| | CONCLUSION | - 142 - |
| | ANNEXE I | - 143 - |
| 1. | Statistiques univariées | - 143 - |
| a. | Age de la personne la plus âgée dans le foyer | - 143 - |
| b. | Ancienneté de la relation du client au niveau de l'Assurance Iard | - 144 - |
| c. | Montant de la cotisation Multirisque habitation | - 146 - |
| d. | Nombre de contrats Auto | - 148 - |
| e. | Nombre de produits détenus | - 150 - |
| f. | PCSP Catégorie Sociale Professionnelle de la personne la plus âgée du foyer | - 151 - |
| g. | Segmentation | - 153 - |
| h. | Montant du dépôt sur le Dav de PEA | - 155 - |
| i. | Montant des mouvements débiteurs domiciliés sur le Dav des particuliers sur 12 mois glissants .. | - 157 - |
| j. | Nombre d'entretiens sur une année mobile | - 159 - |
| k. | Solde moyen du Dav sur 12 mois glissants | - 160 - |
| l. | Nombre de contrats Mrh PNO(Propriétaire Non Occupant) | - 162 - |
| m. | Nombre de contrats IVP | - 164 - |
| n. | Nombre de livrets | - 165 - |
| o. | Montant des émissions permanentes | - 167 - |
| p. | Nombre de crédits OC(ouverture d'un crédit suite à un découvert autorisé pour un besoin temporaire le temps d'une transaction financière) | - 169 - |
| 2. | Sélection de variables après leur classification | - 170 - |
| 3. | Présentation des résultats de la modélisation du score sur l'échantillon d'apprentissage | - 178 - |
| | ANNEXE II | - 183 - |
| 1. | Caractéristiques du preneur d'assurance | - 183 - |
| a. | Variable Marché | - 183 - |
| b. | Variable Usage du véhicule | - 183 - |
| c. | Variable Activité | - 184 - |
| d. | Variable Age | - 185 - |
| e. | Variable Ancienneté du permis | - 187 - |
| f. | Variable Zone | - 189 - |
| g. | Variable Agglomération | - 190 - |
| h. | Variable PEL (détention d'un PEL) | - 192 - |
| i. | Variable Actions (détention d'Actions) | - 192 - |
| j. | Variable SECURICOMPTE (détention d'un Securicompte) | - 193 - |
| 2. | Caractéristiques du véhicule | - 193 - |
| a. | Variable Groupe SRA | - 193 - |
| b. | Variable Classe de prix SRA | - 194 - |
| c. | Variable Classe de réparation SRA | - 196 - |
| d. | Variable Type de carburant | - 197 - |
| e. | Variable Type de carrosserie | - 198 - |
| f. | Variable Tranche de mise en circulation du véhicule | - 199 - |
| 3. | Caractéristiques de la police | - 200 - |
| a. | Variable nombre de conducteurs | - 200 - |
| b. | Variable Fractionnement | - 201 - |
| c. | Variable conduite accompagnée | - 202 - |
| d. | Variable Formule | - 202 - |

| | | |
|----|---|---------|
| e. | Variable Pack + | - 203 - |
| f. | Variable Option Franchise | - 204 - |
| g. | Variable Crédit Bonus | - 204 - |
| h. | Variable Tranche de CRM (Coefficient Réduction Majoration) | - 205 - |
| i. | Variable OPTPCC (Option Protection Corporelle du Conducteur)..... | - 206 - |
| j. | Variable OPTMULTI (Option Multiconducteur)..... | - 206 - |
| k. | Variable OPTCTPRIVE (Option Contenu Privé) | - 207 - |
| l. | Variable GARFERME (Garage Fermé) | - 207 - |
| m. | Variable OPTVEHRPLT (Option Contenu Privé) | - 208 - |
| 4. | Analyse de la distribution du coût moyen d'un sinistre suivant la garantie : | - 210 - |
| a. | concernant la garantie RCM..... | - 210 - |
| b. | concernant la garantie RCC..... | - 210 - |
| c. | concernant la garantie DREC | - 211 - |
| d. | concernant la garantie BDG | - 211 - |
| e. | concernant la garantie DOMM..... | - 212 - |
| f. | concernant la garantie INC..... | - 212 - |
| g. | concernant la garantie VOL | - 213 - |
| 5. | Analyse de la distribution du coût moyen d'un sinistre suivant la garantie : | - 214 - |
| a. | X(i) Montant des sinistres RCM (Responsabilité Civile Matériel) exaequo et a la franchise égale à 0-214 - | |
| b. | X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) non exaequo et a la franchise égale à 0 | - 215 - |
| c. | X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) exaequo et a la franchise égale à 0 - 216 - | |
| d. | X(i) Montant des sinistres DREC (Défense Recours) non exaequo et a la franchise égale à 0... - 217 - | |
| e. | X(i) Montant des sinistres DREC (Défense Recours) exaequo et a la franchise égale à 0..... - 218 - | |
| f. | X(i) Montant des sinistres BDG (Bris De Gglaces) non exaequo et a la franchise égale à 0..... - 219 - | |
| g. | X(i) Montant des sinistres DOMM(Dommage tous accidents) non exaequo et a la franchise égale à 0 - 220 - | |
| h. | X(i) Montant des sinistres VOL/INC(Incendie) non exaequo et a la franchise égale à 0..... - 221 - | |
| 6. | Programme sas permettant le calcul de la prime pure par garantie : | - 268 - |
| | BIBLIOGRAPHIE | - 271 - |

CONTEXTE DE L'ETUDE

En préambule, cette partie a pour objectif de préciser au lecteur les motivations et enjeux de l'étude ainsi que les méthodes retenues.

1. Une nouvelle approche du client

Les produits commercialisés en assurance comme en banque sont en grande partie similaires. De plus, l'octroi éventuel d'une « condition particulière » ne garantit pas systématiquement le retour sur investissement. D'autre part, le marché de la bancassurance arrive à saturation et la négociation devient difficile face à une clientèle de plus en plus avertie. Ces différents constats nécessitent une modification de notre stratégie commerciale.

Depuis quelques années, les sociétés investissent en masse dans la création d'un dataware house leur permettant de passer enfin d'une vision produit à une vision orientée client.

Par cette nouvelle approche globale, la proposition d'offre, de produit ou de service s'adapte et répond aux réels besoins du client et permet ainsi au chargé de portefeuille d'être enfin reconnu.

Il est donc nécessaire de donner à nos commerciaux les moyens d'aborder au mieux la négociation commerciale avec leurs clients. En alliant « geste commercial » à « qualité de la relation », ces gestes commerciaux (tels que la diminution d'un tarif d'assurance, d'un taux crédit conso auto, le rachat de franchise, la rétrocession, ...) sont réalisés de manière plus rationnelle dans la mesure où ils tiennent compte de la contribution du client dans le résultat de la société, au-travers de son équipement en bancassurance, de la qualité de son niveau de risque bancassurance, des marges engrangées depuis son entrée en relation et des marges futures susceptibles d'être engendrées.

Notre objectif est donc de chercher à fidéliser ou à multi équiper les clients qui contribuent ou qui contribueront de manière significative à la mutualisation du risque, ceci afin de garantir l'activité de l'assurance.

2. Les outils utilisés.

Il ne s'agit pas, pour l'instant, d'une approche actuarielle.

Toutefois, une étude réalisée sur le taux d'équipement de nos clients particuliers nous a permis de constater que plus des $\frac{3}{4}$ étaient mono détenteurs en produit d'assurance.

Notre objectif est donc de travailler ce potentiel en créant un contrat de fidélité permettant de proposer à nos « meilleurs » clients une tarification avantageuse, et en prenant le contrat « automobile » comme produit d'appel.

Le client pourra ainsi bénéficier de ce contrat de fidélité « contrat Fidelio », si les conditions suivantes sont respectées : détenir à minima deux contrats d'assurance, i.e. l'auto et l'habitation. Bien entendu la résiliation d'un des deux contrats entraînera de facto la résiliation du « contrat Fidelio ».

En parallèle, un schéma de rétention sera créé. Il déclinera les types de détention d'un contrat Iard en fonction des différents profils clients.

Ceci permettra alors de réaliser une modélisation de la durée de vie d'un client bancassurance et d'apporter ainsi une première réponse à la problématique de l'attrition.

Au final, nous disposerons d'un outil permettant d'anticiper notre action sur le départ éventuel du client en lui proposant un « contrat Fidelio » compétitif et adapté.

Pour réaliser cette nouvelle approche, nous avons besoin de qualifier notre clientèle, qu'elle soit bancassurance ou bancaire seule.

Pour la première, nous utiliserons des indicateurs spécifiques.

Par contre, pour la clientèle bancaire seule, ne détenant pas d'information assurance, nous utiliserons un score permettant de qualifier au mieux le client. Ce score reposera sur des données clients au niveau bancaire comme au niveau assurance, ces clients devant être équipés à minima d'un contrat auto et habitat.

Il s'agira alors de modéliser une variable binaire valant 1 si le ratio S/P du client est nul et 0 sinon. Par construction, cette dernière traduira aussi bien la qualité du risque que la marge du client assuré.

Enfin, dans la modélisation de la fréquence et du coût moyen par garantie des contrats en assurance automobile, seront repris comme critères de tarification les différents indicateurs de qualité du client en bancassurance, ce qui permettra de définir, dans la mesure du possible, l'avantage tarifaire lié au contrat automobile.

Nous pourrions ainsi démontrer ce que nous constatons depuis des années au niveau bancassurance, à savoir qu'« un bon risque bancaire est un bon risque en assurance », notamment à travers la note Bâle II du client, considérée comme la qualité du client au niveau risque bancaire.

Mon mémoire ciblera davantage la partie statistique et actuarielle que la partie produit « contrat Fidelio ».

Dans le cadre de la mise en marché du « contrat Fidelio », nous terminerons par une estimation des marges futures sur 5 ans, tant au niveau bancaire qu'au niveau assurance.

INTRODUCTION

En se différenciant depuis quelques années par la segmentation de leur portefeuille automobile, tout en établissant un tarif commercial en adéquation avec le risque assuré, les Sociétés d'Assurance et les Mutuelles ont pris des parts de marché à la concurrence. Cette segmentation tient compte de l'aspect contrat (choix de la formule, du niveau de la franchise, du mode de paiement, ...), de l'aspect type du véhicule assuré (groupe, classe, tranche de mise en circulation, ...), de l'aspect conducteur principal (âge, ancienneté du permis, CRM, ...) et, depuis l'entrée sur le marché du bancassureur PACIFICA, de la détention d'un PEL.

Toutefois, le marché automobile est aujourd'hui saturé. D'une part, le parc évolue très peu (de l'ordre de 1% chaque année). D'autre part, l'innovation du contrat d'assurance automobile reste limitée à quelques prestations de service, garanties par certaines sociétés d'assistance, telles que Mondiale assistance, pour ne citer que la plus connue. Mais d'un point de vue global, les principales grandes compagnies d'assurance proposent en grande majorité les mêmes garanties avec un tarif relativement proche. De plus, durant de nombreuses années, certaines de ces sociétés ont opté pour une stabilisation des montants de cotisations, voire même pour des baisses de tarif, ce qui a entraîné une érosion des marges. Aujourd'hui, Il est donc nécessaire de s'orienter vers une nouvelle approche de notre clientèle qui d'ailleurs, est de plus en plus courtisée et prédisposée à partir.

Afin de répondre à cette problématique, nous proposerons donc au partenaire commercial, (Monsieur, Madame et ses enfants) de souscrire au « contrat Fidelio ». Ce contrat aura pour objectif de garantir deux contrats minimum, un contrat automobile et un contrat habitation, tout en appliquant une réduction tarifaire sur le produit d'appel automobile. Bien entendu, afin de garantir la pérennité de notre activité en assurance, notre démarche supposera que nous accordions cette offre aux partenaires commerciaux contribuant de manière significative à la mutualisation du risque. L'application de cette réduction tarifaire sera issue d'une étude actuarielle sur la tarification automobile, qui sera fonction de la qualité, du risque et de la marge dégagée sur le partenaire commercial bancassurance. Les indicateurs tiendront compte de la note Bâle II, du PNB⁴, de la fréquence de sinistre et du ratio S/P hors CNAT⁵ depuis l'entrée en relation du partenaire s'il est assuré à la Mraca, ou de la note de score dans le cas contraire. Nous répondrons ainsi aux problématiques de « fidélisation » et d'équipement, d'autant plus que, bien souvent, nos partenaires commerciaux bancaires sont mono détenteurs d'un contrat d'assurance automobile ou habitation.

Avant toute chose, nous nous attarderons à démontrer l'intérêt de ce « contrat Fidelio » en illustrant nos propos par une présentation de cartes projetant le taux d'équipement Auto et Habitation et le taux de pénétration Auto, dans la région Nord-Pas-de-Calais, et en s'appuyant sur la présentation d'un schéma de rétention en fonction de la détention d'un contrat Iard, i.e. estimation par la méthode de Kaplan-Meier de la durée de vie d'un client s'il détient, par exemple, un contrat auto ou habitation ou auto et habitation. Nous mettrons alors en évidence que le taux d'équipement est faible, que plus des ¾ de la population assurée sont mono détenteurs de produit auto ou habitation et que le niveau du taux de pénétration nous laisse à penser que nos parts de marché sont promise à une belle embellie, plus particulièrement, dans

⁴ PNB - Produit Net Bancaire représente pour une banque l'équivalent de la valeur ajoutée créée par l'activité.

⁵ CNAT – sinistre(s) issu de(s) Catastrophe(s) naturelle(s).

le Nord, dans les années à venir.

De plus, le fait que le client détienne un contrat auto et habitation augmente sa durée de vie de 6 années tout en sachant que l'estimateur de Kaplan Meier est sous-estimé puisque la donnée est censurée à droite, d'où la légitimité de l'existence du « contrat Fidelio » dans la stratégie commerciale de la Caisse Régionale Nord-De-France. En capitalisant sur les travaux précédents, nous tenterons de donner une première réponse à la problématique de l'attrition en modélisant la durée de vie d'un client bancassurance. Elle se basera sur une approche paramétrique avec des variables explicatives comme l'équipement en assurance dommage, l'ancienneté de la relation en assurance et l'âge du client. L'objet sera de détecter en amont les clients susceptibles de quitter l'assurance et de leur proposer notre offre.

Même s'il ne fait aucun doute quant à l'intérêt du « contrat Fidelio », il ne faut pas oublier que, dans le cadre de la souscription de ce dernier ou d'une problématique de multi équipement, l'une des principales difficultés sera d'approcher le client dans sa globalité. Pour ce faire, nous déterminerons sa valeur au niveau bancassurance et nous définirons un avantage tarifaire qui suscitera chez ce dernier l'intérêt de souscrire un nouveau contrat ou de continuer à rester fidèle à sa société d'assurance.

Concernant la qualification de nos clients bancaires mais non assurés, nous proposons de déterminer la qualité en assurance en modélisant le ratio S/P nul à l'aide d'un score basé sur nos clients bancaires et assurés à minima en assurance Auto et habitation. Concernant la modélisation en fréquence et en coût moyen des contrats automobile, nous intégrerons les différents indicateurs permettant de définir la qualité d'un client bancassurance. Ainsi, nous pourrions déterminer une prime pure chargée modélisée qui sera à comparer avec la prime pure chargée empirique suivant ces critères de tarification. De plus, nous serons en mesure d'affirmer ce que nous constatons depuis des années au niveau bancassurance « un bon risque bancaire est un bon risque en assurance », notamment à travers la note Bâle II du client considérée comme la qualité du client au niveau du risque bancaire. Nous terminerons par une estimation des marges futures sur 5 ans, aussi bien au niveau bancaire qu'au niveau assurance, dans le cadre, de la mise en marché du « contrat Fidelio ».



Chapitre 1 De la réflexion à la conception du « contrat Fidelio »

1. Introduction

Aujourd'hui encore, nous raisonnons contrat et non client et le client bancaire MRA n'est pas perçu dans sa globalité. Nous nous attachons à des segments pour attribuer une réduction tarifaire et non au fait que le client soit multi équipé et voir même si c'est un bon client en terme de marge et de risque assurance et/ou bancaire.

Demain, il nous faudra absolument avoir une approche de nos clients la plus large possible car ils seront de plus en plus exigeants. Nous raisonnerons toujours contrat afin de garantir notre équilibre technique suivant la famille produit, mais, pour répondre à nos objectifs commerciaux, il nous faudra absolument connaître nos clients bancaires équipés ou non de contrats MRA à l'aide de la création d'une base personne reprenant le **partenaire commercial** (à savoir *Monsieur, Madame et ses enfants*).

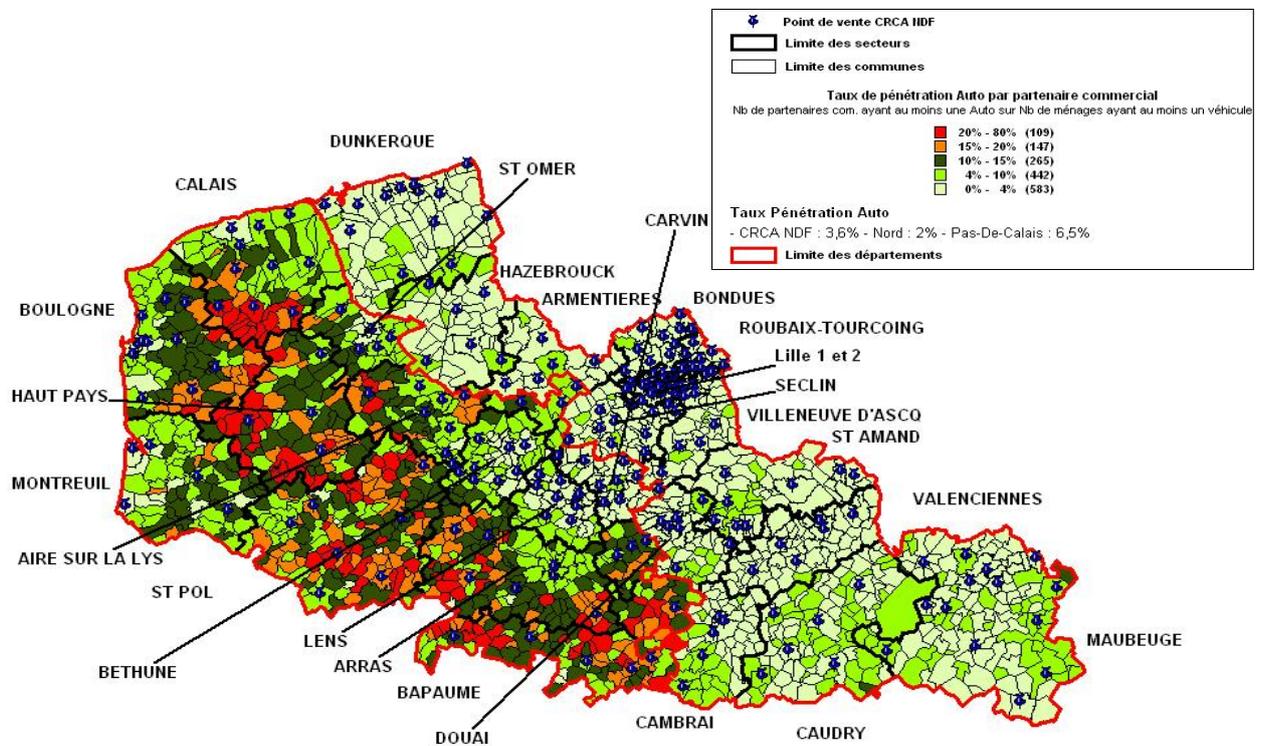
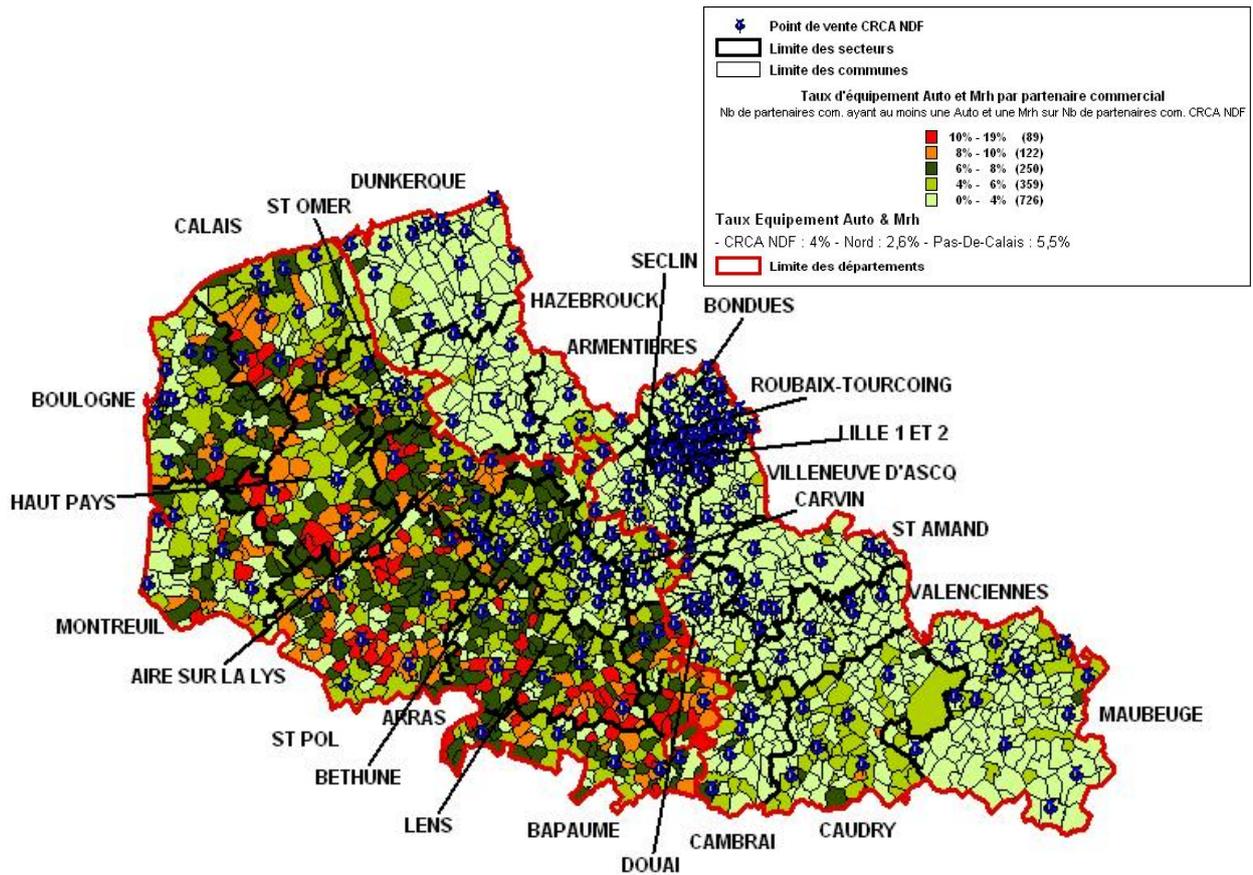
Tout sera basé sur le produit d'appel automobile. Nous proposerons à chaque partenaire commercial CRCA NDF (Caisse Régionale du Crédit Agricole Nord De France) de souscrire le « contrat Fidelio » qui a pour but de les équiper au minimum d'un contrat automobile et habitation en leur accordant une réduction tarifaire sur le produit d'appel automobile en fonction de sa qualité bancassurance.

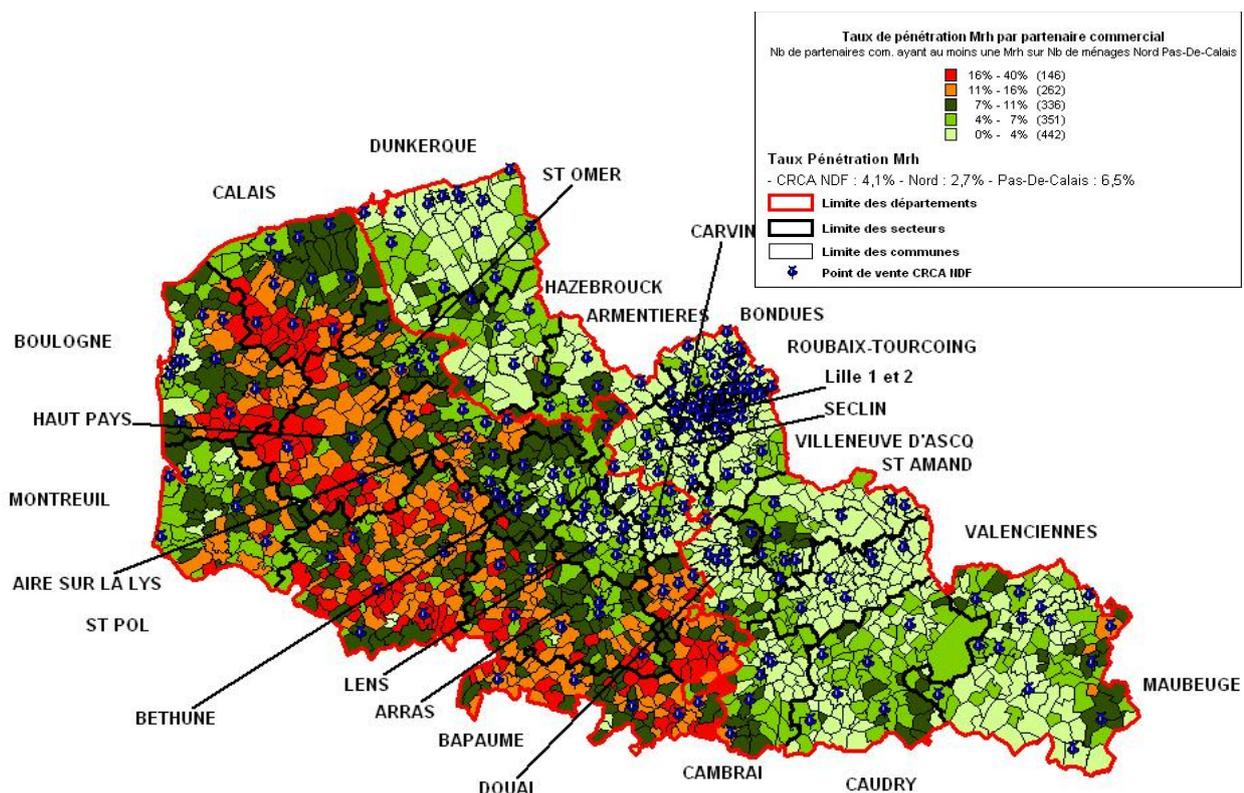
2. Problématique commerciale et intérêts

En fait, l'idée de la conception du « contrat Fidelio » est venu d'un questionnement récurrent de mon entourage qui me demanda clairement « qu'est-ce que tu proposes si je te ramène un ou plusieurs contrat d'assurance ». En clair, il me signifiait que l'effort doit être réalisé dans les deux sens. Le seul problème c'est que des choses ont bien été faites dans ce sens mais en raisonnant d'une part produit et en accordant d'autre part un avantage commerciale qui ne relève à ma connaissance d'aucune étude actuarielle à proprement parler.

Nous allons tenter d'illustrer notre problématique commerciale qui est de fidéliser nos clients bancaires en les équipant au maximum et par conséquent justifier l'intérêt de la mise en place du « contrat Fidelio », à travers la présentation des différentes cartes suivantes :

- ✓ Le taux d'équipement et de pénétration Auto,
- ✓ Le taux d'équipement et de pénétration Mrh,
- ✓ Le taux d'équipement et de pénétration Gav,
- ✓ Le taux d'équipement Auto & Mrh,
- ✓ Le taux d'équipement Auto & Mrh & Gav.





Seules les cartes présentant le taux d'équipement Auto & Mrh et le taux de pénétration Auto et Mrh seront présentées, le reste se trouvant en annexe. Le constat est sans appel, les clients assurés, pour plus des ¾, sont mono détenteurs d'un contrat d'assurance et nous pouvons ajouter que les clients CRCA du Pas-De-Calais sont plus équipés que ceux du Nord, ce qui s'explique principalement par l'ancienneté de la MRACA qui commercialise des produits d'assurance depuis 1986 alors que le Nord a démarré en 1998 avec Pacifica. De plus le département du Pas-De-Calais étant plus rural, il correspond plus à l'image de proximité du Crédit Agricole et la concurrence, hormis Groupama étant moins agressive et moins présente que dans le Nord puisque historiquement le siège national et régional de certaines grandes assurances comme SwissLife et Axa sont présentes. La Macif est aussi très présente dans le Nord.

Ainsi seul 4% des clients de la CRCA NDF sont équipés en assurance Auto et Mrh et 2.5% dans le Nord. Si nous regardons maintenant le taux de pénétration en Auto et en Mrh pour la Mraca, nous avons des chiffres du même ordre que le taux d'équipement, soient respectivement 3.6% et 4.1% et 2% et 2.7% pour le Nord. Nous pouvons donc affirmer sans hésiter qu'il existe un réel potentiel d'équipement pour la CRCA NDF, dans les années à venir, et plus particulièrement pour le Nord. Assurément, la mise en marché du « contrat Fidelio » peut booster le taux d'équipement de la CRCA NDF.

L'intérêt est donc double :

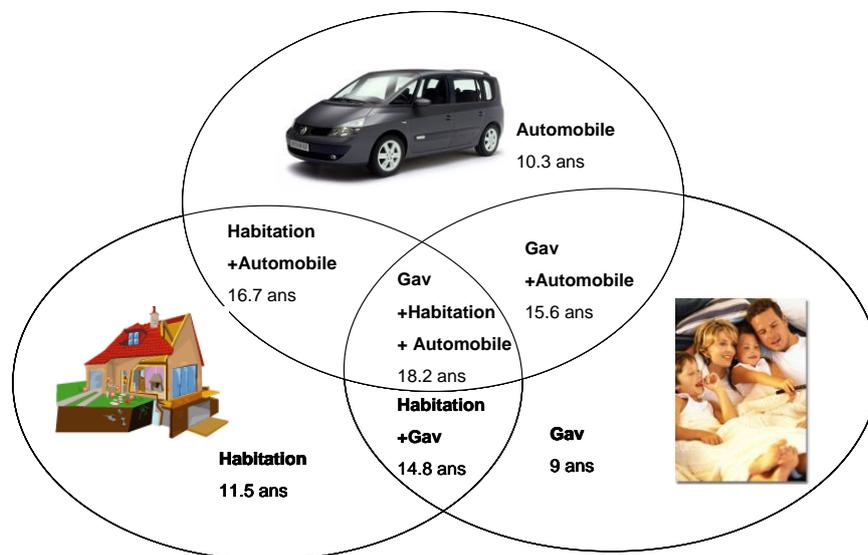
- ✎ Avoir une approche globale en donnant à la disposition du réseau la qualité et la valeur du partenaire commercial notamment pour rationaliser l'attribution des différentes enveloppes commerciales ou autres avantages (exonération de certains frais, rétrocession de certaine facturation, taux préférentiel crédit ou Adi, ristournes MRACA, exonération de franchise...),
- ✎ Orienter ou conforter la politique commerciale impulsée par la direction en garantissant un niveau de risque et de rentabilité.

Nous pourrions aussi tenir compte de la politique commerciale orientée par CASA (Crédit Agricole SA) et ses filiales en accordant plus d'importance aux différents produits phares de l'année.

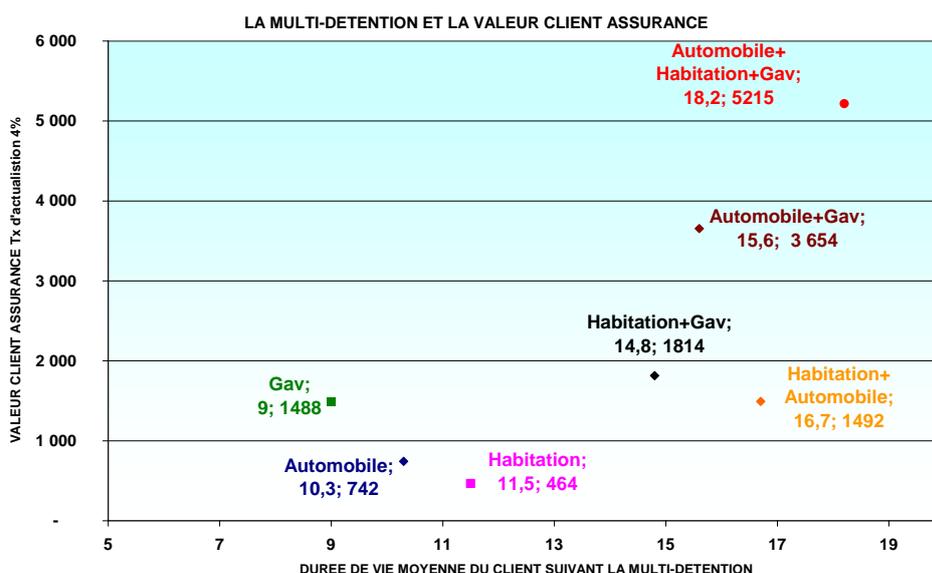
3. Présentation du schéma de rétention

Le schéma de rétention et le graphe « La multi-détention et la Valeur Client² Assurance » ci-dessous confortent la mise en marché du « contrat Fidelio » puisque, le fait que le client détienne un contrat auto et habitation, ceci augmente sa durée de vie d'environ 6 années tout en sachant que l'estimateur de Kaplan Meier est sous-estimé car la donnée est censurée à droite. Une explication sera donnée dans le chapitre suivant quant à la méthode employée pour estimer la durée de vie du client suivant son équipement.

Schéma de rétention en fonction de la possession de contrats ¹



1 : Source : Mraca 2010



² Valeur client – est le produit entre la marge technique et $(1 + \text{taux d'actualisation})^{\text{durée de vie du client}}$.

Les actions de cross selling sont donc d'un grand intérêt pour la bancassurance car, outre leur aspect fidélisant, elles génèrent :

- Du chiffre d'affaires supplémentaire,
- Des nouveaux contrats au prix d'un effort commercial allégé,
- De plus, certaines sociétés ont analysé avec précision le parcours de la multi-détention, c'est-à-dire, selon les segments de clientèle, quel est l'ordre de souscription des contrats (par exemple l'auto, puis l'habitation puis santé ou, dans la bancassurance : l'habitation en premier),
- La capacité de fidélisation de chaque type de contrat.

Au final, cette fidélisation par la multi-détention joue l'inertie. Cependant, pour que le client accepte de souscrire une deuxième puis une troisième assurance, il doit être satisfait des prestations et de sa relation avec son bancassureur. Une politique de multi-détention non soutenue par une recherche de la satisfaction des clients n'aurait pas de sens.

Pour finir cela peut se traduire par une application marketing notamment vers un score d'attrition.

Plutôt que de faire une régression logistique sur le profil des clients assurés en fin de relation en assurance, on cherche les variables explicatives qui accélèrent ou ralentissent l'attrition. Le modèle paramétrique répond à cette problématique et c'est ce que nous chercherons à faire dans le chapitre suivant.

4. Conclusion de la première partie

Nous pouvons conclure sans hésitation que le « contrat Fidelio » a toute sa place dans la stratégie commerciale d'équipement en assurance de la CRCA NDF compte tenu du faible taux d'équipement et de pénétration du marché et plus particulièrement dans le Nord. De plus, le schéma de rétention montre que si nous souhaitons que notre client mono détenteur d'un contrat d'assurance auto ou Mrh augmente sa durée de vie de minimum 6 ans il doit être équipé au minimum d'un contrat d'assurance Auto et Mrh.

Une fois de plus nous renforçons l'idée que la mise en marché du « contrat Fidelio » serait utile pour la réussite d'une campagne commerciale des deux produits phare en assurance sur le marché des particuliers. Mais il faudra entreprendre une politique de multi-détention soutenue par une recherche de la satisfaction des clients sinon l'intérêt du « contrat Fidelio » serait temporaire et l'attrition augmenterait.

Chapitre 2 Modélisation de la durée de vie d'un client bancassurance

1. Objectif

Après une présentation de l'aspect théorique de la modélisation de la durée de vie, nous tenterons de donner une première réponse à la problématique de l'attrition en modélisant la durée de vie d'un client bancassurance qui se base sur une approche paramétrique avec des variables explicatives comme l'équipement en assurance dommage, l'ancienneté de la relation en assurance et l'âge du client. L'objectif est bien de détecter en amont les clients susceptibles de quitter l'assurance et de leur proposer le « contrat Fidelio ».

2. Principes

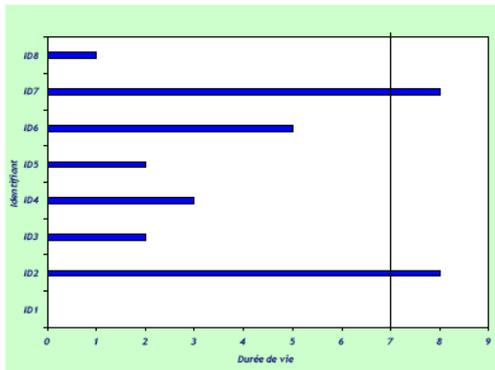
2.1. Notions fondamentales

- Durée de vie : temps T que l'on veut modéliser (choisir l'échelle)
- Fonction de densité associée : $f(x)$, distribution de la durée
→ Probabilité instantanée de « fin de relation en assurance Iard »
- Fonction de répartition : $F(x) = P(T \leq x)$.
→ Probabilité que le temps étudié soit inférieur ou égal à x .
- Fonction de survie : $S(x) = P(T > x)$.
→ Probabilité que le temps étudié soit supérieur à x .
- Taux de risque : $\lambda(x) = \lim_{h \rightarrow 0} \{P(T > x+h | T > x)/h\}$.
→ Probabilité d' « attrition » à l'instant x sachant qu'on était encore en « relation en assurance Iard » à l'instant précédent.

2.2. Relation entre les différentes définitions

- $S(x) = 1 - F(x)$
- $F(x)' = f(x)$
- $$\lambda(x) = \frac{f(x)}{S(x)} = - \frac{d(\log(S(x)))}{dx}$$
- $$S(x) = \exp \left\{ - \int_0^x \lambda(u) du \right\}$$

Durées de vie



La **durée de vie** est :

- Toujours positive
- Toujours inférieure à l'intervalle de temps étudié
- Débute à une date T_0 puis finit :
 - o Soit à une date T , inférieure à la date de fin de l'étude
 - o Soit à une date T , supérieure à la date de fin de l'étude. On dira alors que la donnée est censurée à droite et on utilisera la date de fin d'étude comme date de résiliation fictive

Survie et fonction de répartition

▪ Survie en bleu

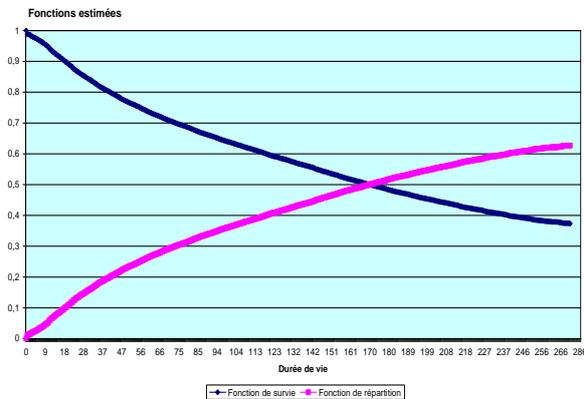
Elle est toujours décroissante, débute à 1 et tend vers 0. La probabilité qu'un client soit toujours actif après un mois est plus grande qu'après 12 mois.

→ Elle décroît en fonction des clients qui partent.

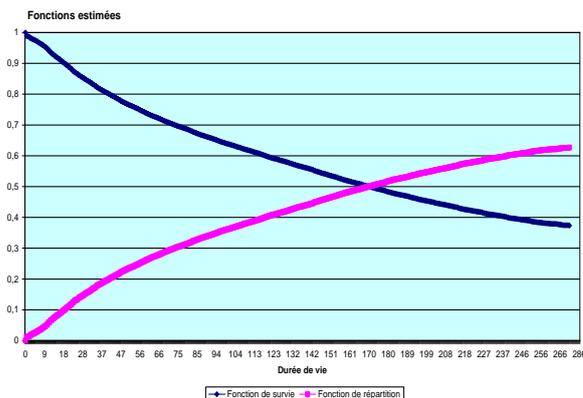
▪ Fonction de répartition en violet

Elle est toujours croissante, débute à 0 et tend vers 1. La probabilité qu'un client parte au bout d'un mois est plus petite qu'au bout de 12 mois.

→ Elle croît en fonction des clients qui partent.
Deux manières de dire la même chose.



2.3. Interprétation Marketing en durée de vie d'un client : fonction de répartition

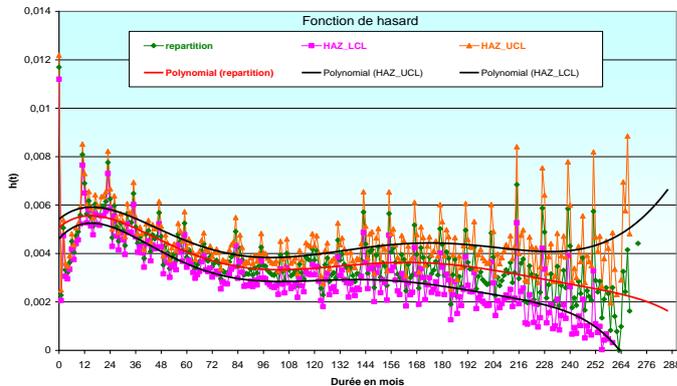


▪ Fonction de répartition

A $t=120$, 120 mois, 40% des clients sont déjà partis, ils ont résilié entre 0 et 120 mois
La durée de vie médiane correspond à $F=50\%$ soit environ 170 mois.
Le premier quartile de durée de vie correspond à $F=25\%$ soit environ 56 mois.

2.4. Taux de risque

Taux de risque associé aux clients : risque en t que survienne l'évènement « fin de la relation en assurance »



- Il peut être supérieur à 1 et donner des indications précieuses
- S'il est constant alors la loi associée est exponentielle
- S'il est croissant alors la survie sera concave, s'il est décroissant alors la survie sera convexe

Il existe des fonctions de risque typiques. La courbe en U étant la plus connue. Elle représente la force de mortalité humaine mais s'applique à d'autres considérations : une phase de début avec un fort taux d'attrition, puis la phase constante avec un taux d'attrition constant et la phase de fin avec un taux d'attrition croissant.

2.5. Spécificité des modèles de durée

Censure et troncature

Les données peuvent être incomplètes. En observant la durée de vie d'un client en assurance, à la date d'étude T , plusieurs possibilités s'offrent à nous :

- On connaît la date d'entrée en relation du client en assurance (date du 1^{er} contrat Iard souscrit), la date de fin de la relation en assurance (dernier contrat Iard résilié) et il n'est donc plus client à la Mraca à T => durée complète de vie du client en assurance connue.
- On connaît la date d'entrée en relation du client en assurance (date du 1^{er} contrat Iard souscrit), et le client est toujours en relation à T => durée partielle de vie du client connue, il peut encore être en relation un certain temps inconnu.
- On ne connaît pas la date de fin de relation et le client n'est plus assuré à la Mraca à T => durée partielle de vie du client connue.
- On ne connaît pas la date de début de la relation du client en assurance et le client est toujours assuré à la Mraca à T => durée partielle de vie du client connue.

2.6. Différence entre troncature et censure

Lorsque l'on fait une troncature, on supprime les données incomplètes de la base de données. Ceci implique une perte d'informations, assumée.

Lorsque l'on définit une censure, on intègre cette donnée dans le modèle de survie. La façon dont elle est intégrée dépendra du type de la censure :

- Donnée **censurée à droite** : Cas le plus fréquent, lorsque la durée qu'on observe est un minorant de la durée réelle
 - La date de fin de la relation du client en assurance est soit postérieure à la date d'étude,
 - Soit inexistante car l'évènement ne s'est pas encore produit : on sait qu'il est encore présent à la fin de l'étude

2.7. Effet d'une censure

- On modifie la contribution à la vraisemblance de la donnée censurée
- La contribution à la vraisemblance du modèle dépendra du type de censure :
 - La densité pour une donnée non censurée
 - La fonction de répartition pour une donnée censurée à gauche
 - La fonction de survie pour une donnée censurée à droite

3. Préparation des données

3.1. Etape très importante : la mise en forme de la base de données

- Qu'est-ce que l'on veut modéliser ?
 - Durée de vie des clients en assurance Iard sur la base de données censurée à droite
- Qualité des données
 - S'assurer que des données n'ont pas été purgées => on étudie le temps d'adhésion en prenant les contrats activés après 1986, il faut s'assurer que les résiliations intervenues depuis plus de 5 ans n'ont pas été purgées du système
 - Regarder la cohérence entre les dates, par exemple date de fin > date de début
 - Traiter les dates incorrectes
 - Traiter les problèmes de dates manquantes
- Choix d'une date de fin d'étude au-delà de laquelle l'évènement n'est plus pris en compte

3.2. La mise en forme de la base de données

- Calcul des durées de vie
- Uniformiser les durées calculées
 - Echelle : en jours, en **mois**, en trimestres...
 - Mettre au même format sachant que l'estimation des coefficients est plus difficile en cas d'égalités, il vaut mieux choisir un pas de mesure fin en l'occurrence **ici le mois**
- Définition des censures
 - Selon la date de fin d'étude
 - Cohérence dans les notations
 - Si la variable s'appelle censure : 1 signifie censure
 - Si la variable s'appelle du nom de l'évènement (resil →résiliation de tous les contrats Iard) : 0 signifie censure
- Ajout de variables extérieures après une analyse minimale

3.3. Remarques importantes

- Ne pas confondre date manquante et donnée censurée. Une date peut être non renseignée, par exemple une date de fin d'adhésion, alors que le client n'est plus dans l'agence. Ceci relève de la qualité des données

- Deux variables sont essentielles à la modélisation des durées de vie :
 - la variable de durée et
 - la variable de censure.

4. Les estimateurs des modèles de durée

Les modèles de durée se prêtent à divers types d'estimation :

4.1. L'estimation fonctionnelle (ou non paramétrique),

qui vise à approximer une ou plusieurs des différentes fonctions caractérisant la distribution observée (fonction de répartition F ou fonction de risque h le plus souvent) sans faire d'hypothèse sur celle-ci.

4.2. L'estimation paramétrique,

qui ayant retenu une forme de distribution donnée (par exemple, la loi de Weibull), cherche à en estimer les paramètres. Un terme correctif peut prendre en compte l'effet de variables exogènes, appelées covariables.

4.3. L'estimation semi-paramétrique,

qui est un compromis entre les deux approches précédentes. En effet, elle n'introduit pas d'hypothèse (autre que la régularité) sur les fonctions de densité et/ou de risque, mais fait des hypothèses sur la manière dont les variables exogènes vont influencer le déroulement du phénomène temporel.

→ Nous n'entamerons pas le troisième point dans ce mémoire.

5. L'estimation non paramétrique

L'approche non paramétrique permet d'estimer la fonction de survie en se basant uniquement sur les données de retour d'expérience, sans suppositions supplémentaires.

Deux méthodes peuvent être utilisées :

L'estimation par la méthode actuarielle,

L'estimation par la méthode de Kaplan-Meier.

Seule la dernière méthode sera développée car plus appropriée.

5.1. Estimateur de Kaplan-Meier

Cet estimateur se base sur l'apparition dans le temps des résiliations rapportées au nombre restants de clients

- Au temps 0- la survie est égale à 1
- Puis c'est un algorithme récurrent ; à chaque résiliation, la survie diminue.
- Cette survie tend vers 0, mais cela est valable pour un nombre de données tendant vers l'infini.

Soit D_i le nombre de résiliations à l'instant t_i , N_i le nombre de clients à risque à l'instant t_i et C_i le nombre de censures à l'instant t_i alors l'estimateur de la fonction de survie de Kaplan Meier à l'instant t_i se définit par :

$$S(t_i) = \prod_{k=1}^i \left(1 - \frac{D_k}{N_k}\right) = S(t_{i-1}) \times \left(1 - \frac{D_i}{N_{i-1} - C_i}\right)$$

La survie est un produit : proba de ne pas avoir résilié tous ces contrats en t=2 = proba de ne pas avoir résilié tous ces contrats en T=1 x (1- proba de résilier tous ces contrats en t=2)

On peut estimer un écart type :

- $\sigma[S(t_i)] = S(t_i) \times \sqrt{\sum_{k=1}^i \frac{D_k}{N_k \times (N_k - D_k)}}$

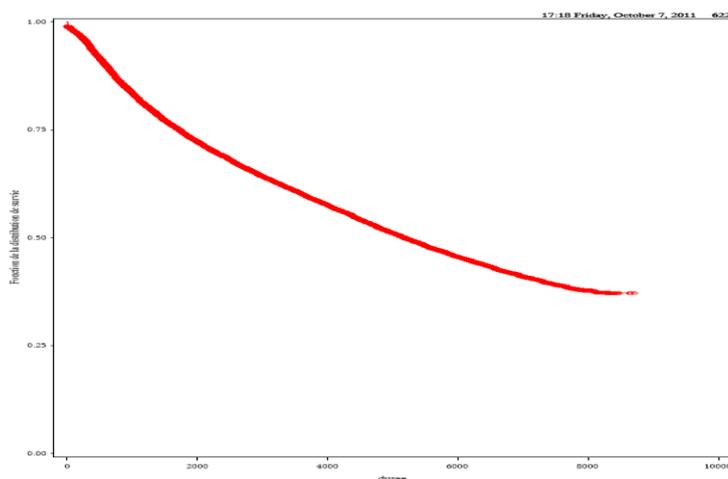
Et en déduire un intervalle de confiance :

- $IC[S(t_i)] = [S(t_i) \pm (Z_{\alpha/2} \times \sigma(S(t_i)))]$

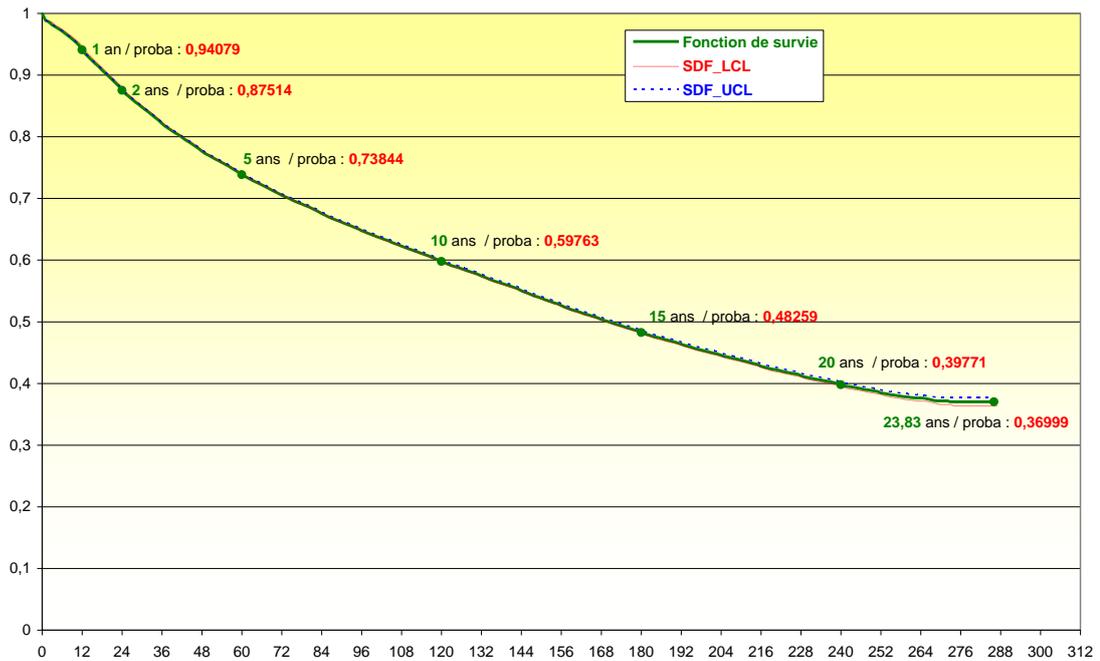
On peut facilement calculer les quartiles. On peut aussi calculer la moyenne de la durée de vie qui correspond à l'intégrale sous la courbe de survie :

- $\mu = \sum_{k=1}^n S(t_{k-1}) \times (t_k - t_{k-1})$ avec $t_0 = 0$
- Médiane : temps t où 50% de la pop est parti (F) ou a survécu (S)

Nous avons donc la représentation graphique de la fonction de survie de Kaplan Meier suivante avec une durée exprimée en jours :



Présentation de la fonction de survie avec une durée exprimée cette fois-ci en mois. Nous pouvons en déduire, par exemple, que la probabilité qu'un client soit toujours en relation avec la bancassurance après 10 ans est égale à 0.60



L'estimation de la durée de vie moyenne d'un client Mraca est de 4947.78 jours, soit 13.6 ans avec un écart type de 9.97 jours et une médiane à 5191 jours, soit 14.21 ans.

Statistiques descriptives pour variable temps duree

| Estimations du quartile | | | | |
|-------------------------|-------------------------|--------------------------------|------------|------------|
| Pourcentage | Valeur estimée du point | Intervalle de confiance à 95 % | | |
| | | Transformation | [Inférieur | Supérieur) |
| 75 | | LOGLOG | . | . |
| 50 | 5191.00 | LOGLOG | 5127.00 | 5248.00 |
| 25 | 1729.00 | LOGLOG | 1706.00 | 1740.00 |

| Moyenne | Erreur type |
|---------|-------------|
| 4947.78 | 9.97 |

Note: The mean survival time and its standard error were underestimated because the largest observation was censored and the estimation was restricted to the largest event time.

| Récapitulatif du nombre de valeurs censurées et non censurées | | | |
|---|-------------|---------|------------------------|
| Total | A échoué | Censuré | Pourcentage censuré |
| 186665 | 61298 | 125367 | 67.16 |

5.2. Tests d'égalité

Remarque générale : SAS affiche la Pvalue qui est toujours la proba de rejeter à tort H_0 , H_0 étant la nullité/égalité.

On peut être en présence d'une variable qui classe les clients. Celle-ci peut influencer considérablement la durée d'adhésion. Dans ce cas, on va estimer la durée de vie du client différemment en fonction de l'appartenance de l'individu à telle ou telle classe.

Les tests servent à savoir si la survie est homogène dans les groupes, s'il y a N groupes, il y a 1 seul test donc le test ne dit pas si le groupe A diffère de D, par contre le graphique permet de le pressentir. Pour conclure à une différence entre A et D, il faut se restreindre à ces deux groupes.

Les tests exposés sont les suivants:

- Test du log_rank
- Test de Gehan-Wilcoxon

Attention ! Ces tests ne sont valables que si les courbes de survie des différentes strates ne se croisent pas, la statistique de test contient des éléments positifs et négatifs qui se compensent, ce qui rend le test inopérant.

5.2.1 Test du log_rank (de Nathan Mantel)

- Principe : est-ce que la durée de vie de la strate diffère de celle de la population totale
- Comparaison de l'attrition observée sur la strate à l'attrition attendue si la strate a la même distribution que la population,
- Prise en compte des censures de la même manière que dans l'estimateur de Kaplan Meier,
- La somme des comparaisons obtenues suit un Chi2 à k degrés de liberté avec k = nombre de strates - 1,
- On peut ensuite en déduire une p-value grâce aux abaques du Chi2,
- Selon le seuil choisi (par ex : 0.05), on en déduit l'acceptation ou le rejet du test,
 - Test accepté => les survies sont identiques
 - Test refusé => les survies sont significativement différentes

5.2.2 Test de Gehan- Wilcoxon

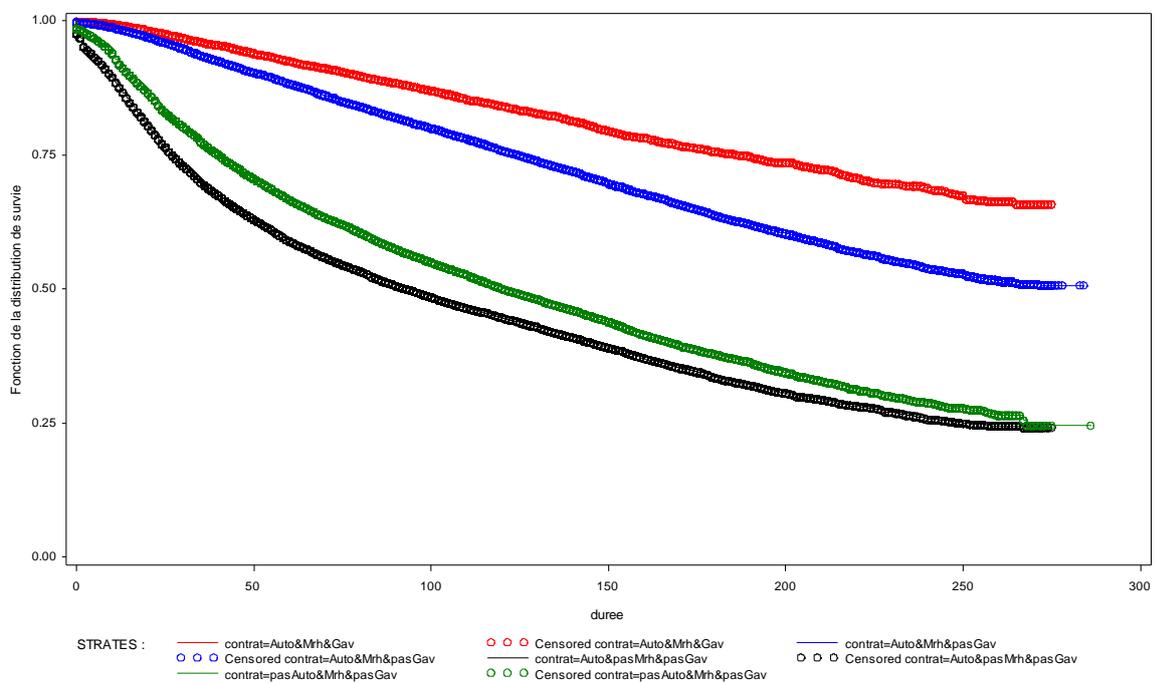
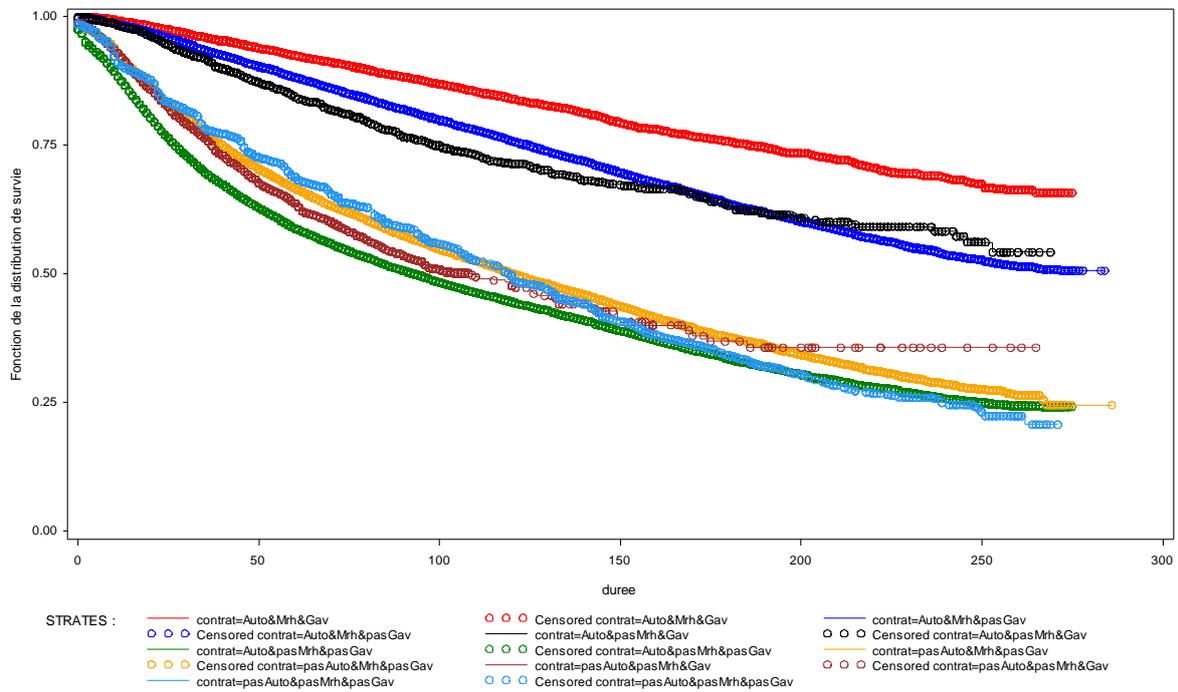
- Identique au précédent sauf la pondération
- Poids plus important donné aux résiliations dans la 2^{ème} année après le premier contrat souscrit : à utiliser pour les durées de vie de courte durée.
- La statistique s'écrit

$$\chi^2 = \frac{\left[\sum_{i=1}^k w_i \left(m_{Pi} - m_i \frac{n_{Pi}}{n_i} \right) \right]^2}{\sum_{i=1}^k w_i^2 m_i \frac{(n_i - m_i) n_{Si} n_{Pi}}{n_i}}$$

- Dans le test du log rank, $w_i = 1$ pour tout i
- A noter que SAS affiche l'expression du numérateur = log rank ²
- Le log rank est la somme des différences entre évènements observés et attendus, plus il est grand, plus le test est concluant
- Si les courbes se croisent, les différences sont positives puis négatives, le test conclura à l'égalité alors que les formes peuvent différer

Voici les résultats obtenus sur les clients Mraca selon une variable à 4 modalités (équipement Auto & Mrh & Gav)

Estimation de la fonction de survie par Kaplan-Meier



Résultats des tests d'égalité :

Procédure LIFETEST

Test de l'homogénéité des courbes de survie pour durée sur les niveaux de discrétisation

| Statistiques de classement | | |
|----------------------------|----------|----------|
| contrat | Log-rang | Wilcoxon |
| Auto&Mrh&Gav | -2461.0 | -2.364E8 |
| Auto&Mrh&pasGav | -9630.4 | -1.033E9 |
| Auto&pasMrh&pasGav | 6040.0 | 6.8377E8 |
| pasAuto&Mrh&pasGav | 6051.4 | 5.8511E8 |

| Matrice de covariance pour la statistique de log-rang | | | | |
|---|--------------|-----------------|--------------------|--------------------|
| contrat | Auto&Mrh&Gav | Auto&Mrh&pasGav | Auto&pasMrh&pasGav | pasAuto&Mrh&pasGav |
| Auto&Mrh&Gav | 3461.9 | -1606.0 | -630.5 | -1225.4 |
| Auto&Mrh&pasGav | -1606.0 | 12475.8 | -3720.8 | -7149.0 |
| Auto&pasMrh&pasGav | -630.5 | -3720.8 | 7530.8 | -3179.5 |
| pasAuto&Mrh&pasGav | -1225.4 | -7149.0 | -3179.5 | 11554.0 |

| Matrice de covariance pour la statistique de Wilcoxon | | | | |
|---|--------------|-----------------|--------------------|--------------------|
| contrat | Auto&Mrh&Gav | Auto&Mrh&pasGav | Auto&pasMrh&pasGav | pasAuto&Mrh&pasGav |
| Auto&Mrh&Gav | 3.308E13 | -1.27E13 | -6.48E12 | -1.39E13 |
| Auto&Mrh&pasGav | -1.27E13 | 1.284E14 | -3.69E13 | -7.88E13 |
| Auto&pasMrh&pasGav | -6.48E12 | -3.69E13 | 8.531E13 | -4.2E13 |
| pasAuto&Mrh&pasGav | -1.39E13 | -7.88E13 | -4.2E13 | 1.346E14 |

| Test d'égalité sur niveaux de discrétisation | | | |
|--|------------|-----|------------|
| Test | Khi-2 | DDL | Pr > Khi-2 |
| Log-rang | 12242.7065 | 3 | <.0001 |
| Wilcoxon | 13006.3566 | 3 | <.0001 |
| -2Log(LR) | 14491.0018 | 3 | <.0001 |

Le premier graphe nous montre que les courbes de survie de certaines strates se croisent. Nous décidons alors de reprendre le test mais avec un nombre restreint de strates d'où le deuxième graphe et le Test de l'homogénéité des courbes de survie pour une durée de vie sur les niveaux de discrétisation.

On peut donc en déduire que les durées de vie entre les 4 strates sont significativement différentes.

Une étude différente pour chaque strate est justifiée.

Remarque : Le troisième test est celui du ratio de vraisemblance, sous l'hypothèse d'une distribution exponentielle, il teste l'égalité des paramètres d'échelle entre les groupes.

5.3. APPROCHE PARAMETRIQUE sans variables explicatives

Principe : La variable aléatoire de durée suit une loi paramétrique connue, il n'y a pas de variables explicatives.

Conséquence : On cherche la forme de la distribution, le raisonnement à l'individu n'a pas de sens, en effet tous auront pour tout t, la même valeur de S, f, F, h.

Il n'y a pas de prédiction individuelle possible car si on le faisait les prévisions suivraient de manière aléatoire la forme de la loi postulée.

Objectif : Trouver la loi paramétrique qui estime le mieux la fonction de survie étudiée

- Description des lois fréquemment utilisées
- Méthodes de sélection d'une loi
- Estimation par l'estimateur du Maximum de vraisemblance
- Application sur la durée de vie d'un client bancassurance Mraca

5.3.1. Lois classiques en modèle de durées

a. Loi exponentielle, un paramètre

- Bon pour un taux de risque constant, i.e. la probabilité de résilier à l'instant t sachant que la personne qui est assurée à l'instant précédent est la même pour tout t.
- Loi à un seul paramètre
 - Paramètre d'échelle : λ (lambda)
 - Sas fournit l'estimation de λ deux fois mais indirectement, dans intercept et weibull scale
 - $\lambda = \exp(-\text{intercept}) = 1/\text{weibull scale}$
- Densité : $f(t) = \lambda.e^{-\lambda t}$
- Fonction de survie : $S(t) = e^{-\lambda t}$
- Taux de risque constant : $h(t) = \lambda$
- Espérance : $E(t) = 1/\lambda$

L'exécution de la Lifereg fournit les résultats suivants :

Le Système SAS
Procédure LIFEREG

| Informations sur le modèle | |
|-----------------------------------|--------------|
| Table | WORK.TELCO2 |
| Variable dépendante | Log(durée) |
| Variable de censure | rsell |
| Valeur(s) de censure | 0 |
| Nombre d'observations | 183032 |
| Valeurs non censurées | 59137 |
| Valeurs censurées à droite | 123895 |
| Valeurs censurées à gauche | 0 |
| Valeurs censurées de l'intervalle | 0 |
| Number of Parameters | 1 |
| Réponse zéro ou négative | 3633 |
| Name of Distribution | Exponential |
| Logarithme de la vraisemblance | -168953.5791 |

| | |
|-----------------------------|--------|
| Nb d'observations lues | 180665 |
| Nb d'observations utilisées | 183032 |

| Statistiques d'ajustement | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 337907.2 |
| AIC (préférer les petites valeurs) | 337909.2 |
| AICC (préférer les petites valeurs) | 337909.2 |
| BIC (préférer les petites valeurs) | 337919.3 |

| Fit Statistics (Unlogged Response) | |
|---|----------|
| -2 log-vraisemblance | 762577.3 |
| Exponential/AIC (préférer les petites valeurs) | 762579.3 |
| Exponential/AICC (préférer les petites valeurs) | 762579.3 |
| Exponential/BIC (préférer les petites valeurs) | 762589.4 |

L'algorithme a convergé.

Procédure LIFEREG

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--------------------------------|----------|----------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance à 95 % | | Pr > Khi-2 |
| Intercept | 1 | 5.4475 | 0.0041 | 5.4395 | 5.4556 | 1754937 <.0001 |
| Scale | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | |
| Echelle de Weibull | 1 | 232.1881 | 0.9548 | 230.3242 | 234.0670 | |
| Forme de Weibull | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | |

| Statistique du multiplicateur de Lagrange | | |
|---|----------|------------|
| Paramètre | Khi-2 | Pr > Khi-2 |
| Echelle | 282.8020 | <.0001 |

b. Loi de Weibull à deux paramètres

- Loi la plus utilisée, monotone Croissante ou décroissante
- Loi à deux paramètres
 - Paramètre d'échelle : η (eta) , $1/\eta = \exp(-\text{intercept})/\exp(\text{scale}) = (1/\text{weibull scale})/\exp(\text{scale})$ dans sas
 - Paramètre de forme : β (beta) weibull shape, si $\beta = 1$ c'est la loi exp

• Densité :
$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta}$$

• Fonction de survie :
$$S(t) = e^{-\left(\frac{t}{\eta}\right)^\beta}$$

• Taux de risque :
$$h(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1}$$

- Si $\beta > 1$, le risque croît avec le temps (usure), Sinon, le risque décroît (rodage)

L'exécution de la Lifereg fournit les résultats suivants :

Le Système SAS 11:22 Sunday, October 9, 2011 3

Procédure LIFEREG

| Informations sur le modèle | |
|-----------------------------------|--------------|
| Table | WORK.TELCO2 |
| Variable dépendante | Log(duree) |
| Variable de censure | zeit |
| Valeur(s) de censure | 0 |
| Nombre d'observations | 183032 |
| Valeurs non censurées | 59137 |
| Valeurs censurées à droite | 123893 |
| Valeurs censurées à gauche | 0 |
| Valeurs censurées de l'intervalle | 0 |
| Number of Parameters | 2 |
| Reponse zéro ou négative | 3633 |
| Name of Distribution | Weibull |
| Logarithme de la vraisemblance | -148890.2639 |

| | |
|-----------------------------|--------|
| Nb d'observations lues | 186665 |
| Nb d'observations utilisées | 183032 |

| Statistiques d'ajustement | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 337600.5 |
| AIC (préférer les petites valeurs) | 337604.5 |
| AICC (préférer les petites valeurs) | 337604.5 |
| BIC (préférer les petites valeurs) | 337624.8 |

| Fit Statistics (Unlogged Response) | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 762270.6 |
| WeibullAIC (préférer les petites va | 762274.6 |
| WeibullAICC (préférer les petites v | 762274.6 |
| WeibullBIC (préférer les petites va | 762294.9 |

L'algorithme a convergé

Procédure LIFEREG

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--------------------------------|----------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance à 95 % | | Pr > Chi-2 |
| Intercept | 1 | 5.4941 | 0.0052 | 5.4839 | 5.5043 | 1118831 |
| Scale | 1 | 1.0596 | 0.0035 | 1.0526 | 1.0665 | |
| Echelle de Weibull | 1 | 243.2551 | 1.2635 | 240.7913 | 245.7442 | |
| Forme de Weibull | 1 | 0.9438 | 0.0032 | 0.9376 | 0.9500 | |

c. Loi lognormal, deux paramètres

- Paramètre de moyenne : μ (mû)
- Paramètre de variance : σ (sigma)

Les expressions de sa densité, survie et fonction de risque sont compliquées.

Loi non monotone, risque croissant puis décroissant
 Cette loi est **utilisée pour sa forme particulière qui permet un pic puis une pente raide sur la fonction de risque.**

Adapté notamment aux phénomènes de répétition : durée entre deux évènements : juste après l'évènement peu probable, ensuite croissance puis décroissance

L'exécution de la Lifereg fournit les résultats suivants :

Procédure LIFEREG

| Informations sur le modèle | |
|-----------------------------------|--------------|
| Table | WORK.TELCO2 |
| Variable dépendante | Log(duree) |
| Variable de censure | resil |
| Valeur(s) de censure | 0 |
| Nombre d'observations | 183032 |
| Valeurs non censurées | 59157 |
| Valeurs censurées à droite | 123895 |
| Valeurs censurées à gauche | 0 |
| Valeurs censurées de l'intervalle | 0 |
| Number of Parameters | 2 |
| Réponse zéro ou négative | 3633 |
| Name of Distribution | Lognormal |
| Logarithme de la vraisemblance | -167477.0204 |

| | |
|-----------------------------|--------|
| Nb d'observations lues | 186665 |
| Nb d'observations utilisées | 183032 |

| Statistiques d'ajustement | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 334954.1 |
| AIC (préférer les petites valeurs) | 334958.1 |
| AICC (préférer les petites valeurs) | 334958.1 |
| BIC (préférer les petites valeurs) | 334978.3 |

| Fit Statistics (Unlogged Response) | |
|--|----------|
| -2 log-vraisemblance | 759624.2 |
| LognormalAIC (préférer les petites valeurs) | 759628.2 |
| LognormalAICC (préférer les petites valeurs) | 759628.2 |
| LognormalBIC (préférer les petites valeurs) | 759646.4 |

L'algorithme a convergé

Procédure LIFEREG

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--------------------------------|--------|---------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance à 95 % | | Pr > Chi-2 |
| Intercept | 1 | 5.1696 | 0.0061 | 5.1577 | 5.1816 | 715761 <.0001 |
| Scale | 1 | 1.6479 | 0.0051 | 1.6380 | 1.6579 | |

d. Autres lois possibles

- Loi Log-logistic (2 paramètres)
- Loi Gamma (3 paramètres ou 2 si scale=shape), le modèle gamma peut prendre toutes les formes y compris le U, seulement il est difficile de le faire converger
- Loi Gamma type erlang 2, 3, 4 (loi croissante avec un pic en début puis décroissante) , $r=2$, dans sas $= 1/\text{shape}^2$ soit $\text{shape} = \text{racine}(1/2) = \text{racine}(2)/2$, $r=4 \Rightarrow \text{shape}=1/2$

Plus la loi a de paramètres :

- Plus il faut de données pour que les estimations soit consistantes
- La loi de survie est mieux estimée
- Le modèle est moins généralisable

L'exécution de la Lifereg fournit les résultats suivants :

Procédure LIFEREG

| Informations sur le modèle | |
|-----------------------------------|--------------|
| Table | WORK.TELCO2 |
| Variable dépendante | Log(duree) |
| Variable de censure | resil |
| Valeur(s) de censure | 0 |
| Nombre d'observations | 183032 |
| Valeurs non censurées | 59137 |
| Valeurs censurées à droite | 123895 |
| Valeurs censurées à gauche | 0 |
| Valeurs censurées de l'intervalle | 0 |
| Number of Parameters | 2 |
| Réponse zéro ou négative | 3633 |
| Name of Distribution | LLogistic |
| Logarithme de la vraisemblance | -168004.4791 |

| | |
|-----------------------------|--------|
| Nb d'observations lues | 186665 |
| Nb d'observations utilisées | 183032 |

| Statistiques d'ajustement | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 336009.0 |
| AIC (préférer les petites valeurs) | 336013.0 |
| AICC (préférer les petites valeurs) | 336013.0 |
| BIC (préférer les petites valeurs) | 336033.2 |

| Fit Statistics (Unlogged Response) | |
|---------------------------------------|----------|
| -2 log-vraisemblance | 760679.1 |
| LLogistic:AIC (préférer les petites) | 760683.1 |
| LLogistic:AICC (préférer les petites) | 760683.1 |
| LLogistic:BIC (préférer les petites) | 760703.3 |

L'algorithme a convergé.

Procédure LIFEREG

| Informations sur le modèle | |
|-----------------------------------|-------------|
| Table | WORK.TELCO2 |
| Variable dépendante | Log(duree) |
| Variable de censure | resil |
| Valeur(s) de censure | 0 |
| Nombre d'observations | 183032 |
| Valeurs non censurées | 59137 |
| Valeurs censurées à droite | 123895 |
| Valeurs censurées à gauche | 0 |
| Valeurs censurées de l'intervalle | 0 |
| Number of Parameters | 3 |
| Réponse zéro ou négative | 3633 |
| Name of Distribution | Gamma |
| Logarithme de la vraisemblance | -167443.001 |

| | |
|-----------------------------|--------|
| Nb d'observations lues | 186665 |
| Nb d'observations utilisées | 183032 |

| Statistiques d'ajustement | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 334886.0 |
| AIC (préférer les petites valeurs) | 334892.0 |
| AICC (préférer les petites valeurs) | 334892.0 |
| BIC (préférer les petites valeurs) | 334922.4 |

| Fit Statistics (Unlogged Response) | |
|--------------------------------------|----------|
| -2 log-vraisemblance | 759556.1 |
| Gamma:AIC (préférer les petites vale | 759562.1 |
| Gamma:AICC (préférer les petites val | 759562.1 |
| Gamma:BIC (préférer les petites vale | 759592.5 |

L'algorithme a convergé.

Procédure LIFEREG

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--------------------------------|--------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance à 95 % | Khi-2 | Pr > Khi-2 |
| Intercept | 1 | 5.1077 | 0.0054 | 5.0971 5.1183 | 893305 | <.0001 |
| Scale | 1 | 0.9093 | 0.0031 | 0.9033 0.9154 | | |

Procédure LIFEREG

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--------------------------------|--------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance à 95 % | Khi-2 | Pr > Khi-2 |
| Intercept | 1 | 5.2230 | 0.0086 | 5.2061 5.2399 | 367820 | <.0001 |
| Scale | 1 | 1.5730 | 0.0104 | 1.5528 1.5934 | | |
| Shape | 1 | 0.1346 | 0.0161 | 0.1031 0.1662 | | |

5.3.2. Comment choisir une loi ?

a. Par la méthode du maximum de vraisemblance

Le principe est de sélectionner la loi la plus vraisemblable parmi un ensemble de lois de la même famille de lois

- On choisit les lois à comparer;
- On calcule le maximum de la log-vraisemblance de chaque loi;

- On sélectionne la loi qui a le maximum de log-vraisemblance le plus élevé
 - Cependant, la comparaison de la vraisemblance n'est possible que dans le cas de modèle de la même famille
- On récupère les paramètres estimés par la maximisation de la logvraisemblance,
- On a ainsi la loi et ses paramètres estimés.

On choisit les lois à comparer :

- Exponential
- Weibull
- Lognormal
- LLogistic
- Gamma
- On calcule le maximum de la log-vraisemblance de chaque loi

i. Calcul du maximum de vraisemblance

La vraisemblance d'un modèle avec censure se calcule différemment d'un modèle sans données censurées. Soit X_1, \dots, X_n les temps de résiliation, soit $f(t, \theta)$ la fonction de densité associée, soient $S(t, \theta)$ et $F(t, \theta)$ les fonctions de survie et de répartition associées et δ_i l'indicatrice de censure ($\delta_i=0$ pour donnée censurée)

Dans le cas de données non censurées, la vraisemblance serait :

$$\blacksquare L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n f(X_i, \theta)$$

En présence de données censurées à droite, cela devient :

$$\blacksquare L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n [f(X_i, \theta)]^{\delta_i} [S(X_i, \theta)]^{1-\delta_i}$$

En présence de données censurées à gauche, cela devient :

$$\blacksquare L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n [f(X_i, \theta)]^{\delta_i} [F(X_i, \theta)]^{1-\delta_i}$$

ii. Interprétation de la vraisemblance

La vraisemblance mesure l'adéquation du modèle à la réalité. Plus elle est importante, plus le modèle est vraisemblable.

En présence de données censurées, la vraisemblance est le produit :

- de la densité pour les points non censurés (proba d'attrition en t_i),

- de la survie pour les points censurés à droite : ces individus n'ont pas mis fin à leur relation, on ne peut donc les prendre en compte par la densité, par contre on sait qu'ils sont toujours en relation jusqu'en t_i ,
- de la répartition pour les points censurés à gauche : ces individus ont mis fin à leur relation en t_i mais t_i est un majorant de la durée réelle inobservée, on peut donc utiliser la fonction de répartition, ils font partie du cumul de l'attrition en t_i .

Les log-vraisemblances de modèles estimés, d'après les résultats énumérés ci-dessus, sont les suivants :

| | | |
|---|----------------------|---------------------|
| ○ | Name of Distribution | Exponential |
| | Log Likelihood | -168953.5791 |
| ○ | Name of Distribution | Weibull |
| | Log Likelihood | -168800.2639 |
| ○ | Name of Distribution | Lognormal |
| | Log Likelihood | -167477.0284 |
| ○ | Name of Distribution | LLogistic |
| | Log Likelihood | -168004.4791 |
| ○ | Name of Distribution | Gamma |
| | Log Likelihood | -167443.001 |

La loi Exponentiel et la loi de Weibull sont, dans l'ordre décroissant, les lois avec la log-vraisemblance la plus élevée.

La statistique du ratio de vraisemblance est généralement utilisée pour comparer des modèles emboîtés ; un modèle est dit emboîté avec un autre modèle si le premier modèle est un cas particulier du second. En d'autres termes le modèle A est emboîté dans le modèle B si le A peut être obtenu en imposant des restrictions aux paramètres du modèle B. A titre illustratif, le modèle Exponentiel est emboîté à la fois au modèle de Weibull et au modèle de Gamma standard. En effet, on obtient un modèle exponentiel en forçant à 1 les paramètres d'échelle et de forme dans le modèle Gamma.

Ainsi, si le modèle A est emboîté dans le modèle B alors on peut évaluer la pertinence de A en prenant 2 fois la différence positive des log-vraisemblances des modèles A et B puis en comparant la valeur obtenue à la valeur du Chi² à 1 degré de liberté (2 ddl dans le cas d'un modèle gamma généralisé).

Seules certaines comparaisons sont possibles, celles qui correspondent à des modèles emboîtés :

| | | |
|------------------------|---|---------|
| Exponential vs Weibull | = | 153.32 |
| Exponential vs Gamma | = | 1510.58 |
| Weibull vs Gamma | = | 1357.26 |
| Lognormal vs Gamma | = | 34.03 |

A l'évidence, les modèles Exponentiel et Weibull sont rejetés. En revanche, les modèles Lognormal et Gamma semblent similaires.

Il est évident que les résultats ne nous permettent pas de conclure sur le choix de la loi à retenir.

Ces limites nous amènent à recourir à la méthode graphique (Hazard Plotting).

b. Par la méthode graphique : « le hazard plotting »

Il existe aussi une méthode graphique : le hazard plotting. En effet, si on compare des lois de familles de lois différentes, la méthode du maximum de vraisemblance n'est plus valable.

- Chaque loi paramétrique a une fonction de survie qui s'exprime en fonction du temps $S(t) = F(\theta, t)$.
- On modifie la fonction de survie (par ex : passage en log) de façon à obtenir une équation linéaire : $g(S(t)) = h(\theta) \cdot \ln(t) + s(\theta)$.
- On estime alors h et s par une régression linéaire et on calcule la statistique de test R^2 associée.
- On sélectionne la loi qui a le R^2 le plus élevé.
- Graphiquement, on sélectionne la loi dont une transformation de la survie en fonction d'une transformation du temps se rapproche le plus possible d'une droite.

i. Calcul des paramètres par le hazard plotting

Les paramètres se déduisent de la régression. Par exemple :

Pour une loi exponentielle, la régression est la suivante :

- $\ln(S(t)) = \lambda t$
- Après résolution graphique, en trouvant la droite se rapprochant le plus du graphe de $\ln(S(t))$, on a :
 - λ , qui est la pente de droite

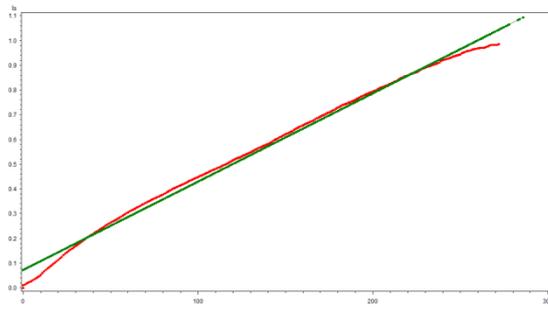
Pour une loi de Weibull à deux paramètres, on a :

- $\ln(-\ln(S(t))) = A \cdot \ln(t) + B = \beta \cdot \ln(t) - \beta \cdot \ln(\eta)$
- Après résolution graphique, en trouvant la droite (en $\ln(t)$) se rapprochant le plus du graphe de $\ln(-\ln(S(t)))$, on a :
 - β , qui est la pente de la droite
 - η , qui se déduit de l'ordonnée à l'origine et de β

Remarque : On préférera le maximum de vraisemblance pour l'estimation des paramètres

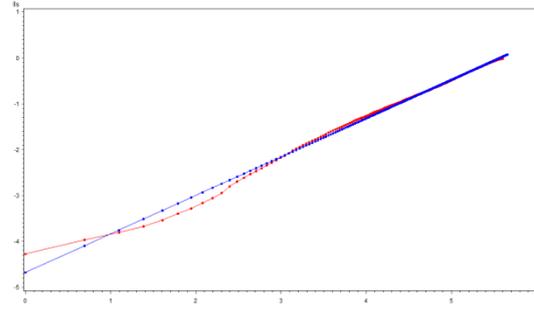
ii. Résultats du hazard plotting

Loi Exponentielle



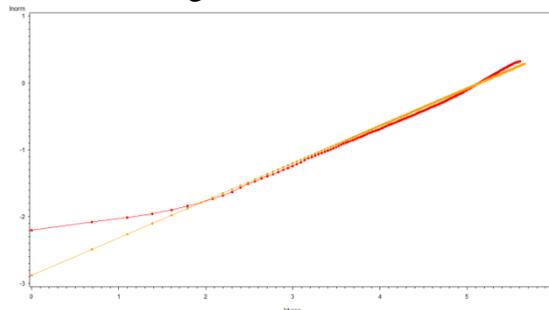
| | | | |
|--------------------|---------|---------------|--------|
| Root MSE | 0.02198 | R carré | 0.9939 |
| Moyenne dépendante | 0.55395 | R car. ajust. | 0.9939 |
| Coeff Var | 3.96771 | | |

Loi Weibull



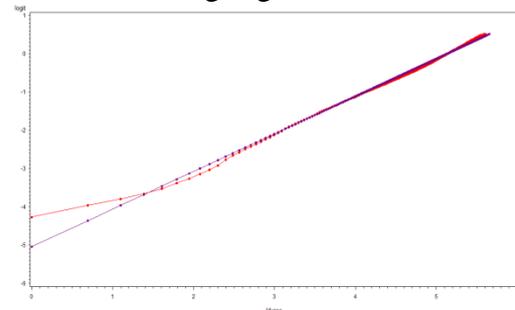
| | | | |
|--------------------|----------|---------------|--------|
| Root MSE | 0.05210 | R carré | 0.9959 |
| Moyenne dépendante | -0.80413 | R car. ajust. | 0.9959 |
| Coeff Var | -6.47961 | | |

Loi LogNormal



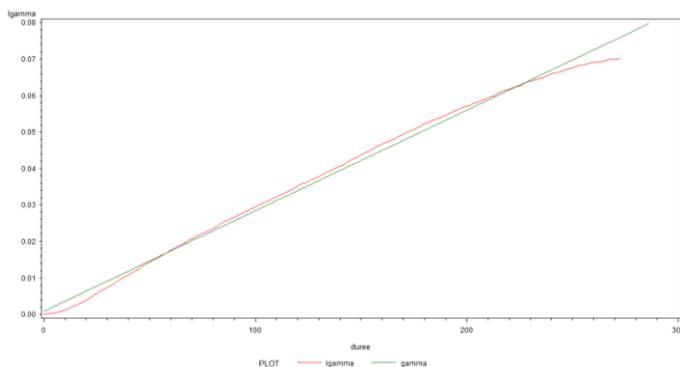
| | | | |
|--------------------|-----------|---------------|--------|
| Root MSE | 0.06893 | R carré | 0.9839 |
| Moyenne dépendante | -0.29773 | R car. ajust. | 0.9839 |
| Coeff Var | -23.15144 | | |

Loi LogLogit



| | | | |
|--------------------|-----------|---------------|--------|
| Root MSE | 0.07102 | R carré | 0.9944 |
| Moyenne dépendante | -0.50956 | R car. ajust. | 0.9944 |
| Coeff Var | -13.93788 | | |

Loi Gamma

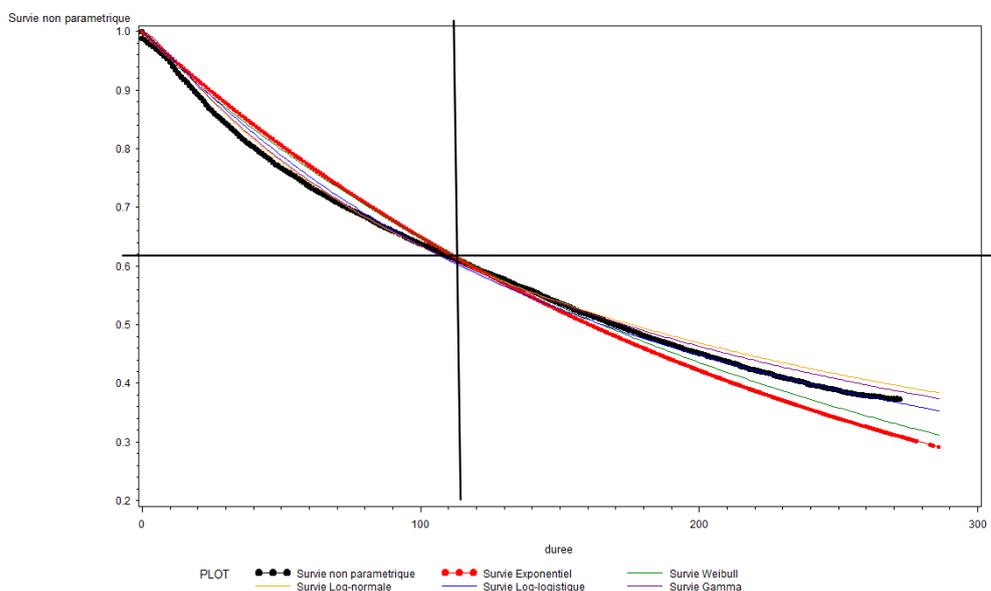


| | | | |
|--------------------|---------|---------------|--------|
| Root MSE | 0.02198 | R carré | 0.9939 |
| Moyenne dépendante | 0.55395 | R car. ajust. | 0.9939 |
| Coeff Var | 3.96771 | | |

→ **Conclusion du Hazard plotting : on choisit la loi de Weibull**

Les deux approches conduisent à un résultat similaire. Le modèle de Weibull paraît donc bien approprié.

c. Représentation en un même graphique de l'estimation paramétrique de KM et celle des différents modèles paramétriques testés.



Cependant on constate que la fonction de survie non paramétrique est mieux représentée par la loi gamma. L'estimation de la médiane en est l'illustration.

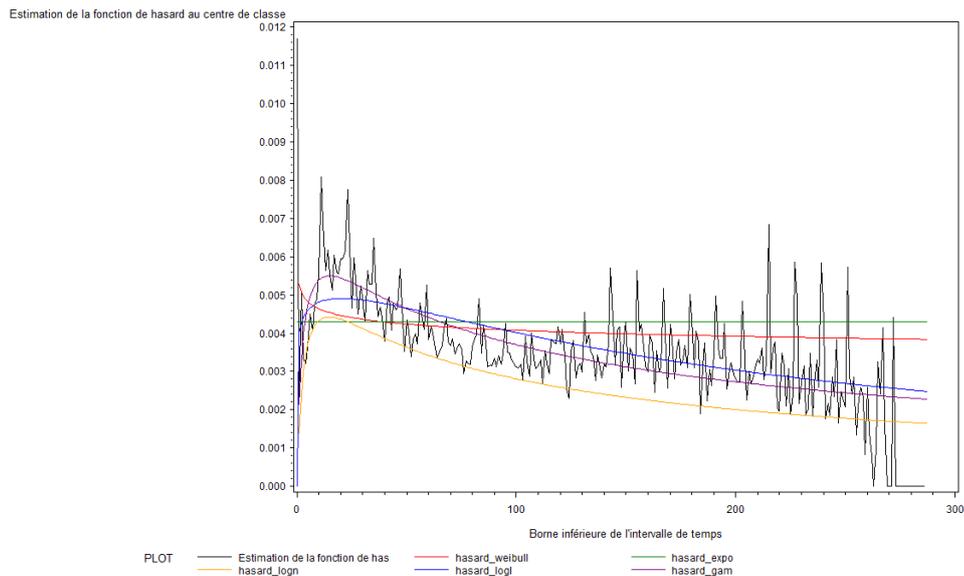
Durée de vie du client bancassurance Mraca :

- Médiane Lifetest : 14.22 ans
- Médiane Lifereg Exponentiel : 14.05 ans
- Médiane Lifereg Weibull : 14.87 ans
- Médiane Lifereg LogNormal : 14.60 ans
- Médiane Lifereg Gamma : 14.08 ans

Nous verrons dans le graphe suivant que le modèle Gamma décrit mieux la fonction de risque

d. Tracé des fonctions de risque issues des estimations non paramétrique et paramétrique

Nous utilisons la table issue de la méthode actuarielle afin de récupérer la fonction de risque non paramétrique



Pour résumer la Méthode de sélection d'une loi, après :

- Estimation non paramétrique par lifetest
- Visualisation de la forme de la survie et du risque
- Estimation paramétrique par lifereg sur plusieurs lois possibles
- Comparaison à l'aide du test du ratio de vraisemblance entre les lois de la même famille
- Hazard Plotting sur le résultat de l'estimation non paramétrique de la survie
- Les R^2 les plus élevés permettent de retenir les lois les plus adaptées
- Visualisation simultanée de la survie non paramétrique (la plus précise) et des survies paramétriques sélectionnées
- Visualisation simultanée du risque non paramétrique (le plus précis, uniquement avec une méthode actuarielle) et des risques paramétriques sélectionnés
- La confrontation des différents résultats permet de sélectionner la loi la plus adaptée

→ Nous considérons que la loi weibull semble la plus appropriée.

5.4. APPROCHE PARAMETRIQUE avec variables explicatives

5.4.1. Intégration de variables explicatives

a. Modèles à durées de vie accélérées

- Prise en compte de variables pouvant influencer la durée de vie
 - Variables qualitatives ou quantitatives
 - Variables dépendantes de la variable temps (t) directement

- Estimation de leur influence
 - Modèle de régression sur un paramètre de la loi estimée
- Test de leur significativité
 - Test du Chi2,
 - Test de significativité global,
 - Test de significativité progressif

Le temps est transformé de la manière suivante :

- Forme fréquente $T_Z = T_0 \cdot e^{(Z \cdot \theta)}$
- T0 suit une loi paramétrique
- L'effet des covariables est multiplicatif sur le temps
- Log linéarisation $\text{Log}(T) = X\beta + \text{Log}(T_0)$

On se retrouve alors avec une régression qui dépend de la loi que suit les durées de vie, estimation par maximisation de la logvraisemblance. Une fois le modèle estimé, on peut ensuite tester la significativité des coefficients estimés.

Les coefficients estimés sont exprimés en log du temps, un coefficient positif indiquera un ralentissement du phénomène (plus long), un coefficient négatif indiquera une accélération.

Le bancassureur se demande si l'équipement de son client en produit d'assurance (Auto, Mrh , Gav, ...), si la tranche d'âge de son client et si la tranche d'ancienneté de son client influent sur sa durée de vie.

- On a déjà estimé la loi de survie : loi de Weibull
- Utilisation d'un modèle de durée de vie d'un client bancassurance
- Estimation de l'influence de l'équipement, de l'âge et de l'ancienneté
- Test de significativité de cette influence

b. les résultats obtenus avec la procédure lifereg

Informations sur les covariables du modèle

| Code_Tr_age | Libellé_Tr_age | Code_Tr_anc_cli | Libellé de la tranche d'ancienneté |
|-------------|--------------------|-----------------|------------------------------------|
| 0 | ≤ 25 ans | 0 | 1 an |
| 1 | Entre 26 et 29 ans | 1 | 2 ans |
| 2 | Entre 30 et 33 ans | 2 | 3 ans |
| 3 | Entre 34 et 37 ans | 3 | 4 ans |
| 4 | Entre 38 et 52 ans | 4 | 5 ans |
| 5 | Entre 53 et 70 ans | 5 | 6 ans |
| 6 | ≥ 71 ans | 6 | 7 ans |
| 7 | | 7 | Entre 8 et 11 ans |
| 8 | | 8 | Entre 12 et 15 ans |
| 9 | | 9 | Plus de 16 ans |

| Code équipement | Libellé équipement |
|-----------------|------------------------------------|
| 0 | pasAuto&pasMrh&Gav |
| 1 | pasAuto&pasMrh&pasGav |
| 2 | Auto&Mrh&Gav |
| 3 | Auto&pasMrh&Gav ou pasAuto&Mrh&Gav |
| 4 | Auto&pasMrh&pasGav |
| 5 | Auto&Mrh&pasGav |
| 6 | pasAuto&Mrh&pasGav |

Procédure LIFEREG

| Informations sur le modèle | |
|-----------------------------------|--------------|
| Table | WORK.TELCO2 |
| Variable dépendante | Log(duree) |
| Variable de censure | resil |
| Valeur(s) de censure | 0 |
| Nombre d'observations | 183032 |
| Valeurs non censurées | 59137 |
| Valeurs censurées à droite | 123895 |
| Valeurs censurées à gauche | 0 |
| Valeurs censurées de l'intervalle | 0 |
| Number of Parameters | 23 |
| Réponse zéro ou négative | 3633 |
| Name of Distribution | Weibull |
| Logarithme de la vraisemblance | -63099.73407 |

| | |
|-----------------------------|--------|
| Nb d'observations lues | 186665 |
| Nb d'observations utilisées | 183032 |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------------------|
| Nom | Niveaux | Valeurs |
| equipement | 7 | 0 1 2 3 4 5 6 |
| tr_anc_cli | 10 | 0 1 2 3 4 5 6 7 8 9 |
| tr_age | 7 | 0 1 2 3 4 8 9 |

| Statistiques d'ajustement | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 126199.5 |
| AIC (préférer les petites valeurs) | 126245.5 |
| AICC (préférer les petites valeurs) | 126245.5 |
| BIC (préférer les petites valeurs) | 126478.2 |

| Fit Statistics (Unlogged Response) | |
|-------------------------------------|----------|
| -2 log-vraisemblance | 550869.6 |
| WeibullAIC (préférer les petites va | 550915.6 |
| WeibullAICC (préférer les petites v | 550915.6 |
| WeibullBIC (préférer les petites va | 551148.3 |

L'algorithme a convergé.

| Analyse des effets de Type III | | | |
|--------------------------------|-----|---------------|------------|
| Effet | DDL | Khi-2 de Wald | Pr > Khi-2 |
| equipement | 6 | 3117.1024 | <.0001 |
| tr_anc_cli | 9 | 1003922.39 | <.0001 |
| tr_age | 6 | 2173.8572 | <.0001 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--------------------------------|---------|---------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance à 95 % | | Khi-2 | Pr > Khi-2 |
| Intercept | | 1 | 5.7723 | 0.0046 | 5.7632 | 5.7814 | 1554395 | <.0001 |
| équipement | 0 | 1 | 0.0505 | 0.0041 | 0.0424 | 0.0585 | 151.33 | <.0001 |
| équipement | 1 | 1 | -0.0770 | 0.0036 | -0.0840 | -0.0701 | 468.23 | <.0001 |
| équipement | 2 | 1 | 0.1537 | 0.0057 | 0.1426 | 0.1648 | 736.36 | <.0001 |
| équipement | 3 | 1 | 0.1303 | 0.0051 | 0.1203 | 0.1403 | 649.47 | <.0001 |
| équipement | 4 | 1 | -0.0278 | 0.0021 | -0.0319 | -0.0237 | 176.92 | <.0001 |
| équipement | 5 | 1 | 0.0533 | 0.0023 | 0.0489 | 0.0577 | 559.85 | <.0001 |
| équipement | 6 | 0 | 0.0000 | . | . | . | . | . |
| tr_anc_cli | 0 | 1 | -3.5784 | 0.0051 | -3.5884 | -3.5685 | 495873 | <.0001 |
| tr_anc_cli | 1 | 1 | -2.7901 | 0.0045 | -2.7989 | -2.7812 | 380212 | <.0001 |
| tr_anc_cli | 2 | 1 | -2.2539 | 0.0045 | -2.2628 | -2.2450 | 247319 | <.0001 |
| tr_anc_cli | 3 | 1 | -1.8895 | 0.0047 | -1.8986 | -1.8804 | 164211 | <.0001 |
| tr_anc_cli | 4 | 1 | -1.6445 | 0.0048 | -1.6540 | -1.6350 | 116249 | <.0001 |
| tr_anc_cli | 5 | 1 | -1.3823 | 0.0050 | -1.3921 | -1.3725 | 77003.9 | <.0001 |
| tr_anc_cli | 6 | 1 | -1.0957 | 0.0052 | -1.1059 | -1.0856 | 44689.8 | <.0001 |
| tr_anc_cli | 7 | 1 | -0.9123 | 0.0045 | -0.9211 | -0.9036 | 41624.2 | <.0001 |
| tr_anc_cli | 8 | 1 | -0.5097 | 0.0047 | -0.5189 | -0.5004 | 11713.2 | <.0001 |
| tr_anc_cli | 9 | 0 | 0.0000 | . | . | . | . | . |
| tr_age | 0 | 1 | -0.1143 | 0.0037 | -0.1216 | -0.1070 | 936.90 | <.0001 |
| tr_age | 1 | 1 | -0.0604 | 0.0037 | -0.0677 | -0.0530 | 260.43 | <.0001 |
| tr_age | 2 | 1 | -0.0576 | 0.0037 | -0.0649 | -0.0503 | 237.13 | <.0001 |
| tr_age | 3 | 1 | -0.0323 | 0.0038 | -0.0398 | -0.0248 | 71.12 | <.0001 |
| tr_age | 4 | 1 | -0.0161 | 0.0032 | -0.0224 | -0.0098 | 25.16 | <.0001 |
| tr_age | 8 | 1 | 0.0196 | 0.0035 | 0.0128 | 0.0263 | 32.16 | <.0001 |
| tr_age | 9 | 0 | 0.0000 | . | . | . | . | . |
| Scale | | 1 | 0.1952 | 0.0007 | 0.1939 | 0.1965 | | |
| Forme de Weibull | | 1 | 5.1228 | 0.0175 | 5.0886 | 5.1572 | | |

Nous pouvons dire, par exemple, que le client de référence pris dans ce modèle est une personne âgée de plus de 70 ans avec une ancienneté de plus de 16 ans à la Mraca et équipée d'une Mrh mais pas d'un contrat Auto ou d'un contrat Gav. La durée de vie en mois est de 321 soit 27 ans. Par exemple, si le client détenait en plus du contrat Mrh un contrat Auto et une Gav alors sa durée de vie augmenterait ceteris paribus de 16.6%, soit 374 mois i.e. 31 ans.

Remarques :

Le modèle est log linéaire, c'est le log de la durée qui est estimée.

Il se trouve que, dans le cas où la loi suivie est exponentielle ou weibull, le modèle est dit également à risque proportionnel et peut s'interpréter en log du risque comme le modèle semi paramétrique

- Risque Proportionnel $h(t/Z, \theta) = h_0(t) \cdot g(\theta, Z)$
 - Dans le temps, le rapport de risque entre deux individus est constant et dépend du rapport de leurs variables explicatives qui est constant
- Loi exponentielle
 - $\log h(t) = \beta_0^* + \beta_1^* x_1 + \dots + \beta_k^* x_k$
 - B_j (estimée dans lifereg en fonction du log temps) = $-B_j^*$ (en fonction du log Risque)
- Loi weibull
 - $\log h(t) = \alpha \log t + \beta_0^* + \beta_1^* x_1 + \dots + \beta_k^* x_k \quad \beta_j^* = -\frac{\beta_j}{\sigma} \quad \forall j = 1, \dots, k \quad \alpha = \frac{1}{\sigma} - 1$
 - Le coefficient de X et scale doivent se modifier

6. Conclusion de la deuxième partie

Après avoir considéré que la loi Weibull semble la plus appropriée, nous avons modélisé la durée de vie d'un client bancassurance avec une approche paramétrique en intégrant des variables explicatives comme l'équipement, l'ancienneté du client dans sa relation avec son assurance et son âge. L'apport de cette modélisation nous indique clairement que si le client mono détenteur détient deux contrats ou plus alors sa durée de vie augmente de manière significative.

Cependant nous pouvons améliorer nettement cette modélisation en intégrant plus de covariables notamment bancaires.

Chapitre 3 Définition des indicateurs représentant la qualité d'un client en bancassurance et Modélisation de la qualité d'un client au niveau Assurance Iard

A. Définition des indicateurs représentant la qualité d'un client en bancassurance

1. Objectif :

Il s'agira ici de qualifier au mieux le partenaire aussi bien au niveau bancaire qu'au niveau assurance. Plus précisément, nous nous attacherons à définir ses indicateurs le plus objectivement possible et d'après l'avis de personnes expertes dans les différents domaines.

2. Définition des données utiles à la construction des différentes variables de qualité au niveau du partenaire commercial :

- Marge annuelle Bancaire au niveau partenaire (y compris marge dégagée sur la Collecte (Epargne Bilan (CAV, CSL, CEL, PEL,...), Epargne Hors Bilan (titres et Assurance Vie (Epargne)), sur l'Assurance Décès Invalidité, sur l'Assurance Non Vie, sur la Facturation, sur la Presse et sur le Crédit (Amortissable, revolving)),

Création d'une variable définissant la qualité du client en termes de marge bancaire

$qualite_PNB = \text{vaut } 1 \text{ si le partenaire commercial a une marge annuelle bancaire supérieure ou égale à } 310 \text{ € et } 0 \text{ sinon.}$

Le seuil des 310 € est en fait la marge annuelle bancaire moyenne au niveau partenaire commercial. Ce choix peut être remis en cause voire ajusté par la direction, par exemple en calculant la moyenne sur les clients actifs (i.e. avec un minimum d'opérations sur le CAV (Compte A Vue)) uniquement.

- Marge annuelle en Assurance Non Vie au niveau partenaire (ratio S/P hors sinistres Catastrophes naturelles calculé depuis la création du partenaire tous produits confondus),

Création d'une variable définissant la qualité du client en termes de marge en Assurance Non Vie

$Qualite_SP = \text{vaut } 1 \text{ si le partenaire commercial a une marge annuelle en Assurance Non Vie inférieure ou égale à } 70\% \text{ et } 0 \text{ sinon}$

Le choix du seuil 70% a été défini en prenant en compte au niveau global les primes acquises et les charges liées à la vente, à la gestion des contrats et des sinistres, à la réassurance et à l'assistance (Chiffres issus de la compta MRACA 2007).

11,43% Commissions versées à la CRCA NDF

3% Frais d'acquisition

4% Frais de gestion
1% Frais d'administration
5% Coût de la réassurance
4% Coût de l'assistance
= **28,43 %** nous avons donc arrondi à **30%**

Niveau de Risque Bancaire définit par la notation Bâle II au niveau partenaire,

Création d'une variable définissant la qualité du client en termes de risque bancaire

qualite_baleII = vaut 1 si le partenaire commercial dispose d'un niveau de risque bancaire acceptable, i.e. ayant une note Bâle II comprise entre A et J et 0 sinon.

C'est un expert Bâle II de la CRCA NDF qui nous a proposé la qualification du partenaire commercial définie ci-dessus.

Niveau de Risque Assurance Non Vie définit par la fréquence globale hors sinistres Catastrophes Naturelles calculée depuis la création du partenaire commercial.

Création d'une variable définissant la qualité du client en termes de risque en Assurance Non Vie

qualite_freq = vaut 1 si le partenaire commercial au niveau risque assurance a une fréquence Hors CNAT inférieure à 10% et 0 sinon

Le seuil des 10% est en fait la fréquence globale moyenne hors sinistres Catastrophes Naturelles au niveau partenaire commercial et depuis sa création. Ce choix peut être remis en cause voire ajusté par la direction, par exemple, en calculant la moyenne sur les trois dernières années uniquement ou en omettant certains sinistres comme un sinistre DDE (Dégâts Des Eaux) survenu dans la dixième année de cotise Mrh ou un sinistre BDG (Bris De Glaces) Auto survenu dans la troisième année de cotise Auto ou un sinistre Dommages tous accidents survenu dans la cinquième année de cotise Auto.

Résumé :

La qualité d'un client bancassurance tient compte à la fois :

- du risque bancaire (notation Bâle II),
- du risque assurance IARD (Fréquence des sinistres Hors sinistre Cat. Nat. depuis la création du partenaire commercial),
- de la marge bancaire annuelle (marge annuelle Crédit, Epargne Bilan et Hors Bilan, Assurance Décès et Invalidité, Assurance Non Vie, Facturation),
- de la Rentabilité en Assurance IARD (ratio S/P hors sinistre Cat. Nat. depuis la création du partenaire commercial).

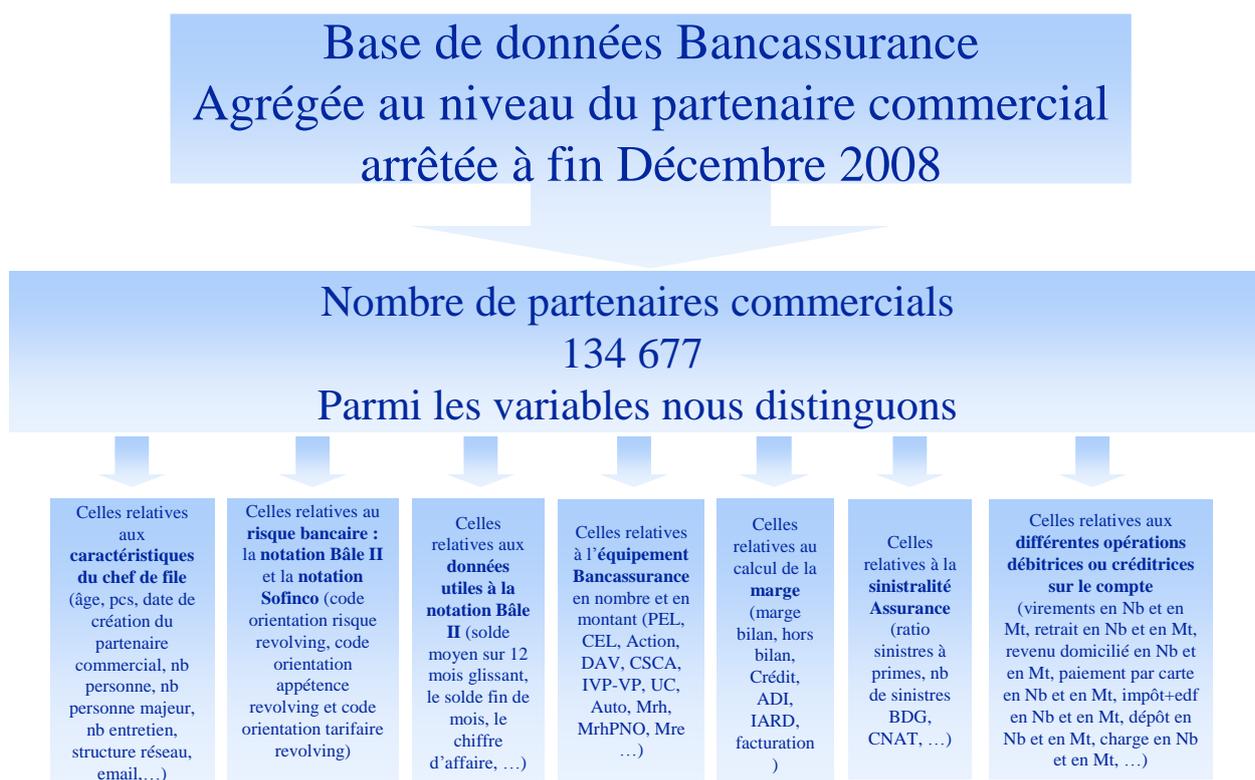
B. Modélisation de la qualité d'un client au niveau Assurance Iard

1. Problématiques de l'étude et Intérêts

La problématique repose principalement sur les clients n'ayant aucun contrat d'assurance Iard ou étant mono détenteurs (i.e. soit un contrat Auto ou un contrat Multirisque Habitation) car il ne sera pas possible ou alors difficile de définir la qualité du client au niveau Assurance Iard.

La solution que je préconise est de baser la conception du score sur les clients détenant à minima un contrat Auto et habitation. L'intérêt de cette modélisation est de pouvoir définir la qualité des clients bancassurance susceptibles de souscrire le « contrat Fidelio ». Bien sûr, il nous faudra, avant toute chose, créer la matrice de travail bancassurance dont la clé primaire est le partenaire commercial et construire le score à l'aide des variables explicatives comme : l'équipement bancassurance, la MIR (Mesure de l'Intensité de la Relation client), la segmentation de notre clientèle, la tranche d'âge,

2. Monitoring



Vous pourrez retrouver en annexe les métadonnées de cette matrice de travail.

3. Construction de la vae (variable à expliquer)

Les différents indicateurs que l'on a pris soin de détailler, dans la section précédente, ne seront pas modélisés compte tenu des règles établies pour définir la qualité d'un client bancassurance, de plus elles restent à l'appréciation de la direction.

Par contre, sur les partenaires ne détenant pas de contrat d'assurance Iard, à la CRCA NDF, nous devons les qualifier sur la partie qualité assurance à la fois sur le niveau risque et sur le niveau marge.

Après différents travaux, il s'avère que la réalisation d'un score sur la partie marge Assurance IARD, en tenant compte du seuil S/P à hauteur de 70, ne donne pas des résultats probants. Il est souvent plus facile de modéliser les comportements extrêmes comme les très bons risques ou comme les très mauvais risques à l'aide de données autres que celles de l'assurance, notamment des données bancaires voire des données géomarketing.

Par conséquent, dans le cadre de la qualification des partenaires sans détention d'assurance Iard ou mono détenteur, nous avons donc redéfini cette qualité client au niveau de la marge assurance IARD comme suit :

- ❑ Marge Globale Assurance Non Vie au niveau partenaire(S/P hors sinistres Catastrophes naturelles calculé depuis la création du partenaire tous produits confondus),

qualite_SP = *vaut 1 si le partenaire commercial a une marge annuelle en Assurance Non Vie inférieure ou égale à 0% et 0 sinon*

De fait, la modélisation au niveau de la fréquence n'as pas de sens puisque nous modélisons ici les clients de bonne voire de très bonne qualité aussi bien au niveau marge que risque en assurance.

4. Exploration des données : examen des distributions

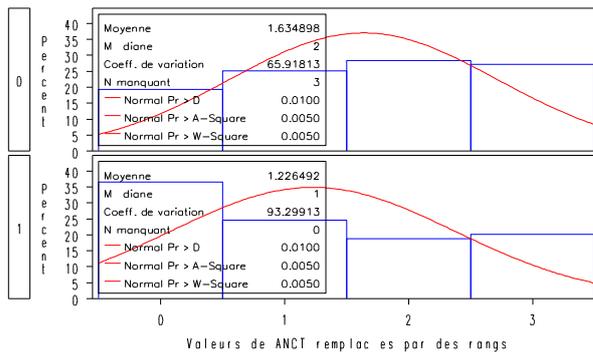
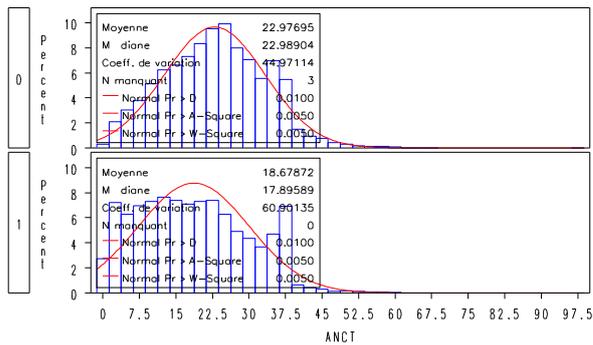
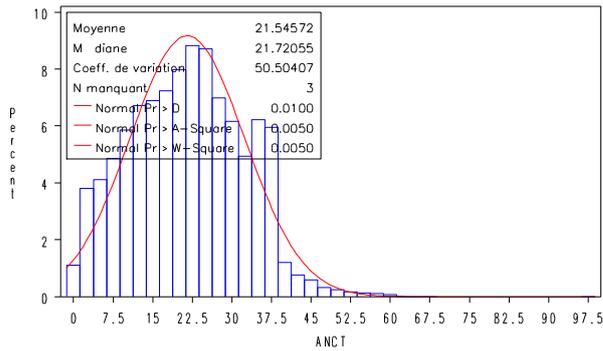
a. Statistiques univariées

L'exploration des données débute avec l'examen des statistiques univariées des variables, qui permet de voir l'allure de leur distribution, d'y détecter d'éventuelles anomalies, des valeurs extrêmes, des valeurs manquantes, d'éventuelles incohérences par rapport à leur définition, et plus simplement d'acquérir un peu de familiarité avec chaque variable.

Pour ce faire nous recourons à la procédure UNIVARIATE de SAS/BASE. Cette procédure possède de nombreuses fonctionnalités.

Nous nous attacherons à explorer une des variables qui ressortent dans la modélisation. Concernant les autres variables les statistiques univariées se trouvent en annexe (cf. Annexe1).

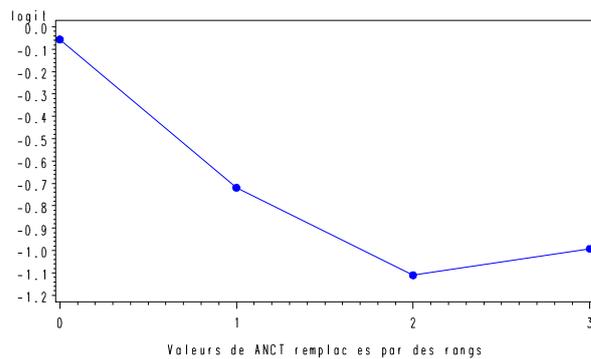
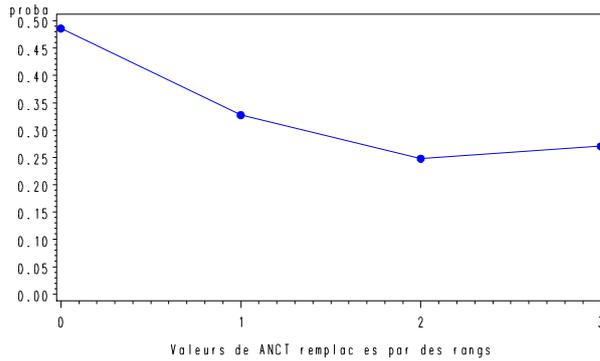
La variable **Ancienneté de l'entrée en relation du client au niveau Bancaire** a une distribution plutôt régulière, dont l'allure n'est pas très éloignée de celle de loi normale. On constate que cette variable est liée à la qualité du client en Assurance Iard.



Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : entre 0 et 13 ans (l'occurrence de référence),
- 1 : entre 14 et 22 ans,
- 2 : entre 23 et 29 ans,
- 3 : supérieure ou égale à 30 ans.

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur de l'ancienneté de l'entrée en relation du client au niveau Bancaire. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



⇒ Ces deux graphes ne suggèrent aucun regroupement de ces modalités. La qualité du client en Assurance Iard est fonction décroissante suivant l'ancienneté de l'entrée en relation du client au niveau Bancaire sur les trois premières tranches.

5. Sélection des variables explicatives corrélées à la variable à expliquer

Plusieurs étapes explicitées ci-dessous visent à identifier un ensemble de variables explicatives susceptibles d'entrer avec pertinence dans un modèle de score. Les résultats des sélections de ces dernières seront ensuite combinés dans la section suivante. La combinaison de ces tests a pour objectif de resserrer les mailles du filet autour des variables potentiellement intéressantes et de n'en oublier aucune, même si toutes ne paraissent pas d'égal intérêt au vu de chaque test effectué.

a. Sélection des variables : liaison avec la variable à expliquer des variables explicatives quantitatives

Dans cette étape, le test non paramétrique de la variance de Kruskal-Wallis va permettre de mesurer l'intensité de la liaison entre la variable cible et les variables explicatives continues.

Nous utiliserons le code sas et la procédure NPAR1WAY pour nous permettre d'avoir en sortie un fichier qui présente les variables explicatives par ordre décroissant de leur valeur de χ^2 Kruskal-Wallis avec la variable de qualité client cible. La liste des variables est restreinte pour des raisons de visibilité.

On remarque que l'ancienneté de la relation du client au niveau de l'Assurance Iard est sensiblement plus liée à la qualité et la valeur du client cible ($\chi^2=5\ 963$) que l'ancienneté de

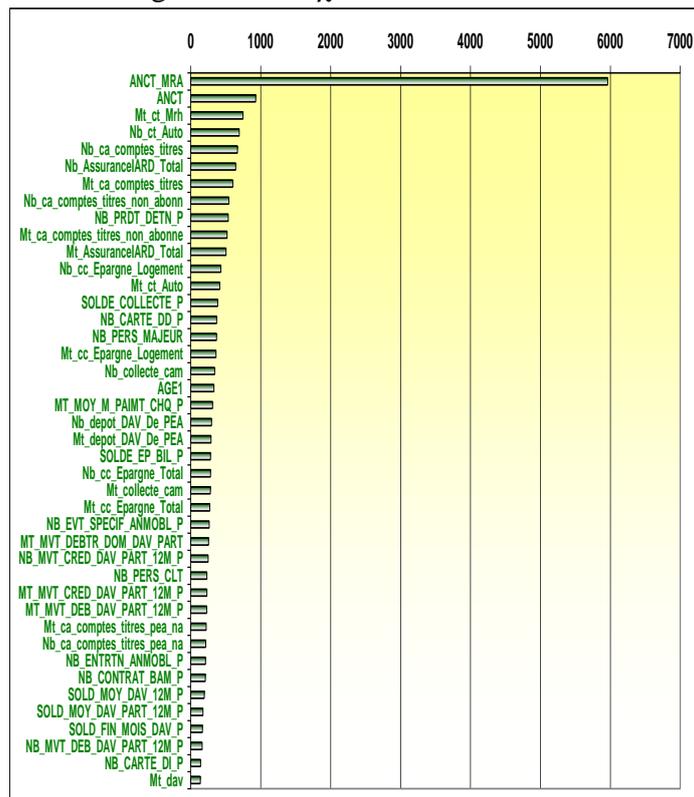
l'entrée en relation du client au niveau Bancaire ($\chi^2=929$). Ceci peut s'expliquer par le fait que l'ancienneté de la relation du client au niveau de l'Assurance Iard permet de donner une idée précise de l'exposition et du comportement du client face au risque.

A ce stade, nous ne fixons pas encore de seuil minimal au χ^2 de Kruskal-Wallis pour la sélection d'une variable, car, comme indiqué au début de la section, cette sélection sera effectuée plus loin, après comparaison avec les résultats des autres tests de liaison réalisés. Néanmoins, il est évident que le χ^2 décroît très vite et qu'une majorité de variables devra être écartée.

On trace l'histogramme des χ^2 de Kruskal-Wallis avec la procédure GCHART.

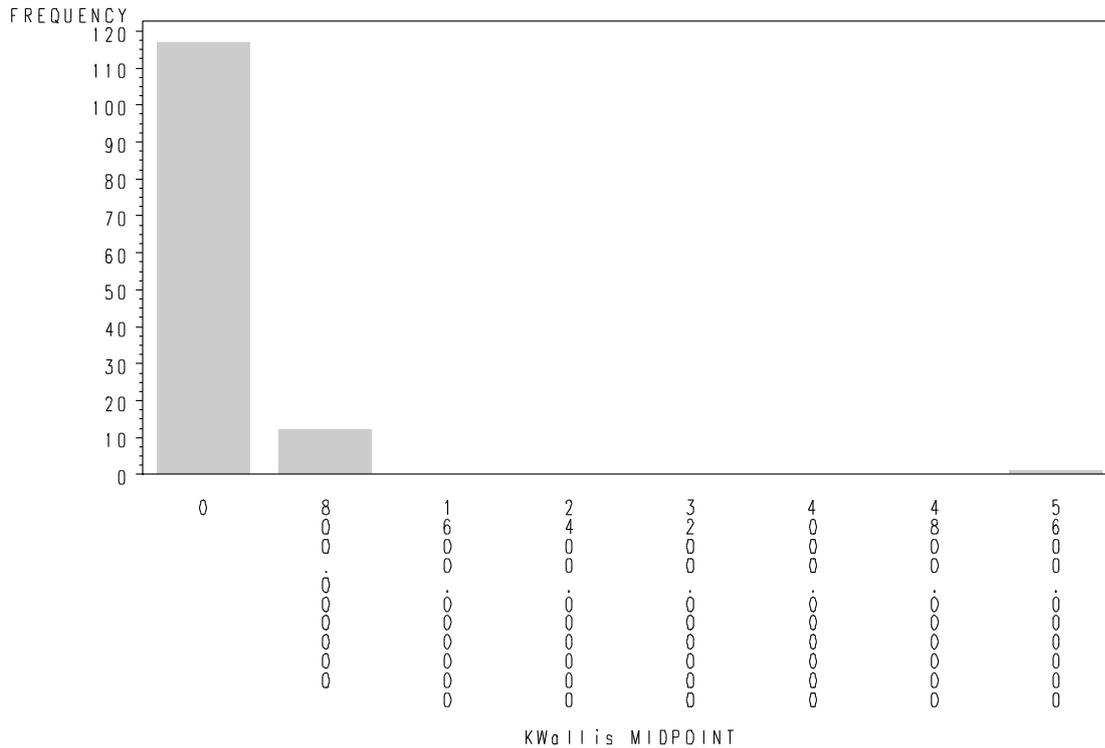
| Variable | KWallis | RgKruskal |
|---------------------------------|---------|-----------|
| ANCT_MRA | 5963 | 1 |
| ANCT | 929 | 2 |
| Mt_ct_Mrh | 747 | 3 |
| Nb_ct_Auto | 693 | 4 |
| Nb_ca_comptes_titres | 667 | 5 |
| Nb_AssuranceIARD_Total | 643 | 6 |
| Mt_ca_comptes_titres | 601 | 7 |
| Nb_ca_comptes_titres_non_abonne | 545 | 8 |
| NB_PRDT_DET_N_P | 537 | 9 |
| Mt_ca_comptes_titres_non_abonne | 521 | 10 |
| Mt_AssuranceIARD_Total | 505 | 11 |
| Nb_cc_Epargne_Logement | 432 | 12 |
| Mt_ct_Auto | 415 | 13 |
| SOLDE_COLLECTE_P | 385 | 14 |
| NB_CARTE_DD_P | 369 | 15 |
| NB_PERS_MAJEUR | 368 | 16 |
| Mt_cc_Epargne_Logement | 362 | 17 |
| Nb_collecte_cam | 342 | 19 |
| AGE1 | 332 | 20 |
| MT_MOY_M_PAINT_CHQ_P | 312 | 21 |
| Nb_depot_DAV_De_PEA | 298 | 22 |
| Mt_depot_DAV_De_PEA | 290 | 23 |
| SOLDE_EP_BIL_P | 283 | 24 |
| Nb_cc_Epargne_Total | 282 | 25 |
| Mt_collecte_cam | 280 | 27 |
| Mt_cc_Epargne_Total | 273 | 28 |
| NB_EVT_SPECIF_ANMOBL_P | 264 | 29 |
| MT_MVT_DEBTR_DOM_DAV_PART_12MP | 255 | 30 |
| NB_MVT_CRED_DAV_PART_12M_P | 249 | 31 |
| NB_PERS_CLT | 232 | 32 |
| MT_MVT_CRED_DAV_PART_12M_P | 230 | 33 |
| MT_MVT_DEB_DAV_PART_12M_P | 228 | 34 |
| Mt_ca_comptes_titres_pea_na | 221 | 35 |
| Nb_ca_comptes_titres_pea_na | 215 | 36 |
| NB_ENTRTN_ANMOBL_P | 211 | 37 |
| NB_CONTRAT_BAM_P | 209 | 38 |
| SOLD_MOY_DAV_12M_P | 193 | 39 |
| SOLD_MOY_DAV_PART_12M_P | 175 | 40 |
| SOLD_FIN_MOIS_DAV_P | 167 | 41 |
| NB_MVT_DEB_DAV_PART_12M_P | 163 | 42 |
| NB_CARTE_DI_P | 140 | 43 |
| Mt_dav | 137 | 44 |

Histogramme des χ^2 de Kruskal-Wallis



Le graphique suivant met lui aussi en évidence qu'une majorité de variables est très peu liée à la variable à expliquer.

Analyse des corrélations entre la variable à expliquer et les variables explicatives



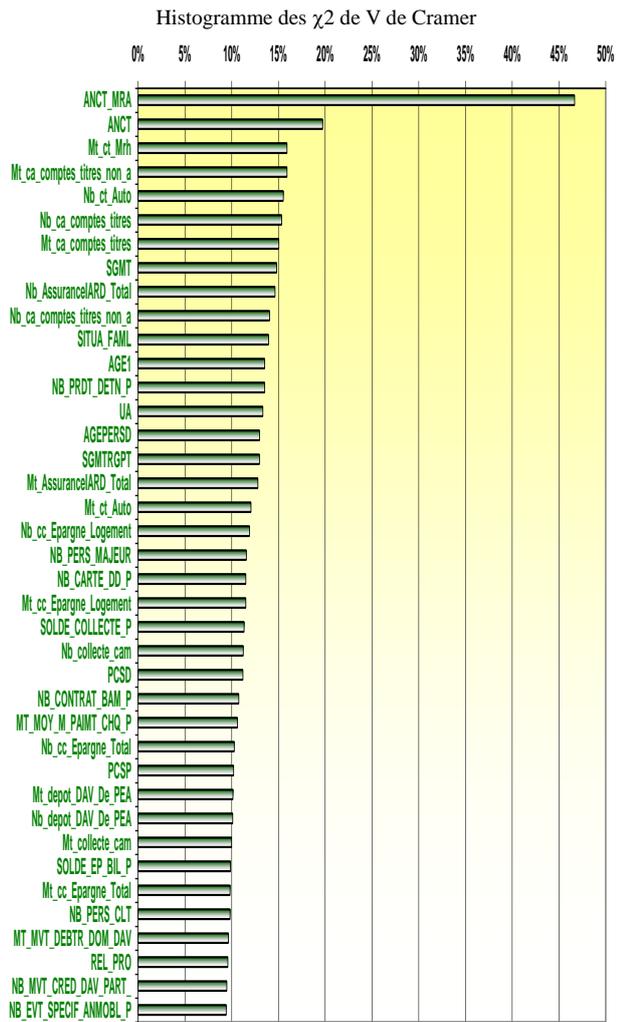
b. Sélection des variables : liaison avec la variable à expliquer des variables explicatives qualitatives ou discrétisées

Dans cette étape, le V de Cramer (χ^2 normalisé) est utilisé pour mesurer l'intensité de la liaison entre la variable cible et les variables qualitatives ou découpées en classes.

Nous avons donc les données sociodémographiques ainsi que les données propres au domaine bancaire telles que la segmentation de la clientèle bancaire, la note d'intensité de la relation bancaire, la note Bâle II, l'affectation à un conseiller, ..., plus les données quantitatives découpées en 4 classes.

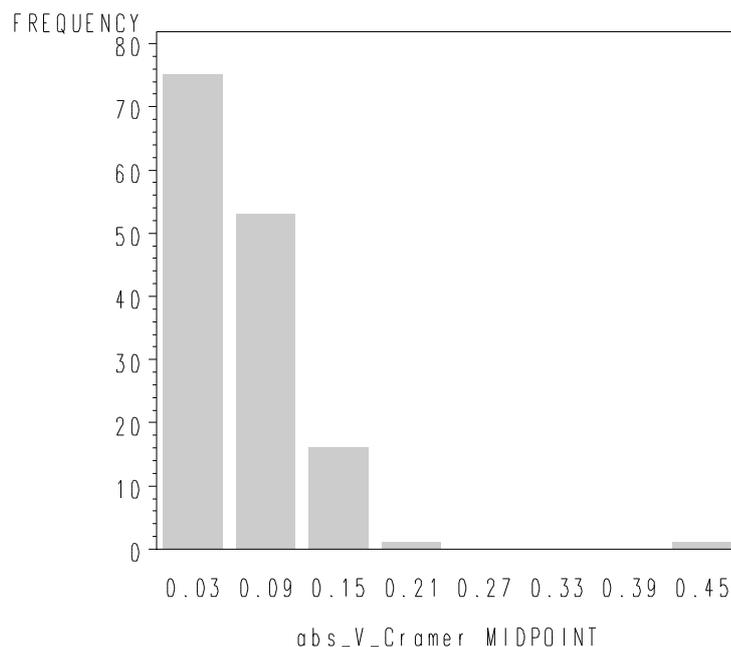
Nous utiliserons le code sas et la procédure FREQ pour nous permettre d'avoir en sortie un fichier qui présente les variables explicatives par ordre décroissant de leur de χ^2 V de Cramer avec la variable de qualité client cible. La liste des variables est restreinte pour des raisons de visibilité.

| Variable | Value | abs V Cramer | RqCramer |
|---------------------------------|-------|--------------|----------|
| ANCT_MRA | 0,47 | 0,47 | 1 |
| ANCT | 0,20 | 0,20 | 2 |
| Mt_ct_Mrh | 0,16 | 0,16 | 3 |
| Mt_ca_comptes_titres_non_abonne | 0,16 | 0,16 | 4 |
| Nb_ct_Auto | -0,16 | 0,16 | 5 |
| Nb_ca_comptes_titres | 0,15 | 0,15 | 6 |
| Mt_ca_comptes_titres | 0,15 | 0,15 | 7 |
| SGMT | 0,15 | 0,15 | 8 |
| Nb_AssuranceIARD_Total | 0,15 | 0,15 | 9 |
| Nb_ca_comptes_titres_non_abonne | 0,14 | 0,14 | 10 |
| SITUA_FAML | 0,14 | 0,14 | 11 |
| AGE1 | 0,14 | 0,14 | 12 |
| NB_PRDT_DET_N_P | 0,14 | 0,14 | 13 |
| UA | 0,13 | 0,13 | 14 |
| AGEPERSD | 0,13 | 0,13 | 15 |
| SGMTRGPT | 0,13 | 0,13 | 16 |
| Mt_AssuranceIARD_Total | 0,13 | 0,13 | 17 |
| Mt_ct_Auto | 0,12 | 0,12 | 18 |
| Nb_cc_Epargne_Logement | 0,12 | 0,12 | 19 |
| NB_PERS_MAJEUR | 0,12 | 0,12 | 20 |
| NB_CARTE_DD_P | -0,12 | 0,12 | 21 |
| Mt_cc_Epargne_Logement | 0,12 | 0,12 | 22 |
| SOLDE_COLLECTE_P | 0,11 | 0,11 | 23 |
| Nb_collecte_cam | 0,11 | 0,11 | 24 |
| PCSD | 0,11 | 0,11 | 25 |
| NB_CONTRAT_BAM_P | 0,11 | 0,11 | 26 |
| MT_MOY_M_PAINT_CHQ_P | 0,11 | 0,11 | 27 |
| Nb_cc_Epargne_Total | 0,10 | 0,10 | 28 |
| PCSP | 0,10 | 0,10 | 29 |
| Mt_depot_DAV_De_PEA | 0,10 | 0,10 | 30 |
| Nb_depot_DAV_De_PEA | -0,10 | 0,10 | 31 |
| Mt_collecte_cam | 0,10 | 0,10 | 32 |
| SOLDE_EP_BIL_P | 0,10 | 0,10 | 33 |
| Mt_cc_Epargne_Total | 0,10 | 0,10 | 34 |
| NB_PERS_CLT | 0,10 | 0,10 | 35 |
| MT_MVT_DEBTR_DOM_DAV_PART_12MP | 0,10 | 0,10 | 36 |
| REL_PRO | 0,10 | 0,10 | 37 |
| NB_MVT_CRED_DAV_PART_12M_P | 0,09 | 0,09 | 38 |
| NB_EVT_SPECIF_ANMOBL_P | 0,09 | 0,09 | 39 |



Un histogramme permet comme précédemment de représenter les V de Cramer.

Analyse des corrélations entre la variable à expliquer et les variables explicatives



Le graphique montre une moins grande dispersion des V de Cramer que des χ^2 de Kruskal-Wallis, ce qui tend à montrer que la discrétisation a amélioré le pouvoir discriminant des variables et augmenté leur V de Cramer (sur discrétisation) par rapport à leur χ^2 de Kruskal-Wallis.

c. Sélection des variables quantitatives par analyse discriminante

Nous allons utiliser la procédure STEPDISC de SAS qui permet, par une analyse discriminante linéaire et une sélection pas à pas, de sélectionner les variables explicatives potentiellement les plus intéressantes, la liste sera récupérée comme dans les sections précédentes.

La procédure STEPDISC réalise une sélection pas à pas des variables, basée par défaut sur la significativité du test F appliqué au lambda de Wilks, et au cours de laquelle des variables peuvent entrer et éventuellement en sortir. Par défaut, la probabilité associée au test F doit être <0.15 . Ce seuil de 15% a été choisi par SAS car il est susceptible de donner généralement de meilleurs résultats qu'un seuil plus ou moins élevé.

Ici, trente cinq variables sont sélectionnées. La variable « Nb dav successions » approche limite des 15%.

| Step | NumberIn | Entered | Removed | PartialRSquare | FValue | ProbF | WilksLambda | ProbLambda | AveSqCanCor | ProbASCC |
|------|----------|---------------------------------|----------------------|----------------|---------|--------|-------------|------------|-------------|----------|
| 1 | 1 | ANCT_MRA | | 0.1908 | 6461.09 | <.0001 | 0.80924464 | <.0001 | 0.19075536 | <.0001 |
| 2 | 2 | Mt_ct_Auto | | 0.0191 | 534.88 | <.0001 | 0.79375470 | <.0001 | 0.20624530 | <.0001 |
| 3 | 3 | Mt_ct_Mrh | | 0.0059 | 161.52 | <.0001 | 0.78910449 | <.0001 | 0.21089551 | <.0001 |
| 4 | 4 | NB_ENTRTN_ANMOBL_P | | 0.0029 | 78.58 | <.0001 | 0.78684847 | <.0001 | 0.21315153 | <.0001 |
| 5 | 5 | NB_MVT_CRED_DAV_PART_12M_P | | 0.0023 | 64.44 | <.0001 | 0.78500276 | <.0001 | 0.21499724 | <.0001 |
| 6 | 6 | Mt_AssuranceIARD_Total | | 0.0013 | 35.39 | <.0001 | 0.78399025 | <.0001 | 0.21600975 | <.0001 |
| 7 | 7 | Nb_ct_PJ | | 0.0010 | 27.77 | <.0001 | 0.78319657 | <.0001 | 0.21680343 | <.0001 |
| 8 | 8 | SOLD_MOY_DAV_PART_12M_P | | 0.0008 | 21.64 | <.0001 | 0.78257851 | <.0001 | 0.21742149 | <.0001 |
| 9 | 9 | Nb_dav | | 0.0005 | 14.19 | 0.0002 | 0.78217357 | <.0001 | 0.21782643 | <.0001 |
| 10 | 10 | NB_CSCA_P | | 0.0004 | 9.90 | 0.0017 | 0.78189120 | <.0001 | 0.21810880 | <.0001 |
| 11 | 11 | NB_MVT_DEB_DAV_PART_12M_P | | 0.0005 | 14.32 | 0.0002 | 0.78148291 | <.0001 | 0.21851709 | <.0001 |
| 12 | 12 | Nb_contrat_OPEN | | 0.0004 | 9.93 | 0.0016 | 0.78119973 | <.0001 | 0.21880027 | <.0001 |
| 13 | 13 | MT_MOY_RETRAIT_M_CARTE_P | | 0.0003 | 8.70 | 0.0032 | 0.78095171 | <.0001 | 0.21904829 | <.0001 |
| 14 | 14 | Nb_ct_Auto | | 0.0003 | 8.02 | 0.0046 | 0.78072305 | <.0001 | 0.21927695 | <.0001 |
| 15 | 15 | Mt_ca_comptes_titres_non_abonne | | 0.0003 | 7.89 | 0.0050 | 0.78049818 | <.0001 | 0.21950182 | <.0001 |
| 16 | 16 | MT_MOY_PAIMT_M_CARTE_P | | 0.0003 | 7.29 | 0.0069 | 0.78029061 | <.0001 | 0.21970939 | <.0001 |
| 17 | 17 | Mt_credit_conso_auto | | 0.0003 | 7.11 | 0.0077 | 0.78008814 | <.0001 | 0.21991186 | <.0001 |
| 18 | 18 | SOLD_FIN_MOIS_DAV_P | | 0.0002 | 6.75 | 0.0094 | 0.77989608 | <.0001 | 0.22010392 | <.0001 |
| 19 | 19 | Nb_ca_comptes_titres | | 0.0002 | 6.34 | 0.0118 | 0.77971552 | <.0001 | 0.22028448 | <.0001 |
| 20 | 20 | Nb_CT_predica_UC | | 0.0002 | 6.12 | 0.0134 | 0.77954149 | <.0001 | 0.22045851 | <.0001 |
| 21 | 21 | Nb_ct_Mrh | | 0.0002 | 5.62 | 0.0178 | 0.77938155 | <.0001 | 0.22061845 | <.0001 |
| 22 | 22 | Nb_AssuranceIARD_Total | | 0.0002 | 4.56 | 0.0327 | 0.77925180 | <.0001 | 0.22074820 | <.0001 |
| 23 | 23 | AGE1 | | 0.0002 | 5.04 | 0.0247 | 0.77910832 | <.0001 | 0.22089168 | <.0001 |
| 24 | 24 | Nb_cc_émissions_permanentes | | 0.0002 | 5.33 | 0.0210 | 0.77895681 | <.0001 | 0.22104319 | <.0001 |
| 25 | 25 | Mt_CT_predica_euro | | 0.0001 | 3.70 | 0.0545 | 0.77885168 | <.0001 | 0.22114832 | <.0001 |
| 26 | 26 | Nb_ct_GAV | | 0.0001 | 3.78 | 0.0518 | 0.77874412 | <.0001 | 0.22125588 | <.0001 |
| 27 | 27 | Mt_ADI_hab | | 0.0001 | 3.62 | 0.0572 | 0.77864125 | <.0001 | 0.22135875 | <.0001 |
| 28 | 28 | Mt_Credit_Total | | 0.0002 | 4.78 | 0.0287 | 0.77850525 | <.0001 | 0.22149475 | <.0001 |
| 29 | 29 | Mt_credit_conso_autres | | 0.0001 | 3.84 | 0.0502 | 0.77839619 | <.0001 | 0.22160381 | <.0001 |
| 30 | 30 | NB_PRDT_DET_N_P | | 0.0001 | 3.90 | 0.0484 | 0.77828540 | <.0001 | 0.22171460 | <.0001 |
| 31 | 31 | Nb_cc_Épargne_Total | | 0.0002 | 6.84 | 0.0089 | 0.77809110 | <.0001 | 0.22190890 | <.0001 |
| 32 | 32 | NB_CARTE_PAIEMENT_P | | 0.0002 | 5.42 | 0.0199 | 0.77793703 | <.0001 | 0.22206297 | <.0001 |
| 33 | 31 | | Nb_ca_comptes_titres | 0.0001 | 1.95 | 0.1630 | 0.77799233 | <.0001 | 0.22200767 | <.0001 |
| 34 | 32 | Nb_depot_Collecte_Monétaire | | 0.0001 | 2.79 | 0.0949 | 0.77791306 | <.0001 | 0.22208694 | <.0001 |
| 35 | 33 | NB_CONNEXION_IMB_6M_P | | 0.0001 | 2.12 | 0.1459 | 0.77785296 | <.0001 | 0.22214704 | <.0001 |
| 36 | 34 | Nb_dav_successions | | 0.0001 | 2.09 | 0.1484 | 0.77779361 | <.0001 | 0.22220639 | <.0001 |

d. Sélection des variables : régressions logistiques sur échantillons bootstrap multiples

Une autre technique de sélection de variables consiste à tenter de s'affranchir de la sensibilité à l'échantillonnage, qui fait que les tests de liaison ne donnent pas exactement les mêmes résultats sur deux échantillons différents de la même population, et que les écarts peuvent être d'autant plus grands que les effectifs des échantillons sont plus petits.

Pour cela on lance le processus de modélisation sur un grand nombre d'échantillons bootstrap (tirages avec remise de N individus parmi les N individus du fichier en entrée), et on regarde sur chaque échantillon comment se comporte chaque variable.

Nous décrivons dans cette section brièvement la macro mettant en œuvre ce processus.

Elle effectuera un grand nombre de tirages bootstrap. Sur chaque échantillon, elle lancera une régression logistique sans sélection de variable et enregistrera la significativité de chaque variable explicative. Pour chaque variable, elle listera enfin le nombre de régressions, sur N, où elle est significative au seuil de 1% (on a au plus 1% de risque de se tromper en la gardant dans le modèle), au seuil de 5% et au seuil de 10%. Elle listera aussi le nombre de fois où une modalité de la variable n'est pas significative au seuil de 5%, c'est-à-dire a une statistique de Wald <3.84. Ceci peut survenir plusieurs fois dans une même régression, et le nombre rapporté ici peut être supérieur à N. La macro indiquera aussi la plus petite statistique de Wald d'une modalité de la variable.

Nous aurons ainsi un diagnostic extrêmement étayé du pouvoir discriminant de chaque variable, mesuré par son apport à un modèle logistique.

Le tableau ci-dessous présente les variables en fonction du nombre de régressions logistiques dans lesquelles leur apport est significatif au seuil de 1%. On note que seules 17 variables sont significatives à ce seuil plus d'une fois sur deux, et 25 au seuil de 5%. Mais il faut se méfier du classement de certaines variables car elles peuvent être nulles sauf dans très peu de cas. Par ailleurs, certaines variables apparaissent avec un nombre élevé d'occurrences significatives à 1%, bien qu'elles aient par ailleurs un nombre de modalités non significatives à ce seuil : la variable peut être sans doute conservée, mais après regroupement d'au moins deux de ces modalités, dont celle(s) non significative(s). C'est notamment le cas de la segmentation de la clientèle bancaire SGMT.

SELECTION DE VARIABLES PAR LOGIT SUR ECHANTILLONS BOOTSTRAP

| Obs | Prédicteur | Occurrences | Signif à 1% | Signif à 5% | Signif à 10% | Non signif à 10% | nb Wald < 3.84 | + petit Wald des modalités |
|-----|----------------------|-------------|-------------|-------------|--------------|------------------|----------------|----------------------------|
| 1 | ANCT_MRA | 200 | 200 | 0 | 0 | 0 | 0 | . |
| 2 | Nb_ct_Auto | 200 | 200 | 0 | 0 | 0 | 0 | . |
| 3 | Mt_ct_Mrh | 200 | 200 | 0 | 0 | 0 | 147 | 0.0000 |
| 4 | SGMT | 200 | 171 | 19 | 7 | 3 | 2104 | 0.0000 |
| 5 | NB_CONNEXION_BAM_6M_ | 200 | 170 | 21 | 3 | 6 | 203 | 0.0000 |
| 6 | SOLD_MOY_DAV_12M_P | 200 | 163 | 27 | 4 | 6 | 357 | 0.0000 |
| 7 | AGEPERSD | 200 | 158 | 18 | 11 | 13 | 256 | 0.0000 |

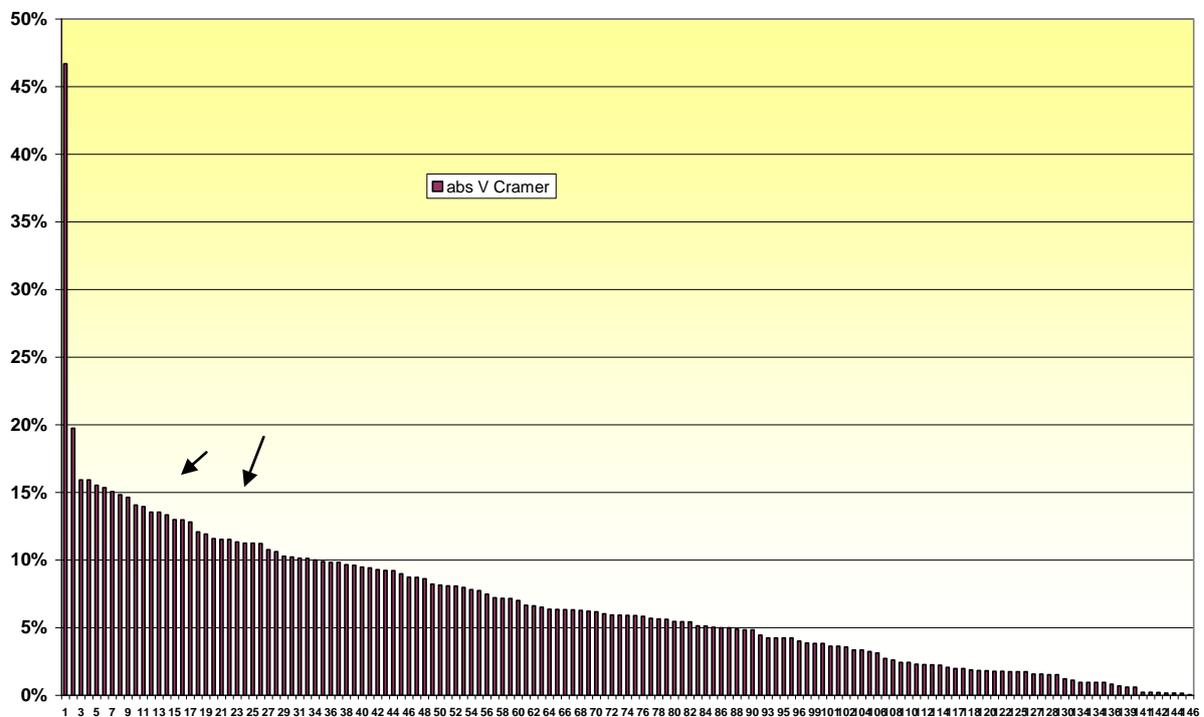
| Obs | Prédicteur | Occurrences | Signif à 1% | Signif à 5% | Signif à 10% | Non signif à 10% | nb Wald < 3.84 | + petit Wald des modalités |
|-----|----------------------|-------------|-------------|-------------|--------------|------------------|----------------|----------------------------|
| 8 | NB_ENTRTRN_ANMOBL_P | 200 | 149 | 29 | 13 | 9 | 376 | 0.0001 |
| 9 | Nb_ct_MrhPNO | 200 | 148 | 25 | 11 | 16 | 27 | 0.0632 |
| 10 | Nb_ct_PJ | 200 | 143 | 32 | 10 | 15 | 25 | 0.5525 |
| 11 | NB_MVT_DEB_DAV_PART_ | 200 | 139 | 37 | 9 | 15 | 219 | 0.0021 |
| 12 | Nb_contrat_OPEN | 200 | 119 | 37 | 15 | 29 | 44 | 0.0020 |
| 13 | BALEII | 200 | 118 | 34 | 16 | 32 | 601 | 0.0000 |
| 14 | UA | 200 | 111 | 44 | 18 | 27 | 2051 | 0.0000 |
| 15 | PCSP | 200 | 110 | 52 | 17 | 21 | 1590 | 0.0002 |
| 16 | Nb_cc_Livrets | 200 | 103 | 49 | 19 | 29 | 290 | 0.0001 |
| 17 | IVP | 200 | 101 | 46 | 17 | 36 | 53 | 0.0001 |
| 18 | SITUA_FAML | 200 | 86 | 43 | 24 | 47 | 711 | 0.0000 |
| 19 | Nb_dav | 200 | 79 | 52 | 16 | 53 | 528 | 0.0000 |
| 20 | MT_MOY_M_REMISE_CHQ_ | 200 | 69 | 59 | 17 | 55 | 467 | 0.0000 |
| 21 | ANCT | 200 | 67 | 48 | 31 | 54 | 481 | 0.0000 |
| 22 | MT_MVT_DEBTR_DOM_DAV | 200 | 66 | 40 | 20 | 74 | 452 | 0.0001 |
| 23 | NB_CARTE_DS_CSCA_P | 200 | 62 | 43 | 30 | 65 | 261 | 0.0000 |
| 24 | Mt_depot_DAV_De_PEA | 200 | 62 | 38 | 24 | 76 | 296 | 0.0000 |
| 25 | NB_CONNEXION_VDTX_6M | 200 | 56 | 53 | 25 | 66 | 91 | 0.0000 |
| 26 | MT_MOY_M_PAIMT_CHQ_P | 200 | 54 | 40 | 32 | 74 | 450 | 0.0001 |
| 27 | MT_MOY_PAIMT_M_CARTE | 200 | 51 | 47 | 27 | 75 | 447 | 0.0000 |
| 28 | Nb_credit_equipement | 200 | 50 | 30 | 21 | 99 | 120 | 0.0000 |
| 29 | NB_CARTE_DI_P | 200 | 47 | 38 | 25 | 90 | 316 | 0.0001 |
| 30 | AGE1 | 200 | 43 | 44 | 21 | 92 | 459 | 0.0000 |
| 31 | NB_CONTRAT_BAM_P | 200 | 41 | 30 | 27 | 102 | 463 | 0.0000 |
| 32 | SOLDE_COLLECTE_P | 200 | 38 | 44 | 20 | 98 | 489 | 0.0000 |
| 33 | SOLDE_LIVRETS_P | 200 | 37 | 45 | 26 | 92 | 523 | 0.0000 |
| 34 | Mt_ca_comptes_titres | 200 | 29 | 42 | 33 | 96 | 489 | 0.0000 |
| 35 | Nb_ct_2ROUES | 200 | 27 | 40 | 25 | 108 | 133 | 0.0000 |
| 36 | Mt_CT_predica_euro | 200 | 27 | 33 | 29 | 111 | 302 | 0.0000 |

| Obs | Prédicteur | Occurrences | Signif à 1% | Signif à 5% | Signif à 10% | Non signif à 10% | nb Wald < 3.84 | + petit Wald des modalités |
|-----|----------------------|-------------|-------------|-------------|--------------|------------------|----------------|----------------------------|
| 37 | Nb_cc_Epargne_Logeme | 200 | 26 | 32 | 18 | 124 | 317 | 0.0000 |
| 38 | NB_EVT_COMNTR_ANMOBL | 200 | 24 | 30 | 30 | 116 | 494 | 0.0000 |
| 39 | MIRE | 200 | 21 | 37 | 29 | 113 | 655 | 0.0000 |
| 40 | AFFECTE | 200 | 20 | 29 | 19 | 132 | 151 | 0.0000 |
| 41 | Nb_ct_GAV | 200 | 19 | 16 | 24 | 141 | 165 | 0.0000 |
| 42 | Mt_cc_emissions_perm | 200 | 16 | 29 | 23 | 132 | 155 | 0.0007 |
| 43 | Mt_CT_predica_UC | 200 | 15 | 34 | 18 | 133 | 151 | 0.0000 |
| 44 | NB_PRDT_DET_N_P | 200 | 13 | 24 | 23 | 140 | 491 | 0.0001 |
| 45 | NB_CHQ_12M_P | 200 | 13 | 18 | 28 | 141 | 525 | 0.0000 |
| 46 | MT_MOY_RETRAIT_M_CAR | 200 | 12 | 21 | 11 | 156 | 542 | 0.0000 |
| 47 | Nb_credit_conso_auto | 200 | 11 | 38 | 19 | 132 | 151 | 0.0008 |
| 48 | Nb_ct_Mrh | 200 | 10 | 26 | 18 | 146 | 164 | 0.0001 |
| 49 | Mt_dav_successions | 200 | 9 | 24 | 17 | 150 | 389 | 0.0004 |
| 50 | Mt_credit_hab | 200 | 9 | 21 | 25 | 145 | 330 | 0.0001 |
| 51 | NB_CONNEXION_WAP_6M_ | 200 | 8 | 11 | 16 | 165 | 181 | 0.0000 |
| 52 | SOLD_FIN_MOIS_DAV_P | 200 | 7 | 21 | 20 | 152 | 528 | 0.0000 |
| 53 | Nb_ADI_conso_autres | 200 | 7 | 14 | 12 | 167 | 179 | 0.0002 |
| 54 | VP | 200 | 5 | 24 | 17 | 154 | 171 | 0.0005 |
| 55 | NB_EVT_SPECIF_ANMOBL | 200 | 5 | 16 | 16 | 163 | 547 | 0.0001 |
| 56 | NB_CARTE_DD_P | 200 | 5 | 13 | 13 | 169 | 182 | 0.0000 |
| 57 | NB_CONNEXION_IMB_6M_ | 200 | 5 | 6 | 13 | 176 | 189 | 0.0000 |
| 58 | Mt_dav_oc_prof | 200 | 3 | 35 | 37 | 125 | 383 | 0.0002 |
| 59 | Nb_Credit_Total | 200 | 3 | 10 | 17 | 170 | 365 | 0.0000 |
| 60 | CRDT_RETARD | 200 | 2 | 10 | 9 | 179 | 188 | 0.0001 |
| 61 | Nb_credit_OC | 200 | 1 | 8 | 14 | 177 | 191 | 0.0000 |
| 62 | Mt_CT_predica_capita | 200 | 0 | 4 | 10 | 186 | 196 | 0.0000 |
| 63 | Nb_credit_tresorerie | 200 | 0 | 0 | 0 | 199 | 199 | 0.0004 |
| 64 | Nb_CT_Produits_Patri | 200 | 0 | 0 | 0 | 198 | 198 | 0.0004 |

présélectionner une soixantaine de variables pour en retenir un tiers à la fin. Sur un effectif de quelques centaines, en l'occurrence ici plus de 900, le modèle pourrait être complexe et comprendre jusqu'à une vingtaine de variables, ce qui nous a incité à en présélectionner une soixantaine.

Nous commençons par trier les variables par rang V de Cramer croissant, c'est-à-dire par valeur décroissante de la valeur absolue du χ^2 de V de Cramer. Nous choisissons ce critère en priorité parce qu'il est calculé pour toutes les variables explicatives. Nous remarquons toutefois que certaines variables n'ont pas de χ^2 de Kruskal-Wallis puisqu'elles sont quantitatives, mais sont très bien placées par le critère de V de Cramer : elles seront donc sélectionnées (la sélection sera indiquée par un fond jaune au niveau de la première colonne).

Nous regardons ensuite le graphique représentant les valeurs décroissantes de la valeur absolue du χ^2 de V de Cramer pour chercher un seuil de sélection de variable.



Les deux premières variables se détachent visiblement, de même que les quatre premières, mais il nous faut évidemment sélectionner plus de variables. On observe ensuite un décrochement assez net après la dix-septième variable : on passe d'un $\chi^2 = 12.81$ à 12.07 , soit une baisse de plus de 2.5 fois la baisse moyenne constatée. Toutefois, nous avons dit que nous cherchions à sélectionner une soixantaine de variables : nous cherchons donc le ou les décrochement(s) suivant(s). Il s'en présente un assez net entre la 27^e variable ($\chi^2=11.21$) et la 28^e variable ($\chi^2=10.77$) , soit une baisse de plus de 1.8 fois la baisse moyenne constatée. Le problème est qu'un bon nombre de variables ont été éliminées puisqu'elles étaient identifiées comme trop corrélées à d'autres. En procédant à ces éliminations sur toutes les variables du fichier, nous sommes tombés à 63 variables, ce qui correspond au nombre souhaité.

Nous trions ensuite la liste des variables par rang RgKruskal croissant, c'est-à-dire par valeur décroissante du χ^2 de Kruskal-Wallis(en valeur absolue). Nous constatons une grande

cohérence entre ce critère et celui du V de Cramer, montrant que la discrétisation respecte la hiérarchie des pouvoirs discriminants. Cette cohérence n'est pas toujours aussi forte, et vient ici de ce que les variables ont une distribution assez régulière et que leur discrétisation ne perturbe pas trop la hiérarchie de leur liaison avec la cible.

Enfin nous examinons la liste des variables selon le troisième critère : le classement par la procédure STEPDISC, le rang de la variable étant dans la colonne « RgDisc » ci-dessous. Un certain nombre de variables repérées par STEPDISC ne sont pas bien classées, ni par le χ^2 de Kruskal-Wallis, ni par le V de Cramer, et ne seront pas sélectionnées. Toutefois, la quatorzième variable sélectionnée par STEPDISC, « Nb_ct_auto » n'est pas trop mal classée selon les deux autres critères soit la 5ème et la 4ème position et sera donc retenue.

Certaines variables, pourtant bien classées, ne sont pas retenues, car elles ont été précédemment jugées inopportunes en raison de leur distribution ou de leur corrélation avec d'autres variables. C'est ainsi le cas de « Nb_ca_comptes_titres » et « Mt_ca_comptes_titres » écartées en raison de leur corrélation avec « Mt_ca_comptes_titres_non_abonne » qui est elle-même plus corrélée à la variable cible. Concernant le rang de sélection sur régressions logistiques sur échantillons bootstrap multiples, nous pouvons en conclure qu'il nous donne une sélection des variables très proche du résultat final de la modélisation du score mais ce n'est pas une surprise en soi par construction.

| Variable explicative | V Cramer | abs V Cramer | Rang V Cramer | Kruskal Wallis | Rang Kruskal Wallis | Rang STEPDISC | Rang logit sur échantillons bootstrap multiples | selection sans la classification | N° classe | selection avec la classification |
|---------------------------------|----------|--------------|---------------|----------------|---------------------|---------------|---|----------------------------------|-----------|----------------------------------|
| ANCT MRA | 47% | 46,67% | 1 | 5 963 | 1 | | | X | | 28 |
| ANCT | 20% | 19,74% | 2 | 929 | 2 | 1 | | X | | 28 |
| Mt ct Mrh | 16% | 15,92% | 3 | 747 | 3 | 3 | | X | | 33 |
| Mt ca comptes titres non abonne | 16% | 15,92% | 4 | 521 | 10 | 16 | | X | | 15 |
| Nb ct Auto | -16% | 15,53% | 5 | 693 | 4 | 14 | 2 | X | | 16 |
| Nb ca comptes titres | 15% | 15,35% | 6 | 667 | 5 | 1 | | | | 15 |
| Mt ca comptes titres | 15% | 15,06% | 7 | 601 | 7 | | 34 | | | 15 |
| SGMT | 15% | 14,82% | 8 | 585 | 8 | 4 | | X | | X |
| Nb AssurancesIARD Total | 15% | 14,64% | 9 | 643 | 6 | 22 | | | | 31 |
| Nb ca comptes titres non abonne | 14% | 14,06% | 10 | 545 | 8 | | | | | 15 |
| SITUA_FAMIL | 14% | 13,95% | 11 | | | | 18 | X | | X |
| AGE1 | 14% | 13,54% | 12 | 332 | 20 | 23 | 30 | X | | 29 |
| NB PRDT DETN P | 14% | 13,53% | 13 | 537 | 9 | 30 | 44 | X | | 2 |
| UA | 13% | 13,34% | 14 | | | | 14 | X | | X |
| AGEPERSD | 13% | 12,98% | 15 | | | | 7 | X | | X |
| SENTRGPT | 13% | 12,98% | 16 | | | | | | | |
| Mt AssurancesIARD Total | 13% | 12,81% | 17 | 505 | 11 | 6 | | | | 16 |
| Mt ct Auto | 12% | 12,07% | 18 | 415 | 13 | 2 | | | | 16 |
| Nb cc Epargne Logement | 12% | 11,91% | 19 | 432 | 12 | | 37 | X | | X |
| NB PERS_MAJORS | 12% | 11,57% | 20 | 368 | 16 | | | | | 32 |
| NB CARTE DD P | -12% | 11,52% | 21 | 369 | 19 | | 56 | X | | 33 |
| Mt cc Epargne Logement | 12% | 11,52% | 22 | 362 | 17 | | | | | 1 |
| SOLDE COLLECTE P | 11% | 11,32% | 23 | 385 | 14 | | 32 | X | | 1 |
| Nb collecte cam | 11% | 11,29% | 24 | 342 | 18 | | | | | 30 |
| RSD | 11% | 11,21% | 25 | | | | | | | |
| NB CONTRAT BAM P | 11% | 10,77% | 27 | 209 | 38 | | 31 | X | | 13 |
| MT MOY M PAIANT CHO P | 11% | 10,62% | 28 | 312 | 21 | | 26 | X | | 2 |
| NB cc Epargne Total | 10% | 10,28% | 29 | 282 | 25 | 31 | | | | 30 |
| PCSP | 10% | 10,22% | 30 | | | | 15 | X | | X |
| Mt depot DAV De PEA | 10% | 10,13% | 31 | 290 | 23 | | 24 | X | | 3 |
| Nb depot DAV De PEA | -10% | 10,11% | 32 | 298 | 3 | | | | | 3 |
| Mt collecte cam | 10% | 10,00% | 34 | 280 | 27 | | | | | 1 |
| SOLDE EP Bil P | 10% | 9,89% | 35 | 283 | 24 | | | | | 1 |
| Mt cc Epargne Total | 10% | 9,83% | 36 | 273 | 28 | | | | | 1 |
| NB PERS_CLT | 10% | 9,83% | 37 | 232 | 32 | | | | | 32 |
| MT MVT DEBTR DOM DAV PART 12MP | 10% | 9,65% | 38 | 255 | 30 | | 22 | X | | 2 |
| REC_PRO | 10% | 9,60% | 39 | | | | | | | |
| NB MVT CRED DAV PART 12M P | 9% | 9,49% | 40 | 249 | 31 | 5 | | | | 2 |
| NB EVT SPECIF ANMOBL P | 9% | 9,42% | 41 | 264 | 29 | | 55 | X | | 1 |
| Mt CT predica euro | 9% | 9,28% | 42 | 118 | 51 | 25 | 36 | X | | 9 |
| MT MVT CRED DAV PART 12M P | 9% | 9,22% | 43 | 230 | 33 | | | | | 2 |
| MT MVT DEB DAV PART 12M P | 9% | 9,21% | 44 | 228 | 34 | | | | | 2 |
| Mt ca comptes titres pes na | 9% | 9,08% | 45 | 221 | 41 | | | | | 3 |
| Nb ca comptes titres pes na | -9% | 8,72% | 46 | 215 | 36 | | | | | 3 |
| NB ENTRTN ANMOBL P | 9% | 8,71% | 47 | 211 | 37 | 4 | 8 | X | | 36 |
| SOLD MOY DAV 12M P | 9% | 8,61% | 48 | 193 | 39 | | 6 | X | | 8 |
| SOLD MOY DAV PART 12M P | 8% | 8,21% | 49 | 175 | 43 | 8 | | | | 4 |
| Mt credit hab | 8% | 8,15% | 50 | 48 | 83 | | | X | | 9 |
| Mt CT Total | 8% | 8,08% | 51 | 128 | 47 | | | | | 9 |
| NB CT Predica | 8% | 8,06% | 52 | 128 | 48 | | | | | 9 |
| NB CARTE DI P | 8% | 7,97% | 53 | 140 | 43 | | 29 | X | | 13 |
| SOLD FIN MOIS DAV P | 8% | 7,90% | 54 | 167 | 41 | 18 | 52 | X | | 8 |
| NB MVT DEB DAV PART 12M P | 8% | 7,73% | 55 | 163 | 42 | 11 | 11 | X | | 2 |
| MT MVT CRDTR DOM DAV PART 12MP | 7% | 7,47% | 56 | 109 | 53 | | | | | 2 |
| MARCHE | 7% | 7,21% | 57 | | | | | | | |
| Mt div | 7% | 7,17% | 58 | 137 | 44 | | | | | 8 |
| Mt div clientele | 7% | 7,16% | 59 | 136 | 45 | | | | | 8 |
| Mt div cc prive | 7% | 7,00% | 60 | 131 | 46 | | | | | 8 |
| Nb ct MrhPNO | -7% | 6,65% | 61 | 123 | 49 | | 9 | X | | 33 |
| NB CARTE DS CSCA P | 7% | 6,61% | 62 | 97 | 57 | | 23 | X | | 13 |
| NB CT predica euro | 7% | 6,51% | 63 | 87 | 61 | | | | | 4 |
| Nb credit hab | 6% | 6,36% | 64 | 121 | 50 | | | | | 4 |
| IYP | -6% | 6,35% | 65 | | | | 17 | X | | X |
| NB CHO 12M P | 6% | 6,33% | 66 | 116 | 52 | | 45 | X | | 25 |
| NB CSCA | 6% | 6,32% | 67 | 86 | 63 | 10 | | | | 13 |
| Nb ct Mrh | -6% | 6,27% | 68 | 103 | 54 | 21 | 48 | X | | 33 |
| Mt ADI hab | 6% | 6,22% | 69 | 88 | 76 | 27 | | | | 4 |
| Nb cc Livrets | 6% | 6,17% | 70 | 94 | 58 | | 16 | X | | 30 |
| SOLDE LIVRETS P | 6% | 6,03% | 71 | 93 | 59 | | 33 | X | | 30 |
| Nb CT Total | 6% | 5,93% | 72 | 85 | 63 | | | | | 9 |
| NB CT Predica | 6% | 5,93% | 73 | 85 | 63 | | | | | 9 |
| Mt cc Livrets | 6% | 5,90% | 74 | 88 | 60 | | | | | 30 |
| Nb ADI hab | 6% | 5,90% | 75 | 104 | 55 | | | | | 4 |
| Nb Credit Total | 6% | 5,84% | 76 | 88 | 75 | | 58 | X | | 4 |
| BALEIF | 6% | 5,69% | 77 | | | | 13 | X | | X |
| NB REMISE 12M P | 6% | 5,65% | 78 | 98 | 56 | | | | | 25 |
| NB EVT COMNTR ANMOBL P | 6% | 5,62% | 79 | 67 | 70 | | 38 | X | | 1 |
| Nb dav | 5% | 5,46% | 80 | 85 | 64 | 8 | 19 | X | | 11 |
| Nb dav clientele | 5% | 5,44% | 81 | 72 | 67 | | 12 | X | | 11 |
| MT MOY PAIANT M CARTE P | 5% | 5,43% | 82 | 52 | 78 | 16 | 27 | X | | 2 |
| Nb dav cc prive | 5% | 5,14% | 83 | 65 | 72 | | | | | 11 |
| Nb ADI Total | 5% | 5,11% | 84 | 57 | 77 | | | | | 4 |
| Nb ca comptes titres pes abonne | -5% | 5,03% | 85 | 71 | 68 | | | | | 14 |
| Mt ca comptes titres pes abonne | -5% | 4,99% | 86 | 69 | 69 | | | | | 14 |
| MT MOY M REMISE CHO P | 5% | 4,99% | 87 | 72 | 66 | | 20 | X | | 25 |
| Nb ct PJ | -5% | 4,88% | 88 | 66 | 71 | 7 | 10 | X | | 31 |
| Nb credit OC | -5% | 4,85% | 89 | 64 | 73 | | 61 | X | | 19 |
| MRDUS_OG | -5% | 4,85% | 90 | 64 | 74 | | | | | 19 |
| MIRE | 4% | 4,45% | 91 | | | | 39 | X | | X |
| Nb ADI equipment | -4% | 4,24% | 93 | 49 | 79 | | | | | 5 |
| Mt ADI equipment | -4% | 4,24% | 93 | 49 | 81 | | | | | 5 |
| Nb credit equipment | -4% | 4,24% | 93 | 49 | 80 | | 28 | X | | 5 |
| Mt credit equipment | -4% | 4,24% | 93 | 49 | 82 | | | | | 5 |
| AFFECTE | 4% | 4,01% | 96 | | | | 40 | X | | X |
| ASSU_VIE | 4% | 3,86% | 97 | | | | 53 | | | |
| Nb ADI conso autres | 4% | 3,84% | 99 | 40 | 84 | | | X | | 6 |
| Mt ADI conso autres | 4% | 3,84% | 99 | 38 | 85 | | | | | 6 |
| Mt credit conso autres | 4% | 3,63% | 101 | 38 | 86 | 29 | | | | 6 |
| Nb credit conso autres | 4% | 3,63% | 101 | 36 | 88 | | | | | 6 |
| VP | 4% | 3,56% | 102 | | | | 54 | X | | X |
| Mt ADI Total | 3% | 3,34% | 104 | 30 | 90 | | | | | 4 |
| Mt ADI conso | -3% | 3,34% | 104 | 30 | 90 | | | | | 4 |
| NB ENF CLI | -3% | 3,23% | 105 | 36 | 87 | | | | | 11 |
| Mt Credit Total | -3% | 3,13% | 106 | 16 | 95 | 28 | | | | 4 |
| NB CONNEXION VDTX 6M P | -3% | 2,71% | 107 | 20 | 91 | | 25 | X | | 37 |
| MT MOY RETRAIT M CARTE P | 3% | 2,60% | 108 | 19 | 92 | 13 | 46 | X | | 13 |
| Mt CT predica capitalisation | -2% | 2,43% | 110 | 16 | 93 | | 62 | X | | 21 |
| Nb CT predica capitalisation | -2% | 2,43% | 110 | 16 | 94 | | | | | 21 |
| Nb dav cc prof | 2% | 2,30% | 111 | 10 | 103 | | 58 | X | | 23 |
| NB ENF | -2% | 2,27% | 112 | 14 | 97 | | | | | 23 |
| Nb ct ZROUES | -2% | 2,26% | 113 | 14 | 96 | | | | | 34 |
| NB CARTE PAIEMENT P | 2% | 2,22% | 114 | 5 | 112 | 32 | | | | 13 |
| Mt cc émissions permanentes | -2% | 2,06% | 115 | 13 | 99 | | 42 | X | | 27 |
| Mt capital OPEN | 2% | 1,97% | 117 | 9 | 107 | | | | | 12 |
| Nb contrat OPEN | 2% | 1,97% | 117 | 10 | 101 | 12 | 12 | X | | 12 |
| NB CONNEXION BAM 6M P | 2% | 1,86% | 118 | 5 | 111 | | 9 | X | | 28 |
| Nb cc émissions permanentes | -2% | 1,82% | 119 | 9 | 105 | 24 | | | | 27 |
| NB CONNEXION WAP 6M P | 2% | 1,81% | 120 | 9 | 106 | | 51 | X | | 35 |
| Nb credit conso auto | -2% | 1,77% | 122 | 8 | 108 | 17 | | | | 7 |
| Nb credit conso auto | -2% | 1,77% | 122 | 10 | 102 | | 47 | X | | X |
| NB CONNEXION WEB 6M P | 2% | 1,74% | 123 | 3 | 119 | | | | | 26 |
| Nb CT predica UC | -2% | 1,74% | 125 | 10 | 104 | 20 | | | | 28 |
| Nb ADI predica UC | -2% | 1,74% | 125 | 14 | 98 | | 43 | X | | 28 |
| Nb ADI conso auto | 2% | 1,56% | 127 | 8 | 109 | | | | | 7 |
| Mt ADI conso auto | -2% | 1,56% | 127 | 11 | 108 | | | | | 7 |
| Mt dav successions | 2% | 1,52% | 128 | 4 | 114 | | 49 | X | | 24 |
| NB CONNEXION IMB 6M P | 2% | 1,52% | 128 | 6 | 110 | 35 | | X | | 35 |
| Nb dav successions | -1% | 1,22% | 130 | 4 | 113 | 36 | | | | 24 |
| NB CARTE OPEN P | -1% | 1,11% | 131 | 1 | 123 | | | | | 12 |
| Nb credit tresorie | -1% | 0,95% | 134 | 2 | 117 | | 63 | X | | 10 |
| Mt credit tresorie | -1% | 0,95% | 134 | 2 | 119 | | | | | 10 |
| NB CT Produits Patrimoniaux | -1% | 0,95% | 134 | 2 | 117 | | 64 | X | | 17 |
| Mt CT Produits Patrimoniaux | -1% | 0,95% | 134 | 2 | 118 | | | | | 17 |
| CRDT RETARD | -1% | 0,92% | 136 | | | | 60 | X | | X |
| Nb ct GAV | 1% | 0,69% | 137 | 1 | 121 | 26 | 41 | X | | 31 |
| Nb ADI tresoria | -1% | 0,60% | 139 | 1 | 122 | | | | | 10 |
| Nb ADI tresoria | -1% | 0,60% | 139 | 1 | 123 | | | | | 10 |
| Nb credit engagement signature | 0% | 0,23% | 141 | 0 | 124 | | | | | 18 |
| Mt credit engagement signature | 0% | 0,23% | 141 | 0 | 124 | | | | | 18 |
| Nb ca comptes titres abonnees | 0% | 0,21% | 142 | 0 | 126 | | | | | 22 |
| Mt depot Collecte Monetaire | 0% | 0,16% | 143 | 0 | 127 | | | | | 20 |
| Nb depot Collecte Monetaire | 0% | 0,16% | 144 | 0 | 128 | 34 | | | | 20 |
| Mt ca comptes titres abonnees | 0% | 0,14% | 145 | 0 | 129 | | | | | 22 |
| NB CARTE RETRAIT P | 0% | 0,04% | 148 | 0 | 130 | | | | | 34 |

g. Sélection de variables après leur classification

Nous avons remarqué, dans la section précédente, que la sélection des variables les plus discriminantes selon un critère tel que V de Cramer ou le χ^2 de Kruskal-Wallis fait souvent retenir, dans un premier temps, des variables très corrélées entre elles, telles que « Nb_ca_comptes_titres » et « Mt_ca_comptes_titres_non_abonne », parmi lesquelles il faut ensuite opérer une sélection supplémentaire pour éviter de voir la qualité des modélisations ultérieures dégradée par de fortes corrélations des variables du modèle. Ce qui fait que sur 147 variables au départ, seules 63 seront retenues pour démarrer la modélisation du score.

La méthode de sélection présentée n'apparaît alors pas complètement appropriée car elle laisse au statisticien un travail résiduel important de sélection manuelle des variables.

Une façon à la fois rigoureuse, élégante et pratique de résoudre cette difficulté consiste à opérer une classification des variables sélectionnées, afin de mettre en évidence des groupes homogènes de variables, puis à aller chercher, dans chaque groupe, des variables, parfois moyennement discriminantes dans l'absolu mais suffisamment peu liées aux variables explicatives des autres groupes pour être potentiellement intéressantes et dignes d'être sélectionnées. Il vaut mieux souvent sélectionner la variable la plus discriminante d'un groupe de variables, même si son pouvoir discriminant est globalement moyen, que de sélectionner la variable la moins discriminante de son groupe, fût-elle globalement très discriminante. L'idée est donc, plutôt que de sélectionner les N variables globalement les plus discriminantes, de sélectionner la (ou les deux variables) plus discriminante(s) de chaque groupe.

Nous allons utiliser la procédure VARCLUS de sas qui nous permet d'implémenter une méthode aboutie de classification des variables. Elle consiste en une classification descendante hiérarchique combinée à une ACP avec rotation.

La procédure VARCLUS résume son processus de classification sous la forme d'un tableau (le tableau détaillé se trouve en Annexe II), montrant l'affinement progressif de la classification, avec une proportion de la variance expliquée en augmentation régulière et importante, de 15.83% à 37.21%. On voit que, hormis le saut de 1 à 3 classes, l'évolution de la variance expliquée ne montre pas de point d'inflexion susceptible de suggérer un nombre de classes inférieur à 37. Qui plus est, on constate que c'est seulement à partir de 37 classes que le R2 minimum pour une variable, signalant quand il est bas une variable mal classée, fait un bond de 0.0273 à 0.1691. Tous ces éléments incitent à conserver le nombre de 37 classes, sans augmenter le seuil de seconde valeur propre de la procédure VARCLUS afin de diminuer le nombre de classes.

1

No cluster meets the criterion for splitting.

| Nombre de classifications | Variation totale expliquée par classifications | Proportion de variation expliquée par classifications | Proportion minimum expliquée par une classification | Valeur propre seconde maximum dans une classification | R carré minimum pour une variable | 1-R**2 Ratio maximum pour une variable |
|---------------------------|--|---|---|---|-----------------------------------|--|
| 1 | 20.583449 | 0.1583 | 0.1583 | 13.626600 | 0.0000 | |
| ... | ... | ... | ... | ... | ... | ... |
| 36 | 103.930670 | 0.7995 | 0.3721 | 1.004864 | 0.0273 | 0.9787 |
| 37 | 104.917154 | 0.8071 | 0.3721 | 0.965550 | 0.1691 | 0.8844 |

Le tableau suivant indique la répartition des variables analysées suivant les 37 classes finalement obtenues (« cluster 1 » à « cluster 37 »).

La colonne « R2 with own cluster » (« R carré avec propre classification ») indique le coefficient de corrélation entre la variable et sa classe. La colonne « R2 with next closest » (« R carré avec le plus proche») indique le coefficient de corrélation entre la variable et la classe (autre que la sienne) la plus proche. La colonne suivante est le rapport entre 1-R2 (own cluster) et 1-R2 (next closest), qui sera d'autant plus petit que la variable sera bien classée et pourrait même être >1 si la variable était particulièrement mal classée.

La totalité du tableau se trouvant en Annexe II

| 37 Clusters | | R carré avec | | 1-R**2 Ratio |
|----------------|------------------------|--------------------------|-------------------|-----------------|
| Classification | Variable | Propre classification | Le plus proche | |
| Cluster 1 | Mt_cc_Epargne_Logement | 0.7355 | 0.3537 | 0.4092 |
| | Mt_cc_Epargne_Total | 0.9308 | 0.6116 | 0.1782 |
| | Mt_collecte_cam | 0.9354 | 0.6110 | 0.1661 |
| | SOLDE_COLLECTE_P | 0.8125 | 0.5254 | 0.3951 |
| | SOLDE_EP_BIL_P | 0.9126 | 0.5907 | 0.2135 |
| | Nb_cc_Epargne_Logement | 0.6782 | 0.3648 | 0.5066 |
| | NB_EVT_COMNTR_ANMOBL_P | 0.3308 | 0.1890 | 0.8251 |
| | NB_EVT_SPECIF_ANMOBL_P | 0.2741 | 0.1792 | 0.8844 |
| ... | ... | ... | ... | ... |
| Cluster 35 | NB_CONNEXION_IMB_6M_P | 0.5464 | 0.0052 | 0.4560 |
| | NB_CONNEXION_WAP_6M_P | 0.5464 | 0.0069 | 0.4567 |
| Cluster 36 | NB_ENTRTN_ANMOBL_P | 1.0000 | 0.0908 | 0.0000 |
| Cluster 37 | NB_CONNEXION_VDTX_6M_P | 1.0000 | 0.0067 | 0.0000 |

On récupère en sortie du fichier OUTSTAT les enregistrements de type GROUP, qui nous fournit, pour chaque valeur du nombre _NCL_ de classes de variables retenues, la classe à laquelle appartient chacune des variables analysées. On retrouve dans les deux dernières colonnes respectivement le numéro de la classe d'appartenance de la variable « N°classe » ainsi que la sélection définie par une croix « sélection avec la classification ». Nous passons alors de 63 à 45 variables sélectionnées.

6. Modélisation du score

a. Construction des échantillons d'apprentissage et de validation

Afin d'obtenir ultérieurement des mesures non biaisées des performances des modèles que nous construirons, nous séparons par tirage aléatoire la population étudiée en un échantillon pour l'apprentissage des modèles et un échantillon pour leur validation. Les performances des modèles sont bien sûr à mesurer et à comparer sur l'échantillon de validation.

Nous réalisons un tirage aléatoire simple selon la loi uniforme entre 0 et 1 effectué par la fonction RANUNI de SAS.

On a choisi d'utiliser deux tiers des observations pour l'apprentissage et un tiers pour la validation.

b. Calcul de l'aire sous la courbe ROC

L'aire sous la courbe ROC (Receiver Operating Characteristic), en abrégé AUC (Area Under Curve), d'un modèle de score sera calculée à l'aide d'une macro sas en se basant sur la statistique de Wilcoxon et son lien avec AUC. Cette courbe est une méthode universelle de représentation du pouvoir discriminant d'un modèle prédictif binaire. Pour rappel l'aire sous la courbe ROC est la probabilité pour que le score d'un individu pris au hasard parmi les « cible = 1 » (client à fort potentiel de marge en Assurance) soit supérieur au score d'un individu pris au hasard parmi les « cible = 0 » (pas un client à fort potentiel de marge en Assurance). Cette propriété explique le lien avec la statistique de Wilcoxon.

c. Régression logistique pas à pas sur l'ensemble des variables

La régression logistique est de loin la méthode de modélisation la plus répandue car la généralité, l'interprétabilité et la robustesse sont trois atouts majeurs de cette technique fondamentale. C'est donc logiquement celle qui sera utilisée. Il s'agit ici de régression logistique binaire, puisque la variable à expliquer n'a que deux modalités. Nous modéliserons à l'aide de la procédure LOGISTIC de sas.

Dans la régression logistique binaire, on considère une variable cible $Y = 0$ ou 1 , et p variables explicatives X_i continues, binaires ou qualitatives. Dans les notations qui suivent, on rassemble les p variables X_j dans un vecteur $X = (X_1, X_2, \dots, X_p)$.

L'objectif de la régression logistique est celui de toute régression : modéliser l'espérance conditionnelle $E[Y/X=x]$. On veut connaître la valeur moyenne de Y pour toute valeur de X . Pour une valeur Y valant 0 ou 1 (loi bernoulli), cette valeur moyenne est la probabilité que $Y = 1$. On a donc :

$$E[Y/X=x] = \text{Prob}[Y=1/X=x].$$

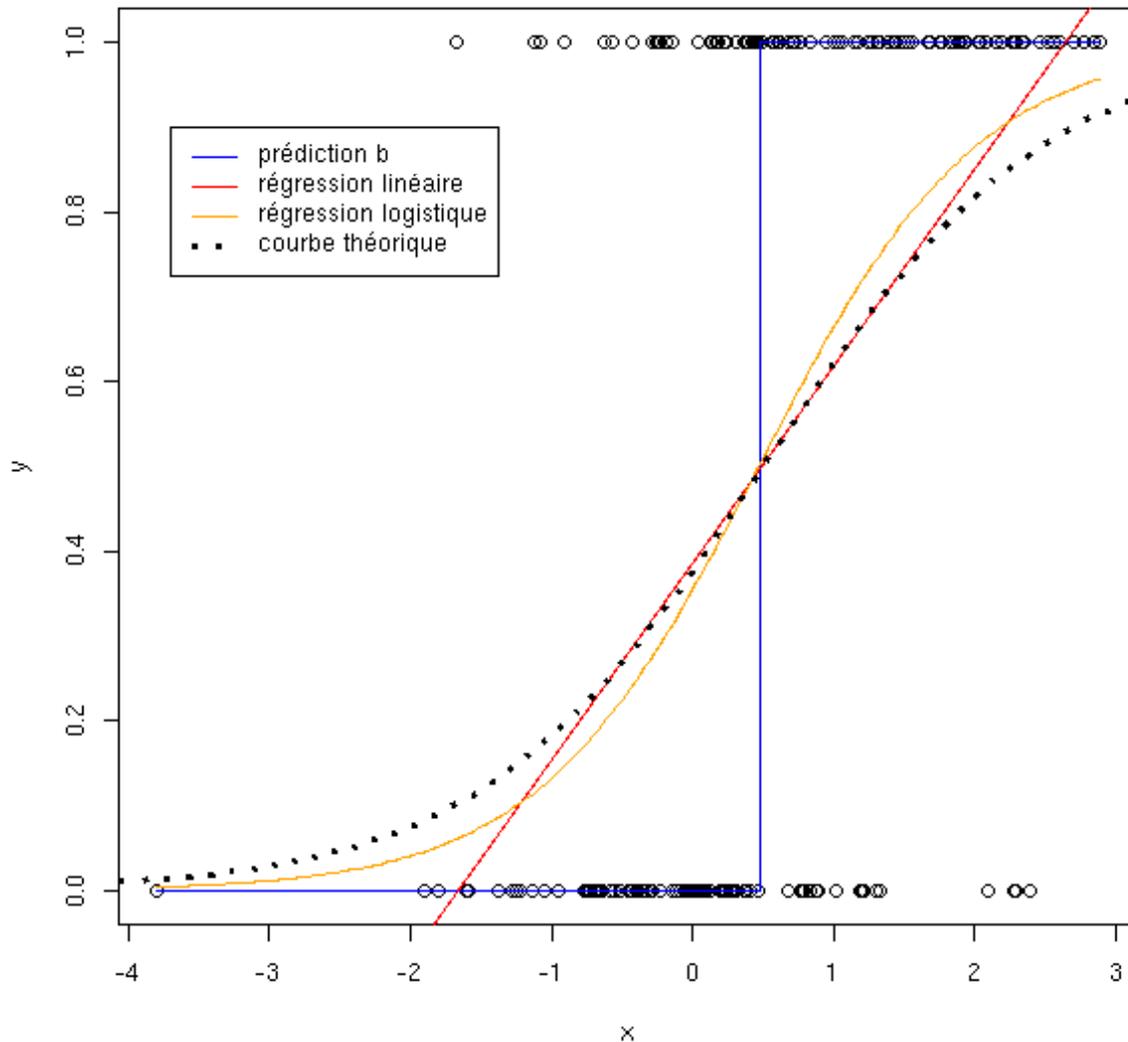
En régression linéaire, on cherche à faire passer un hyperplan au milieu du nuage des points $(x_1, x_2, \dots, x_p, y)$, de sorte que l'ensemble des valeurs moyennes de Y pour toutes les valeurs de X est approché par cet hyperplan, d'équation :

$$E[Y/X=x] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p.$$

Cette approximation ne convient évidemment plus lorsque $Y = 0$ ou 1 , puisque le terme $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ est non borné alors que $\text{Prob}[Y=1/X=x]$ est dans l'intervalle $[0,1]$.

En réalité, dans la solution favorable (celle qui nous intéresse) où X parvient à discriminer Y , le nuage de points (x,y) a une allure dont l'exemple typique est celui de la figure ci-dessous : quand x est petit, on a plus souvent $y=0$, et quand x est grand, on a plus souvent $y=1$. Pour simplifier, nous nous sommes placés ici dans le cas très classique d'une variable x multinormale et homoscedastique : $x \sim N(0,1)$ sur l'ensemble des points tels que $y=0$ et $x \sim N(1,1)$ sur l'ensemble des points tels que $y=1$.

Comparaison des régressions linéaire et logistique



Visiblement, les valeurs $\text{Prob}[Y=1/X=x]$ quand x varie suivent la courbe théorique représentée par les croix « + » sur la figure. C'est une courbe en S et non une droite. Plus précisément, avec les hypothèses faites ici sur la distribution de x , la formule de Bayes montre que la courbe de la fonction $x \rightarrow \text{Prob}[Y=1/X=x]$ est donnée par $f_{N(1,1)}(x) / [f_{N(1,1)}(x) + f_{N(0,1)}(x)]$, où $f_{N(\mu,\sigma)}$ est la fonction de densité de la $N(\mu,\sigma)$. Si l'on suit l'expression de cette courbe, on peut écrire $\pi(x) = \text{Prob}[Y=1/X=x]$ sous la forme :

$$\pi(x) = \frac{\exp(\beta_0 + \sum \beta_j x_j)}{1 + \exp(\beta_0 + \sum \beta_j x_j)}$$

Equation équivalente à

$$\text{Log} \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

La fonction $f(p) = \log(p/1-p)$ est appelée logit. C'est un cas particulier des fonctions de lien rencontrées dans les régressions logistiques et les modèles linéaires généralisés. Dans ce type de modèle, ce n'est donc pas l'espérance $\pi(x) = E(Y/X=x)$ qui est écrite comme combinaison

linéaire des variables explicatives, mais $f(\pi(x))$, f étant la fonction de lien. Dans la régression logistique la plus courante, on modélise le logit de l'espérance conditionnelle comme combinaison linéaire des variables explicatives. Cette écriture est cohérente avec la règle bayésienne de l'analyse discriminante et le calcul de la probabilité a posteriori qui en découle dans le cas d'une distribution normale de X/Y avec égalité des variances et égalité des probabilités a priori.

d. les odds-ratios

L'odds-ratio d'une variable explicative mesure l'évolution du rapport des probabilités d'apparition de l'évènement $Y = 1$ contre $Y = 0$ (odds = « cote » des parieurs) lorsque X_i passe de x à $x+1$. Dans ce cas, $\text{logit}(\pi(x))$ augmente du coefficient β_i de X_i et donc la cote $\pi(x) / [1-\pi(x)]$ est multipliée par $\exp(\beta_i)$. Ceci s'écrit :

$$\text{OR} = \frac{\pi(x+1) / [1-\pi(x+1)]}{\pi(x) / [1-\pi(x)]} = \exp(\beta_i)$$

Il nous faut bien faire la différence entre le risque relatif $\pi(x+1) / \pi(x)$ et l'odds-ratio sauf quand $\pi(x)$ est petit (détection des phénomènes rares).

Si X_i est binaire 0/1, la formule de l'odds-ratio devient :

$$\text{OR} = \frac{\text{Prob}(Y=1 / X_i=1) / \text{Prob}(Y=0 / X_i=1)}{\text{Prob}(Y=1 / X_i=0) / \text{Prob}(Y=0 / X_i=0)} = \exp(\beta_i)$$

Une variable X binaire a un seul odds-ratio. Si l'on s'intéresse au client à fort potentiel de marge en Assurance ($Y = 1$), un odds-ratio de 1.281 pour la variable «NB_ENTRTN_ANMOBL_P (=2 à 3 : ≥ 3 et 0 à 1 : ≤ 2) signifiera que le rapport d'un client à fort et à faible potentiel de marge en assurance est 1.281 plus important pour les clients ayant eu plus de 3 entretiens sur une année mobile que pour des clients ayant eu moins de 3 entretiens sur une année mobile.

e. Estimation des paramètres

Les paramètres à estimer dans un modèle logistique logit sont les coefficients de la combinaison linéaire exprimant le logit de la probabilité $\text{Prob}[Y=1/X=x]$. Nous nous trouvons dans la situation où il faut modéliser des données observées $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ dans lesquelles chaque y_i vaut 0 ou 1, et x_i est le vecteur de variables explicatives de la i ème observation. La notion fondamentale qui va nous servir à estimer les paramètres d'un modèle logit, est celle de vraisemblance. Celle-ci est, par définition, la probabilité d'obtenir les données observées $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$. Si $y_i=1$, la probabilité d'obtenir (x_i, y_i) sera par définition $\text{Prob}[Y=1/X=x_i] = \pi(x_i)$, où $\pi(x_i)$ est exprimé par :

$$\pi(x_i) = \frac{\exp(\beta_0 + \sum \beta_j x_i^j)}{1 + \exp(\beta_0 + \sum \beta_j x_i^j)}$$

Si $y_i=0$, la probabilité d'obtenir (x_i, y_i) sera $\text{Prob}[Y=0/X=x_i] = 1 - \pi(x_i)$. On peut unifier ces deux cas en écrivant que la probabilité d'obtenir (x_i, y_i) est :

$$\pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

Quelque soit y_i .

Maintenant, afin de poursuivre les calculs, il faut ajouter une hypothèse fondamentale au cadre de la régression logistique : les observations (x_i, y_i) sont indépendantes. Cette hypothèse

ne pourra être levée que dans une généralisation de la régression logistique (la régression logistique sur données corrélées). Cette hypothèse d'indépendance permet d'écrire que la probabilité d'obtenir les données observées $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ est le produit des probabilités :

$$\prod_{i: 1 \text{ à } n} \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

En remplaçant $\pi(x_i)$ par l'expression qui est la sienne dans le modèle logistique, on voit que la vraisemblance vaut :

$$\prod_{i: 1 \text{ à } n} \frac{\exp(\beta_0 + \sum \beta_j x_i^j)^{y_i}}{1 + \exp(\beta_0 + \sum \beta_j x_i^j)} \left(1 - \frac{\exp(\beta_0 + \sum \beta_j x_i^j)}{1 + \exp(\beta_0 + \sum \beta_j x_i^j)}\right)^{1-y_i}$$

C'est une fonction $L(\beta_0, \beta_1, \dots, \beta_p)$ des coefficients β_j qui doit évidemment être maximisée : il faut trouver les coefficients tels que $L(\beta_0, \beta_1, \dots, \beta_p)$ soit le plus proche possible de 1, car cela signifiera que le modèle s'ajuste le mieux possible aux données observées.

En réalité, la vraisemblance ne peut valoir 1, et le modèle ne peut correspondre parfaitement aux données observées, que dans un seul cas : lorsque le modèle contient autant de coefficients qu'il y a d'observations (x_i, y_i) distinctes. Un tel modèle est qualifié de saturé. Pour prendre une analogie, le modèle linéaire simple est saturé quand le nuage de points se réduit à deux points, et à la droite ajuste parfaitement le nuage formé des deux points.

La détermination du meilleur modèle logistique passera donc par une recherche des coefficients qui maximisent la vraisemblance. C'est un problème qui n'a pas de solution analytique, c'est-à-dire de solution s'exprimant directement à partir des données initiales, à l'instar des fonctions discriminantes de l'analyse discriminante que l'on obtient en inversant la matrice des covariances. Ici, la solution optimale $(\beta^0, \beta^1, \dots, \beta^p)$ sera trouvée par une méthode numérique itérative, les plus populaires de ces méthodes numériques étant les algorithmes de Newton-Raphson et de Fisher. On note $(\beta^0, \beta^1, \dots, \beta^p)$ l'estimation des coefficients par maximisation de la vraisemblance, qui conduit à une estimation $\pi^{\wedge}(x_i)$ de la probabilité conditionnelle $\pi(x_i)$ d'avoir $Y = 1$ sachant que $x = x_i$. C'est la valeur prédite par le modèle logistique. Il est intéressant de noter que l'on a :

$$\sum_{i: 1 \text{ à } n} y_i = \sum_{i: 1 \text{ à } n} \pi^{\wedge}(x_i)$$

Autrement dit, la somme des valeurs observées de Y est égale à la somme des valeurs prédites.

Comme en régression linéaire, on n'atteint jamais les véritables coefficients β_j (sauf avec un modèle saturé) et les estimateurs β^{\wedge}_j présentent une certaine variance. Un estimateur sera évidemment d'autant plus fiable que sa variance sera plus faible, et l'on privilégie les variables X_j pour lesquelles cette variance est faible. Plus précisément, le test de student suggère de définir ce que l'on appelle statistique de wald d'une variable X_j , qui vaut $(\beta_j / \sigma(\beta^{\wedge}_j))^2$. Sous hypothèse nulle H_0 que $\beta^{\wedge}_j = 0$, le ratio $(\beta_j / \sigma(\beta^{\wedge}_j))$ suit une loi normale standard. La statistique de wald permet de tester la significativité du coefficient estimé β^{\wedge}_j et l'apport de la variable X_j . Cet apport n'est réel que si la statistique de wald est > 4 , plus précisément si elle est $> 3.84 (= (1.96)^2)$. En effet, dire que $((\beta_j / \sigma(\beta^{\wedge}_j))^2)$ équivaut à dire que β_j se trouve à plus de deux écarts-types de 0, ce qui entraîne que son intervalle de confiance à 95% ne contient pas 0, et que l'intervalle de confiance à 95% de l'odds-ratio $\exp(\beta^{\wedge}_j)$ ne contient pas 1. Cet odds-ratio doit être significativement supérieur à 1 (influence positive de la variable) ou significativement inférieur à 1 (influence négative).

Il faut prendre garde qu'un odds-ratio peut-être très éloigné de 1 tout en ayant un intervalle de confiance si large qu'il contient la valeur 1 et que la variable est finalement sans grand intérêt. Nous verrons que cela se produit en présence de variables colinéaires et que l'absence de colinéarité est une hypothèse essentielle à vérifier très scrupuleusement.

f. Déviance et mesure de qualité d'un modèle

Soient n le nombre total d'individus (ou observations), k le nombre de variables explicatives d'un modèle ajusté donné (chaque variable qualitative à m modalités comptant pour $m-1$ dans le calcul de k), $L(\beta_k)$ la vraisemblance de ce modèle, $L(\beta_0)$ la vraisemblance du modèle réduit à la constante, et $L(\beta_{\max})$ la vraisemblance du modèle saturé. $L(\beta_{\max})$ est la vraisemblance maximale, à laquelle se compare la vraisemblance $L(\beta_k)$ de tout modèle. D'où la notion de déviance :

$$D(\beta_k) = -2 [\log L(\beta_k) - \log L(\beta_{\max})] = \log [L(\beta_{\max}) / L(\beta_k)]^2.$$

Elle est analogue à la somme des carrés résiduels SCE de la régression linéaire, et elle est calculée comme une somme sur l'ensemble des n individus. Elle est égale à la somme des carrés des résidus des déviations individuelles. Dans le cas d'une variable à expliquer 0/1, la vraisemblance du modèle saturé vaut 1 et on a donc :

$$D(\beta_k) = -2 [\log (\text{vraisemblance du modèle ajusté})],$$

Soit, d'après la formule de la vraisemblance vue dans la section précédente :

$$D(\beta_k) = -2 \sum_{i=1}^n [y_i \log (\pi(x_i)) + (1-y_i) \log (1 - \pi(x_i))]$$

Cette déviance se lit dans la sortie de SAS reproduite dans un des tableaux de la section suivante dans la colonne « intercept and covariates » (« constantes » et « covariables »).

Le résidu de la déviance de la $i^{\text{ème}}$ observation vaut :

$$\pm \sqrt{[y_i \log (\pi(x_i)) + (1-y_i) \log (1 - \pi(x_i))]}$$

Le signe de l'expression étant positif si la valeur observée y_i est plus grande que la valeur prédite $\pi(x_i)$, négatif sinon.

Le but de la régression logistique est de maximiser la vraisemblance $L(\beta_k)$ du modèle ajusté ou la log-vraisemblance $\text{Log}[L(\beta_k)]$, ce qui équivaut à minimiser la déviance $D(\beta_k)$.

Pour k variables explicatives données, on cherche les coefficients qui maximisent la vraisemblance $L(\beta_k)$, comme indiqué dans la section précédente. Une fois ces k coefficients trouvés, il faut savoir si l'on améliore le modèle en lui ajoutant / variables explicatives. Pour cela, on calcule la différence des déviations :

$$D(\beta_k) - D(\beta_{k+1}) = -2 [\log L(\beta_k) - \log L(\beta_{\max})] + 2 [\log L(\beta_k) - \log L(\beta_{\max})] = -2 [\log L(\beta_k) - \log L(\beta_{k+1})]$$

Qui est positive. On peut alors utiliser un résultat fondamental : sous l'hypothèse H_0 de la nullité des / nouveaux coefficients, la différence des déviations précédente suit une loi du χ^2 à / degrés de liberté.

En particulier, l'hypothèse nulle démontrant que le modèle s'ajuste bien aux données peut se reformuler en disant que les p paramètres du modèle (en incluant la constante) sont suffisants pour l'ajustement, et que les paramètres supplémentaires $p+1$, $p+2$, ..., n pour arriver au nombre de paramètres du modèle saturé sont tous nuls. Sous cette hypothèse nulle, la déviance du modèle ajusté suit une loi du χ^2 à $n-p$ degrés de liberté.

La notion de vraisemblance est donc essentielle pour comparer un modèle de régression logistique à un sous-modèle. C'est la plus utilisée et elle peut intervenir dans toute régression pas à pas : ascendante, descendante ou mixte.

Quand on hésite à ajouter une variable à un modèle à k variables, un bon critère de décision, avant de comparer les deux modèles à l'aide d'une courbe ROC, consiste à comparer la valeur $D(\beta_k) - D(\beta_{k+1})$ à la valeur théorique du χ^2 à 1 degré de liberté au seuil de 1% : $D(\beta_k) - D(\beta_{k+1}) > 6.63$ on peut ajouter la variable. Ce qu'il faut savoir, c'est donc, non pas si la vraisemblance a augmenté quand on a ajouté une variable, car elle augmente toujours, mais si elle a suffisamment augmenté.

Comme cas particulier de ce qui précède, la différence :

$$D(\beta_0) - D(\beta_k) = -2 [\log L(\beta_0) - \log L(\beta_k)]$$

Suit une loi du χ^2 à k degrés de liberté sous l'hypothèse H_0 de la nullité de tous les coefficients $\beta^1, \beta^2, \dots, \beta^k$. On rejette H_0 si la différence $D(\beta_0) - D(\beta_k)$ dépasse le seuil critique du χ^2 à k degrés de liberté. Plus la différence précédente est grande, meilleur est le modèle. Dit autrement, on veut que

$-2 \log [L(\beta_k)]$ soit la plus petite possible. On retrouve cette exigence sous une forme équivalente dans deux autres critères classiques :

- ✚ Le critère d'Akaïké AIC = $-2 \log [L(\beta_k)] + 2(k+1)$

- ✚ Le critère de Schwartz SC = $-2 \log [L(\beta_k)] + 2(k+1) \log(n)$, souvent noté BIC (Bayesian information criterion).

Ce qui distingue ces critères travaux est que BIC, plus que AIC, pénalise les modèles complexes (dès que $n > \exp(2)$, soit $n > 3.74$). BIC est préférable quand n est grand. Ces deux critères permettent de comparer deux modèles, le meilleur étant celui pour lequel AIC et SC sont les plus bas.

Comme l'ont montré certains travaux, pour la sélection des variables, ces critères que nous venons de voir, fondés sur la notion de vraisemblance, semblent préférables à l'autre critère courant qui est celui de la statistique de Wald (section précédente), au moins dans deux cas quand le nombre d'observations est faible ou quand les coefficients β_i sont grands.

g. Présentation des résultats

i. Présentation des résultats sur l'échantillon d'apprentissage

Nous avons choisi la procédure LOGISTIC de sas et après exécution sur l'échantillon d'apprentissage nous présentons les résultats en sortie de sas.

Les tableaux qui suivent viennent rappeler le nom de la variable à expliquer, le nombre d'observations traitées pour la modélisation ou simplement lues (pour application du modèle).

score de fidélité sur l'échantillon d'apprentissage

The LOGISTIC Procedure

| Informations sur le modèle | |
|----------------------------|---------------------|
| Data Set | WORK.APPRENT |
| Response Variable | qualite_SP_hs_CNAT0 |
| Number of Response Levels | 2 |
| Model | binary logit |
| Optimization Technique | Fisher's scoring |

| | |
|-----------------------------|-------|
| Number of Observations Read | 18229 |
| Number of Observations Used | 18212 |

| Profil de réponse | | |
|-------------------|---------------------|------------------|
| Valeur ordonnée | qualite_SP_hs_CNAT0 | Fréquence totale |
| 1 | 0 | 12123 |
| 2 | 1 | 6089 |

Probability modeled is qualite_SP_hs_CNAT0=1.

Note: 17 observations were deleted due to missing values for the response or explanatory variables.

Information sur les occurrences de références pour chaque variable retenue dans le modèle sont désignées par -1.

| Informations sur le niveau de classe | | | | |
|--------------------------------------|--------|-----------------------|----|----|
| Classe | Valeur | Variables de création | | |
| ANCT_MRA | 0 | 1 | 0 | 0 |
| | 1 | -1 | -1 | -1 |
| | 2 | 0 | 1 | 0 |
| | 3 | 0 | 0 | 1 |
| ANCT | 0 à 2 | -1 | | |
| | 3 | 1 | | |
| Nb_ct_Auto | 1 | -1 | | |
| | 3 | 1 | | |
| Mt_ct_Mrh | 0 | 1 | 0 | 0 |
| | 1 | 0 | 1 | 0 |
| | 2 | -1 | -1 | -1 |
| | 3 | 0 | 0 | 1 |
| NB_CONNEXION_BAM_6M_P | 1 | -1 | | |
| | 2 et 3 | 1 | | |
| NB_PRDT_DET_N_P | 0 et 2 | -1 | | |
| | 3 | 1 | | |
| IVP | 1 | 1 | | |
| | 2 | -1 | | |

Il s'agira de modéliser la valeur de la variable à expliquer qualite_SP_hs_CNAT0=1.

La procédure de sélection STEPWISE pas à pas peut commencer. Il faut signaler que l'ordre dans lequel on entre les variables dans le modèle, c'est-à-dire de la plus corrélée à la moins corrélée avec la variable cible, a son importance. Le détail du résultat de la procédure est en Annexe I.

Après la septième itération, plus aucune variable n'est suffisamment significative pour entrer dans le modèle, et le processus de sélection pas à pas s'interrompt. Un tableau rappelle l'ordre d'entrée des variables, avec une dernière colonne indiquant la significativité de la variable en termes de probabilité du χ^2 .

| Récapitulatif sur la sélection séquentielle | | | | | | | | |
|---|----------------------|----------|-----|-------------|----------------|---------------|------------|---|
| Étape | Effet | | D F | Nombre dans | Khi 2 du score | Khi 2 de Wald | Pr > Khi 2 | Libellé de variable |
| | Saisi | Supprimé | | | | | | |
| 1 | ANCT_MRA | | 3 | 1 | 3931.9193 | | <.0001 | Valeurs de ANCT_MRA remplacées par des rangs |
| 2 | Nb_ct_Auto | | 1 | 2 | 219.2914 | | <.0001 | Valeurs de Nb_ct_Auto remplacées par des rangs |
| 3 | Mt_ct_Mrh | | 3 | 3 | 157.4769 | | <.0001 | Valeurs de Mt_ct_Mrh remplacées par des rangs |
| 4 | SGMT | | 3 | 4 | 64.0364 | | <.0001 | |
| 5 | NB_CONNEXION_BAM_6M_ | | 1 | 5 | 24.6502 | | <.0001 | Valeurs de NB_CONNEXION_BAM_6M_P remplacées par des rangs |
| 6 | NB_PRDT_DET_N_P | | 1 | 6 | 16.1766 | | <.0001 | Valeurs de NB_PRDT_DET_N_P remplacées par des rangs |
| 7 | Nb_cc_Livrets | | 2 | 7 | 20.6652 | | <.0001 | Valeurs de Nb_cc_Livrets remplacées par des rangs |
| 8 | Nb_credit_OC | | 1 | 8 | 10.8115 | | 0.0010 | Valeurs de Nb_credit_OC remplacées par des rangs |
| 9 | IVP | | 1 | 9 | 8.5914 | | 0.0034 | |
| 10 | Mt_cc_émissions_perm | | 1 | 10 | 8.6568 | | 0.0033 | Valeurs de Mt_cc_émissions_permanentes remplacées par des rangs |

| Analyse des effets Type 3 | | | |
|---------------------------|----|---------------|------------|
| Effet | DF | Khi 2 de Wald | Pr > Khi 2 |
| ANCT | 1 | 9.1452 | 0.0025 |
| ANCT_MRA | 3 | 2927.3243 | <.0001 |
| Mt_ct_Mrh | 3 | 116.0056 | <.0001 |
| Nb_ct_Auto | 1 | 102.3259 | <.0001 |

| Analyse des effets Type 3 | | | |
|---------------------------|----|---------------|------------|
| Effet | DF | Khi 2 de Wald | Pr > Khi 2 |
| NB_PRDT_DET_N_P | 1 | 26.3955 | <.0001 |
| IVP | 1 | 10.8990 | 0.0010 |
| NB_CONNEXION_BAM_6M_ | 1 | 28.6775 | <.0001 |

Le tableau précédent dévolu à « l'analyse des effets de type 3 » est affiché par SAS en présence de variables qualitatives (c'est-à-dire de l'instruction CLASS). Cette analyse est effectuée, pour chaque variable, en comparant le sous-modèle excluant la variable au modèle incluant cette variable et les autres, afin de tester, soit la significativité de la déviance, soit ici (χ^2 de Wald) l'hypothèse nulle que cette variable est sans effet dans le modèle, pourvu que les autres variables y soient. L'analyse de type 3 sur le χ^2 de Wald est plus rapide à calculer que celle sur la log-vraisemblance, mais elle est moins fiable lorsque l'échantillon est petit.

Le tableau suivant est le plus utile pour le scoring, puisqu'il contient les coefficients de chaque variable dans la régression logistique (ou plus précisément les estimateurs de ces coefficients). Il s'agit de la colonne « Estimation ». Chaque estimateur est soumis à une certaine incertitude mesurée par son écart-type (« Erreur std ») et le ratio (Estimation/Erreur)² est d'autant plus élevé que l'estimateur est significativement différent de 0. Ce ratio est appelé « χ^2 de Wald » ou « statistique de Wald » et doit être supérieur à 6.63, au seuil de confiance de 99%, pour que l'estimateur du coefficient soit significativement différent de 0. Quand on examine les sorties d'une régression logistique, il faut particulièrement s'attacher aux χ^2 de Wald et s'assurer qu'ils sont suffisamment élevés, les plus bas correspondant aux variables les moins fiables. Ici, nous constatons que la plus petite valeur du χ^2 de Wald est 8.5914, qu'elle est proche du seuil critique 6.63 pour la variable « IVP ») mais n'est donc pas sujette à caution.

| Analyse des estimations de la vraisemblance maximum | | | | | | |
|---|---|----|------------|------------|---------------|------------|
| Paramètre | | DF | Estimation | Erreur std | Khi 2 de Wald | Pr > Khi 2 |
| Intercept | | 1 | -0.9350 | 0.0300 | 971.0393 | <.0001 |
| ANCT | 3 | 1 | 0.0706 | 0.0233 | 9.1452 | 0.0025 |
| ANCT_MRA | 0 | 1 | 1.5641 | 0.0326 | 2298.9236 | <.0001 |
| ANCT_MRA | 2 | 1 | -0.5895 | 0.0328 | 323.8377 | <.0001 |
| ANCT_MRA | 3 | 1 | -1.4677 | 0.0415 | 1250.6506 | <.0001 |
| Mt_ct_Mrh | 0 | 1 | 0.2300 | 0.0316 | 52.9007 | <.0001 |
| Mt_ct_Mrh | 1 | 1 | 0.1362 | 0.0311 | 19.1476 | <.0001 |
| Mt_ct_Mrh | 3 | 1 | -0.3343 | 0.0345 | 93.6524 | <.0001 |
| Nb_ct_Auto | 3 | 1 | -0.2204 | 0.0218 | 102.3259 | <.0001 |

| Analyse des estimations de la vraisemblance maximum | | | | | | |
|---|--------|----|------------|------------|---------------|------------|
| Paramètre | | DF | Estimation | Erreur std | Khi 2 de Wald | Pr > Khi 2 |
| NB_PRDT_DET_N_P | 3 | 1 | -0.1196 | 0.0233 | 26.3955 | <.0001 |
| IVP | 1 | 1 | 0.0709 | 0.0215 | 10.8990 | 0.0010 |
| NB_CONNEXION_BAM_6M_ | 2 et 3 | 1 | -0.1068 | 0.0199 | 28.6775 | <.0001 |

Le tableau suivant contient les odds-ratios des variables. Ils sont supérieurs à 1 pour les variables influant positivement sur la probabilité à prédire, inférieurs à 1 sinon, et leur intervalle de confiance ne doit pas contenir 1 ce qui est le cas ici. Cette dernière condition est équivalente à celle formulée plus haut sur le χ^2 de Wald. Un odds-ratio de 1.152 pour une variable comme la variable IVP signifie que le rapport des clients à fort potentiel de marge assurance aux clients sans potentiel de marge assurance est multiplié par 1.152 si ce dernier détient une prévoyance IVP.

| Estimations des rapports de cotes | | | |
|-----------------------------------|----------------|----------------------------------|-------|
| Effet | Point Estimate | 95% Limites de confiance de Wald | |
| ANCT 3 vs 0 à 2 | 1.152 | 1.051 | 1.262 |
| ANCT_MRA 0 vs 1 | 2.918 | 2.664 | 3.197 |
| ANCT_MRA 2 vs 1 | 0.339 | 0.309 | 0.372 |
| ANCT_MRA 3 vs 1 | 0.141 | 0.125 | 0.158 |
| Mt_ct_Mrh 0 vs 2 | 1.299 | 1.175 | 1.437 |
| Mt_ct_Mrh 1 vs 2 | 1.183 | 1.070 | 1.308 |
| Mt_ct_Mrh 3 vs 2 | 0.739 | 0.665 | 0.821 |
| Nb_ct_Auto 3 vs 1 | 0.644 | 0.591 | 0.701 |
| NB_PRDT_DET_N_P 3 vs 0 et 2 | 0.787 | 0.719 | 0.862 |
| IVP 1 vs 2 | 1.152 | 1.059 | 1.253 |
| NB_CONNEXION_BAM_6M_ 2 et 3 vs 1 | 0.808 | 0.747 | 0.873 |

Un dernier tableau fournit les résultats des tests de concordance. Il s'agit de compter les paires concordantes, une paire d'observation étant concordante lorsque l'une vérifie cible=0, l'autre cible=1, et que la probabilité estimée que cible=1 est plus grande pour l'observation telle que cible=1. On dérive plusieurs indicateurs de ces comptages, rassemblés dans ce tableau, dont nous ne regardons que le coin inférieur droit.

| Association des probabilités prédites et des réponses observées | | | |
|---|----------|-----------|-------|
| Percent Concordant | 78.8 | Somers' D | 0.584 |
| Percent Discordant | 20.4 | Gamma | 0.589 |
| Percent Tied | 0.7 | Tau-a | 0.260 |
| Pairs | 73816947 | C | 0.792 |

L'aire sous la courbe ROC sur l'échantillon d'apprentissage est fournie par l'indicateur « c » du tableau précédent. Elle vaut donc ici 0.792. Pour avoir l'aire sous la courbe ROC sur l'échantillon de validation, il nous faut lancer une macro.

Outre l'aire sous la courbe ROC, on s'intéresse parfois au D de Sommers, plus connu dans le cadre de l'évaluation de modèles sous le nom « d'indice de Gini ». Il est égal à deux fois (AUC-0.5), autrement dit à deux fois l'aire comprise entre la courbe ROC et la diagonale.

score de fidelite

The NPAR1WAY Procedure

| Wilcoxon Scores (Rank Sums) pour la variable proba3 Classée par variable qualite_SP_hs_CNAT0 | | | | | |
|---|-------|------------------|------------------|--------------------|-------------|
| qualite_SP_hs_CNAT0 | Nb | Somme des scores | Attendue sous H0 | Écart-type sous H0 | Score moyen |
| 0 | 12123 | 88839517.5 | 110398100 | 334703.462 | 7328.1793 |
| 1 | 6089 | 77008060.5 | 55449478.5 | 334703.462 | 12647.0784 |
| Les scores moyens ont été utilisés pour les liens. | | | | | |

| Wilcoxon Two-Sample Test | |
|--------------------------|---------------|
| Statistic | 77008060.5000 |
| Normal Approximation | |
| Z | 64.4110 |
| One-Sided Pr > Z | <.0001 |
| Two-Sided Pr > Z | <.0001 |
| t Approximation | |
| One-Sided Pr > Z | <.0001 |
| Two-Sided Pr > Z | <.0001 |

| Kruskal-Wallis Test | |
|---------------------|-----------|
| Chi-Square | 4148.7759 |
| DF | 1 |
| Pr > Chi-Square | <.0001 |

Aire sous la courbe ROC

| AUC | Gini |
|---------|---------|
| 0.79205 | 0.58411 |

La dernière sortie montre que l'AUC vaut 0.79205 sur l'échantillon de validation, contre 0.79188 sur l'échantillon d'apprentissage. L'écart entre l'AUC en apprentissage et en validation étant très faible, nous pouvons alors conclure au fait qu'il n'y a pas de signe de sur apprentissage et le modèle se généralise bien car pas trop complexe. Les dix variables explicatives ne sont donc pas nombreuses et favorisent sa généralisation à d'autres échantillons.

Pour nous conforter dans notre choix de ces variables, nous décidons de vérifier l'absence de colinéarité des variables du modèle prises deux à deux.

Analyse des corrélations entre les variables explicatives du dernier modèle retenu

| value | ve2 | | | | | | |
|-----------------------|------|----------|-------|-----------|-----------------------|------------|-----------------|
| ve1 | ANCT | ANCT_MRA | IVP | Mt_ct_Mrh | NB_CONNEXION_BAM_6M_P | Nb_ct_Auto | NB_PRDT_DET_N_P |
| ANCT | 1 | 0,26 | 0,28 | 0,16 | 0,16 | 0,08 | 0,08 |
| ANCT_MRA | 0,26 | 1 | 0,12 | 0,11 | 0,07 | 0,14 | 0,08 |
| IVP | 0,28 | 0,12 | 1 | 0,18 | 0,10 | -0,04 | 0,16 |
| Mt_ct_Mrh | 0,16 | 0,11 | 0,18 | 1 | 0,03 | 0,21 | 0,19 |
| NB_CONNEXION_BAM_6M_P | 0,16 | 0,07 | 0,10 | 0,03 | 1 | 0,15 | 0,15 |
| Nb_ct_Auto | 0,08 | 0,14 | -0,04 | 0,21 | 0,15 | 1 | 0,32 |
| NB_PRDT_DET_N_P | 0,08 | 0,08 | 0,16 | 0,19 | 0,15 | 0,32 | 1 |

Les coefficients de corrélation de toutes les paires de variables sont tous inférieurs à 0.32, nous pouvons donc conclure à l'absence de colinéarité entre les variables du modèle. Mais cela ne suffit pas, car il peut exister une relation linéaire entre trois variables. Nous devons donc vérifier l'absence de multi colinéarité entre les variables, ce qui peut se faire de deux façons.

Soit on calcule un indice appelé « tolérance » ou son inverse, le VIF (« variance inflation factor »). La tolérance d'une variable est la proportion de la variance de cette variable qui n'est pas expliquée par les autres (c'est-à-dire $1-R^2$, où R est la corrélation entre la variable testée et les autres variables explicatives) ; on considère souvent qu'elle doit être supérieure à 0.2 ou au moins 0.1 ($VIF \leq 10$).

Soit on examine la matrice des corrélations et on calcule ses indices de conditionnement (de Belsey, Kuh et Welsch), définis comme les racines carrées des ratios de la plus grande valeur propre sur chacune des valeurs propres.

$$\eta k^2 = \mu_{\max} / \mu_k$$

On considère après Belsey que la multi colinéarité est modérée si des indices η_k sont >10 , et forte si des indices η_k sont > 30 . Si cette situation survient, on regarde si l'on peut relier la valeur propre correspondante à une forte contribution ($>50\%$) de la composante principale (vecteur propre associé à la valeur propre) à la variance de deux ou plusieurs variables, c'est-à-dire si deux ou plusieurs colonnes dans le tableau ci-dessous contiennent des variables supérieures à 0.5. Dans notre cas, ce tableau montre que nous n'avons pas de problème de multi colinéarité, celle existant entre les variables 2,3, 7 et 8 restant modérée.

Dans l'ordre, le tableau 1 et le tableau 2 sont issus de la PROC REG de sas avec les options en sortie de TOL VIF COLLIN

| Résultats estimés des paramètres | | | | | | | |
|----------------------------------|----|---------------------|------------|------------------|---------|-----------|-----------------------|
| Variable | DF | Résultat estimé des | Erreur std | Valeur du test t | Pr > t | Tolérance | Inflation de variance |
| Intercept | 1 | 0.83595 | 0.01343 | 62.23 | <.0001 | . | 0 |
| ANCT | 1 | 0.00085924 | 0.00027350 | 3.14 | 0.0017 | 0.72815 | 1.37334 |
| ANCT_MRA | 1 | -0.03229 | 0.00046906 | -68.85 | <.0001 | 0.78679 | 1.27098 |
| Mt_ct_Mrh | 1 | -0.0002911 | 0.00002681 | -10.83 | <.0001 | 0.80243 | 1.24621 |
| Nb_ct_Auto | 1 | -0.04696 | 0.00459 | -10.24 | <.0001 | 0.86437 | 1.15691 |
| NB_PRDT_DET_N_P | 1 | -0.00207 | 0.00032584 | -6.37 | <.0001 | 0.76647 | 1.30468 |
| IVP | 1 | -0.01823 | 0.00625 | -2.92 | 0.0035 | 0.87933 | 1.13723 |
| NB_CONNEXION_BAM_6M_P | 1 | -0.0000583 | 0.00003838 | -1.52 | 0.1286 | 0.95180 | 1.05064 |

| Collinearity Diagnostics | | | | | | | | | | |
|--------------------------|---------------|--------------------|-------------------------|--------|----------|-----------|------------|----------------|--------|-----------------------|
| Nombre | Valeur propre | Index de condition | Proportion de variation | | | | | | | |
| | | | Intercept | ANCT | ANCT MRA | Mt ct Mrh | Nb ct Auto | NB PRDT DETN P | IVP | NB CONNEXION BAM 6M P |
| 1 | 6,3692 | 1,0000 | 0,0008 | 0,0031 | 0,0044 | 0,0032 | 0,0030 | 0,0028 | 0,0012 | 0,0034 |
| 2 | 0,8712 | 2,7038 | 0,0001 | 0,0040 | 0,0040 | 0,0005 | 0,0000 | 0,0001 | 0,0003 | 0,9019 |
| 3 | 0,2402 | 5,1495 | 0,0027 | 0,0781 | 0,5083 | 0,0616 | 0,0493 | 0,0459 | 0,0013 | 0,0572 |
| 4 | 0,1634 | 6,2426 | 0,0018 | 0,2391 | 0,2259 | 0,0793 | 0,2463 | 0,0681 | 0,0279 | 0,0235 |
| 5 | 0,1333 | 6,9116 | 0,0303 | 0,0000 | 0,0483 | 0,4015 | 0,1263 | 0,1585 | 0,0803 | 0,0001 |
| 6 | 0,1062 | 7,7443 | 0,0000 | 0,2954 | 0,1362 | 0,4147 | 0,0136 | 0,4591 | 0,0137 | 0,0017 |
| 7 | 0,0931 | 8,2713 | 0,0401 | 0,3779 | 0,0725 | 0,0319 | 0,5232 | 0,1416 | 0,0826 | 0,0101 |
| 8 | 0,0234 | 16,5073 | 0,9241 | 0,0023 | 0,0005 | 0,0074 | 0,0384 | 0,1239 | 0,7929 | 0,0021 |

Dans le tableau 1, nous avons la tolérance qui est très supérieure à 0.2, et les indices de conditionnement du tableau 2 sont tous inférieurs à 10 sauf le dernier mais nettement inférieur à 30. L'absence de multi colinéarité entre les prédicteurs est ici avérée et corrobore le calcul des coefficients de corrélation effectué plus haut.

ii. La table de classification d'un modèle

La table de classification d'un modèle est la matrice de confusion obtenue pour tout un ensemble de seuils possibles de la fonction de score. Le seuil est la valeur de la fonction de score au-delà de laquelle on considère que l'évènement est prédit, et en deçà de laquelle on considère que c'est un non-évènement qui est prédit. Sous SAS, il est possible de fixer un autre seuil que celui pris par défaut (0.50) avec l'option PPROB.

L'exemple présenté ci-dessous est celui du modèle d'apprentissage. L'évènement à prédire est le client à fort potentiel de marge. Fixer un seuil à 0.5 signifie que l'on considère comme client à fort potentiel de marge en assurance IARD tout client dont la probabilité dépasse 0.5.

Voici les significations des colonnes de la table de classification regroupées sous l'étiquette « Pourcentages ».

La colonne « Correct » contient le pourcentage de bonnes prédictions, ici égal à $(3066+10376)/18\ 214 = 73.8\%$.

La colonne « sensibilité » contient le pourcentage d'évènements prédits comme tels, ici égal au quotient de 3066(évènements prédits) sur 6 089(évènements observés), soit 50.4%.

La colonne « Spécificité » contient le pourcentage de non-évènements prédits comme tels, ici égal au quotient de 10 376 (non-évènements prédits) sur 12 123(non-évènements observés), soit 85.6%.

Symétriquement, la colonne « POS fausse » contient le pourcentage de faux positifs, c'est-à-dire la proportion d'évènements à tort (prédits à tort comme client à fort potentiel de marge en assurance IARD) parmi l'ensemble des évènements prédits. Il y a ici 1 699 prédits à tort comme clients à fort potentiel de marge en assurance IARD parmi les 4 908 clients prédits à fort potentiel de marge en assurance IARD. D'où un pourcentage de $1747/4813=36.3\%$.

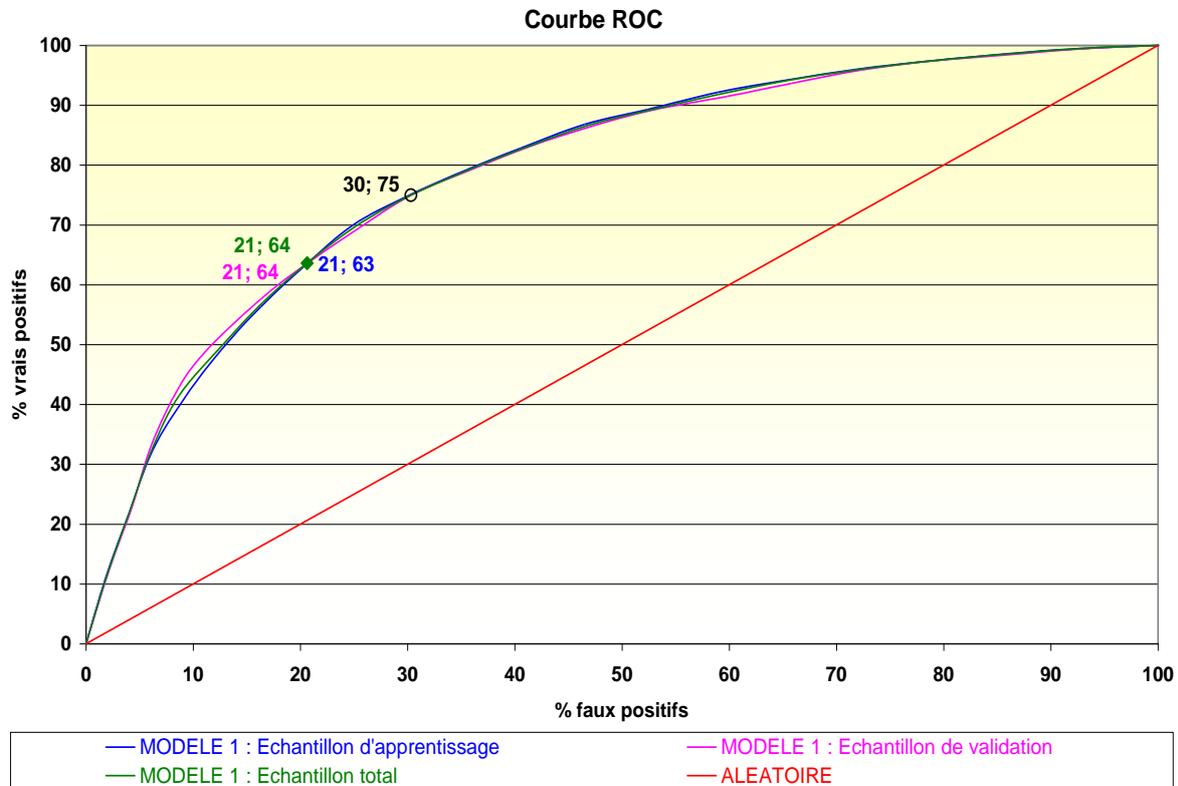
Enfin, la colonne « NEG fausse » contient le pourcentage de faux négatifs, c'est-à-dire la proportion de non-évènements prédits. Il y a ici 3023 clients à tort comme n'étant pas à fort potentiel de marge en assurance IARD parmi 13306 clients n'étant pas à fort potentiel de marge en assurance IARD. D'où un pourcentage de $3023/13399 = 22.56\%$.

Cette table de classification permet de fixer un seuil de score dans le but de minimiser le taux d'erreur global, ou la proportion de faux positifs ou de faux négatifs, selon l'objectif recherché. En l'occurrence ici, il est préférable de minimiser la proportion de faux positifs mais les deux autres pourcentages sont aussi importants pour l'intérêt du modèle, c'est pourquoi le seuil de 0.5 est maintenu.

| Table de classification | | | | | | | | | |
|-------------------------|-------------|---------------|-------------|---------------|--------------|-------------|-------------|-------------|-------------|
| Niveau de prob. | Correct | | Incorrect | | Pourcentages | | | | |
| | Événement | Non-événement | Événement | Non-événement | Correct | Sensibilité | Spécificité | POS fausse | NEG fausse |
| 0.020 | 6089 | 0 | 12123 | 0 | 33.4 | 100.0 | 0.0 | 66.6 | . |
| 0.040 | 6085 | 235 | 11888 | 4 | 34.7 | 99.9 | 1.9 | 66.1 | 1.7 |
| 0.060 | 6056 | 905 | 11218 | 33 | 38.2 | 99.5 | 7.5 | 64.9 | 3.5 |
| 0.080 | 5996 | 1822 | 10301 | 93 | 42.9 | 98.5 | 15.0 | 63.2 | 4.9 |
| 0.100 | 5899 | 2944 | 9179 | 190 | 48.6 | 96.9 | 24.3 | 60.9 | 6.1 |
| 0.120 | 5817 | 3669 | 8454 | 272 | 52.1 | 95.5 | 30.3 | 59.2 | 6.9 |
| 0.140 | 5623 | 4910 | 7213 | 466 | 57.8 | 92.3 | 40.5 | 56.2 | 8.7 |
| 0.160 | 5507 | 5468 | 6655 | 582 | 60.3 | 90.4 | 45.1 | 54.7 | 9.6 |
| 0.180 | 5403 | 5937 | 6186 | 686 | 62.3 | 88.7 | 49.0 | 53.4 | 10.4 |
| 0.200 | 5334 | 6243 | 5880 | 755 | 63.6 | 87.6 | 51.5 | 52.4 | 10.8 |
| 0.220 | 5246 | 6630 | 5493 | 843 | 65.2 | 86.2 | 54.7 | 51.2 | 11.3 |
| 0.240 | 5091 | 7009 | 5114 | 998 | 66.4 | 83.6 | 57.8 | 50.1 | 12.5 |
| 0.260 | 4918 | 7579 | 4544 | 1171 | 68.6 | 80.8 | 62.5 | 48.0 | 13.4 |
| 0.280 | 4777 | 7929 | 4194 | 1312 | 69.8 | 78.5 | 65.4 | 46.8 | 14.2 |
| 0.300 | 4542 | 8289 | 3834 | 1547 | 70.5 | 74.6 | 68.4 | 45.8 | 15.7 |
| 0.320 | 4478 | 8648 | 3475 | 1611 | 72.1 | 73.5 | 71.3 | 43.7 | 15.7 |
| 0.340 | 4395 | 8834 | 3289 | 1694 | 72.6 | 72.2 | 72.9 | 42.8 | 16.1 |
| 0.360 | 4330 | 8968 | 3155 | 1759 | 73.0 | 71.1 | 74.0 | 42.2 | 16.4 |
| 0.380 | 4262 | 9104 | 3019 | 1827 | 73.4 | 70.0 | 75.1 | 41.5 | 16.7 |
| 0.400 | 4044 | 9363 | 2760 | 2045 | 73.6 | 66.4 | 77.2 | 40.6 | 17.9 |
| 0.420 | 3866 | 9626 | 2497 | 2223 | 74.1 | 63.5 | 79.4 | 39.2 | 18.8 |
| 0.440 | 3736 | 9745 | 2378 | 2353 | 74.0 | 61.4 | 80.4 | 38.9 | 19.4 |
| 0.460 | 3532 | 10027 | 2096 | 2557 | 74.5 | 58.0 | 82.7 | 37.2 | 20.3 |
| 0.480 | 3396 | 10167 | 1956 | 2693 | 74.5 | 55.8 | 83.9 | 36.5 | 20.9 |
| 0.500 | 3066 | 10376 | 1747 | 3023 | 73.8 | 50.4 | 85.6 | 36.3 | 22.6 |
| 0.520 | 2976 | 10606 | 1517 | 3113 | 74.6 | 48.9 | 87.5 | 33.8 | 22.7 |
| 0.540 | 2591 | 10936 | 1187 | 3498 | 74.3 | 42.6 | 90.2 | 31.4 | 24.2 |
| 0.560 | 2423 | 11105 | 1018 | 3666 | 74.3 | 39.8 | 91.6 | 29.6 | 24.8 |
| 0.580 | 2340 | 11150 | 973 | 3749 | 74.1 | 38.4 | 92.0 | 29.4 | 25.2 |
| 0.600 | 2254 | 11202 | 921 | 3835 | 73.9 | 37.0 | 92.4 | 29.0 | 25.5 |
| 0.620 | 2163 | 11253 | 870 | 3926 | 73.7 | 35.5 | 92.8 | 28.7 | 25.9 |
| 0.640 | 2072 | 11305 | 818 | 4017 | 73.5 | 34.0 | 93.3 | 28.3 | 26.2 |
| 0.660 | 1882 | 11402 | 721 | 4207 | 72.9 | 30.9 | 94.1 | 27.7 | 27.0 |
| 0.680 | 1713 | 11489 | 634 | 4376 | 72.5 | 28.1 | 94.8 | 27.0 | 27.6 |
| 0.700 | 1603 | 11537 | 586 | 4486 | 72.2 | 26.3 | 95.2 | 26.8 | 28.0 |
| 0.720 | 1334 | 11633 | 490 | 4755 | 71.2 | 21.9 | 96.0 | 26.9 | 29.0 |
| 0.740 | 1005 | 11731 | 392 | 5084 | 69.9 | 16.5 | 96.8 | 28.1 | 30.2 |
| 0.760 | 685 | 11896 | 227 | 5404 | 69.1 | 11.2 | 98.1 | 24.9 | 31.2 |
| 0.780 | 244 | 12053 | 70 | 5845 | 67.5 | 4.0 | 99.4 | 22.3 | 32.7 |
| 0.800 | 3 | 12123 | 0 | 6086 | 66.6 | 0.0 | 100.0 | 0.0 | 33.4 |
| 0.820 | 0 | 12123 | 0 | 6089 | 66.6 | 0.0 | 100.0 | . | 33.4 |

iii. Présentation des Courbes ROC et des Courbes LIFT sur l'échantillon d'apprentissage, de validation et total

Courbe ROC



On visualise le pouvoir discriminant du score à l'aide de la courbe ROC, acronyme pour « Receiver Operating Characteristic », terme issu du traitement du signal. Elle représente sur l'axe Y la proportion d'évènements (clients qui est à fort potentiel de marge assurance) détectés comme tels car leur score est supérieur à s , en fonction (sur l'axe X) de la proportion de faux évènements, c'est-à-dire de non-évènements détectés comme évènements à cause de leur score supérieur à s (clients qui n'est pas à fort potentiel de marge assurance), lorsque l'on fait varier le seuil s de séparation du score. Précisément on définit deux fonctions de s :

- La sensibilité $\alpha(s)$ = probabilité de bien détecter un évènement au seuil s = $\text{Prob}(\text{score}(x) \geq s/x = \text{évènement})$,

- La sensibilité $\beta(s)$ = probabilité de bien détecter un non-évènement au seuil s = $\text{Prob}(\text{score}(x) < s/x = \text{non-évènement})$,

Et l'on peut dire que la proportion de faux évènements parmi les non-évènements est $1 - \beta(s) = \text{Proba}(\text{score}(x) \geq s/x = \text{non-évènement})$.

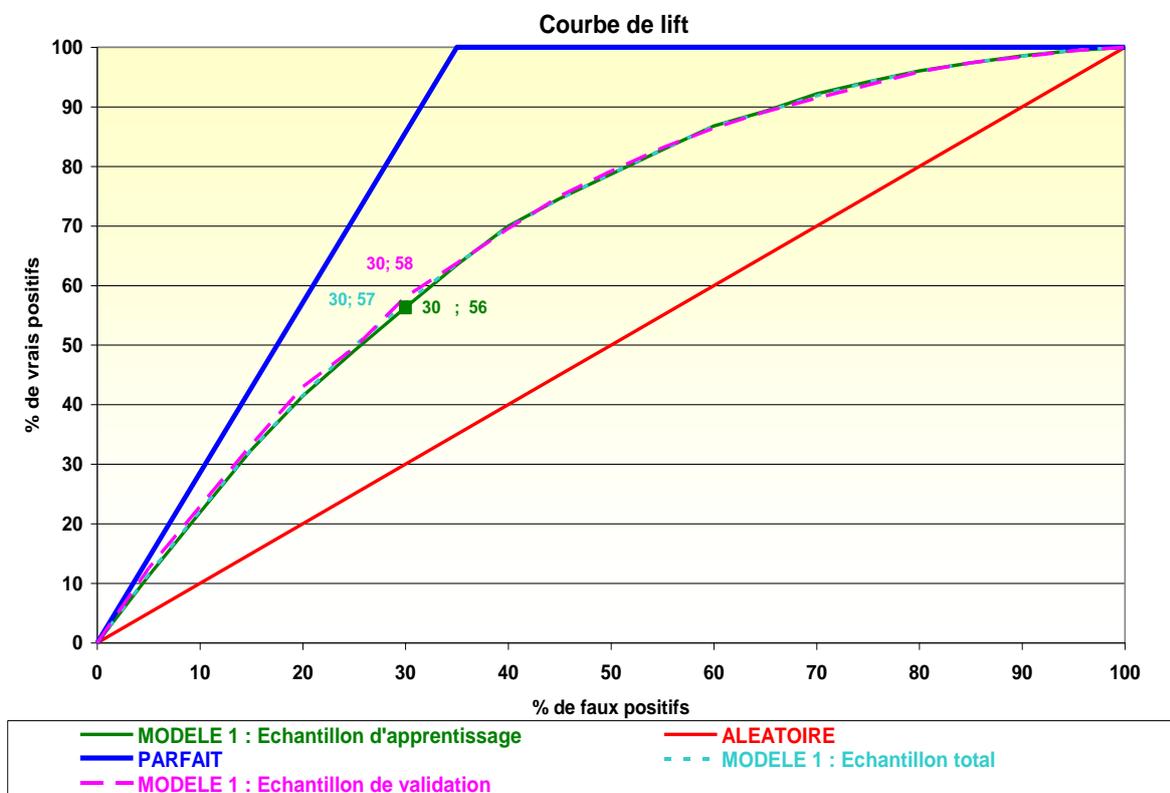
La courbe ROC représente donc $\alpha(s)$ en fonction $1 - \beta(s)$, pour des valeurs de s allant au maximum (on considère tous les individus comme non-évènements, d'où $\alpha(s) = 1 - \beta(s) = 0$) au minimum (on considère tous les individus comme évènements, d'où $\alpha(s) = 1 - \beta(s) = 1$).

Si cette courbe coïncide avec la diagonale, c'est que le modèle n'est pas plus performant qu'un modèle aléatoire. Plus la courbe ROC s'approche du coin supérieur gauche du carré de la figure ci-dessus, meilleur est le modèle, car il permet de capturer le plus possible de vrais évènements avec le moins possible de faux évènements.

La courbe ROC de la régression logistique de la figure ci-dessous passe par le point $x=0.3$ et $y=0.75$. Ce point correspond à un seuil s qui est tel que : si l'on considère « événements » tous les individus dont le score est $> s$, on a détecté 30% de faux événements (30% de non-événements ont un score $> s$) et 75% de vrais événements (90% d'événements ont un score $> s$). Attention : 0.3 ne correspond pas à 30% de la population totale et il ne faut pas confondre la courbe ROC avec la courbe LIFT définie plus bas !

Globalement, on peut comparer les performances de deux modèles en comparant les aires A sous leurs courbes ROC respectives. Ce que l'on peut dire c'est qu'un modèle est d'autant plus performant, en termes de capacité à séparer les vrais événements des faux événements, que son aire sous la courbe ROC est plus proche de 1. Un modèle est d'autant moins performant que son aire est plus proche de 0, et un modèle dont l'aire vaut 0.5 ne classe pas mieux que le hasard.

Courbe LIFT



Une variante de la courbe ROC est la courbe LIFT, depuis longtemps utilisée en marketing. Elle est aussi utilisée en économétrie sous le nom de courbe de Lorentz ou de puissance. Elle représente la proportion $\alpha(s)$ des événements détectés, c'est-à-dire la sensibilité, en fonction de la proportion des individus sélectionnés, c'est-à-dire ayant un score $> s$. On voit que la courbe LIFT a une même ordonnée $\alpha(s)$ que la courbe ROC mais une abscisse différente. Dès que $\alpha(s) > (1 - \beta(s))$, cette abscisse sera plus grande que l'abscisse $1 - \beta(s)$ de la courbe ROC, puisqu'il s'agit de la proportion de tous les individus « non-événement » ayant un score $> s$. Puisque, pour une même ordonnée, la courbe LIFT a généralement une abscisse plus grande que la courbe ROC, la courbe LIFT est sous la courbe ROC.

Cette abscisse, la proportion d'individus « événement » de score $> s$, se décompose en la somme de deux termes :

- ✚ La proportion d'individus « évènement » de score $> s$
 - $\text{Prob}(x=\text{évènement}).\text{Prob}(\text{score}(x)>s / x=\text{évènement})$
 - $p.\alpha(s)$;
- ✚ La proportion d'individus « non-évènement » de score $> s$
 - $\text{Prob}(x=\text{non-évènement}).\text{Prob}(\text{score}(x)>s / x=\text{non-évènement})$
 - $(1-p).(1-\beta(s))$.

Les courbes LIFT des 3 modèles de scores testés sont très proches et 30% des clients ayant le meilleur score représentent en moyenne 57% des clients à fort potentiel de marge en assurance Iard.

iv. Calcul d'une grille de score

Lorsqu'un modèle de score a été élaboré par régression logistique sur des variables qualitatives ou discrétisées, avec par conséquent un coefficient par modalité de variable (éventuellement nul pour la modalité de référence), on peut présenter la formule de calcul du score sous une forme très simple, si l'on ne se soucie pas d'obtenir exactement la probabilité a posteriori fournie par la régression logistique. Pour cela, il faut :

- ✚ Substituer le logit(combinaison linéaire des indicatrices des modalités) à la probabilité $\text{EXP}(\text{logit})/\{1+\text{exp}(\text{logit})\}$ comme valeur du score,
- ✚ Et normaliser le logit de sorte qu'il soit compris entre 0 et 100 (ou une autre valeur que 100, mille par exemple).

Dans cette normalisation du logit, les coefficients de la régression logistique sont remplacés par de nouveaux coefficients, appelés « nombre de points », associés chacun à une modalité. Par exemple, au lieu d'associer 1.5909 à la modalité « entre 0 et 4 ans » de « ANCT_MRA », (i.e. l'ancienneté de la relation du client au niveau de l'Assurance Iard), on lui associe 73 points. Le nombre de points associé à chaque modalité est déterminé de sorte que tout individu ait un nombre total de points compris entre 0 et 100, ces deux bornes étant atteignables, au moins en principe.

Cette façon de calculer un score est courante en crédit scoring, où l'on additionne les points liés à chaque modalité, pour avoir un nombre total de points qui représente le score de l'individu. Telle tranche d'ancienneté d'assurance ajoute tant de points, telle tranche de montant d'assurance Mrh ajoute tant de points, etc. Cet ensemble de points est appelé « grille de score ». Il est très explicite et montre parfaitement l'impact de chaque modalité.

Ce nombre de points est linéairement parfaitement corrélé (corrélation de Pearson = 1) au logit, mais non au « vrai » score logistique, c'est-à-dire à la probabilité $\text{EXP}(\text{logit})/\{1+\text{EXP}(\text{logit})\}$. En revanche, il est parfaitement corrélé au score logistique en termes de rangs (corrélation de spearman = 1), et son pouvoir discriminant est exactement le même, puisque les rangs, et donc le classement des individus, sont conservés par la fonction croissante $\text{EXP}(x)/\{1+\text{EXP}(x)\}$. En conséquence, l'aire sous la courbe ROC de la grille de score est égale à celle du score logistique.

Ce calcul du score représente la somme de points. On note $c(j,i)$ le coefficient du modèle associé à la modalité i de la variable j .

Pour chaque variable j , on cherche le coefficient $c(j,i)$ le plus petit, noté $\text{min}(j)$, le coefficient $c(j,k)$, le plus grand, et l'on calcule $\text{Deltamax}(j)$ la différence des deux : on obtient le plus grand écart entre deux coefficients d'une même variable.

Puis on calcule le Poids total de la somme sur j de tous les Deltamax .

Enfin, à chaque modalité i de la variable j est associé un nombre de points :

$$N(j,i) = 100 \times (c(j,i) - \text{min}(j)) / \text{Poids_total}.$$

En résumé, on a :

$$N(j,i) = 100 \times \frac{c(j,i) - \min_k (c(j,k))}{\sum_l [\max_m c(l,m) - \min_m c(l,m)]}$$

Présentation des résultats de calcul du nombre de points N(j,i) de la variable j associé à chaque modalité i, en l'arrondissant à l'entier le plus proche pour plus de simplicité et de lisibilité.

| Variable | ClassVal0 | Estimate | nbpoints |
|-----------------------|-----------|----------|----------|
| Intercept | | -0.9276 | 0 |
| ANCT | 1 | 0 | 0 |
| ANCT | 3 | 0.0672 | 2 |
| ANCT_MRA | 3 | -1.4525 | 0 |
| ANCT_MRA | 2 | -0.5963 | 21 |
| ANCT_MRA | 1 | 0 | 35 |
| ANCT_MRA | 0 | 1.5909 | 73 |
| IVP | 1 | 0 | 0 |
| IVP | 1 | 0.0763 | 2 |
| Mt_ct_Mrh | 3 | -0.3275 | 0 |
| Mt_ct_Mrh | 2 | 0 | 8 |
| Mt_ct_Mrh | 1 | 0.1440 | 11 |
| Mt_ct_Mrh | 0 | 0.2135 | 13 |
| NB_CONNEXION_BAM_6M_P | 2 et 3 | -0.1106 | 0 |
| NB_CONNEXION_BAM_6M_P | 0 à 1 | 0 | 3 |
| NB_PRDT_DET_N_P | 3 | -0.0885 | 0 |
| NB_PRDT_DET_N_P | 1 | 0 | 2 |
| Nb_ct_Auto | 3 | -0.2269 | 0 |
| Nb_ct_Auto | 1 | 0 | 5 |

On voit qu'à chaque modalité est associé un nombre de points ≥ 0 et que la somme des points que peut avoir un client scoré est au maximum à 100, ce maximum correspondant à la plus grande probabilité d'être un client à potentiel de marge en assurance Iard. Ce sera le cas d'un client répondant à l'un ou à plusieurs des critères suivant :

- + Ancieneté de l'entrée en relation du client au niveau Bancaire ≥ 30 ans (2 points)
- + Ancieneté de la relation du client au niveau de l'Assurance Iard entre 8 et 14 ans (21 points)
- + Ancieneté de la relation du client au niveau de l'Assurance Iard entre 5 et 7 ans (35 points)
- + Détention d'un contrat IVP (Prévoyance Décès Invalidité) oui (2 points)
- + Montant de la cotisation Hors Taxes Multirisque habitation entre 201 € et 268 € (8 points)
- + Montant de la cotisation Hors Taxes Multirisque habitation entre 151 € et 200 € (11 points)
- + Montant de la cotisation Hors Taxes Multirisque habitation entre 35 € et 150 € (13 points)
- + Nombre de connexions à la BAM (Banque à Accès Multiple) sur 6 mois glissants 0 (3 points)
- + Nombre de produits détenus entre 0 et 24 (2 points)
- + Nombre de contrats Auto 1 (5 points).

Dans les scores de comportement ou d'appétence du client on voit que la détention de certains produits a plus de poids que le profil sociodémographique. Ceci est une règle assez générale.

On applique ensuite très simplement la grille de score au fichier contenant déjà le score directement issu de la procédure logistique. On peut d'ailleurs l'appliquer à n'importe quel fichier, et l'implémenter dans n'importe quel langage informatique tant la formule est simple.

**Nombre de points
octroyés**

**Ancienneté de l'entrée en relation du
client au niveau Bancaire**

| | |
|-------------------|---|
| entre 0 et 29 ans | 0 |
| ≥ 30 ans | 2 |

**Nombre de connexions à la BAM
(Banque à Accès Multiple) sur 6 mois
glissants**

| | |
|-----|---|
| ≥ 1 | 0 |
| 0 | 3 |

Nombre de produits détenus

| | |
|---------------|---|
| ≥ 25 | 0 |
| entre 0 et 24 | 2 |

**Ancienneté de la relation du client au
niveau de l'Assurance Iard**

| | |
|-------------------|----|
| ≥ 15 ans | 0 |
| entre 8 et 14 ans | 21 |
| entre 5 et 7 ans | 35 |
| entre 0 et 4 ans | 73 |

Nombre de contrats Auto

| | |
|-----|---|
| ≥ 2 | 0 |
| 1 | 5 |

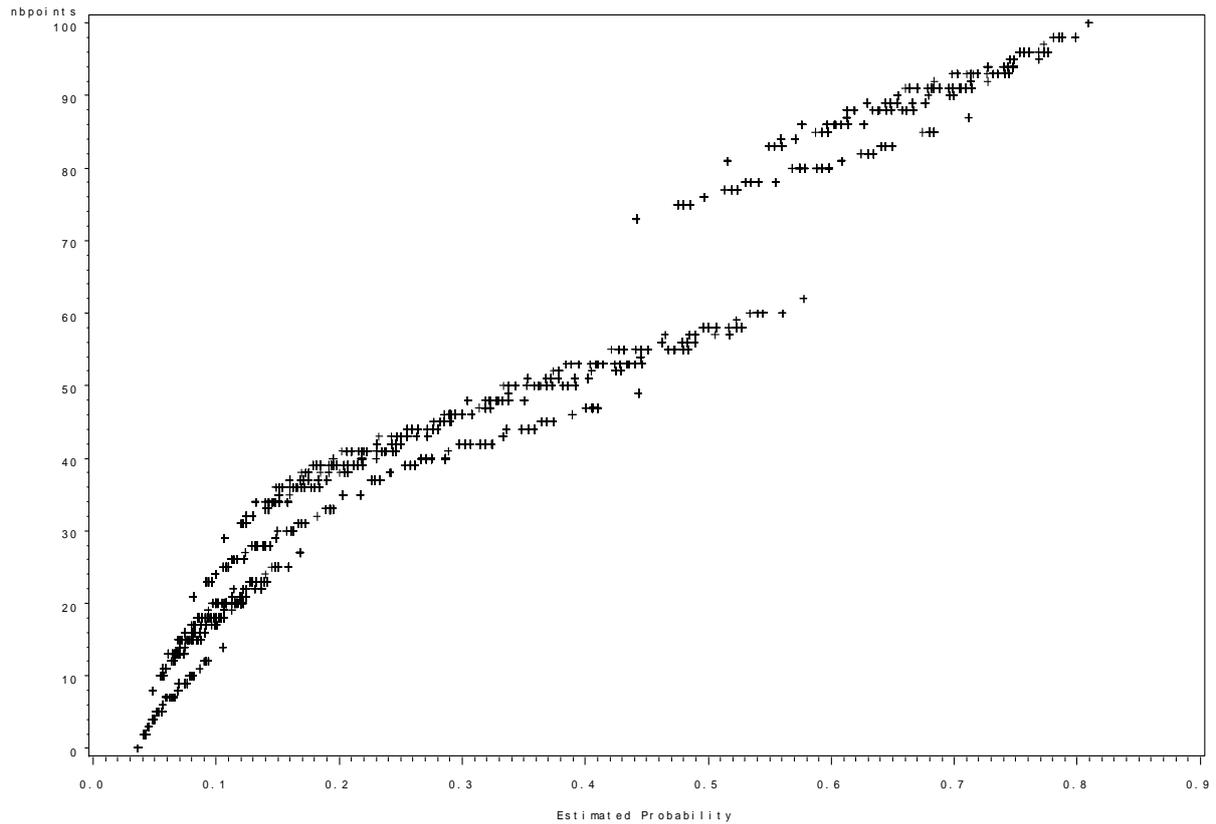
**Montant de la cotisation Hors Taxes
Multirisque habitation**

| | |
|----------------------|----|
| ≥ 269 € | 0 |
| entre 201 € et 268 € | 8 |
| entre 151 € et 200 € | 11 |
| entre 35 € et 150 € | 13 |

**Détention d'un contrat IVP (Prévoyance
Décès Invalidité)**

| | |
|-----|---|
| oui | 2 |
| non | 0 |

Nous utilisons la procédure GPLLOT de sas afin de vérifier le lien entre le score logistique et le nombre de points issu de la grille.



La forme de la courbe montre une corrélation non linéaire entre le score logit et le nombre de points, corrélation toutefois quasiment parfaite au niveau des rangs, c'est-à-dire de l'indicateur de spearman. C'est ce que confirme le calcul de corrélation.

| Coefficients de corrélation de Pearson Prob > r under H0: Rho=0 Nombre d'observations | | | |
|---|---------|-------------|----------|
| | proba3 | score_logit | nbpoints |
| proba3 | 1.00000 | 0.96775 | 0.97764 |
| Estimated Probability | | <.0001 | <.0001 |
| | 27412 | 27412 | 27412 |
| score_logit | 0.96775 | 1.00000 | 0.97763 |
| | <.0001 | | <.0001 |
| | 27412 | 27437 | 27437 |
| nbpoints | 0.97764 | 0.97763 | 1.00000 |
| | <.0001 | <.0001 | |
| | 27412 | 27437 | 27437 |

| Coefficients de corrélation de Spearman Prob > r under H0: Rho=0 Nombre d'observations | | | |
|--|---------|-------------|----------|
| | proba3 | score_logit | Nbpoints |
| proba3 | 1.00000 | 0.99448 | 0.99363 |
| Estimated Probability | | <.0001 | <.0001 |
| | 27412 | 27412 | 27412 |
| score_logit | 0.99448 | 1.00000 | 0.99811 |
| | <.0001 | | <.0001 |
| | 27412 | 27437 | 27437 |
| nbpoints | 0.99363 | 0.99811 | 1.00000 |
| | <.0001 | <.0001 | |
| | 27412 | 27437 | 27437 |

Si l'on calcule l'aire sous la courbe ROC de la grille de score, sur l'échantillon de validation, on obtient comme prévu un résultat (quasiment) identique à l'aire du score logistique : 0.78953 contre 0.78951, l'écart de deux dix millièmes étant imputable au fait que le nombre de points est toujours calculé, tandis que le score logistique n'est pas calculé pour 7 observations.

| Wilcoxon Scores (Rank Sums) pour la variable score_logit Classée par variable qualite_SP_hs_CNAT0 | | | | | |
|--|------|------------------|------------------|--------------------|-------------|
| qualite_SP_hs_CNAT0 | Nb | Somme des scores | Attendue sous H0 | Écart-type sous H0 | Score moyen |
| 0 | 6169 | 22977124.5 | 28405160.5 | 119941.770 | 3724.61088 |
| 1 | 3039 | 19421111.5 | 13993075.5 | 119941.770 | 6390.62570 |
| Les scores moyens ont été utilisés pour les liens. | | | | | |

Aire sous la courbe ROC

| AUC | Gini |
|---------|---------|
| 0.78953 | 0.57906 |

Enfin on découpe les deux scores en déciles à l'aide de la procédure rank de sas, puis l'on croise ces déciles, qui montrent un classement quasi identique des observations puisque la diagonale est surreprésentée.

La procédure FREQ

| Fréquence | Table de nbpoints par score_logit | | | | | | | | | | |
|--------------|--|--|------|------|------|------|------|------|------|------|-------|
| | nbpoints(Valeurs de nbpoints remplacées par des rangs) | score_logit(Valeurs de score_logit remplacées par des rangs) | | | | | | | | | Total |
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| 0 | 2353 | 281 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 2637 |
| 1 | 338 | 2371 | 56 | 0 | 0 | 0 | 8 | 3 | 0 | 0 | 2776 |
| 2 | 0 | 157 | 2653 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 2818 |
| 3 | 0 | 0 | 16 | 2575 | 33 | 0 | 0 | 0 | 0 | 0 | 2624 |
| 4 | 0 | 0 | 0 | 154 | 2596 | 56 | 0 | 0 | 0 | 0 | 2806 |
| 5 | 0 | 0 | 0 | 0 | 184 | 2441 | 353 | 0 | 0 | 0 | 2978 |
| 6 | 0 | 0 | 0 | 0 | 0 | 182 | 1976 | 0 | 0 | 0 | 2158 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 469 | 2665 | 11 | 0 | 3145 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 2627 | 5 | 2642 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 114 | 2739 | 2853 |
| Total | 2691 | 2809 | 2725 | 2729 | 2813 | 2680 | 2808 | 2686 | 2752 | 2744 | 27437 |

Nous allons maintenant procéder à un regroupement en tranches des nombres de points fournis par la grille de score, c'est-à-dire le découpage en notes discrètes, plus faciles pour la bancassurance qui souhaite tirer parti de son score en quelques valeurs qui nous indiqueront si le client a ou non un fort potentiel de marge en assurance Iard.

Le premier moyen qui vient à l'esprit consiste à utiliser le découpage de la grille en déciles effectué auparavant à l'aide de la procédure rank. On obtient ainsi simplement dix notes, de 0 à 9 ce dernière correspondante au plus fort taux de bon client assurance en termes de marge. On croise ensuite ces notes avec la variable « qualite_SP_hs_CNAT0 » pour avoir le taux de qualité du client associé à chaque note. On s'attend bien sûr à ce que ce taux soit strictement croissant, ce qui est le cas.

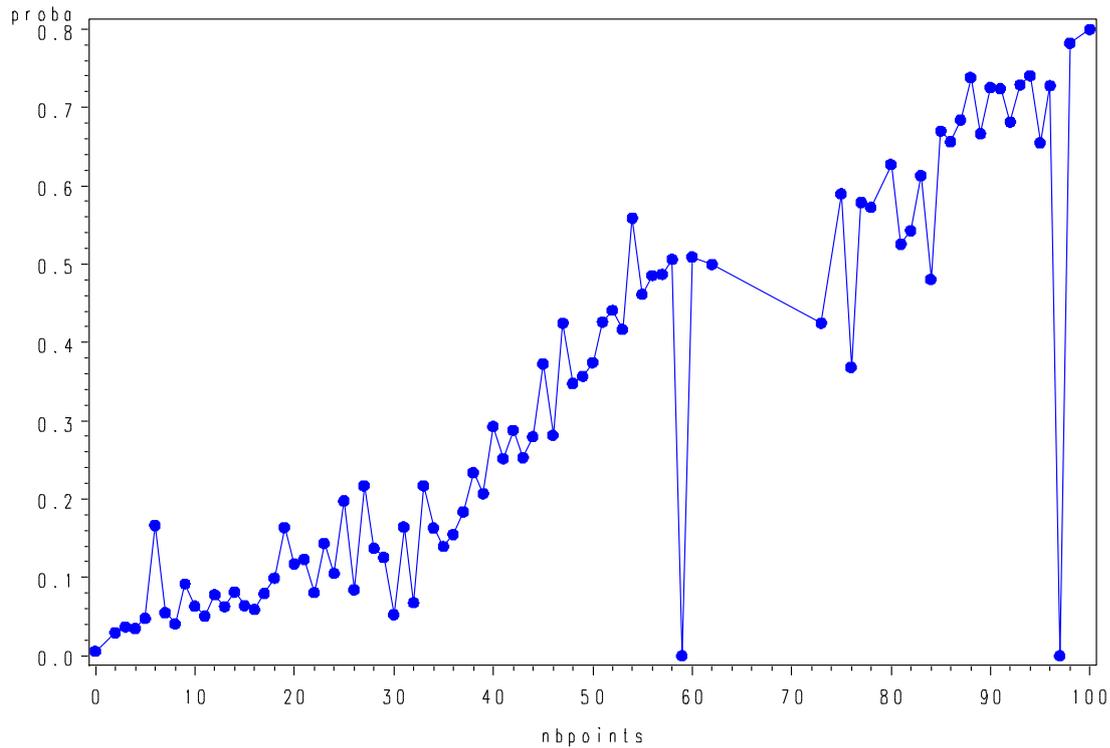
| Fréquence Pourcentage Pourct. en ligne Pourct. en col. | Table de nbpoints par qualite_SP_hs_CNAT0 | | | |
|---|---|---------------------------------|--|-----------------|
| | nbpoints | qualite_SP_hs_CNAT0 | | Total |
| | | 0 | 1 | |
| | 0 à 10 points | 2044 7.45 95.20 11.17 | 103 0.38 4.80 1.13 | 2147 7.83 |
| | 11 à 20 points | 2984 10.88 91.37 16.30 | 282 1.03 8.63 3.09 | 3266 11.90 |
| | 21 à 30 points | 2464 8.98 86.27 13.46 | 392 1.43 13.73 4.29 | 2856 10.41 |
| | 31 à 40 points | 3115 11.35 80.85 17.02 | 738 2.69 19.15 8.08 | 3853 14.04 |
| | 41 à 50 points | 3695 13.47 68.83 20.19 | 1673 6.10 31.17 18.31 | 5368 19.56 |
| | 51 à 60 points | 2271 8.28 53.12 12.41 | 2004 7.30 46.88 21.94 | 4275 15.58 |
| | >60 points | 1729 6.30 30.48 9.45 | 3943 14.37 69.52 43.16 | 5672 20.67 |
| | Total | 18302 66.71 | 9135 33.29 | 27437 100.00 |

Nous faisons un petit rappel sur la variable à expliquer « qualite_SP_hs_CNAT0 » qui correspond au ratio Sinistres / Primes hors Catastrophes naturelles égale à 0. Nous décidons de croiser ces notes avec la variable « qualite_SP_hs_CNAT » qui correspond cette fois ci au ratio Sinistres / Primes hors Catastrophes naturelles inférieur ou égale à 70%, ce qui permet de se rasséréner sur un mauvais classement de la grille de score.

| Fréquence Pourcentage Pourct. en ligne Pourct. en col. | Table de nbpoints par qualite_SP_hs_CNAT | | | |
|---|--|---------------------------------|-----------------|-------|
| | nbpoints | qualite_SP_hs_CNAT | | Total |
| | | 0 | 1 | |
| 0 à 10 points | 337 1.23 15.70 7.66 | 1810 6.60 84.30 7.86 | 2147 7.83 | |
| 11 à 20 points | 468 1.71 14.33 10.64 | 2798 10.20 85.67 12.14 | 3266 11.90 | |
| 21 à 30 points | 432 1.57 15.13 9.82 | 2424 8.83 84.87 10.52 | 2856 10.41 | |
| 31 à 40 points | 667 2.43 17.31 15.17 | 3186 11.61 82.69 13.83 | 3853 14.04 | |
| 41 à 50 points | 929 3.39 17.31 21.12 | 4439 16.18 82.69 19.27 | 5368 19.56 | |
| 51 à 60 points | 761 2.77 17.80 17.30 | 3514 12.81 82.20 15.25 | 4275 15.58 | |
| >60 points | 804 2.93 14.17 18.28 | 4868 17.74 85.83 21.13 | 5672 20.67 | |
| Total | 4398 16.03 | 23039 83.97 | 27437 100.00 | |

La courbe suivante permet de représenter le taux de la qualité du client en assurance Iard en fonction du nombre de points. Elle confirme que ce taux est bien globalement croissant en fonction du nombre de points, mais qu'il existe malgré tout des fluctuations de ce taux, bien visibles sur la courbe. Ces fluctuations sont en partie gommées par le découpage en déciles. Mais la courbe montre autre chose : le taux de qualité du client en assurance Iard n'est pas une fonction linéaire du nombre de points. Ce taux commence par évoluer longtemps dans une bande basse, jusqu'à environ 40 points avec un taux de qualité du client en Assurance Iard entre 0% et 20%. Puis on passe dans une deuxième tranche 41 à 60 points avec un taux de qualité du client en Assurance Iard entre 20% et 50%. On termine par une troisième tranche entre 61 et 84 points avec un taux de qualité du client en Assurance Iard entre 51% et 64%. Enfin une quatrième tranche entre 85 et 100 points avec un taux de qualité du client en Assurance Iard entre 65% et 80%.

score de fidelite



C. Conclusion de la deuxième partie

L'intérêt de la création de ce score est de pouvoir qualifier nos clients bancaires mono détenteurs ou non assurés en Assurance Iard. Mais il faudra toutefois pour scorer ces derniers disposer de données non présentes dans le système d'information, i.e. l'ancienneté de la relation du client en Assurance Iard, le nombre de contrat Auto et le montant de la cotisation Mrh Hors Taxe.

Nous sommes donc en mesure d'avoir une approche globale, quel que soit le client CRCA NDF, à travers la qualité de ce dernier au niveau de la banque et de l'assurance en termes de risque et de marge.

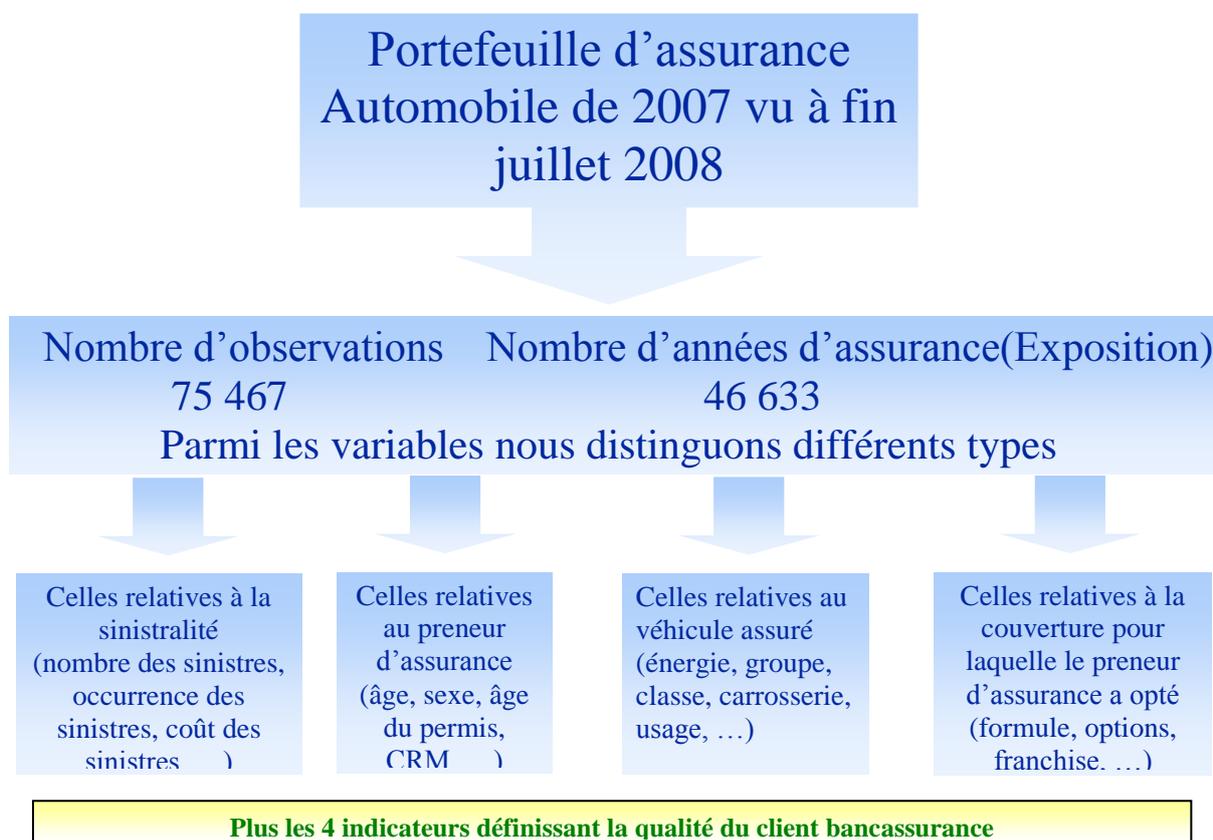
Chapitre 4 Modélisation de la fréquence et du coût moyen

1. Objectif

Après une description succincte de la matrice de travail et des difficultés rencontrées lors de sa création, nous vous proposons une présentation graphique des variables décrivant la sinistralité (fréquence et coût moyen des sinistres) suivant les différents critères de tarification tels que les caractéristiques du preneur d'assurance, du véhicule assuré et de la police d'assurance et l'exposition au risque. Nous analyserons ensuite la distribution de la fréquence et du coût moyen des sinistres par garantie et nous déterminerons la loi la plus en adéquation avec cette dernière. Nous pourrions alors enchaîner sur la modélisation et présenter des résultats. Les différents indicateurs permettant d'apprécier la qualité d'un client bancassurance, seront repris dans la modélisation au même titre que les critères de tarification classiques en assurance automobile. Ainsi nous pourrions démontrer statistiquement et actuariellement qu'ils ressortent clairement comme des variables significatives dans les deux modèles utiles à la construction de la tarification en assurance automobile. Il sera alors aisé de déterminer une prime pure chargée modélisée que nous comparerons à la prime pure chargée empirique suivant les critères de tarification permettant d'apprécier la qualité du client bancassurance et être force de proposition quant à la réduction tarifaire automobile que nous pourrions proposer, dans le cadre de la souscription du « contrat Fidelio ».

2. Monitoring du portefeuille et intégrité des données

a. Monitoring du portefeuille



Avant de débiter la modélisation du nombre de sinistres et du coût moyen d'un sinistre par garantie, il est essentiel de bien connaître le portefeuille sur lequel on travaille. Il faut dès lors prendre le temps de décrire en détail les différentes variables tarifaires et d'examiner la composition du portefeuille à analyser.

Nous travaillons sur un fichier de polices d'assurance. Un tel fichier compte autant de lignes que de polices. Il résume l'information disponible en début de période à propos de chacun des contrats et décrit la sinistralité relative à ceux-ci.

En marge du fichier de polices d'assurance, l'assureur dispose également d'un fichier sinistres où sont renseignées toutes les caractéristiques des sinistres produits par les polices en portefeuille (circonstances dans lesquelles ces sinistres ont eu lieu, personnes impliquées, présence de dommages corporels, ...).

Ces deux fichiers sont mis en relation grâce au numéro de police et à la période.

La phase de constitution de la base de données est cruciale pour le tarif qui découlera de l'analyse :

Comment arriver à un tarif correct en utilisant des données erronées, incomplètes ou obsolètes ?

La phase d'extraction des données et leur nettoyage représentent une part considérable du temps consacré à l'étude.

Petit aparté :

La plupart des compagnies sont à présent conscientes de la nécessité de disposer de données aussi nombreuses que possible et de bonne qualité. La constitution et la maintenance des bases de données figurent d'ailleurs parmi les préoccupations les plus importantes des grands groupes financiers.

La MRACA dispose d'une base de données de très bonne qualité après toilettage bien sûr. A noter que le CRM (Customer Relationship Management) et, de fait, le datawarehouse de la caisse régionale Nord De France est retenu, dans le cadre de notre projet groupe, qui consistera principalement à basculer vers un seul SI (Système d'Information) et à la mise en place de la GRC (Gestion de la Relation Client), au niveau national.

b. Intégrité des données

Nous allons tenter d'énumérer les différents problèmes rencontrés lors de la création de la matrice de travail.

➤ Nettoyage de la base pour les polices (contrat/avenant) ayant :

- une année de début de l'avenant \geq à l'année de fin effet de l'avenant,
- une année de fin effet de l'avenant non renseignée alors que cette police est résiliée.

➤ transformation de la base en données de panel :

- calcul de la période d'après les dates de début et de fin effet de l'avenant,
- calcul de la durée suivant la période d'après les dates de début et de fin effet de l'avenant.

➤ Concernant les caractéristiques de l'assuré :

- calcul de l'historique du CRM (Coefficient Réduction Majoration) suivant la période d'existence de la police car nous avons, dans la base, la précédente et la dernière valeur du CRM,
- calcul de l'âge et de l'ancienneté du permis suivant la période (d'après la date de naissance et la date de permis du conducteur principal).

➤ Concernant les caractéristiques du véhicule :

- calcul de l'ancienneté du véhicule suivant la période (d'après l'année de construction du véhicule).

3. Les variables décrivant la sinistralité

a. Le nombre de sinistre NIN

Il s'agit du nombre de sinistres déclarés par l'assuré à la compagnie et ayant donné lieu à récupération (i.e., de sorte qu'il y ait eu paiement de la part de l'assureur car sinon ce ne sont pas à proprement parlé des sinistres).

Ce n'est donc pas le nombre de sinistres causés par l'assuré sur l'année. L'assuré peut en effet estimer (à tort ou à raison) avoir intérêt à dédommager lui-même le tiers lésé en cas de préjudice mineur (comme une griffe dans la portière d'un véhicule ancien, par exemple, pour lequel le jeune propriétaire préférera peut être un billet de 100 €, synonyme d'agapes et de réjouissances estudiantines, à une remise en état du véhicule). Il est évident que le nombre de sinistres dédommagés directement par l'assuré est fonction de la politique de tarification *a posteriori* de l'assureur (comprenez la manière dont les sinistres déclarés et ayant donné lieu à une indemnisation par la compagnie sont pénalisés). Il convient donc d'être particulièrement vigilant à cet égard lorsqu'on envisage de modifier les mécanismes de personnalisation *a posteriori* du montant des primes.

En RC AUTOMOBILE, NSIN mérite une attention toute particulière car les coûts des sinistres ne prêtent souvent pas à une segmentation poussée. De plus, NSIN jouera un rôle central dans la personnalisation *a posteriori* des montants des primes (la plupart des systèmes commerciaux, tels les mécanismes bonus-malus, n'intégrant que le nombre des sinistres dans la formule de réévaluation de la prime en cours de contrat).

Le gros avantage de NSIN est d'être généralement connu avec précision par la compagnie. Si on excepte les sinistres survenus en fin de période, qui ne seront sans doute déclarés à l'assureur qu'au début de la période suivante, les sinistres sont généralement renseignés rapidement à la compagnie, soit en application d'une clause contractuelle imposant un délai de déclaration sous peine de déchéance de la garantie, soit que l'assuré désire être dédommagé aussi vite que possible.

| période | 2007 | | | | De 1997 à 2007 | | | |
|--------------|-----------------------|---------------------------|--------------------------|------|----------------|---------------------------|--------------------------|------|
| | Nombre k de sinistres | Nombre de polices observé | Nombre de polices prédit | khi2 | signe | Nombre de polices observé | Nombre de polices prédit | khi2 |
| 0 | 85 028 | 84 445 | 4 | + | 614 056 | 609 374 | 36 | + |
| 1 | 10 266 | 11 313 | 97 | - | 83 990 | 92 424 | 770 | - |
| 2 | 1 124 | 758 | 177 | + | 9 964 | 7 009 | 1 246 | + |
| 3 | 118 | 34 | 209 | + | 1 039 | 354 | 1 323 | + |
| 4 | 9 | 1 | 55 | + | 106 | 13 | 638 | + |
| 5 | 5 | 0 | 813 | + | 18 | 0 | 759 | + |
| 6 | 1 | 0 | 1 473 | + | 2 | 0 | 384 | + |
| Total | 96 551 | 96 551 | 2 828 | | 709 175 | 709 175 | 5 156 | |

Tableau 1. Sinistralité observée dans le portefeuille suivant le nombre de police et ajustement par une loi de Poisson

| période | 2007 | | | | De 1997 à 2007 | | | |
|--------------|-----------------------|---------------------|--------------------|------|----------------|---------------------|--------------------|------|
| | Nombre k de sinistres | Exposition observée | Exposition prédite | khi2 | signe | Exposition observée | Exposition prédite | khi2 |
| 0 | 50 474 | 50 095 | 3 | + | 375 202 | 372 200 | 24 | + |
| 1 | 8 084 | 8 746 | 50 | - | 66 686 | 71 985 | 390 | - |
| 2 | 970 | 764 | 56 | + | 8 668 | 6 961 | 419 | + |
| 3 | 108 | 44 | 92 | + | 945 | 449 | 549 | + |
| 4 | 9 | 2 | 25 | + | 98 | 22 | 269 | + |
| 5 | 5 | 0 | 322 | + | 17 | 1 | 296 | + |
| 6 | 1 | 0 | 505 | + | 2 | 0 | 144 | + |
| Total | 59 651 | 59 651 | 1 053 | | 451 617 | 451 617 | 2 091 | |

Tableau 2. Sinistralité observée dans le portefeuille suivant la durée d'exposition de la police et ajustement par une loi de Poisson

- ❑ Pour le calcul de la fréquence nous tenons compte de la durée d'exposition.
- ❑ La fréquence moyenne des sinistres pour le portefeuille 2007 est de 17,46 % par an.
- ❑ Les tableaux 1. et 2. nous apprennent que le nombre maximum de sinistres déclarés par un assuré depuis 1997 vaut 6.
- ❑ Plus en détail, de 1997 à 2007,
 - ✓ 614 056(soit 86,5874 %) assurés n'ont déclaré aucun sinistre,
 - ✓ 83 990 assurés(soit 11,8433 %) en ont déclaré 1,
 - ✓ 9 964 assurés(soit 1,4050 %) en ont déclaré 2,
 - ✓ 1 039 assurés (soit 0,1465 %) en ont déclaré 3,
 - ✓ 106 assurés (soit 0,0149 %) en ont déclaré 4,
 - ✓ 18 assurés (soit 0,0025 %) en ont déclaré 5,
 - ✓ 1 assuré (soit 0,0003 %) en ont déclaré 6.

- Le Tableau 2. décrit l'ajustement de la distribution observée de NSIN par une loi de Poisson de paramètre λ identique pour toutes les polices. L'estimateur du maximum de vraisemblance du paramètre est $\hat{\lambda} = 0.1746$. On constate que l'ajustement est très mauvais, et rejette sans aucune hésitation par un test khi-carré (valeur observée de la statistique khi-carré valant 1 053, pour une p-valeur inférieure à 10^{-4}). Si on observe la succession des signes de la séquence :

$$\Pr[NSIN = k] - \Pr[Poi(\hat{\lambda}) = k], k \in N,$$

donnée dans la dernière colonne du Tableau 2., on voit clairement apparaître la succession de signes +, -, +, +, Cette séquence n'est pas fortuite. Elle est due à une propriété très importante des mélanges de Poisson connue comme le "Shaked's Two Crossings Theorem".

La succession de signes conforte l'hypothèse d'une loi de Poisson mélange NSIN au niveau du portefeuille. Ceci indique que le portefeuille est hétérogène et justifie la différenciation *a priori* des assurés.

| NSIN de 1997 à 2007 | Total Nb police | Total % npol | Total exposition | Total % exposition | NSIN | npi | P [Xi = k] | n^pi | khi1 | Signe |
|---------------------|-----------------|----------------|------------------|--------------------|--------------|----------------|--------------|----------------|--------------|-------|
| 0 | 614 056 | 86,59% | 375 202 | 83,08% | 0 | 375 202 | 0,8241 | 372 200 | 24 | + |
| 1 | 83 990 | 11,84% | 66 686 | 14,77% | 1 | 66 686 | 0,1594 | 71 985 | 390 | - |
| 2 | 9 964 | 1,41% | 8 668 | 1,92% | 2 | 8 668 | 0,0154 | 6 961 | 419 | + |
| 3 | 1 039 | 0,15% | 945 | 0,21% | 3 | 945 | 0,0010 | 449 | 549 | + |
| 4 | 106 | 0,01% | 98 | 0,02% | 4 | 98 | 0,0000 | 22 | 269 | + |
| 5 | 18 | 0,00% | 17 | 0,00% | 5 | 17 | 0,0000 | 1 | 296 | + |
| 6 | 2 | 0,00% | 2 | 0,00% | 6 | 2 | 0,0000 | 0 | 144 | + |
| Total | 709 175 | 100,00% | 451 617 | 100,00% | Total | 451 617 | 1 | 451 617 | 2 091 | |

| | |
|--|--|
| Espérance $E[X]=$ 19,34% Moment d'ordre 2 $E[X^2]=$ 24,78% Variance $V[X]=$ 21,04% La variance est supérieure à l'espérance suivant le nombre de polices $\hat{\lambda} =$ 15,17% suivant la durée d'exposition $\hat{\lambda} =$ 19,34% | Formule de calcul $P [Xi = k] = (\hat{\lambda} / k!) * e^{-\hat{\lambda}}$ $khi2 = [npi - n^pi]^2 / n^pi$ $n^pi =$ durée d'exposition totale * P [Xi = k] $n =$ nombre de classes = 6 et $p =$ nombre de paramètres = 1 n - p - 1 = 4 $khi2$ n - p - 1, 95% = 9,488 $\alpha = 5\%$ si $khi2 >$ khi n - p - 1 alors on rejette la loi de Poisson donc ici on rejette la loi de Poisson si $khi2 <$ 0,05 alors on regroupe la tranche |
|--|--|

b. L'occurrence des sinistres IND

Il s'agit d'une variable binaire indiquant si l'assuré a déclaré au moins un sinistre sur l'année, i.e.

$$\text{IND} = 1[\text{NSIN} \geq 1] = \begin{cases} 1, & \text{si } \text{NSIN} \geq 1, \\ 0, & \text{sinon.} \end{cases}$$

En 2007, sur les 96 551 polices du portefeuille, 85 028 (soit 88,07 %) n'ont déclaré aucun sinistre et 11 523 (soit 11,93 %) ont fait jouer la garantie de la compagnie durant l'année. De 1997 à 2007, sur les 709 175 polices du portefeuille, 614 056 (soit 86,59 %) n'ont déclaré aucun sinistre et 95 119 (soit 13,41 %) ont fait jouer la garantie de la compagnie durant l'année.

c. Le coût des sinistres CTOT

☐ Il s'agit de la charge des sinistres pour l'année 2007 et de 1997 à 2007, c'est-à-dire du coût total (en euro) mis par l'assuré à charge de la compagnie, ou encore, la somme des paiements nets de recours, des réserves nettes des prévisions recours hors frais de gestion de sinistres survenus en 2007 ou de 1997 à 2007.

☐ Le risque monétaire a donc deux composantes :

- ✚ la composante occurrence IND qui nous renseigne si l'assuré a fait jouer la garantie de la compagnie au moins une fois sur la période considérée,
- ✚ la composante coût total des sinistres déclarés (nulle si $\text{IND}=0$).

☐ On peut donc représenter CTOT comme $\text{CTOT} = \text{IND} * \text{CTOT}+$
Où $\text{CTOT}+$ a même loi que CTOT sachant que $\text{CTOT}+ > 0$ ou encore sachant que $\text{IND}=1$.
La variable $\text{CTOT}+$ est strictement positive, alors que CTOT est le plus souvent nulle (nulle dans 88,07 % des cas en 2007).

Si on distingue le coût de chacun des sinistres, on obtient

$\text{CTOT} = \sum_{k=1}^{\text{NSIN}} \text{C}_k$ où C_k est le coût supposé strictement positif) du k-ième sinistre (avec une convention $\text{CTOT}=0$ si $\text{NSIN}=0$).

☐ Le coût moyen \hat{C} d'un sinistre ayant touché une police est alors donné par :
 $\hat{C} = \text{CTOT}+ / \text{NSIN}$, si $\text{NSIN} > 0$, et $\hat{C} = 0$ si $\text{NSIN} = 0$.

☐ Lors de l'étude du coût des sinistres, il est important de tenir compte du nombre de sinistres engendrés par l'assuré. Ainsi, on analysera le plus souvent \hat{C} en introduisant un poids NSIN.

☐ Une autre variable souvent utilisée est le ratio S/P par catégories de polices (à savoir la charge totale des sinistres causés par cette catégorie d'assurés divisée par le montant des primes encaissées). Cette variable dépend donc du tarif en vigueur lors de la récolte des observations. Elle introduit donc l'ancien tarif dans l'élaboration du nouveau et cet usage est à déconseiller.

☐ Nous savons que seuls 11 523 assurés sur 96 551 on fait jouer la garantie de l'assureur durant l'année 2007. Intéressons-nous à la sinistralité de ces assurés (c'est-à-dire à la variable $\text{CTOT}+$).

d. Le coût des sinistres CTOT de 1997 à 2007

Le coefficient d'asymétrie vaut 107, ce qui traduit une asymétrie gauche. Plus précisément, **25%** des polices qui ont donné lieu au paiement d'indemnités par la compagnie ont causé des sinistres dont le montant était inférieur à **190 €**, **50%** inférieur à **385 €**, **75%** inférieur à **1 172 €**. La moyenne de CTOT+ de 1997 à 2007 s'élève à **1 303 €**. La prime pure empirique vaut donc :

$$E[^{CTOT+}]E[^{IND}] = E[^{CTOT}] = 1\,303 * 20,71\% (93\,521 / 451\,617) = 270 \text{ €}$$

Le coefficient de variation de CTOT+ vaut **10**, pour un écart-type de **12 990 €**.

Les quatre polices ayant causé les sinistres dont les montants sont plus élevés sont **919 677 €**, **996 346 €**, **1 501 198 €**, **1 777 877 €** et **2 141 077 €**.

Il est souvent nécessaire d'écarter les sinistres afin d'analyser leurs coûts. Une approche classique consiste à les plafonner à un quantile, par exemple **q99% = 10 191 €**. La théorie des valeurs extrêmes fournit les outils adéquats pour traiter les sinistres graves.

4. La mesure de l'exposition au risque DUR

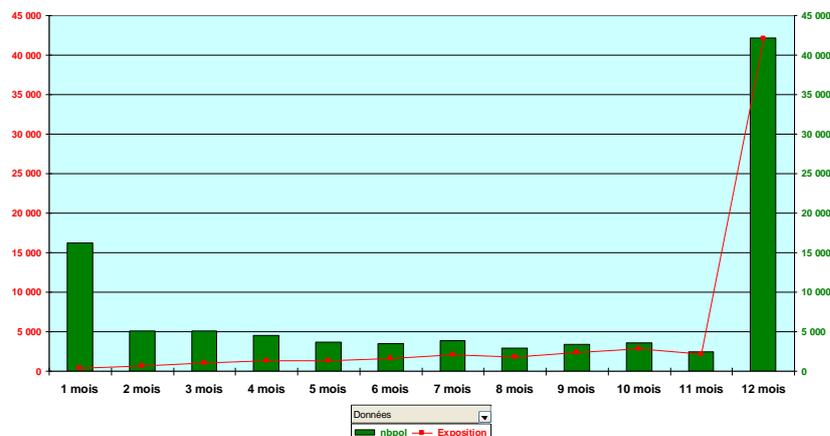
a. La mesure de l'exposition au risque : la variable DUR

□ Il s'agit du nombre de jours où la police a été en vigueur durant l'année 2007. Nous nous en servons pour mesurer l'exposition au risque. Elle permet de tenir compte du fait qu'un sinistre déclaré par une police en vigueur durant 1 mois est un plus mauvais signe pour l'assureur qu'un sinistre relatif à une police en vigueur toute l'année.

□ Dans notre portefeuille, la durée moyenne de couverture est de 225 jours et la médiane est de 272 jours. Certaines polices n'ont été en vigueur que pendant une seule journée, soit 7,62% du portefeuille 2007. Près de 41% du portefeuille ont une durée de 1 an.

□ Notons au passage que l'exposition au risque devrait être mesurée par le nombre de kilomètres parcourus plutôt que par le nombre de jours où la police était en vigueur (le véhicule pourrait fort bien rester au garage pendant certaines périodes, n'étant donc pas soumis au risque).

b. Répartition du nombre de polices et de l'exposition issue du portefeuille 2007 suivant de la durée de la couverture



5. Définition des variables à expliquer

a. Variables calculées

La fréquence et le coût moyen sont calculés comme suit :

✚ Fréquence

$$\text{FREQ} = \frac{\sum_i (\text{Nombre de sinistres relatifs à la police } i)}{\sum_i (\text{durée d'exposition au risque pour la police } i)}$$

✚ CM Coût Moyen

CMHR Coût Moyen Hors Réassurance (écrêtement des sinistres à 300 k€)

CMHR100 Coût Moyen Hors Réassurance (écrêtement des sinistres à 100 k€)

$$\text{CM} = \frac{\sum_i (\text{Montant des sinistres actualisé relatifs à la police } i)}{\sum_i (\text{Nombre de sinistres relatifs à la police } i)}$$

6. Caractéristiques du preneur d'assurance

Seule la variable sexe sera présentée, dans cette section, le reste se trouvant en annexe II.

a. Variable Sexe

Cette variable qualitative binaire donne le sexe du souscripteur, i.e.

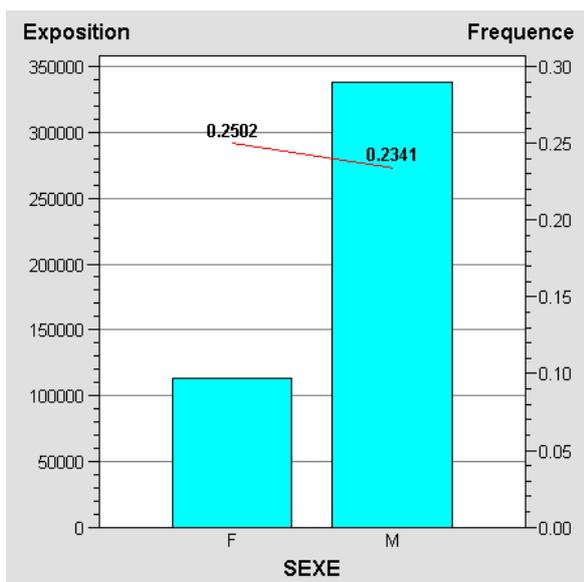
SEXE = $\begin{cases} 0, & \text{si l'assuré est un homme (niveau de référence),} \\ 1, & \text{si l'assuré est une femme.} \end{cases}$

Le portefeuille est constitué de 113 280 femmes toutes années confondues de 1997 à 2007 (soit 25 % du portefeuille) et 338 337 hommes toutes années confondues (soit 75% du portefeuille). On peut constater, au niveau des graphes suivants que :

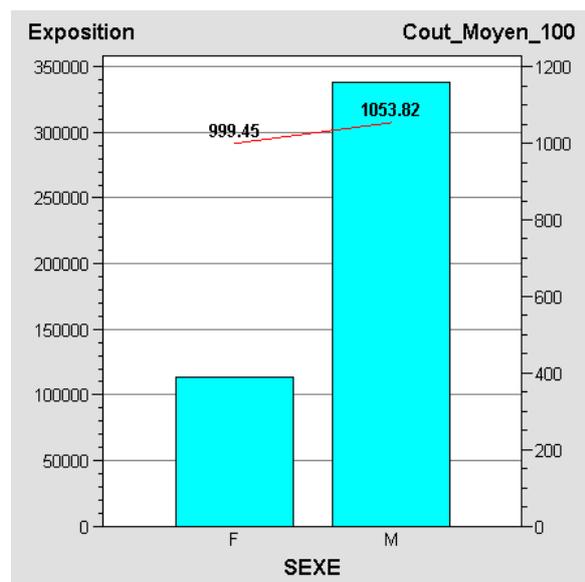
- la fréquence est plus élevée pour les femmes,
- le coût moyen est plus élevé pour les hommes.

Pour donner une vision plus objective du coût moyen suivant le sexe, nous devons écrêter les sinistres en prenant comme crête 100 K€.

Analyse de la fréquence
suivant le Sexe

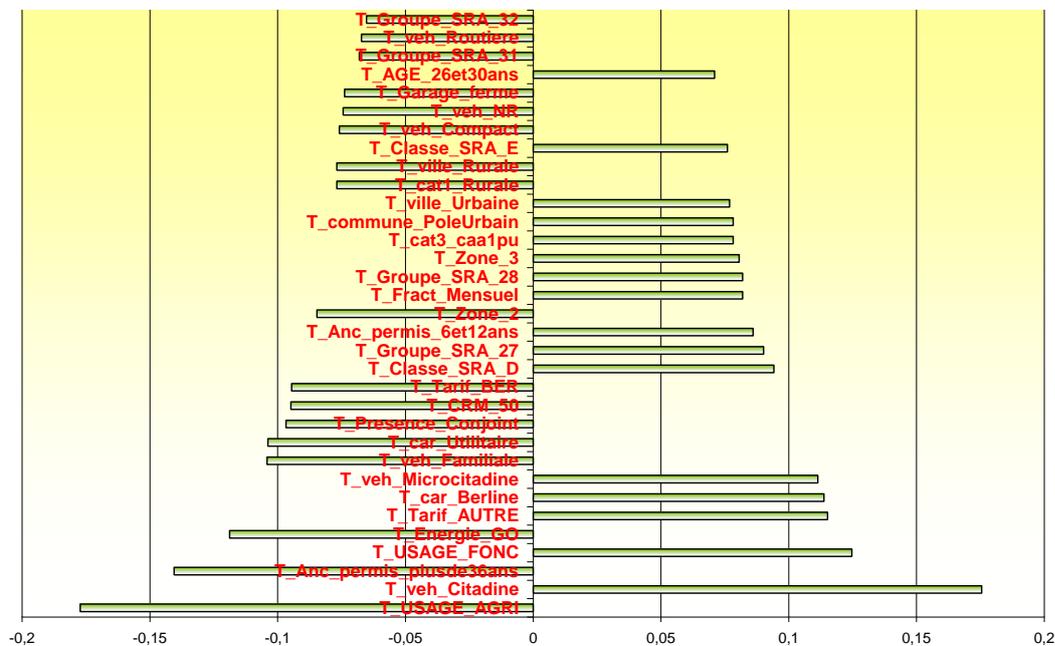


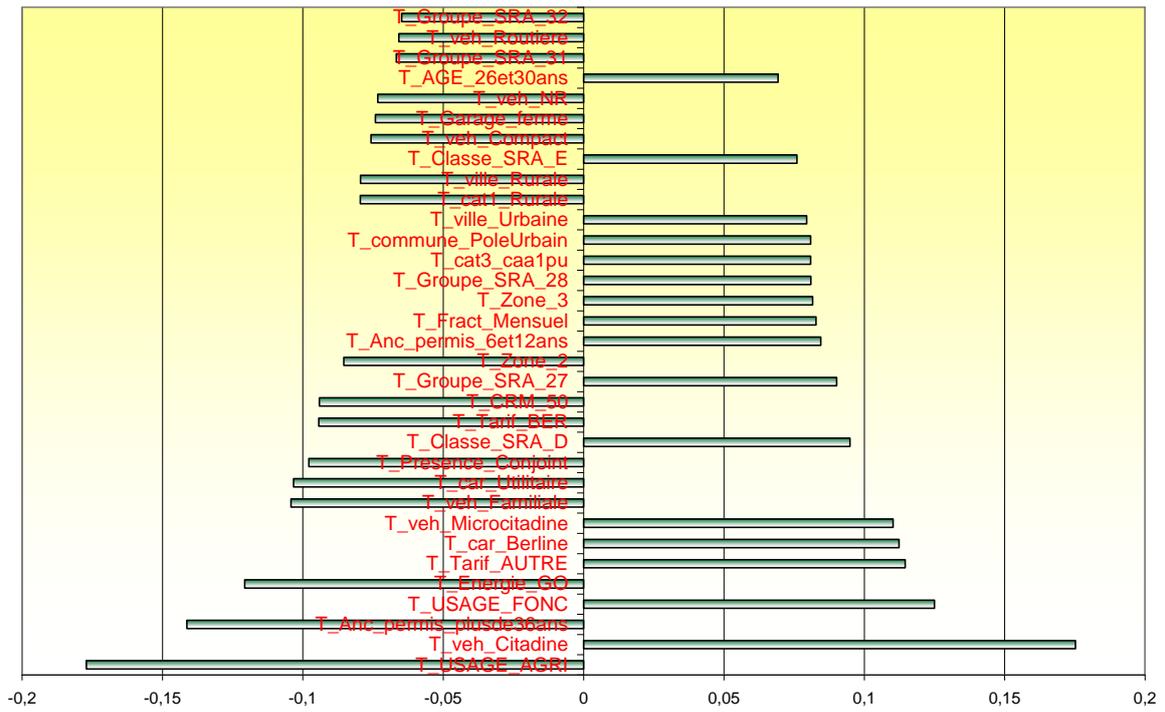
Analyse du coût moyen hors sinistres
supérieur à 100 K€ suivant le sexe



Remarque : La Cour de justice de l'Union européenne a tranché. A compter du 21 décembre 2012, les compagnies d'assurance ne pourront plus attribuer des primes ou fixer leurs prix en fonction du sexe de leurs assurés. « Prendre en compte le sexe des individus assurés comme un facteur de risque dans les contrats d'assurance constitue une discrimination », a rappelé l'autorité. Il est évident que les assureurs vont tenter de trouver une solution à cette problématique. Une première réponse peut être donnée en réalisant une analyse des corrélations entre la variable sexe et les autres critères de tarification à l'aide d'un test de V Cramer. Elle nous permet d'affirmer que la suppression du facteur de risque sexe peut être compensé, par exemple, par l'intégration du type de véhicule dans la tarification. Si nous regardons le graphe, ci-dessous, nous pouvons dire qu'il existe une corrélation positive entre la variable sexe et le type de véhicule citadine voire micro citadine et une corrélation négative entre la variable sexe et le type de véhicule familiale voire routière ou un utilitaire.

Les deux graphes projetant la valeur d'une statistique sont issus du test de V Cramer réalisé à l'aide de la proc freq de sas sur la base donnée utile à la modélisation de la fréquence RC Matériel IDA et Hors IDA. Il permet de définir la corrélation entre la variable sexe et les autres critères de tarification. Les résultats sont similaires sur les bases de données utiles à la modélisation Dommage IDA, Dommage hors IDA, Bris de glaces, Vol/Inc et RC Corporelle mais ils ne seront pas présentés ici.





7. Caractéristiques du véhicule

Seule la variable type de carburant sera présentée, dans cette section, le reste se trouvant en annexe II.

a. Variable Type de véhicule

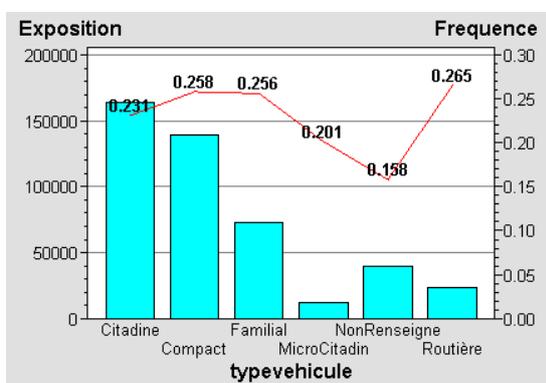
La variable qualitative type de véhicule est définie par 6 modalités. Dans notre portefeuille,

- **11 980** véhicules de type Micro Citadine (soit **3 %**) : Twingo, C1, 106, ...,
- **164 365** véhicules de type Citadine (soit **36 %**) : Clio, C2, 207, ..., (**niveau de référence**)
- **139 710** véhicules de type Compact (soit **31 %**) : Mégane, C3, 307, ...,
- **72 943** véhicules de type Familiale (soit **16 %**) : Laguna, C4, 407, ...,
- **23 450** véhicules de type Routière (soit **5 %**) : Safrane, C5, 607, ...,
- **39 169** véhicules de type Non Renseigné (soit **9 %**) : mauvaise saisie du code auto ou saisie de code provisoire ou non renseigné ancien contrat Monceau.

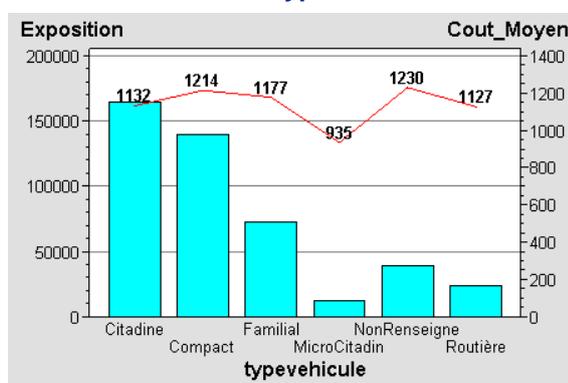
Afin d'apprécier l'influence du Type du véhicule sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant le Type du véhicule.

On constate que la fréquence, est élevée pour les routières, les familiales et les compacts. Seule les compacts et les non renseignés ont coût moyen au-dessus de la moyenne. Il est à noter toutefois que la dégradation du coût moyen est moins significative que la fréquence. Les résultats sur les Microcitadines sont plus que remarquables, c'est donc un segment à ne pas négliger.

**Analyse de la fréquence
suivant type de véhicule**



**Analyse du coût moyen
suivant le type de véhicule**



8. Caractéristiques de la police

Seule la variable présence d'un conducteur novice occasionnel sera présentée, dans cette section, le reste se trouvant en annexe II.

a. Variable présence d'un conducteur novice occasionnel

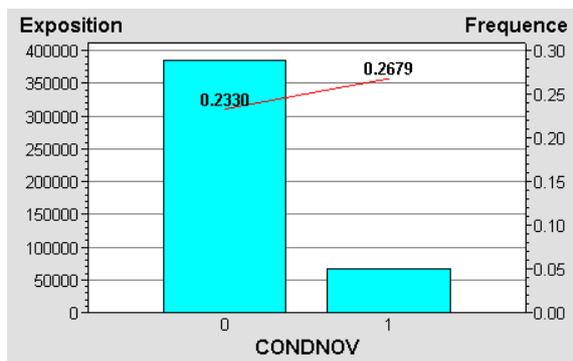
Cette variable qualitative binaire nous renseigne sur la présence d'un conducteur novice occasionnel déclaré au contrat, i.e.

CONDNOV = $\begin{cases} 0, & \text{si pas de présence d'un conducteur novice occasionnel } \mathbf{384\ 944} \text{ (soit } \mathbf{85\%}) \\ & \text{(niveau de référence),} \\ 1, & \text{si présence d'un conducteur novice occasionnel, } \mathbf{66\ 673} \text{ (soit } \mathbf{15\%}). \end{cases}$

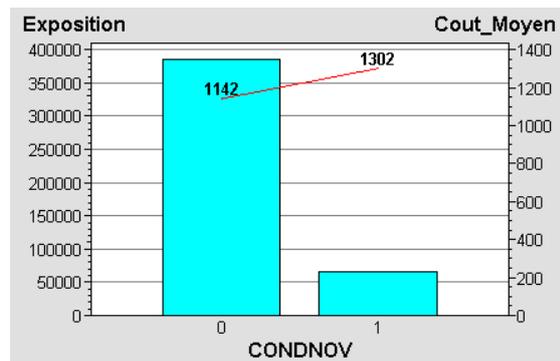
Afin d'apprécier l'influence de la variable CONDNOV sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la variable CONDNOV.

On peut noter au niveau des graphes suivants, aussi bien, en terme de fréquence que de coût moyen, une sur sinistralité significative pour les contrats ayant des conducteurs novices occasionnels déclarés.

Analyse de la fréquence suivant la variable conducteur novice occasionnel



Analyse du coût moyen suivant la variable conducteur novice occasionnel

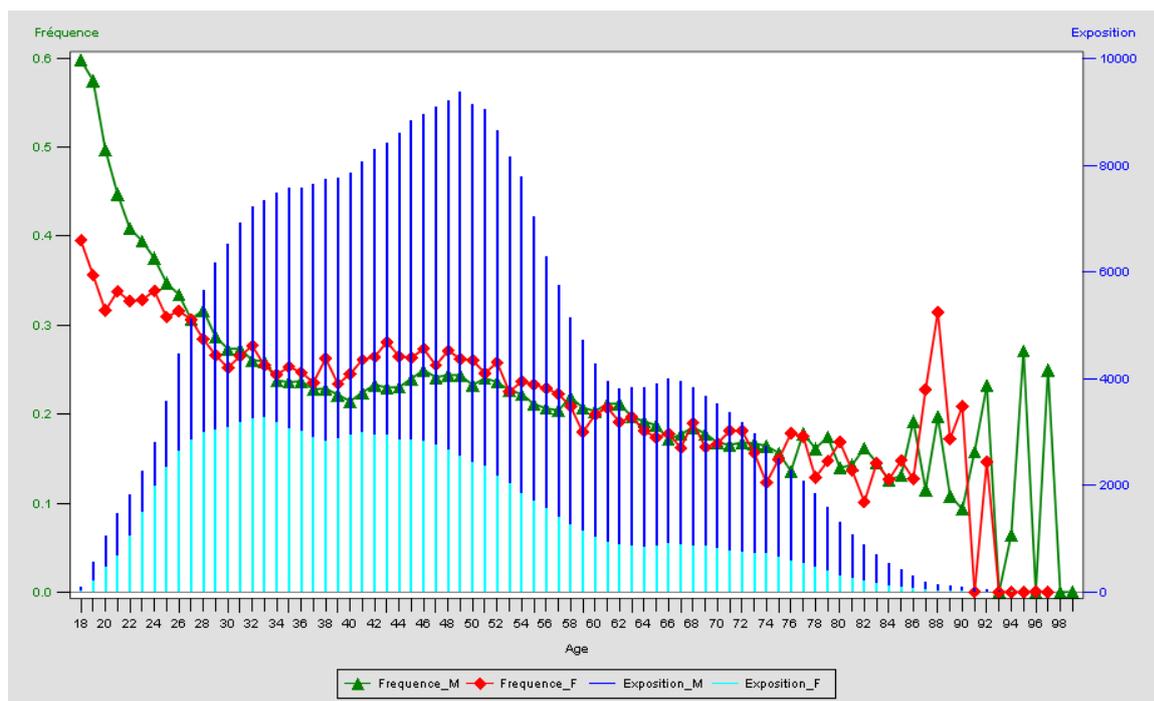


9. Interaction entre variables tarifaires

Souvent, on constate une interaction entre l'âge et le sexe de l'assuré. Par interaction, il faut entendre ici une influence différente de l'âge de l'assuré sur la fréquence de sinistre, selon qu'on est un homme ou une femme.

Dans notre portefeuille, si nous distinguons les hommes et les femmes, nous obtenons les fréquences de sinistre par âge décrites dans les graphes suivants.

On constate une sur-sinistralité des jeunes hommes par rapport aux jeunes femmes jusqu'à 30 ans, et ensuite la tendance s'inverse jusqu'à 57 ans et pour le reste les fréquences sont assez semblables. Nous ne poursuivons pas ici la modélisation des interactions.



10. Premier tri parmi les variables tarifaires

a. Test du khi-carré d'indépendance

Une manière efficace d'opérer un premier tri parmi les variables à notre disposition consiste à effectuer des tests khi-carrés sur base de tables de contingence.

Pour des raisons d'effectifs, il vaut mieux travailler avec IND (0 : pas de sinistre ou 1 : un ou plusieurs sinistres) qu'avec NSIN (nombre de sinistres).

En effet, si nous croisons NSIN avec SEXE, beaucoup d'effectifs attendus sont inférieurs à 5%; on ne peut donc pas valablement effectuer de test khi-carré sur la base d'une telle table de contingence décrite dans le tableau ci-dessous (les effectifs attendus sous l'hypothèse d'indépendance sont indiqués entre parenthèse), sur laquelle nous pouvons baser le test khi-carré.

La valeur observée de la statistique khi-carrée d'indépendance vaut 63.6799, ce qui donne une p-valeur inférieure à 10^{-4} (<0.0001). On constate donc une forte association entre le sexe et le fait de causer ou pas de sinistre.

b. Table de contingence croisant IND et SEXE

La valeur observée de la statistique khi-carrée d'indépendance =

$$(\text{Nb observé} - \text{Nb théorique})^2 / \text{Nb théorique}$$

à comparer avec la statistique du Khi-deux à $4-1-1=2$ degrés de liberté

| IND | | SEXE | | | Total |
|-------------------------------|------------------|---------|---------|---------|---------|
| | | Inconnu | Femme | Homme | |
| 0 : Pas de sinistre | Nb observé | 148 | 93 344 | 281 789 | 375 282 |
| | Nb théorique | 132 | 94 133 | 281 017 | 83,10% |
| | Écart | 16 | -788 | 772 | |
| | Cell. Khi-2 | 2 | 7 | 2 | |
| | Pourcentage | 0,03% | 20,67% | 62,40% | |
| | Pourct. en ligne | 0,04% | 24,87% | 75,09% | |
| | Pourct. en col. | 93,44% | 82,40% | 83,33% | |
| 1 : Un ou plusieurs sinistres | Nb observé | 10 | 19 936 | 56 389 | 76 335 |
| | Nb théorique | 27 | 19 147 | 57 161 | 16,90% |
| | Écart | -16 | 788 | -772 | |
| | Cell. Khi-2 | 10 | 32 | 10 | |
| | Pourcentage | 0,00% | 4,41% | 12,49% | |
| | Pourct. en ligne | 0,01% | 26,12% | 73,87% | |
| | Pourct. en col. | 6,56% | 17,60% | 16,67% | |
| Total | | 159 | 113 280 | 338 179 | 451 617 |
| | | 0,04% | 25,08% | 74,88% | 100,00% |

Le rejet de l'hypothèse d'indépendance indique que le sexe influence la variable IND et donc également la variable NSIN. Le non rejet par contre ne permet pas de conclure, d'une part pour des raisons liées à l'erreur de second espèce (laquelle n'est pas contrôlée) et d'autre part car le fait de ne pas influencer IND n'implique pas qu'on n'influence pas NSIN (car la variable explicative peut influencer la répartition dans les catégories 1,2,... sinistres).

En croisant successivement toutes les variables tarifaires avec IND, on obtient les résultats qui figurent dans les graphes suivants :

c. Dépendances vraie et apparente

Notez que le rejet de l'hypothèse d'indépendance entre SEXE et IND indique que le sexe de l'assuré influence la probabilité d'avoir au moins un sinistre sur la période. Il peut néanmoins s'agir d'une dépendance vraie, auquel cas la probabilité d'occurrence dépend vraiment du sexe de l'assuré, ou d'une dépendance apparente, auquel cas la probabilité d'avoir au moins un sinistre dépend d'une variable corrélée avec le sexe (qu'elle soit cachée, comme l'agressivité au volant par exemple, ou observable, comme l'âge de l'assuré, si les structures d'âge diffèrent entre hommes et femmes). Dans ce dernier cas de figure, la dépendance entre le sexe de l'assuré et le fait d'avoir ou pas de sinistres disparaîtrait si on ne tenait pas compte de la troisième variable.

11. Analyse des fréquences de sinistre

11.1. Modèle de régression de poisson et de negative binomiale pour le nombre des sinistres

Dans la modélisation des processus de comptage, ici de la fréquence des sinistres, deux sortes de modèle sont couramment mis en œuvre; le modèle de poisson et le modèle binomial négatif.

Rappelons les définitions et les propriétés de ces modèles pour bien comprendre par la suite l'emploi des modèles.

Modèle de Poisson

Dans un modèle de Poisson, la probabilité pour qu'une variable aléatoire Y (nombre de sinistres responsables déclarés) prenne la valeur y_i ($y_i = 0, 1, 2, \dots$) pour un assuré i est donnée par :

$$P(Y = y_i / X_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad (1)$$

où le paramètre λ_i dépend du vecteur X_i des caractéristiques (régresseurs) de l'assuré i par une équation log linéaire, à savoir : $\ln \lambda_i = X_i' \beta$ où β est le vecteur des coefficients à estimer.

On vérifie aisément que dans la loi de Poisson (1), l'espérance est égale à la variance.

$$E(Y_i / X_i) = Var(Y_i / X_i) = \lambda_i = e^{X_i' \beta} \quad (2)$$

Cette hypothèse d'équidispersion (homogénéité du portefeuille par rapport au risque) est très restrictive. Mais dans la pratique, du fait d'une abondance de valeurs nulles et de la présence de quelques valeurs extrêmes, la variance est supérieure à la moyenne. Dans ce cas, on parle d'une sur-dispersion de la variable Y. Cette situation implique une sous-estimation des écarts types et on rejette trop souvent l'hypothèse nulle de non significativité des coefficients β du modèle. D'où l'idée d'utiliser un modèle de comptage alternatif, basé sur la loi binomiale négative, qui prend en compte cette sur-dispersion par l'introduction d'un paramètre supplémentaire (α) qui permet, en outre, de capter l'hétérogénéité inobservée de la variable endogène (qui peut impliquer la sur-dispersion observée).

Dans un modèle binomial négatif, on définit la probabilité pour que Y prenne la valeur y_i par:

$$P(Y = y_i / X_i) = \frac{\Gamma(y_i + \nu)}{\Gamma(y_i + 1) \cdot \Gamma(\nu)} \cdot \left(\frac{\nu}{\nu + \lambda_i} \right)^\nu \left(\frac{\lambda_i}{\nu + \lambda_i} \right)^{y_i} \quad (3)$$

En posant $\nu = 1/\alpha$, l'espérance et la variance s'expriment ainsi :

$$E(y_i / X_i) = \lambda_i = e^{X_i \beta}, \quad Var(y_i / X_i) = \lambda_i (1 + \alpha \lambda_i) \quad (4)$$

La variance est donc différente de l'espérance et le paramètre α traduit une sur dispersion (ou une sous dispersion) des données. Si $\alpha = 0$, le modèle binomial se réduit au modèle de Poisson. Si $\alpha > 0$, le modèle de Poisson est rejeté au profit du modèle négatif binomial. La sur-dispersion peut être testée soit par le ratio $D / (n-p)$, où D désigne la déviance, n le nombre d'observations et p le nombre de paramètres dans le modèle, soit par le ratio $X^2 / (n-p)$, où X^2 correspond à la statistique du chi-deux de Pearson. La déviance est définie comme 2 fois la différence entre le maximum possible de la log vraisemblance et le maximum atteint sur le modèle estimé.

Le X^2 de Pearson correspond à la somme des carrés des écarts à la moyenne.

Si ces ratios sont supérieurs à 1, les données présentent une sur dispersion (et une sous dispersion si ces ratios sont inférieurs à 1).

11.2. Explication de l'effet des paramètres sur la prime

Soit N_i le nombre de sinistres déclarés par l'assuré i durant les années 1997 à 2007 ou durant l'année 2007, $i = 1, 2, \dots, n$.

Nous notons d_i l'exposition au risque pour l'individu i.

Le modèle de Poisson suppose que la loi conditionnelle de N_i sachant x_i est de Poisson. Il suffit dès lors de spécifier sa moyenne $E[N_i|x_i]$. Comme cette dernière est strictement positive, on retient généralement une moyenne de forme exponentielle linéaire, i.e.

$$E[N_i|x_i] = d_i * \exp(\beta^t x_i), \quad i = 1, 2, \dots, n,$$

où β est un vecteur de coefficients de régressions inconnus. Nous travaillons donc dans le modèle

$$N_i \sim \text{Poi}(d_i * \exp(\beta^t x_i)), \quad i = 1, 2, \dots, n.$$

Lorsque toutes les variables sont catégorielles, chaque assuré est donc représenté par un vecteur x_i dont les composantes valent 0 ou 1. Dans ce cas, la fréquence annuelle $\lambda_i = \exp(\beta^t x_i)$ apparaît comme un produit de coefficients de majoration ou de réduction par rapport à la fréquence de l'individu de référence du portefeuille. Plus précisément,

$$\lambda_i = \exp(\beta^t x_i) = \exp(\beta_0) + \sum_j \beta_j x_{ij}$$

Dès lors, $\exp(\beta_0)$ est la fréquence annuelle de l'individu de référence du portefeuille tandis que chacun des facteurs $\exp(\beta_j)$ traduit l'influence d'un critère de segmentation (si $\beta_j > 0$, les individus présentant cette caractéristique subiront une majoration de prime par rapport à la prime de référence $\exp(\beta_0)$, tandis que $\beta_j < 0$, indiquera une réduction de prime).

Les paramètres possèdent ainsi l'interprétation suivante : si on considère le caractère codé par la jème variable binaire,

$$E[Ni \mid \text{caractère présent}] = \exp(\beta_j)$$

$$E[Ni \mid \text{caractère absent}]$$

Sur base de cette dernière équation, $\exp(\beta_j)$ est donc le facteur par lequel il faut multiplier la fréquence de sinistres des individus ne présentant pas le caractère codé par la jème variable binaire pour obtenir des individus présentant ce caractère.

Le modèle négative binomiale amène aux mêmes conclusions.

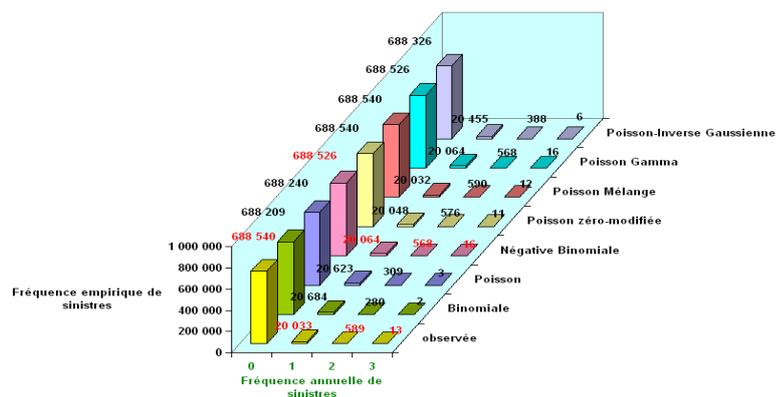
11.3. Présentation des différents modèles de comptage sur un échantillon représentant le nombre de sinistres de 1997 à 2007 vu à fin juillet 2008

Nous présenterons uniquement le modèle de comptage concernant la garantie RCM (Responsabilité Civile Matériel) et concernant les autres garanties nous trouverons la présentation en Annexe.

a. concernant la garantie RCM

| Modèle de comptage concernant la Garantie RCM | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | |
|---|----------|------------------------------------|-----------------|---------------------|--------------|----------|--------------|----------|--------------------|----------|-----------------------|----------|-----------------|-------------|---------------|----------|----------------------------|----------|
| | | | $p^n/n! \alpha$ | $\beta^n/n! \alpha$ | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | |
| | | | 0,00699 | 0,00300 | 1 | 10 | 1 | 0,0300 | 2 | 2 | 1 | 1 | 3 | 2 | 2 | 2 | 2 | 2 |
| | | | α | α | α | α | λ | 0,0300 | p | r | λ | 0,0575 | λ_1 | λ_2 | p | r | μ | β |
| classes de fréquence annuelle | observée | fréquence relative empirique f_j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux |
| 0 | 688 540 | 97,03% | 688 178,18 | 688 209,26 | 0,19 | 0,16 | 688 240,22 | 0,13 | 688 526,35 | 0,00 | 688 540,00 | 0,00 | 688 540,20 | 0,00 | 688 526,35 | 0,00 | 688 325,66 | 0,07 |
| 1 | 20 033 | 2,82% | 20 745,17 | 20 683,75 | 24,45 | 20,47 | 20 622,70 | 16,86 | 20 064,15 | 0,05 | 20 047,76 | 0,01 | 20 032,20 | 0,00 | 20 064,15 | 0,05 | 20 455,06 | 8,71 |
| 2 | 689 | 0,08% | 250,15 | 279,74 | 459,02 | 341,90 | 308,97 | 253,79 | 568,11 | 0,77 | 576,04 | 0,29 | 590,19 | 0,00 | 568,11 | 0,77 | 387,98 | 104,15 |
| 3 | 13 | 0,00% | 1,51 | 2,24 | 67,57 | 51,62 | 3,09 | 31,66 | 16,39 | 0,70 | 11,19 | 0,29 | 12,41 | 0,03 | 16,39 | 0,70 | 6,29 | 7,15 |
| nb occurrence | 709 175 | 100,00% | 709 175 | 709 175 | 571,23 | 414,16 | 709 175 | 302,63 | 709 175 | 1,52 | 709 175 | 0,59 | 709 175 | 0,03 | 709 175 | 1,52 | 709 175 | 120 |
| λ | 0,0300 | 0,0300 | 0,0300 | | 0,0300 | | 0,0300 | | 0,0300 | | 0,0300 | | 0,0300 | | 0,0300 | | 0,0300 | |
| λ^2 | 5,99 | 5,99 | 5,99 | | 5,99 | | 5,99 | | 5,99 | | 5,99 | | 5,99 | | 5,99 | | 5,99 | |
| λ^3 | 0,0317 | 0,0317 | 0,0317 | | 0,0317 | | 0,0317 | | 0,0317 | | 0,0317 | | 0,0317 | | 0,0317 | | 0,0317 | |
| λ^4 | 0,0308 | 0,0308 | 0,0308 | | 0,0308 | | 0,0308 | | 0,0308 | | 0,0308 | | 0,0308 | | 0,0308 | | 0,0308 | |
| λ^5 | 5,86 | 5,86 | 5,86 | | 5,86 | | 5,86 | | 5,86 | | 5,86 | | 5,86 | | 5,86 | | 5,86 | |
| λ^6 | 0,0354 | 0,0354 | 0,0354 | | 0,0354 | | 0,0354 | | 0,0354 | | 0,0354 | | 0,0354 | | 0,0354 | | 0,0354 | |
| λ^7 | 0,0430 | 0,0430 | 0,0430 | | 0,0430 | | 0,0430 | | 0,0430 | | 0,0430 | | 0,0430 | | 0,0430 | | 0,0430 | |
| λ^8 | 0,0326 | 0,0326 | 0,0326 | | 0,0326 | | 0,0326 | | 0,0326 | | 0,0326 | | 0,0326 | | 0,0326 | | 0,0326 | |
| λ^9 | 0,0399 | 0,0399 | 0,0399 | | 0,0399 | | 0,0399 | | 0,0399 | | 0,0399 | | 0,0399 | | 0,0399 | | 0,0399 | |
| λ^{10} | 6,02 | 6,02 | 6,02 | | 6,02 | | 6,02 | | 6,02 | | 6,02 | | 6,02 | | 6,02 | | 6,02 | |
| λ^{11} | 37,86 | 37,86 | 37,86 | | 37,86 | | 37,86 | | 37,86 | | 37,86 | | 37,86 | | 37,86 | | 37,86 | |
| λ^{12} | 97,03% | 97,03% | 97,03% | | 97,03% | | 97,03% | | 97,03% | | 97,03% | | 97,03% | | 97,03% | | 97,03% | |

Diagramme en bâtons sur données MRACA concernant la garantie RCM Automobile



Conclusions :

Compte tenu des résultats exposés ci-dessus, il ressort très clairement que les lois de Poisson, Négative Binomiale et Inverse Gaussienne pour toutes les garanties sont proches de leur distribution empirique respective sauf pour les garanties BDG et DOMM, où la loi de Poisson mélange donne très nettement de meilleurs résultats.

12. Analyse des coûts des sinistres

1. Analyse par garantie

Nous allons travailler sur l'échantillon des montants de sinistres par garantie.

Dans de nombreux ouvrages « mathématiques » (ou actuariels) traitant de l'assurance non vie, les modèles proposés pour l'estimation du coût des sinistres font référence à trois lois de probabilité : la loi Gamma, la loi de Weibull et la loi log-normal (celle qui semble faire autorité).

Nous sommes donc partis sur cette base et nous avons tenté de modéliser la distribution du coût des sinistres par l'une de ces lois et d'autres comme la loi de Pareto ou la loi Exponentielle. Un simple test du Chi2 montre que toute tentative est vaine. Après un examen plus approfondi de notre échantillon, il apparaît plusieurs points « d'accumulation », reflets d'un nombre important d'ex-aequo. Il est donc clair qu'une distribution, possédant autant d'ex-aequo, ne peut être modélisée par une distribution continue.

En fait, les modèles proposés dans la majorité des ouvrages ne tiennent pas compte de la réalité d'une compagnie d'assurance. En effet, ils n'utilisent que des échantillons de sinistres entièrement réglés. Or, dans notre échantillon, de nombreux sinistres ne sont pas réglés en totalité et d'autres ne sont même qu'à « l'état de provision ». De plus, les forfaits IDA constituent aussi des points d'accumulation.

Le montant des ex-aequo correspondent aux montants des provisions (issues de la grille des coûts moyens forfaitaires par garantie de la société) des sinistres non ouverts et des forfaits **IDA(1 204 €)**, ce qui explique ces nombreux points d'accumulation et nous amène à penser que la distribution de notre échantillon correspond à un mélange de lois : une loi continue pour modéliser la distribution (sans doublons) du coût des sinistres et une loi discrète pour modéliser la distribution des points d'accumulation correspondant aux provisions des sinistres non encore ouverts.

2. Présentation des deux lois les plus usitées pour la modélisation du coût individuel de sinistre

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

- le modèle Gamma sur les coûts individuels Y_i
- le modèle log-normal sur les coûts individuels Y_i , ou plutôt un modèle Gaussien sur le logarithme des coûts, $\log(Y_i)$, la loi lognormale n'appartenant pas à la famille exponentielle.

Pour rappel la loi Gamma admet pour densité

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha) \beta^{-\alpha}}$$

d'espérance $\alpha/\beta = \exp(\beta^t \mathbf{x}i)$, et de variance $\alpha/\beta^2 = (\exp(\beta^t \mathbf{x}i))^2/\alpha$. Et la loi lognormale,

$$f(x; \mu, \sigma) = \frac{e^{-(\ln x - \mu)^2 / (2\sigma^2)}}{x\sigma\sqrt{2\pi}}$$

d'espérance $e^{\mu + \sigma^2/2}$ et de variance $(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$.avec $\mu = \beta^t \mathbf{x}i$.

3. Utilisation des Q-Q plots pour déterminer la loi de distribution des montants de sinistre et présentation des statistiques permettant d'avoir une vision globale de la répartition des montants de sinistre par garantie.

a. Q-Q plots

Très opérationnelle, cette technique d'analyse exploratoire des données présente l'avantage de permettre un examen visuel de la queue de distribution. Des compléments seront trouvés dans D'Agostino et al., 1986, ou BP.

Cette technique repose sur le résultat suivant :

Proposition Si (X_1, \dots, X_n) est un n-échantillon i.i.d de F, on a

$E[F(X(i))] = i / (n+1)$ et $V[F(X(i))] = O(1/n)$ pour $i = 1, \dots, n$.

La conséquence pratique de ce résultat consiste à assimiler, pour $n \geq 30$, la réalisation $F(X(i))$ à son espérance mathématique

$E[F(X(i))] = i / (n+1)$.

On en déduit $F(x(i)) \approx i / (n+1)$ pour $i = 1, \dots, n$ puis $x(i) \approx F_{-1}(i / (n+1))$

(approximation moins bonne à l'extrémité qu'au centre).

Si l'hypothèse $F \in \mathcal{F} = \{ F_\theta : \theta \in \Theta \}$ est vérifiée, il existe $\theta \in \Theta$ tel que $F = F_\theta$ et donc tel que

$F_\theta(x(i)) \approx i / (n+1)$ où $x(i) \approx F_{-1}(i / (n+1))$ pour $i = 1, \dots, n$.

Une transformation judicieuse dépendant de la famille F doit alors conduire à un quasi alignement de n points liés aux $x(i)$.

Un jugement, subjectif, sur l'alignement de ces points permet de conclure $F \in \mathcal{F}$ ou d'infirmer cette hypothèse.

Exemple : Famille exponentielle $\mathcal{F} = [\varepsilon(\lambda) : \lambda > 0]$ comme Fonction de répartition d'une loi exponentielle

$F_\lambda(x) = 1 - \exp(-\lambda x)$, $F_\lambda(x(i)) \approx i / (n+1)$ est équivalent à $\lambda x \approx - \ln(1 - i / (n+1))$

Si $F \in \mathcal{F}$, les n points $(x(i), - \ln(1 - i / (n+1)))$ $i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 passant par 0.

Remarque : Cette procédure met en relation les quantiles théoriques $F_{-1}(i / (n+1))$ et empiriques $x(i)$. D'où la dénomination Q-Q plot. Des procédures Q-Q plots simples ne sont pas disponibles pour les distributions Burr, Benktander I et II et Pareto étendue.

Le tableau suivant donne les procédures Q-Q plots pour différentes distributions standards.

| Familles de distributions | Procédures d'adéquation |
|---|---|
| Exponentielle $\varepsilon(\lambda), \lambda > 0$ | $(x(i), -\ln(1 - i / (n+1))) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 passant par 0. |
| Gamma $\gamma(v, \beta) v, \beta > 0$ | Par essais successifs, on cherche s'il existe $v > 0$ tel que $(x(i), \chi^2_{2v}(i / (n+1))) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 . |
| Log Normale Log N(μ, σ^2), $\mu \in \mathbb{R}, \sigma > 0$ | $(\ln(x(i)), \Phi^{-1}(i / (n+1))) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 . |
| Weibull $W(\tau, v) \tau, v > 0$ | $(\ln(x(i)), \ln[-\ln(1 - i / (n+1))]) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 . |
| Loi de Pareto type 2 P(a, α) a donné, $\alpha > 0$ | $(\ln(x(i)/a), -\ln(1 - i / (n+1))) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 passant par 0. |
| Pareto P(a, α) a donné, $\alpha > 0$ | Par essais successifs, on cherche s'il existe $a > 0$ tel que $(\ln(1 + x(i)/a), -\ln(1 - i / (n+1))) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 passant par 0. |
| Log Gamma $\gamma(v, \beta) v, \beta > 0$ | Par essais successifs, on cherche s'il existe $v > 0$ tel que $(\ln(x(i)), \chi^2_{2v}(i / (n+1))) i = 1, \dots, n$ sont « sensiblement » alignés sur une droite de pente > 0 passant par 0. |

b. Présentation de l'échantillon des montants de sinistres de 1997 à 2007 vu à fin juillet 2008 suivant les différentes garanties, telles que :

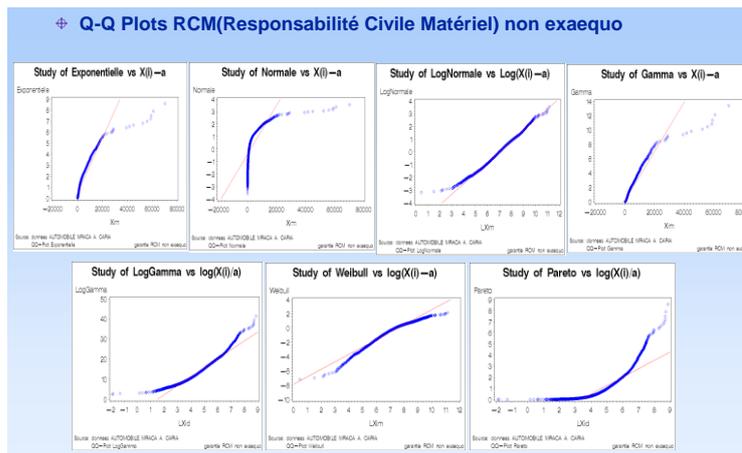
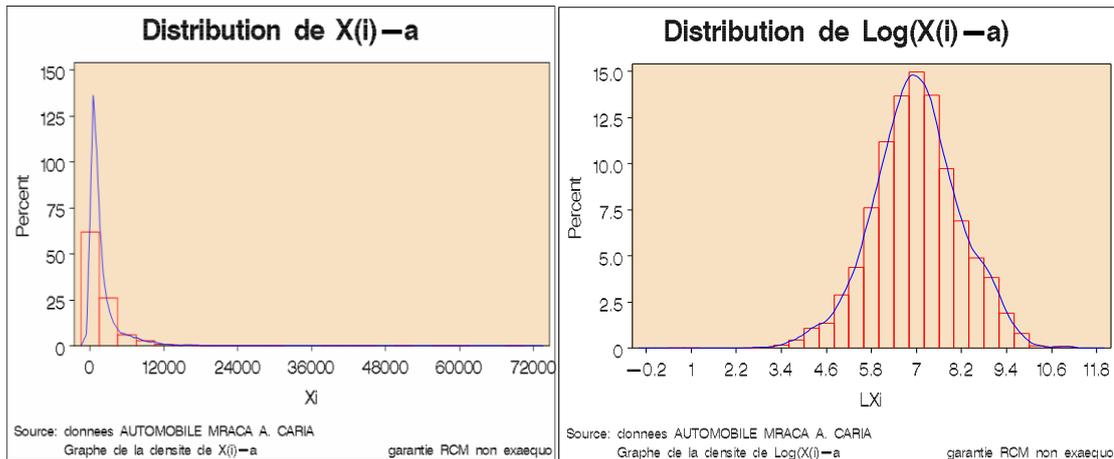
- ⊗ RCM(Responsabilité Civile Matériel),
- ⊗ RCC(Responsabilité Civile Corporelle),
- ⊗ DOMM(Dommage tous accidents),
- ⊗ BDG(Bris De Glaces),
- ⊗ VOL(Vol ou tentative de Vol),
- ⊗ INC(Incendie),
- ⊗ AUTRES(Autres garanties).

| Garantie RCM | | | | | | | | | | | | | | | |
|-------------------------|-------------|------------|-------------|-------------|-------------|-------------|--------------------|--------------------|------------|---------|---------|-----------|---------|-----------|------------|
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile inférieur | Quartile supérieur | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 28 586 | 143 | 442 | 564 | 1 204 | 2 266 | 7 797 | 884 | 1 204 | 320 | 1 098 | 1 250 | 306 714 | -3 049 | 309 763 | 2 426 |
| Garantie RCC | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile inférieur | Quartile supérieur | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 2 660 | 16 | 102 | 172 | 19 497 | 50 000 | 183 417 | 407 | 6 098 | 5 691 | 1 515 | 12 574 | 2 135 000 | -1 358 | 2 136 358 | 72 804 |
| Garantie DREC | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile inférieur | Quartile supérieur | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 10 419 | -1 019 | -765 | -620 | 1 034 | 1 546 | 3 067 | -331 | 370 | 701 | 53 | 143 | 18 683 | -8 825 | 27 508 | 835 |
| Garantie Dommage | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile | Quartile | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 60 011 | -919 | -663 | -470 | 2 690 | 4 008 | 7 633 | 39 | 1 281 | 1 242 | 469 | 914 | 27 424 | -6 470 | 33 894 | 1 686 |
| Garantie BDG | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile | Quartile | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 54 399 | 25 | 56 | 73 | 391 | 459 | 637 | 109 | 303 | 194 | 217 | 224 | 2 607 | -974 | 3 581 | 136 |
| Garantie VOL | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile | Quartile | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 6 128 | 30 | 71 | 122 | 3 096 | 5 163 | 11 089 | 319 | 1 376 | 1 056 | 725 | 1 363 | 32 079 | -1 103 | 33 182 | 2 172 |
| Garantie INC | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile | Quartile | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 537 | 38 | 61 | 100 | 3 837 | 7 200 | 14 558 | 431 | 2 076 | 1 645 | 978 | 1 891 | 33 354 | 20 | 33 334 | 3 057 |
| Garantie AUTRE | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile | Quartile | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 970 | -137 | -137 | -128 | 173 | 418 | 6 098 | -102 | 126 | 228 | 2 | 175 | 10 160 | -267 | 10 427 | 938 |
| Global | | | | | | | | | | | | | | | |
| Nb | 1er centile | 5e centile | 10e centile | 90e centile | 95e centile | 99e centile | Quartile | Quartile | Intervalle | Médiane | Moyenne | Maximum | Minimum | Étendue | Écart-type |
| 135 851 | -802 | -418 | 5 | 2 249 | 3 671 | 9 493 | 110 | 1 068 | 958 | 317 | 1 083 | 2 138 538 | -3 525 | 2 142 063 | 10 640 |

4. Analyse de la distribution du coût moyen d'un sinistre suivant la garantie :

Nous présenterons uniquement la distribution du montant des sinistres de la garantie RCM sans la présence d'exaquo, l'analyse de la distribution concernant les autres garanties est présentée en Annexe.

a. X(i) Montant des sinistres RCM (Responsabilité Civile Matériel) non exaquo et a la franchise égale à 0

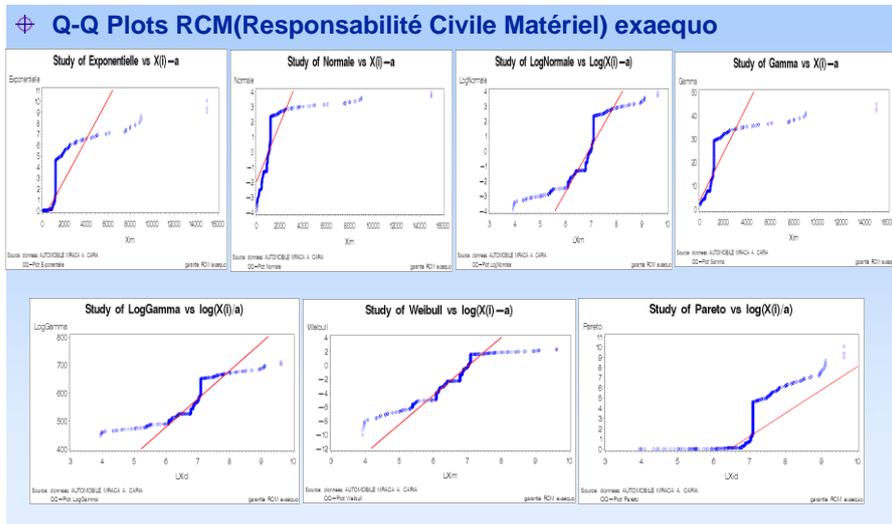
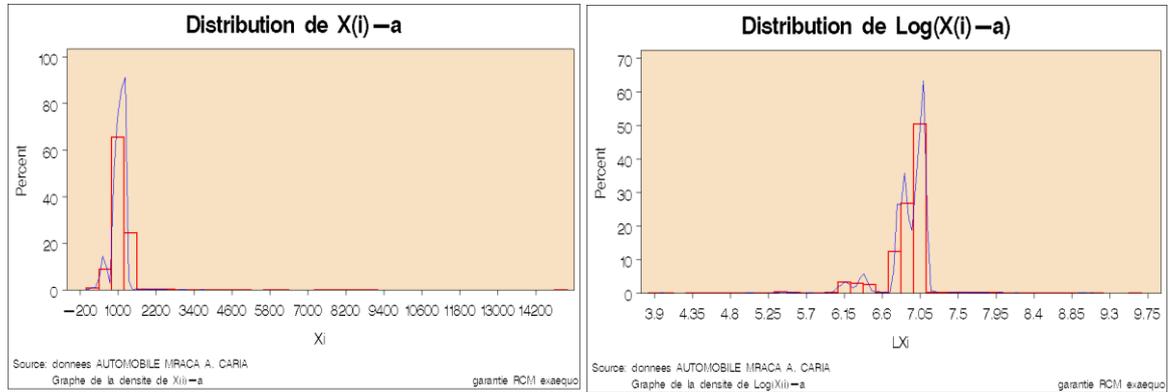


↓ Résultats Statistiques RCM(Responsabilité Civile Matériel) non exaquo

| Loi | Empirique | Exponentielle | Normales | LogNormales | Gamma | LogGamma | Weibull | Pareto |
|--|------------|---------------|---|---------------|-----------|------------|-----------|---------------------|
| Paramètre 1 | θ | λ | μ | μ | ν | α | α | α |
| Estimation du paramètre 1 | 0,0004 | 2,281 9130 | 7,027 | 0,8994 | 33,8876 | 0,8758 | 0,1412 | |
| Paramètre 2 | σ | σ | σ | σ | β | β | β | β |
| Estimation du paramètre 2 | 0,0155 | 3 980,7973 | 1,1514 | 0,0004 | 4,7846 | 2103,0529 | 0,0155 | |
| rsque relatif paramètre1 | | 0,0273 | 0,0025 | 0,0191 | 0,0218 | 0,0128 | | |
| rsque relatif paramètre2 | | 87,0445 | 0,0219 | 0,0251 | 0,0219 | 0,0187 | | |
| rsque relatif_moy | | 0,0270 | 0,0229 | 0,0163 | 0,000007 | 0,0177 | | |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6448 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 4 183 | 4 183 | 4 183 | 4 183 | 4 183 | 4 183 | 4 183 | 4 183 |
| moy | 2 283 | 2 283 | 2 283 | 2 312 | 2 283 | 2 324 | 2 248 | infini |
| var | 15 980 536 | 5 207 127 | 15 946 747 | 14 772 194 | 5 789 433 | 84 600 000 | 6 630 041 | infini |
| sigma | 3 981 | 2 282 | 3 981 | 3 843 | 2 406 | 9 198 | 2 573 | infini |
| Minimum | 2 | | | | | | | |
| Maximum | 133 266 | | | | | | | |
| moy_borne_inf | | 2 214 | 2 162 | 2 208 | 2 210 | 2 824 | 2 170 | - |
| moy_borne_sup | | 2 362 | 2 404 | 2 416 | 2 366 | 2 824 | 2 326 | - |
| CV | 2 | 1 | 2 | 0 | 1 | 0 | 1 | 0,17 |
| G1 | 332 | 2 | 0 | 10 | 2 | 0 | 0 | infini |
| G2 | 13 | 6 | 0 | 344 | 7 | 0 | 0 | 0,00 |
| bp | 0,62 | 0,52 | 1,00 | 0,62 | 0,54 | 1,00 | 0,57 | 17 849,89 |
| G | | 0,50 | 2,00 | 0,85 | 0,52 | 0,10 | 0,97 | -1,39 |
| Q25 | 595 | 650 | 403 | 580 | 586 | 507 | 507 | 0 |
| mediane | 1 306 | 1 292 | 2 292 | 1 191 | 1 413 | 1 111 | 1 284 | 136 |
| Q75 | 2 390 | 3 163 | 4 967 | 2 590 | 3 163 | 2 594 | 3 054 | 18 375 |
| Q90 | 5 096 | 5 254 | 7 384 | 5 210 | 5 393 | 5 883 | 5 450 | 12 098 047 |
| Q95 | 9 672 | 9 636 | 9 630 | 7 916 | 7 088 | 9 854 | 7 361 | 1 639 528 051 |
| Q99 | 15 634 | 10 669 | 11 543 | 17 348 | 11 089 | 27 339 | 12 027 | 146 352 740 000 000 |
| Dn | | 0,12 | 0,28 | 0,03 | 0,11 | 0,05 | 0,08 | 0,47 |
| W2n | | 19,26 | 111,79 | 0,98 | 15,04 | 3,26 | 10,44 | 291,55 |
| A2n | | 96,94 | 569,67 | 6,07 | 80,72 | 19,66 | 65,47 | 1 365,28 |
| P_KS | | 0,03 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 7,55 | 18,34 | 1,73 | 5,86 | 3,22 | 5,37 | 30,45 |
| W2n_T | | 19,26 | 112,66 | 0,98 | 15,05 | 3,26 | 10,46 | 291,62 |
| A2n_T | | 96,85 | 569,67 | 6,07 | 80,72 | 19,66 | 65,67 | 1 365,28 |
| P value KS | | P<=0,010 | P<=0,010 | 0,001<P<=0,01 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value_CVM | | P<=0,005 | P<=0,005 | 0,001<P<=0,01 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0,00000004 | 0 | 0,00000 | 0 | 0 |
| kh2_theo | | 380 | 7 382 | 48 | 409 | 110 | 403 | 15 256 |
| la distribution vérifie ou pas, la règle de Pareto ? | NON | NON | NON | NON | OUI | NON | OUI | |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| Prob_charge_negative | | | 0,283159411 | | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

b. X(i) Montant des sinistres RCM (Responsabilité Civile Matériel) exaequo et a la franchise égale à 0

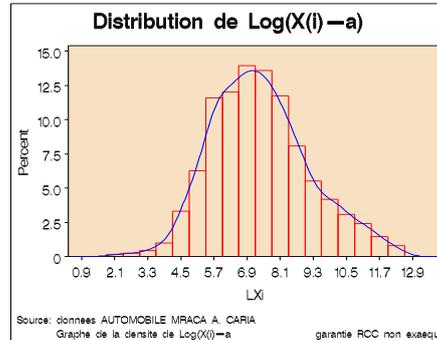
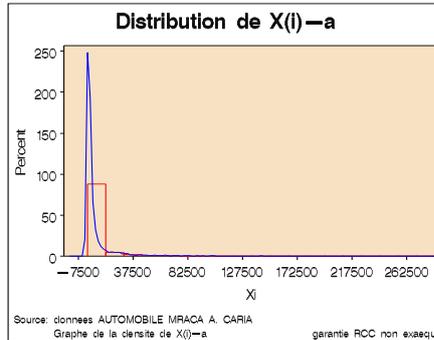


↓ Résultats statistiques RCM(Responsabilité Civile Matériel) exaequo

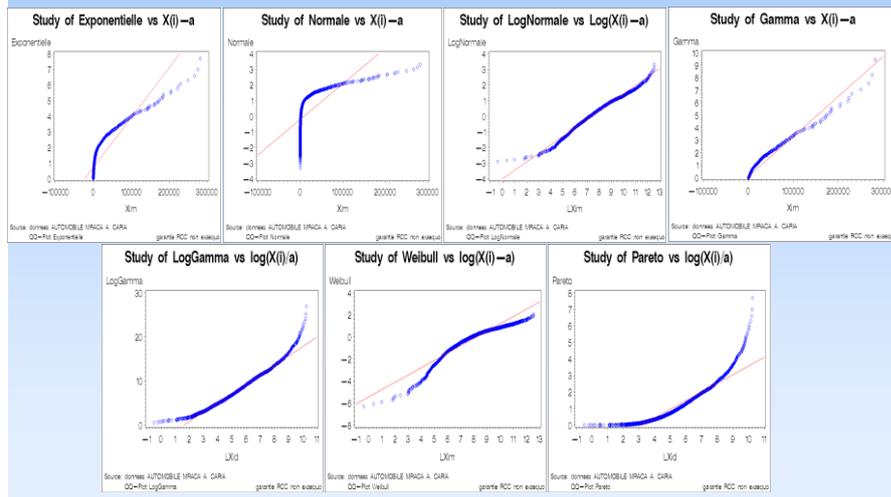
| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|--|-----------|--------------------|---|--------------------|--------------------|--------------------|--------------------|---------------------|
| Paramètre 1 | | 0 | μ | μ | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0009 | 1 170,3940 | 7 0195 | 11,1475 | 205,7287 | 4,2523 | 0,1425 |
| Paramètre 2 | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | 310,6574 | 0,4022 | 0,0095 | 29,3082 | 1751,6853 | |
| risque relatif paramètre1 | | 0,0076 | 0,0020 | 0,0004 | 0,0106 | 0,0108 | 0,0063 | 0,0076 |
| risque relatif paramètre2 | | 0,0076 | 3,3520 | 0,0108 | 0,0108 | 0,0108 | 0,0019 | |
| risque relatif moy | | 0,0076 | 0,0020 | 0,0032 | 0,0023 | 0,000003 | 0,0020 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 |
| moy | 1 171 | 1 171 | 1 171 | 1 213 | 1 171 | 1 265 | 1 594 | infini |
| var | 96 514 | 1 369 822 | 96 508 | 258 075 | 122 982 | 467 975 | 178 954 | infini |
| sigma | 311 | 1 170 | 311 | 508 | 351 | 684 | 423 | infini |
| Minimum | 0 | | | | | | | |
| Maximum | 10 000 | | | | | | | |
| moy_borne_inf | | 1 154 | 1 167 | 1 206 | 1 166 | 1 265 | 1 588 | - |
| moy_borne_sup | | 1 189 | 1 176 | 1 221 | 1 177 | 1 265 | 1 601 | - |
| CV | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0,07 |
| G1 | 302 | 2 | 0 | 1 | 1 | 0 | 0 | infini |
| G2 | 11 | 6 | 0 | 3 | 1 | 0 | 0 | 0,00 |
| Dp | 0,23 | 0,52 | 1,00 | 0,33 | 0,29 | 1,00 | 0,27 | 16 123,15 |
| G | | 0,50 | 2,00 | 0,47 | 0,17 | | 0,75 | -1,49 |
| Q25 | 1 204 | 337 | 961 | 853 | 919 | 799 | 1 307 | 8 |
| mediane | 1 217 | 811 | 1 170 | 1 118 | 1 136 | 1 106 | 1 607 | 130 |
| Q75 | 1 237 | 1 623 | 1 380 | 1 467 | 1 384 | 1 546 | 1 892 | 16 833 |
| Q90 | 1 269 | 2 695 | 1 569 | 1 872 | 1 636 | 2 108 | 2 131 | 10 459 237 |
| Q95 | 1 270 | 3 506 | 1 681 | 2 167 | 1 800 | 2 549 | 2 267 | 1 357 000 656 |
| Q99 | 1 270 | 5 390 | 1 893 | 2 850 | 2 136 | 3 670 | 2 509 | 109 395 640 000 000 |
| Dn | | 0,53 | 0,37 | 0,43 | 0,41 | 0,43 | 0,77 | 0,58 |
| W2n | 1 318,57 | 773,16 | 890,42 | 890,42 | 846,14 | 918,03 | 2 791,40 | 1 636,54 |
| A2n | 6 116,06 | 3 635,43 | 4 143,50 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,31 | 7 482,10 |
| P_KS | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | 69,29 | 48,39 | 55,89 | 54,17 | 54,17 | 56,59 | 100,57 | 76,42 |
| W2n_T | 1 318,58 | 776,10 | 890,48 | 890,48 | 846,19 | 918,08 | 2 795,66 | 1 636,64 |
| A2n_T | 6 116,27 | 3 635,59 | 4 143,50 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,55 | 7 482,10 |
| P_value_KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0,00000000 | 0 | 0,000000 | 0 | 0 |
| kh2_theo | | 670 647 | 147 229 | 244 002 | 176 438 | 302 037 | 330 630 | 4 857 713 |
| la distribution vérifie ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | OUI | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 8,13973E-05 | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous estimation au niveau des queues de distribution.

c. X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) non exæquo et a la franchise égale à 0



Q-Q Plots RCC(Responsabilité Civile Corporelle) non exæquo

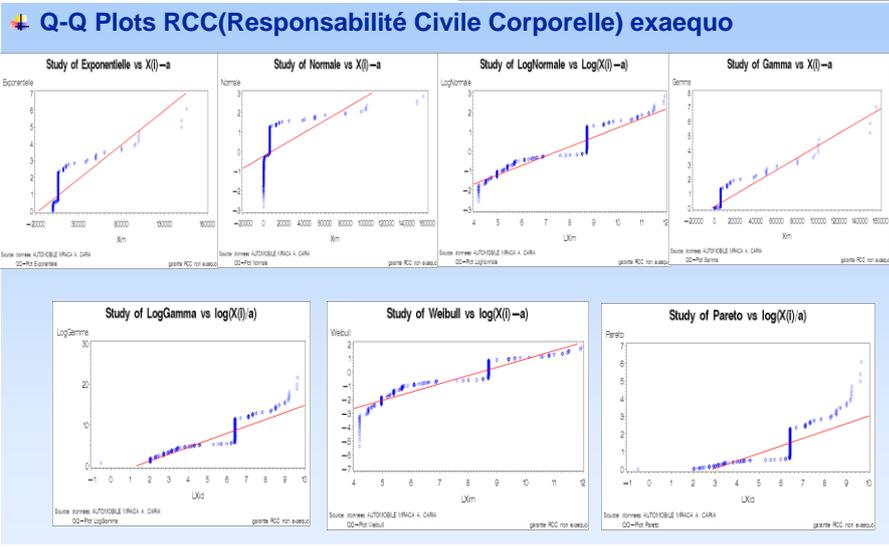
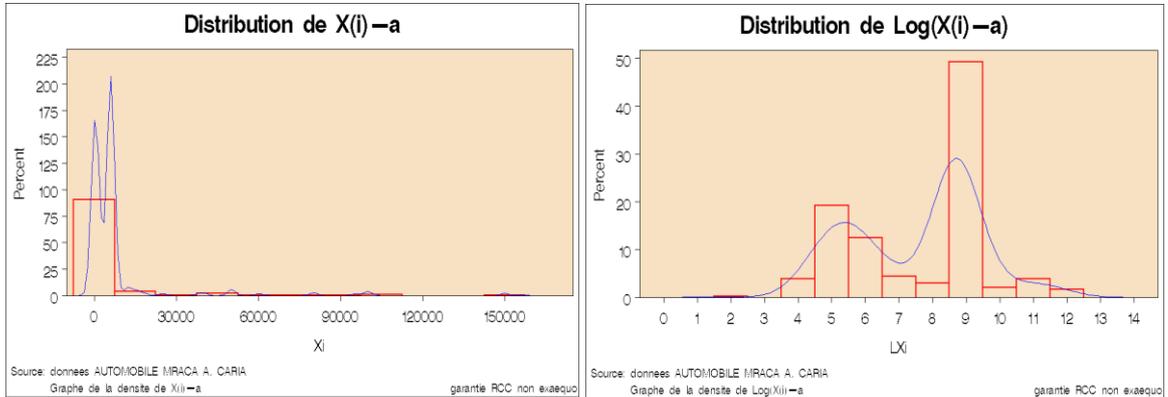


↓ Résultats statistiques RCC(Responsabilité Civile Corporelle) non exæquo

| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|--|---------------|---------------|---|---------------|-------------|---------------------------|-------------|---------------------|
| Paramètre 1 | | 0 | μ | μ | | | | |
| Estimation du paramètre 1 | | 0,0001 | 11 439,4727 | 7,2934 | 0,3463 | 15,9787 | 0,5008 | 0,1351 |
| Paramètre 2 | | | | | | | | |
| Estimation du paramètre 2 | | | 37 973,7950 | 1,8421 | 0,00003 | 2,1595 | 4271,0173 | |
| risque_relatif_parametre1 | | 0,0218 | 0,0723 | 0,0064 | 0,0245 | 0,0395 | 0,0180 | 0,0218 |
| risque_relatif_parametre2 | | | 1 169 3936 | 0,0308 | 0,0444 | 0,0310 | 0,0461 | |
| risque_relatif_moy | | 0,0218 | 0,0723 | 0,0669 | 0,0370 | 0,030032 | 0,0470 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 |
| moy | 11 440 | 11 440 | 11 440 | 8 921 | 11 440 | 20 687 | 8 617 | infini |
| var | 1 442 693 169 | 130 861 535 | 1 442 009 104 | 2 288 968 905 | 377 911 303 | 1 210 000 000 000 000 000 | 360 908 388 | infini |
| sigma | 37 983 | 11 439 | 37 974 | 47 843 | 19 440 | 1 100 000 000 | 18 998 | infini |
| Minimum | 4 | | | | | | | |
| Maximum | 400 000 | | | | | | | |
| moy_borne_inf | | 10 952 | 9 820 | 7 770 | 10 611 | 20 687 | 7 732 | 0,95 |
| moy_borne_sup | | 11 929 | 13 061 | 10 073 | 12 270 | 20 687 | 9 302 | 0,84 |
| CV | 3 | 1 | 3 | 0 | 2 | 0 | 2 | 0,25 |
| G1 KURTOSIS | 42 | 2 | 0 | 170 | 3 | 1 | 0 | infini |
| G2 SKEWNESS | 6 | 6 | 0 | 840 446 | 17 | 0 | 0 | 0,00 |
| Dp | 0,89 | 0,52 | 1,00 | 0,81 | 0,73 | 1,00 | 0,78 | 29 715,62 |
| G | | 0,50 | 2,00 | 1,24 | 0,71 | 0,14 | 1,00 | -1,37 |
| Q25 | 406 | 3 291 | -14 173 | 472 | 437 | 438 | 365 | 8 |
| mediane | 1 359 | 7 929 | 11 439 | 1 635 | 3 458 | 1 402 | 2 055 | 169 |
| Q75 | 5 039 | 15 859 | 37 052 | 5 684 | 13 623 | 4 167 | 8 199 | 26 502 |
| Q90 | 19 392 | 26 340 | 60 105 | 17 329 | 33 089 | 18 932 | 22 582 | 25 083 872 |
| Q95 | 59 095 | 34 270 | 73 901 | 33 840 | 49 923 | 43 647 | 38 190 | 4 234 613 435 |
| Q99 | 236 697 | 52 681 | 99 760 | 118 751 | 92 935 | 235 936 | 90 118 | 629 200 610 000 000 |
| Dn | | 0,43 | 0,38 | 0,04 | 0,19 | 0,02 | 0,11 | 0,44 |
| Wzn | | 178 111 | 110 895 | 1 401 | 30 222 | 0,14 | 9 17 | 116 63 |
| Azn | | 1 153 62 | 530 87 | 10 21 | 154 93 | 1,65 | 60 42 | 557 43 |
| P_KS | | 0,04 | 0,02 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 19 85 | 17 53 | 2,03 | 9 85 | 0 92 | 5 20 | 20 05 |
| Wzn_T | | 178 12 | 112 06 | 1 40 | 30 23 | 0,14 | 9 21 | 116 69 |
| Azn_T | | 1 153 95 | 531 06 | 10 21 | 154 93 | 1,65 | 60 88 | 557 43 |
| P_value_KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,25 | P<=0,010 | P<=0,001 |
| P_value_CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,25 | P<=0,010 | P<=0,001 |
| P_value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | 0,10-P<=0,150 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 6,91E-13 | 0 | 0,00295 | 0 | 0 |
| kh2_theo | | 4 333 | 48 774 | 72 | 921 | 22 | 403 | 3 928 |
| la distribution vérifie ou pas, la règle de Pareto ? | OUI | NON | OUI | OUI | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 0,391603229 | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

d. X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) exaequo et a la franchise égale à 0

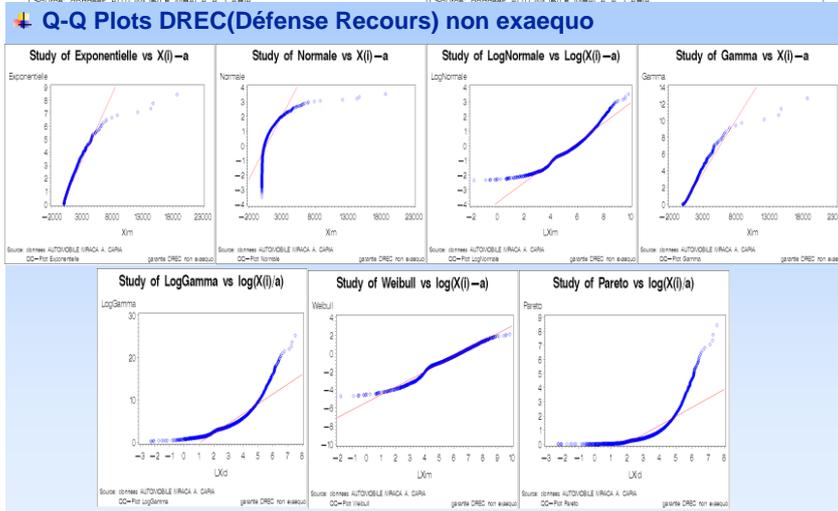
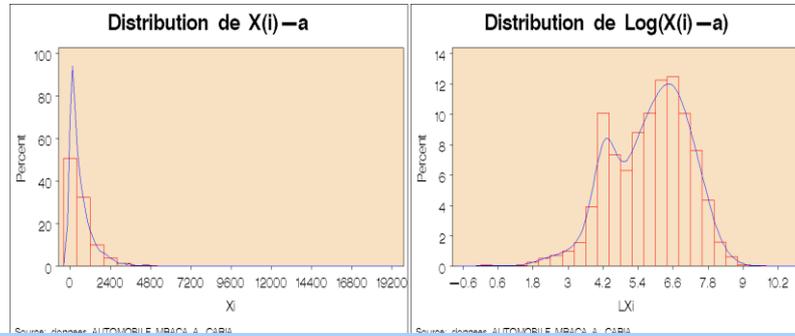


| Paramètre 1 | Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----|-----------|---------------|------------|------------|----------|----------|-----------|---------------------|
| Estimation du paramètre 1 | | | 0,0009 | 1 170,3940 | 7,0195 | 11,1475 | 205,7287 | 4,2523 | 0,1425 |
| Paramètre 2 | | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | | 310,6574 | 0,4022 | 0,0095 | 29,3082 | 1751,6853 | |
| risque relatif paramètre1 | | | 0,0076 | 0,0020 | 0,0004 | 0,0106 | 0,0108 | 0,0063 | 0,0076 |
| risque relatif paramètre2 | | | | 3,3520 | 0,0108 | 0,0109 | 0,0108 | 0,0019 | |
| risque relatif moy | | | 0,0076 | 0,0020 | 0,0032 | 0,0023 | 0,000003 | 0,0020 | infini |
| Constante K1 utile pour Weibull | | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | | 0,6449 | |
| a | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 |
| moy | | 1 171 | 1 171 | 1 171 | 1 213 | 1 171 | 1 285 | 1 594 | infini |
| var | | 96 514 | 1 369 822 | 96 508 | 258 075 | 122 882 | 467 975 | 178 854 | infini |
| sigma | | 311 | 1 170 | 311 | 508 | 351 | 684 | 423 | infini |
| Minimum | | 6 | | | | | | | |
| Maximum | | 10 000 | | | | | | | |
| moy_borne_inf | | | 1 154 | 1 167 | 1 206 | 1 166 | 1 265 | 1 588 | - |
| moy_borne_sup | | | 1 189 | 1 176 | 1 221 | 1 177 | 1 265 | 1 601 | |
| CV | | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0,07 |
| G1 | | 302 | 2 | 0 | 1 | 1 | 0 | 0 | infini |
| G2 | | 11 | 6 | 0 | 3 | 1 | 0 | 0 | 0,00 |
| Dp | | 0,23 | 0,52 | 1,00 | 0,33 | 0,29 | 1,00 | 0,27 | 16 123,15 |
| G | | | 0,50 | 2,00 | 0,47 | 0,17 | 0,75 | 0,75 | -1,40 |
| Q25 | | 1 204 | 337 | 961 | 853 | 919 | 799 | 1 307 | 8 |
| mediane | | 1 217 | 811 | 1 170 | 1 118 | 1 136 | 1 106 | 1 607 | 130 |
| Q75 | | 1 237 | 1 623 | 1 380 | 1 467 | 1 384 | 1 545 | 1 892 | 16 833 |
| Q90 | | 1 269 | 2 695 | 1 569 | 1 872 | 1 636 | 2 108 | 2 131 | 10 459 237 |
| Q95 | | 1 270 | 3 906 | 1 681 | 2 167 | 1 800 | 2 549 | 2 267 | 1 357 000 656 |
| Q99 | | 1 270 | 5 390 | 1 883 | 2 850 | 2 136 | 3 670 | 2 509 | 109 395 640 000 000 |
| Dn | | | 0,53 | 0,37 | 0,43 | 0,41 | 0,43 | 0,77 | 0,58 |
| W2n | | | 1 318,57 | 773,16 | 890,42 | 846,14 | 918,03 | 2 791,40 | 1 636,54 |
| A2n | | | 6 116,06 | 3 635,43 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,31 | 7 482,10 |
| P_KS | | | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | | 69,29 | 48,39 | 55,89 | 54,17 | 56,59 | 100,57 | 76,42 |
| W2n_T | | | 1 318,58 | 776,10 | 890,48 | 846,19 | 918,08 | 2 795,66 | 1 636,64 |
| A2n_T | | | 6 116,27 | 3 635,59 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,65 | 7 482,10 |
| P value KS | | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| kh22 theo | | | 670 647 | 147 229 | 244 002 | 176 438 | 302 037 | 330 630 | 4 857 713 |
| la distribution vérifiée ou pas, la règle de Pareto ? | | NON | NON | OUI | NON | NON | OUI | OUI | OUI |

non adéquation du modèle Normal aux données
Prob_charge_negative 8,13973E-05

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

e. X(i) Montant des sinistres DREC (Défense Recours) non exæquo et a la franchise égale à 0

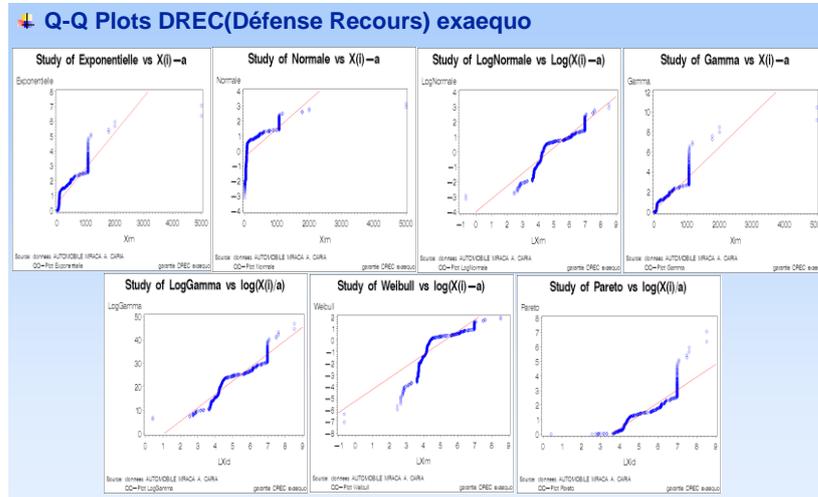
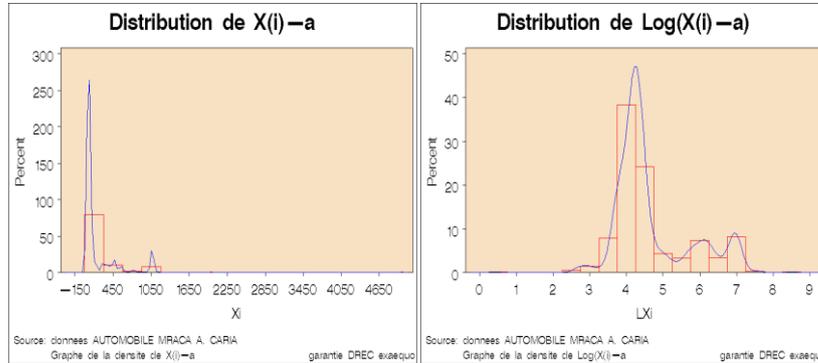


↓ Résultats statistiques DREC(Défense Recours) non exæquo

| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|--|-----------|---------------|---|-------------|-----------|----------------|-------------|-----------------|
| Paramètre 1 | | θ | μ | μ | λ | ν | τ | α |
| Estimation du paramètre 1 | | 0,0015 | 674,3866 | 5,7105 | 0,7457 | 13,7041 | 0,8037 | 0,1751 |
| Paramètre 2 | | | σ | σ | β | β | α | α |
| Estimation du paramètre 2 | | | 1 101,4028 | 1,3857 | 0,0011 | 2,3998 | 587,8711 | |
| risque_relatif_parametre1 | | 0,0142 | 0,0232 | 0,0034 | 0,0172 | 0,0199 | 0,0117 | 0,0142 |
| risque_relatif_parametre2 | | | 22,1390 | 0,0201 | 0,0238 | 0,0202 | 0,0178 | infini |
| risque_relatif_moy | | 0,0142 | 0,0232 | 0,0275 | 0,0164 | 0,0000 | 0,422784335 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,644834067 | |
| Constante K2 utile pour Weibull | | | | | | | | |
| n | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| moy | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 |
| var | 675 | 675 | 675 | 790 | 675 | 1 616 | 665 | infini |
| sigma | 1 213 333 | 454 811 | 1 213 088 | 3 623 856 | 609 881 | 46 400 000 000 | 693 387 | infini |
| Minimum | 1 102 | 674 | 1 101 | 1 904 | 781 | 215 407 | 833 | infini |
| Maximum | 0 | | | | | | | |
| moy_borne_inf | 34 587 | 657 | 645 | 747 | 654 | 1 616 | 642 | 1 |
| moy_borne_sup | | 694 | 706 | 833 | 697 | 1 616 | 688 | 1 |
| CV | 2 | 1 | 2 | 0 | 0 | 0 | 1 | 0 |
| G1 KURTOSIS | 239 | 2 | 0 | 21 | 2 | 1 | 0 | infini |
| G2 SKENNESS | 10 | 6 | 0 | 2 596 | 6 | 0 | 0 | 0 |
| Dp | 0,62 | 0,52 | 1,00 | 0,71 | 0,57 | 1,00 | 0,60 | 1 961,13 |
| G | | 0,50 | 2,00 | 0,98 | 0,56 | 0,15 | 0,98 | -1,54 |
| Q25 | 102 | 194 | -68 | 119 | 137 | 100 | 125 | 5 |
| mediane | 332 | 467 | 674 | 302 | 407 | 263 | 373 | 52 |
| Q75 | 674 | 936 | 1 417 | 769 | 930 | 792 | 683 | 2 742 |
| D90 | 1 652 | 1 553 | 2 086 | 1 784 | 1 868 | 2 333 | 1 660 | 513 467 |
| D95 | 2 325 | 2 020 | 2 486 | 2 951 | 2 244 | 4 731 | 2 302 | 26 887 349 |
| D99 | 4 592 | 3 106 | 3 237 | 7 587 | 3 612 | 19 866 | 3 951 | 263 648 247 387 |
| Dn | | 0,14 | 0,27 | 0,05 | 0,07 | 0,06 | 0,06 | 0,43 |
| Wzn | | 29,60 | 95,45 | 3,07 | 8,08 | 7,45 | 4,15 | 273,95 |
| Azn | | 159,54 | 510,25 | 19,91 | 47,01 | 63,99 | 29,71 | 1 311,79 |
| P_KS | | 0,02 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 9,62 | 19,01 | 3,38 | 4,30 | 4,52 | 4,40 | 30,31 |
| Wzn_T | | 29,60 | 95,13 | 3,07 | 8,08 | 7,46 | 4,16 | 274,01 |
| Azn_T | | 159,56 | 510,33 | 19,91 | 47,01 | 53,99 | 29,79 | 1 311,79 |
| P_value_KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| hiz_theo | | 1 011 | 11 736 | 398 | 676 | 676 | 559 | 10 127 |
| la distribution vérifie ou pas, la règle de Pareto 2 | NON | NON | OUI | NON | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 0,269867134 | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

f. X(i) Montant des sinistres DREC (Défense Recours) exæquo et a la franchise égale à 0

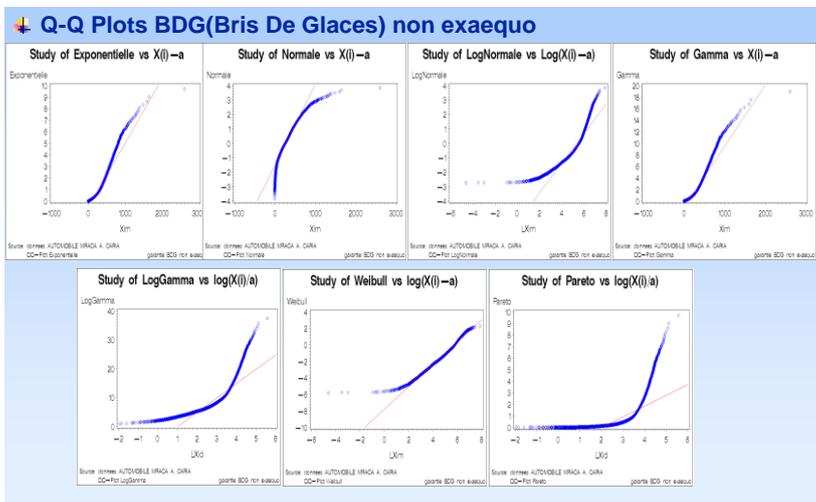
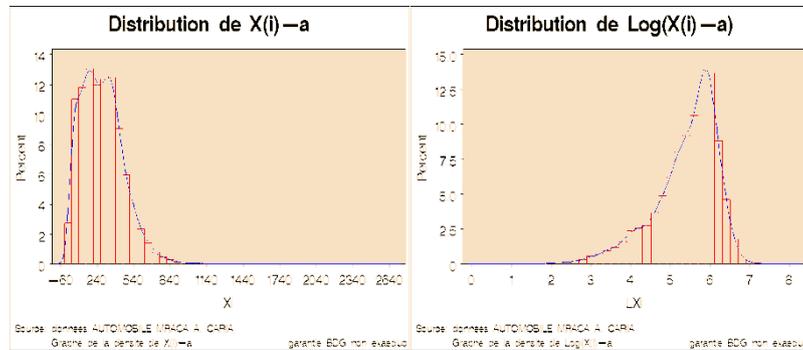


| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|--|--|--|--|--|--|--|--|--|
| Paramètre 1 | | θ | μ | ν | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0015 | 674,3966 | 5,7105 | 0,7457 | 13,7041 | 0,8037 | 0,1751 |
| Paramètre 2 | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | 1 101,4028 | 1,3857 | 0,00111 | 2,3988 | 587,8711 | |
| risque relatif paramètre1 | | 0,0142 | 0,0232 | 0,0034 | 0,0172 | 0,0199 | 0,0117 | 0,0142 |
| risque relatif paramètre2 | | | 22 1380 | 0,0201 | 0,0238 | 0,0202 | 0,0187 | |
| risque relatif_moy | | 0,0142 | 0,0232 | 0,0275 | 0,0164 | 0,0000 | 0,0178 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 |
| moy | 675 | 675 | 675 | 790 | 1 616 | 665 | infini | |
| var | 1 213 333 | 454 811 | 1 213 088 | 3 623 856 | 609 881 | 46 400 000 000 | 693 387 | infini |
| sigma | 1 102 | 674 | 1 101 | 1 904 | 781 | 215 407 | 833 | infini |
| Minimum | 0 | | | | | | | |
| Maximum | 34 587 | | | | | | | |
| moy_borne_inf | | 657 | 645 | 747 | 654 | 1 616 | 642 | 1 |
| moy_borne_sup | | 694 | 706 | 833 | 697 | 1 616 | 688 | 0 |
| CV | 2 | 1 | 2 | 0 | 1 | 0 | 1 | 0 |
| G1 KURTOSIS | 239 | 2 | 0 | 21 | 2 | 1 | 0 | infini |
| G2 SKEWNESS | 10 | 6 | 0 | 2 936 | 8 | 0 | 0 | 0 |
| Dp | 0,62 | 0,52 | 1,00 | 0,71 | 0,57 | 1,00 | 0,60 | 1 961,13 |
| G | | 0,50 | 2,00 | 0,98 | 0,56 | 0,15 | 0,98 | -1,54 |
| Q25 | 102 | 194 | -68 | 119 | 137 | 100 | 125 | 5 |
| mediane | 332 | 467 | 674 | 302 | 407 | 263 | 373 | 52 |
| Q75 | 874 | 935 | 1 417 | 769 | 930 | 782 | 883 | 2 742 |
| Q90 | 1 652 | 1 553 | 2 086 | 1 784 | 1 668 | 2 333 | 1 660 | 513 467 |
| Q95 | 2 325 | 2 020 | 2 486 | 2 951 | 2 244 | 4 731 | 2 302 | 26 887 349 |
| Q99 | 4 592 | 3 106 | 3 237 | 7 587 | 3 612 | 19 866 | 3 931 | 263 648 247 387 |
| Dn | | 0,14 | 0,27 | 0,05 | 0,07 | 0,06 | 0,06 | 0,43 |
| W2n | | 29,60 | 96,45 | 3,07 | 8,08 | 7,45 | 4,15 | 273,95 |
| A2n | | 159,54 | 510,25 | 19,91 | 47,01 | 53,99 | 29,79 | 1 311,79 |
| P_KS | | 0,02 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 9,62 | 19,01 | 3,38 | 4,90 | 4,52 | 4,40 | 30,31 |
| W2n_T | | 29,60 | 96,13 | 3,07 | 8,08 | 7,46 | 4,16 | 274,01 |
| A2n_T | | 159,56 | 510,33 | 19,91 | 47,01 | 53,99 | 29,79 | 1 311,79 |
| P_value_KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| khi2_theo | | 1 011 | 11 736 | 308 | 676 | 428 | 559 | 10 127 |
| la distribution vérifie ou pas, la règle de Pareto ? | la distribution ne vérifie pas, la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution vérifie la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution vérifie la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution vérifie la règle de Pareto |

non adéquation du modèle Normal aux données
Prob_charge_negative 0,27

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

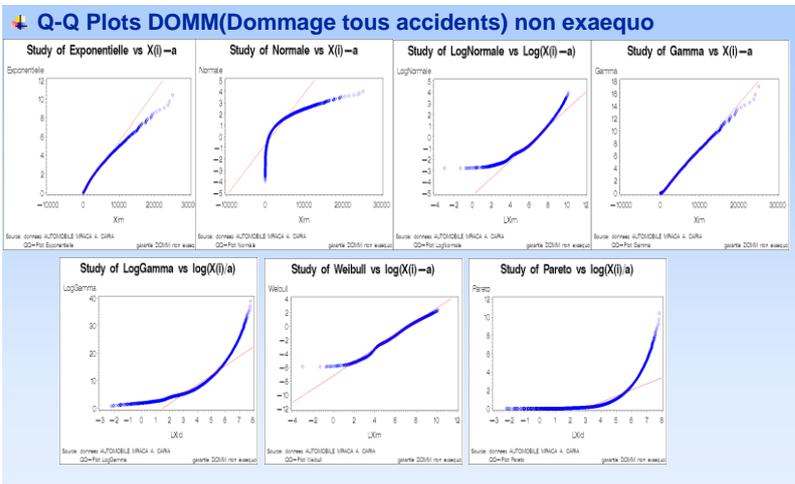
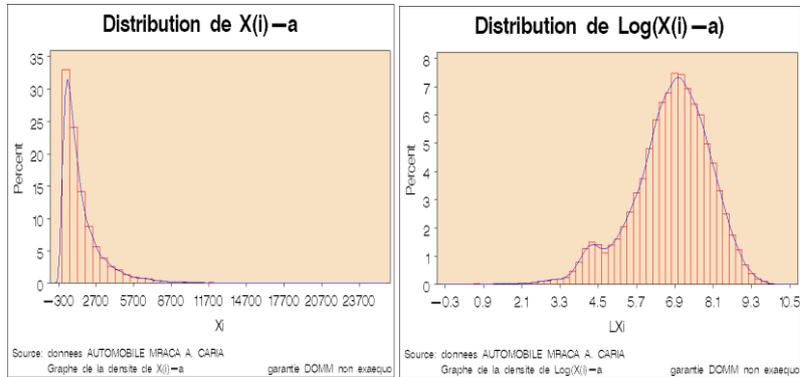
g. X(i) Montant des sinistres BDG (Bris De Glaces) non exæquo et a la franchise égale à 0



| ❑ Résultats statistiques BDG(Bris De Glaces) non exæquo | | | | | | | | |
|---|-----------|--------------------|---|--------------------|--------------------|--------------------|--------------------|--------------------|
| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
| Paramètre 1 | | θ | μ | μ | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0036 | 274,7454 | 5,4179 | 2,6840 | 56,1327 | 1,7845 | 0,1846 |
| Paramètre 2 | | | | | | | | |
| Estimation du paramètre 2 | | | 161,6321 | 0,6857 | 0,00977 | 10,3607 | 309,1898 | |
| risque_relatif_parametre1 | | 0,0049 | 0,0028 | 0,0006 | 0,0065 | 0,0069 | 0,0040 | 0,0049 |
| risque_relatif_parametre2 | | | 1,1059 | 0,0069 | 0,0071 | 0,0069 | 0,0029 | |
| risque_relatif_moy | | 0,0049 | 0,0028 | 0,0037 | 0,0030 | 0,000012 | 0,0028 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| α | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 |
| moy | 276 | 276 | 276 | 269 | 276 | 299 | 276 | infini |
| var | 25 603 | 75 485 | 25 603 | 48 798 | 28 124 | 80 399 | 25 402 | infini |
| sigma | 161 | 275 | 161 | 221 | 168 | 284 | 159 | infini |
| Minimum | 2 | | | | | | | |
| Maximum | 2 982 | | | | | | | |
| moy_borne_inf | | 273 | 274 | 264 | 274 | 299 | 275 | 0,77 |
| moy_borne_sup | | 278 | 277 | 269 | 277 | 299 | 276 | 0,77 |
| CV | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,13 |
| G1 KURTOSIS | 5 | 2 | 0 | 3 | 1 | 0 | 0 | infini |
| G2 SKEWNESS | 1 | 6 | 0 | 16 | 2 | 0 | 0 | 0,00 |
| Dp | 0,37 | 0,52 | 1,00 | 0,44 | 0,39 | 1,00 | 0,38 | 1 224,49 |
| G | | 0,50 | 2,00 | 0,61 | 0,33 | 0,08 | 0,88 | -1,59 |
| Q25 | 140 | 79 | 166 | 142 | 151 | 136 | 154 | 5 |
| mediane | 270 | 190 | 275 | 225 | 241 | 216 | 252 | 43 |
| Q75 | 365 | 381 | 363 | 358 | 362 | 360 | 371 | 1 828 |
| Q90 | 470 | 633 | 481 | 543 | 500 | 580 | 493 | 261 737 |
| Q95 | 554 | 823 | 539 | 696 | 596 | 790 | 572 | 11 189 352 |
| Q99 | 755 | 1 265 | 648 | 1 111 | 805 | 1 395 | 728 | 68 506 229 264 |
| Dn | | 0,19 | 0,06 | 0,11 | 0,07 | 0,12 | 0,05 | 0,50 |
| W2n | | 593,52 | 44,85 | 128,22 | 66,33 | 169,06 | 26,99 | 3 203,81 |
| A2n | | 3 284,00 | 380,35 | 663,42 | 285,29 | 691,54 | 159,73 | 14 923,50 |
| P_KS | | 0,01 | 0,03 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn.T | | 36,46 | 11,91 | 22,70 | 15,06 | 25,62 | 9,65 | 102,60 |
| W2n.T | | 593,52 | 44,96 | 128,22 | 66,33 | 169,07 | 27,02 | 3 203,88 |
| A2n.T | | 3 284,05 | 380,35 | 663,42 | 285,29 | 691,54 | 159,89 | 14 923,50 |
| P value KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| khiz theo | | 16 989 | 4 374 | 6 440 | 3 517 | 8 350 | 2 617 | 241 512 |
| la distribution vérifie ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | | 0,043023704 | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

h. X(i) Montant des sinistres DOMM(Dommage tous accidents) non exaequo et a la franchise égale à 0

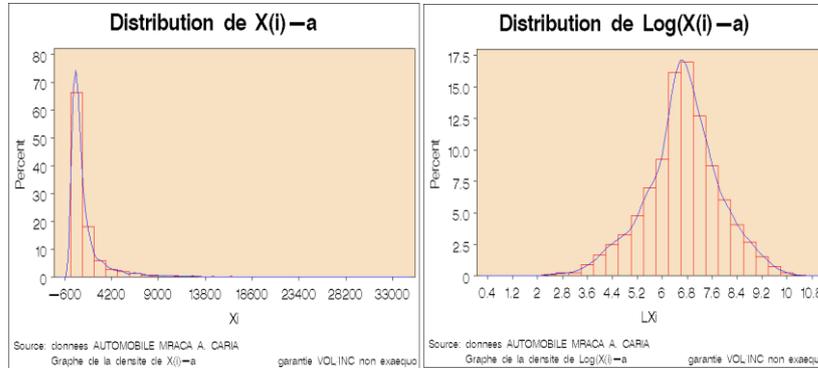


↓ Résultats statistiques DOMM(Dommage tous accidents) non exaequo

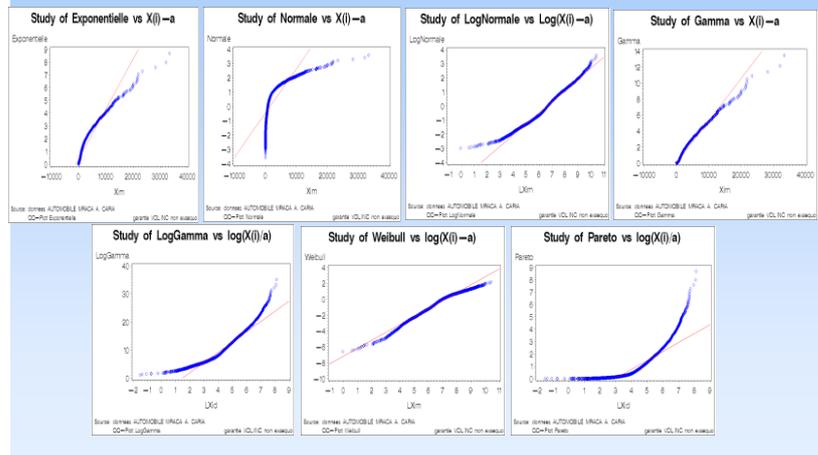
| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----------|--------------------|---|--------------------|--------------------|--------------------|--------------------|--------------------|
| Paramètre 1 | | 0 | μ | μ | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0006 | 1 646,8184 | 6 7826 | 0,9328 | 23 7268 | 0,9304 | 0,1497 |
| Paramètre 2 | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | 1 986,8324 | 1 2983 | 0,00057 | 3 4382 | 1638 6216 | |
| risque relatif paramètre1 | | 0,0060 | 0,0072 | 0,0011 | 0,0074 | 0,0084 | 0,0060 | 0,0069 |
| risque relatif paramètre2 | | | 16 8904 | 0,0085 | 0,0097 | 0,0085 | 0,0084 | |
| risque relatif moy | | 0,0060 | 0,0072 | 0,0101 | 0,0062 | 0,000003 | 0,0064 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| λ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 |
| moy | 1 646 | 1 646 | 1 646 | 1 646 | 1 646 | 1 646 | 1 646 | infini |
| var | 3 946 950 | 2 712 340 | 3 946 708 | 14 885 257 | 2 907 792 | 538 080 000 | 3 121 862 | infini |
| sigma | 1 987 | 1 647 | 1 987 | 3 632 | 1 705 | 23 195 | 1 767 | infini |
| Minimum | | | | | | | | |
| Maximum | 32 000 | | | | | | | |
| moy_borne_inf | | 1 629 | 1 625 | 1 910 | 1 629 | 2 948 | 1 623 | 0,83 |
| moy_borne_sup | | 1 667 | 1 671 | 1 967 | 1 668 | 2 948 | 1 664 | 0,82 |
| CV | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,22 |
| G1 KURTOSIS | 17 | 2 | 0 | 14 | 2 | 0 | 0 | infini |
| G2 SKEWNESS | 3 | 6 | 0 | 860 | 6 | 0 | 0 | 0,00 |
| dp | 0,57 | 0,52 | 1,00 | 0,66 | 0,53 | 1,00 | 0,55 | 9 356,14 |
| σ | | 0,50 | 2 120 | 0,91 | 0,51 | 0,12 | 0,36 | -1,43 |
| Q25 | 461 | 474 | 307 | 378 | 440 | 330 | 416 | 7 |
| mediane | 1 013 | 1 142 | 1 647 | 882 | 1 109 | 802 | 1 071 | 103 |
| Q75 | 2 059 | 2 283 | 2 987 | 2 062 | 2 283 | 2 129 | 2 267 | 10 526 |
| Q90 | 3 828 | 3 792 | 4 193 | 4 426 | 3 899 | 5 525 | 3 899 | 4 796 950 |
| Q95 | 6 296 | 4 934 | 4 915 | 6 991 | 5 057 | 10 134 | 5 166 | 492 180 459 |
| Q99 | 9 883 | 7 584 | 6 269 | 16 481 | 7 856 | 34 026 | 8 201 | 23 009 772 000 000 |
| Dn | | 0,05 | 0,20 | 0,06 | 0,04 | 0,10 | 0,03 | 0,43 |
| W2h | | 21,12 | 389,89 | 33,31 | 14,75 | 86,54 | 10,80 | 1 822,38 |
| A2h | | 110,85 | 2 150,46 | 193,35 | 62,93 | 502,30 | 67,65 | 8 655,87 |
| P_KS | | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 8,02 | 33,99 | 10,31 | 7,02 | 16,33 | 5,72 | 72,89 |
| W2h_T | | 21,12 | 391,06 | 33,31 | 14,75 | 86,54 | 10,81 | 1 822,44 |
| A2h_T | | 110,86 | 2 150,52 | 193,35 | 62,93 | 502,30 | 67,73 | 8 655,87 |
| P value KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| kh2_theo | | 390 | 22 523 | 1 242 | 394 | 2 418 | 407 | 86 196 |
| la distribution vérifiée ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 0,203409796 | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

i. X(i) Montant des sinistres VOL/INC(Incendie) non exæquo et a la franchise égale à 0



Q-Q Plots VOL / INC(Incendie) non exæquo



| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----------|---------------|----------|------------|----------|----------|----------|-----------------|
| Paramètre 1 | | θ | μ | μ | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0028 | 355,4086 | 5,7051 | 3,1344 | 70,9172 | 2,0277 | 0,1753 |
| Paramètre 2 | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | 183,4918 | 0,6435 | 0,00882 | 12,4305 | 401,1241 | |
| risque relatif paramètre1 | | 0,0136 | 0,0070 | 0,0015 | 0,0184 | 0,0192 | 0,0113 | 0,0136 |
| risque relatif paramètre2 | | | 3,5395 | 0,0193 | 0,0199 | 0,0193 | 0,0071 | |
| risque relatif moy | | 0,0136 | 0,0070 | 0,0096 | 0,0077 | 0,0000 | 0,0071 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 5 375 | 5 375 | 5 375 | 5 375 | 5 375 | 5 375 | 5 375 | 5 375 |
| moy | 357 | 356 | 356 | 371 | 356 | 384 | 356 | infini |
| var | 33 675 | 126 315 | 33 669 | 70 050 | 40 299 | 106 188 | 33 669 | infini |
| sigma | 184 | 355 | 183 | 265 | 201 | 326 | 183 | infini |
| Minimum | 1 | | | | | | | |
| Maximum | 1 609 | | | | | | | |
| moy borne inf | | 347 | 352 | 364 | 351 | 384 | 351 | 1 |
| moy borne sup | | 366 | 361 | 377 | 362 | 384 | 361 | 1 |
| CV | 0 | 1 | 1 | 3 | 1 | 0 | 1 | 0 |
| G1 KURTOSIS | 1 | 2 | 0 | 0 | 1 | 0 | 0 | infini |
| G2 SKEWNESS | 1 | 6 | 0 | 13 | 2 | 0 | 0 | 0 |
| Dp | 0,35 | 0,52 | 1,00 | 0,42 | 0,38 | 1,00 | 0,36 | 1944,18 |
| G | | 0,50 | 2,00 | 0,59 | 0,31 | 0,07 | 0,86 | -1,54 |
| Q25 | 229 | 102 | 232 | 195 | 208 | 188 | 217 | 5 |
| mediane | 346 | 246 | 355 | 300 | 318 | 292 | 335 | 52 |
| Q75 | 460 | 493 | 479 | 464 | 463 | 467 | 471 | 2 722 |
| Q90 | 597 | 818 | 591 | 685 | 625 | 727 | 605 | 507 128 |
| Q95 | 676 | 1 065 | 657 | 866 | 737 | 957 | 689 | 26 456 297 |
| Q99 | 856 | 1 637 | 782 | 1 342 | 979 | 1 633 | 852 | 257 178 763 772 |
| Dn | 0,24 | 0,05 | 0,14 | 0,10 | 0,10 | 0,07 | 0,07 | 0,50 |
| W2n | 105,48 | 3,60 | 25,89 | 11,81 | 33,63 | 4,74 | 424,36 | |
| A2n | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | -5 326,37 |
| P_KS | 0,02 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_KS | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_CVM | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| P_AD | 17,51 | 3,57 | 10,55 | 7,56 | 11,83 | 5,11 | 37,08 | |
| Dn_T | 105,48 | 3,62 | 25,90 | 11,81 | 33,64 | 4,75 | 424,44 | |
| W2n_T | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | -5 326,37 |
| A2n_T | | | | | | | | |
| P value KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | P>0,25 | P>0,25 | P>0,25 | P>0,25 | P>0,25 | P>0,25 | P>0,25 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| khi2 theo | | 3 032 | 414 | 1 173 | 673 | 1 433 | 435 | 39 236 |
| la distribution vérifiée ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |

non adéquation du modèle Normal aux données
Prob_charge_negative 0,03

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

Ainsi, ont été testées :

- pour modéliser le nombre annuel de sinistres par police des lois discrètes : loi de Poisson, loi binomiale négative,
- pour modéliser les coûts par sinistre des lois continues : loi gamma, loi log-normale et loi de Weibull.

Dans la suite des travaux, nous retiendrons la loi Negative Binomiale pour la modélisation de la fréquence de sinistres et la loi Gamma pour la modélisation du coût des sinistres.

13. Modélisation de la fréquence et du coût moyen d'un sinistre suivant la garantie

13.1. Analyse de la distribution du coût moyen d'un sinistre suivant la garantie :

a. Méthodologie

Nous avons modélisé toutes les garanties sinistres en fréquence et en coût. Pour certaines garanties nous devons tenir compte de quelques spécificités, i.e. séparer l'IDA du non IDA. Pour la garantie RC (Responsabilité Civile) : nous séparons la RC corporelle de la RC matérielle et au sein de la garantie RC matérielle nous séparons l'IDA du non IDA car il y a des sinistres au forfait que nous n'avons pas besoin de modéliser. Pour la garantie RC matérielle IDA nous n'avons qu'à modéliser la fréquence que nous multiplierons au forfait de l'année correspondante (1 204 €).

Pour le corporel nous allons entreprendre 2 approches : nous écrêtons et modéliserons ou tout simplement nous prendrons un coût moyen écrêté.

Concernant la RC corporelle, nous avons à séparer la garantie sinistre avance qui est comme son nom l'indique une avance consentie au client en attendant les recours.

Suite à une étude de la distribution de la charge, nous avons décidé d'écrêter la RC corporelle, la RC matérielle et l'AVA(avance), d'où respectivement les seuils et le pourcentage de la charge non modélisée, à ajouter à la prime pure :

- RC corporelle : 100 000 € et 3.3581667292,
- RC matérielle hors IDA : 30 000 € et 1.1561307064,
- AVA : 10 000 € et 2.3328445016,

Nous considérons les trois coûts moyens qui seront repris dans la modélisation, notamment pour la RC Corporelle, les avances et le forfait IDA, respectivement : 10 465 €, 3542 € et 1 204 €.

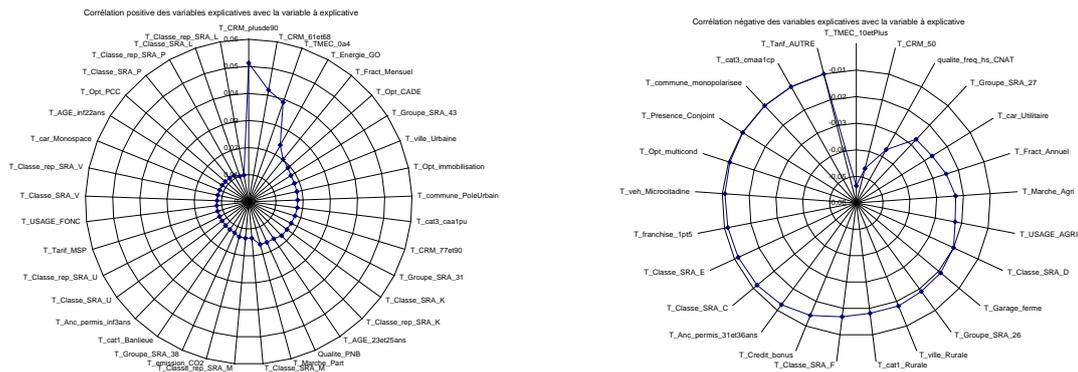
Nous pouvons nous référer sur une partie du programme sas (cf. Annexe II 6.) de la tarification automobile où l'on reconstitue les primes pures avec toutes les garanties modélisées en fréquence et en coût moyen et on réintègre donc l'écrêtement et on tient compte des forfaits IDA.

b. Présentation des résultats

Nous nous attacherons à présenter les résultats de la modélisation en fréquence et en coût moyen de la garantie Dommage hors forfait IDA, le reste se trouvant en annexe.

i. Modèle de régression de la Negative Binomiale pour le nombre de sinistres suivant la garantie Dommage hors forfait IDA

Nous commençons par un premier tri en regardant la corrélation entre la variable à expliquer (au moins un sinistre Dommage hors forfait IDA) et les variables explicatives. Après avoir discrétiser toutes les variables explicatives quantitatives, nous utilisons le test de V de Cramer pour identifier les corrélations de ces dernières avec la variables à expliquer. Nous illustrons une partie des résultats à travers la présentation de deux Graphes des corrélations positives ou négatives :



Puis modélisation de chaque variable à l'aide de la procédure Genmod, nous vérifions que le modèle converge mais aussi que la P-value est inférieure à 5% pour la constante et la variable explicative prise en compte. Nous récupérons de plus le Khi-2 de chaque variable prise séparément ce qui nous permettra de trier les variables explicatives par ordre décroissant et nous les intégrerons dans cet ordre dans le modèle final.

La procédure GENMOD de SAS permet donc de réaliser la régression de la Negative Binomiale du nombre de sinistres Dommage Hors forfait Ida.

Certaines variables explicatives ont été retirées du modèle car non significatives ou car elles sont corrélées avec d'autres variables.

Nous procédons aussi à un regroupement de certaines variables, notamment les variables de Classe du véhicule en l'occurrence ici la classe de réparation SRA.

Nous arrivons au modèle retenu, lequel est décrit ci-dessous.

ii. Quelques explications des différentes sorties issues de la Proc Genmod de sas

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | grisedo |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Ldusse |

Algorithm converged.

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_Classe_rep_SRA_G_a | 2 | 0 1 |
| T_Classe_rep_SRA_J | 2 | 0 1 |
| T_Classe_rep_SRA_K | 2 | 0 1 |
| T_Classe_rep_SRA_L | 2 | 0 1 |
| T_Classe_rep_SRA_M_a | 2 | 0 1 |
| T_Fract_Annuel | 2 | 0 1 |
| T_TMEC_10etPlus | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_car_Utilitaire | 2 | 0 1 |
| qualite_baleII | 2 | 0 1 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--|---------------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | 1 | -2.8252 | 0.2734 | -3.3610 -2.2893 | 106.79 | <.0001 |
| T_Classe_rep_SRA_G_a | 0 | -0.2062 | 0.0600 | -0.3238 -0.0887 | 11.83 | 0.0006 |
| T_Classe_rep_SRA_G_a | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_rep_SRA_J | 0 | -0.2915 | 0.0745 | -0.4376 -0.1454 | 15.29 | <.0001 |
| T_Classe_rep_SRA_J | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_rep_SRA_K | 0 | -0.3585 | 0.0723 | -0.5003 -0.2168 | 24.57 | <.0001 |
| T_Classe_rep_SRA_K | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_rep_SRA_L | 0 | -0.2716 | 0.0727 | -0.4141 -0.1291 | 13.95 | 0.0002 |
| T_Classe_rep_SRA_L | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_rep_SRA_M_a | 0 | -0.3001 | 0.0679 | -0.4332 -0.1671 | 19.56 | <.0001 |
| T_Classe_rep_SRA_M_a | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Fract_Annuel | 0 | 0.2510 | 0.0416 | 0.1695 0.3325 | 36.44 | <.0001 |
| T_Fract_Annuel | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_TMEC_10etPlus | 0 | 0.7747 | 0.0656 | 0.6461 0.9033 | 139.45 | <.0001 |
| T_TMEC_10etPlus | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 0.2117 | 0.0401 | 0.1331 0.2903 | 27.84 | <.0001 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_car_Utilitaire | 0 | 0.5816 | 0.1158 | 0.3546 0.8087 | 25.20 | <.0001 |
| T_car_Utilitaire | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| qualite_baleII | 0 | 0.1668 | 0.0555 | 0.0581 0.2756 | 9.05 | 0.0026 |
| qualite_baleII | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| Dispersion | 1 | 0.1975 | 0.0870 | 0.0832 0.4685 | . | . |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 46E3 | 15262.9187 | 0.3297 |
| Scaled Deviance | 46E3 | 15262.9187 | 0.3297 |
| Pearson Chi-Square | 46E3 | 58552.2787 | 1.2647 |
| Scaled Pearson X2 | 46E3 | 58552.2787 | 1.2647 |
| Log Likelihood | | -10756.8748 | |
| Full Log Likelihood | | -10854.9619 | |
| AIC (smaller is better) | | 21733.9238 | |
| AICC (smaller is better) | | 21733.9305 | |
| BIC (smaller is better) | | 21838.8403 | |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|--------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -21868.042 | | | |
| T_Classe_rep_SRA_G_a | -21850.544 | 1 | 9.50 | 0.0021 |
| T_Classe_rep_SRA_J | -21856.899 | 1 | 1.64 | 0.1997 |
| T_Classe_rep_SRA_K | -21844.656 | 1 | 12.24 | 0.0005 |
| T_Classe_rep_SRA_L | -21834.345 | 1 | 10.31 | 0.0013 |
| T_Classe_rep_SRA_M_a | -21786.371 | 1 | 47.97 | <.0001 |
| T_Fract_Annuel | -21720.445 | 1 | 65.93 | <.0001 |
| T_TMEC_10etPlus | -21579.223 | 1 | 141.22 | <.0001 |
| T_TMEC_5a9 | -21547.862 | 1 | 31.36 | <.0001 |
| T_car_Utilitaire | -21522.463 | 1 | 25.40 | <.0001 |
| qualite_baleII | -21513.750 | 1 | 8.71 | 0.0032 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|--------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Classe_rep_SRA_G_a | 1 | 12.09 | 0.0005 |
| T_Classe_rep_SRA_J | 1 | 15.07 | 0.0001 |
| T_Classe_rep_SRA_K | 1 | 24.29 | <.0001 |
| T_Classe_rep_SRA_L | 1 | 13.85 | 0.0002 |
| T_Classe_rep_SRA_M_a | 1 | 19.66 | <.0001 |
| T_Fract_Annuel | 1 | 37.61 | <.0001 |
| T_TMEC_10etPlus | 1 | 160.76 | <.0001 |
| T_TMEC_5a9 | 1 | 27.98 | <.0001 |
| T_car_Utilitaire | 1 | 29.92 | <.0001 |
| qualite_baleII | 1 | 8.71 | 0.0032 |

iii. Quelques explications sur la table de résultats " Model Information"

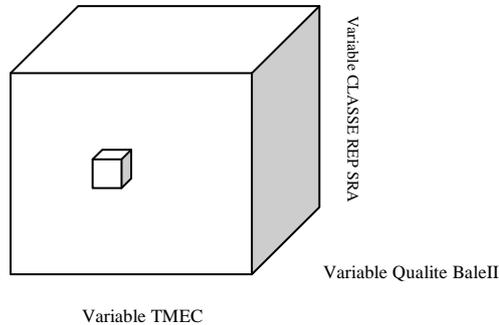
1. Sur la table de résultats " Model Information"

Pas d'observation de sinistre nul. La fonction de lien pour ce type de modélisation (utilisant la loi Negative Binomial) étant la fonction logarithme népérien. D'où :

Observations Used 46 307
Missing Values 0

2. Sur la table de résultats "Criteria For Assessing Goodness of Fit"

⇒ Calcul du nombre de degrés de liberté :



- T_Classe_rep_SRA_G_a_I ⇒ 2 modalités
- T_Classe_rep_SRA_J ⇒ 2 modalités
- T_Classe_rep_SRA_K ⇒ 2 modalités
- T_Classe_rep_SRA_L ⇒ 2 modalités
- T_Classe_rep_SRA_M_a_V ⇒ 2 modalités
- T_Fract_Annuel ⇒ 2 modalités
- T_TMEC_5a9 ⇒ 2 modalités
- T_TMEC_10etPlus ⇒ 2 modalités
- T_car_Utilitaire ⇒ 2 modalités
- qualite_baleII ⇒ 2 modalités

Ainsi le modèle "INTERCEPT T_Classe_rep_SRA_G_a_I T_Classe_rep_SRA_J T_Classe_rep_SRA_K T_Classe_rep_SRA_L T_Classe_rep_SRA_M_a_V T_Fract_Annuel T_TMEC_5a9 T_TMEC_10etPlus T_car_Utilitaire qualite_baleII" compte $(2-1)+(2-1)+(2-1)+(2-1)+(2-1)+(2-1)+(2-1)+(2-1)+(2-1)+(2-1)+1=11$ variables. D'où, le nombre de degrés de liberté est $46\ 307 - 11$, c'est à dire 46 296.

⇒ Sur la donnée "Value/DF"

cf. ANTONIADIS "Sachant que l'espérance d'une loi de Chi² est son nombre de degrés de libertés.... , l'usage est souvent de comparer les statistiques avec le nombre de degrés de liberté. Le modèle peut être jugé **satisfaisant** pour un rapport : **Déviante normalisée/DF < 1.**" Ainsi, le modèle Dommage Hors Ida peut être jugé satisfaisant car le rapport est < 1 : $15262.9187 / 46\ 296 = 0.3297$

⇒ Sur la sur-dispersion

La sur-dispersion est un phénomène qui apparaît quelques fois dans la modélisation par une loi de Poisson ou Negative Binomial. Si l'estimation de la dispersion calculée comme le **Chi² de Pearson sur le degré de liberté est supérieur à 1**, on parle de **sur-dispersion** (parallèlement si ce rapport est inférieur à 1, on parle de sous-dispersion).

Ainsi, pour le modèle Dommage Hors forfait Ida, la surdispersion est donnée par le Scaled Pearson X^2 / DDL soit $58552.2787 / 46\ 296 = 1.2647$.

iv. Quelques explications sur la table de résultats "LR Statistics For Type 1 Analysis" and "LR Statistics For Type 3 Analysis"

⇒ Sur l'analyse TYPE 1

Chaque entrée dans la colonne "Deviance" représente la déviance pour le modèle contenant l'effet de la variable de cette ligne et celles de toutes les lignes précédentes.

Exemple : La "Deviance" correspondant à " T_Classe_rep_SRA_G_a_I " est la déviance du modèle "INTERCEPT T_Classe_rep_SRA_G_a_I "

Ainsi, plus de termes explicatifs sont inclus dans le modèle, plus la déviance diminue.

La valeur "Chisquare" pour " T_Classe_rep_SRA_G_a_I " est donc deux fois la différence entre les vraisemblances des modèles "INTERCEPT T_Classe_rep_SRA_G_a_I T_Classe_rep_SRA_J " et "INTERCEPT T_Classe_rep_SRA_G_a_I "

Ainsi, on diminue la déviance de manière significative i.e. avec une P_value < 5% sauf pour T_Classe_rep_SRA_J.

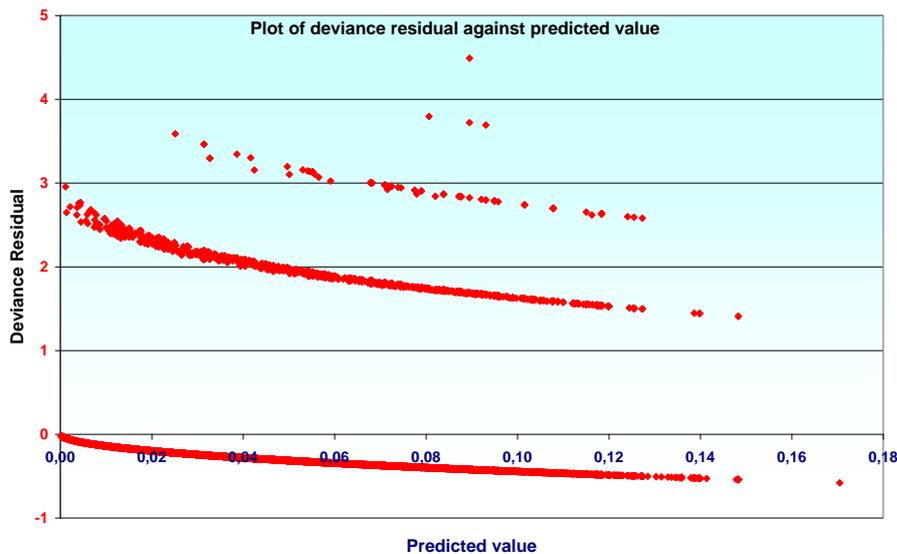
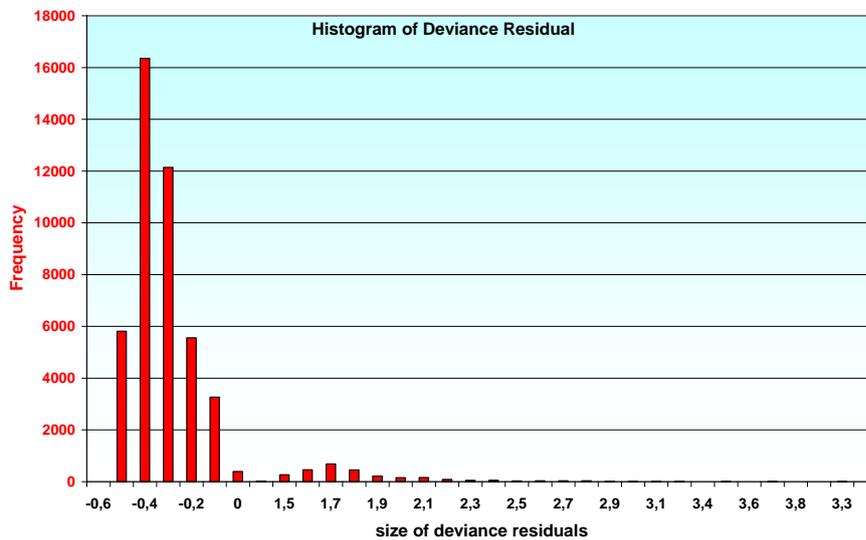
➤ Sur l'analyse TYPE 3

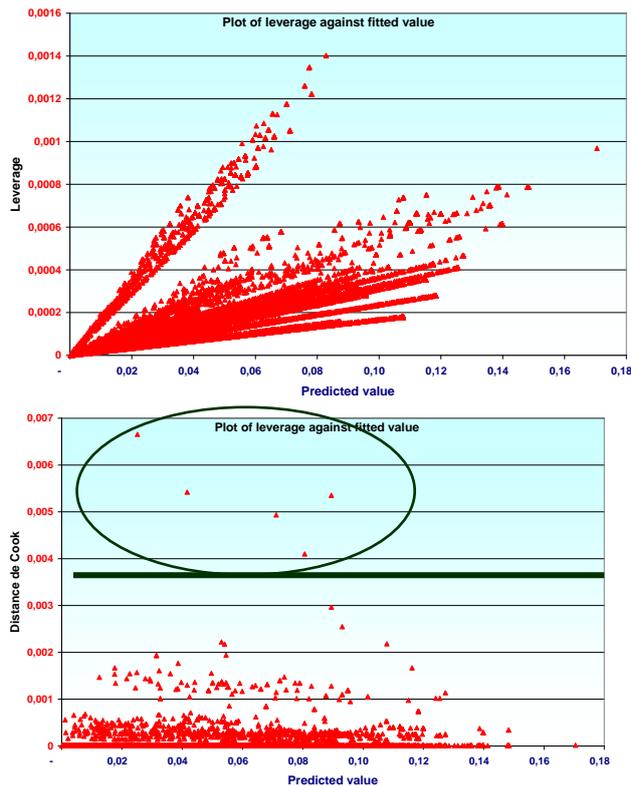
La valeur "Chisquare" pour " T_Classe_rep_SRA_G_a_I " est deux fois la différence entre les vraisemblances des modèles "INTERCEPT T_Classe_rep_SRA_G_a_I T_Classe_rep_SRA_J " et "INTERCEPT T_Classe_rep_SRA_J "

Ainsi, les tests sont significatifs i.e. avec une P_value < 5%.

v. Observation de la distribution des résidus en fonction des paramètres.

Le modèle converge mais si nous analysons le graphe des résidus individuels en fonction des valeurs prédites, nous n'y constatons aucune structure particulière, mais des valeurs assez élevées de certains résidus, qui remettent en question la justesse du modèle.





Les deux derniers graphes imposent une application d'une clause pour retirer les points avec une distance de Cook supérieure à 0.004 car nous sommes en présence d'un fort effet de levier et d'une grande distance de Cook pour certaines valeurs prédites.

vi. Modèle de régression de la loi Gamma pour le coût moyen des sinistres suivant la garantie Dommage hors forfait IDA

Modélisation de chaque variable à l'aide de la procédure Genmod, nous vérifions que le modèle converge mais aussi que la P_value est inférieure à 5% pour la constante et la variable explicative prise en compte. Nous récupérons de plus le Khi-2 de chaque variable prise séparément, ce qui nous permettra de trier les variables explicatives par ordre décroissant et nous les intégrerons dans cet ordre dans le modèle final.

La procédure GENMOD de SAS permet donc de réaliser la régression de la loi Gamma du coût moyen des sinistres Dommage Hors forfait Ida.

Certaines variables explicatives ont été retirées du modèle car non significatives ou car elles sont corrélées avec d'autres variables.

Nous procédons aussi à un regroupement de certaines variables notamment les variables de Classe du véhicule en l'occurrence ici la classe de prix SRA.

Nous arrivons au modèle retenu, lequel est décrit ci-dessous.

vii. Les sorties SAS issues de la Proc Genmod

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Chardo |

Algorithm converged.

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| qualite_baleII | 2 | 0 1 |
| T_Opt_multicond | 2 | 0 1 |
| T_Opt_PCC | 2 | 0 1 |
| T_USAGE_AGRI | 2 | 0 1 |
| T_Anc_peremis_0a5ans | 2 | 0 1 |
| T_Anc_peremis_6et12an | 2 | 0 1 |
| T_Anc_peremis_13et16a | 2 | 0 1 |
| T_TMEC_10etPlus | 2 | 0 1 |
| T_Fract_Mensuel | 2 | 0 1 |
| T_Classe_SRA_A_a_I | 2 | 0 1 |
| T_Classe_SRA_J | 2 | 0 1 |
| T_Classe_SRA_M | 2 | 0 1 |
| T_Classe_SRA_N_a_V | 2 | 0 1 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--|---------------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | 1 | 8.7400 | 0.2191 | 8.3106 9.1693 | 1591.88 | <.0001 |
| qualite_baleII | 0 1 | 0.2103 | 0.0523 | 0.1077 0.3129 | 16.15 | <.0001 |
| qualite_baleII | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Opt_multicond | 0 1 | -0.2140 | 0.0461 | -0.3083 -0.1197 | 19.79 | <.0001 |
| T_Opt_multicond | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Opt_PCC | 0 1 | -0.0883 | 0.0451 | -0.1768 0.0001 | 3.84 | 0.0502 |
| T_Opt_PCC | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_USAGE_AGRI | 0 1 | -0.2310 | 0.0705 | -0.3693 -0.0928 | 10.73 | 0.0011 |
| T_USAGE_AGRI | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Anc_peremis_0a5ans | 0 1 | -0.4288 | 0.0894 | -0.6039 -0.2536 | 23.02 | <.0001 |
| T_Anc_peremis_0a5ans | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Anc_peremis_6et12an | 0 1 | -0.1392 | 0.0526 | -0.2423 -0.0360 | 7.00 | 0.0082 |
| T_Anc_peremis_6et12an | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Anc_peremis_13et16a | 0 1 | -0.1807 | 0.0648 | -0.3078 -0.0536 | 7.76 | 0.0053 |
| T_Anc_peremis_13et16a | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_TMEC_10etPlus | 0 1 | 0.3569 | 0.0612 | 0.2369 0.4768 | 33.99 | <.0001 |
| T_TMEC_10etPlus | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Fract_Mensuel | 0 1 | -0.1420 | 0.0399 | -0.2203 -0.0638 | 12.66 | 0.0004 |
| T_Fract_Mensuel | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_SRA_A_a_I | 0 1 | -0.1398 | 0.0451 | -0.2282 -0.0515 | 9.62 | 0.0019 |
| T_Classe_SRA_A_a_I | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_SRA_J | 0 1 | -0.2203 | 0.0649 | -0.3474 -0.0931 | 11.53 | 0.0007 |
| T_Classe_SRA_J | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_SRA_M | 0 1 | -0.1982 | 0.0718 | -0.3390 -0.0575 | 7.62 | 0.0058 |
| T_Classe_SRA_M | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_SRA_N_a_V | 0 1 | -0.2455 | 0.0705 | -0.3837 -0.1073 | 12.13 | 0.0005 |
| T_Classe_SRA_N_a_V | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . |
| Scale | 1 | 0.9882 | 0.0220 | 0.9460 1.0324 | . | . |

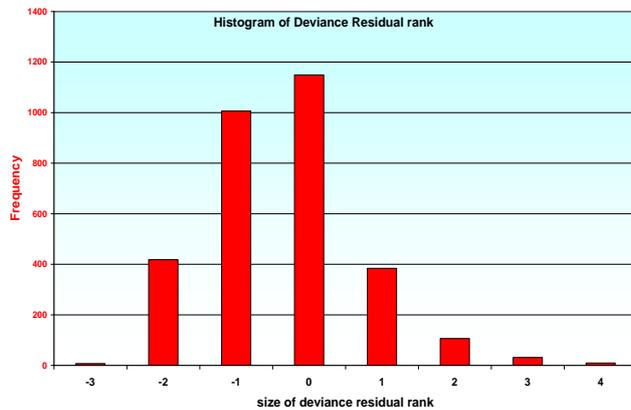
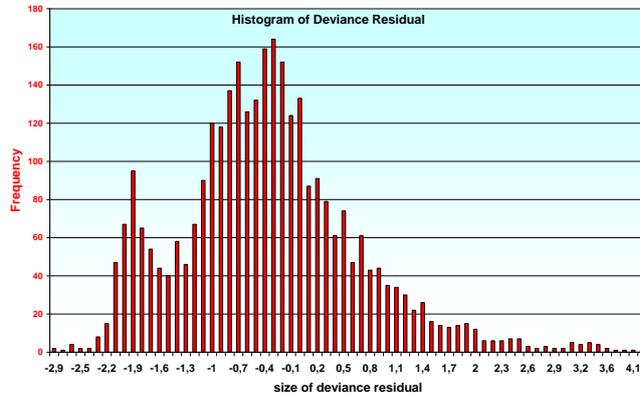
| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 3096 | 3638.1274 | 1.1751 |
| Scaled Deviance | 3096 | 3595.2907 | 1.1613 |
| Pearson Chi-Square | 3096 | 4833.3456 | 1.5612 |
| Scaled Pearson X2 | 3096 | 4776.4359 | 1.5428 |
| Log Likelihood | | -25688.9360 | |
| Full Log Likelihood | | -25688.9360 | |
| AIC (smaller is better) | | 51407.8720 | |
| AICC (smaller is better) | | 51408.0271 | |
| BIC (smaller is better) | | 51498.5077 | |

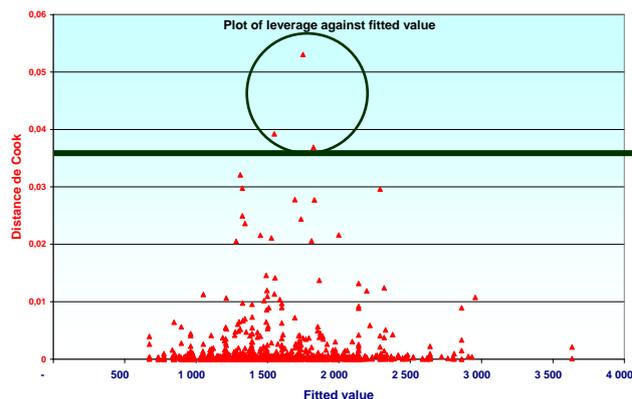
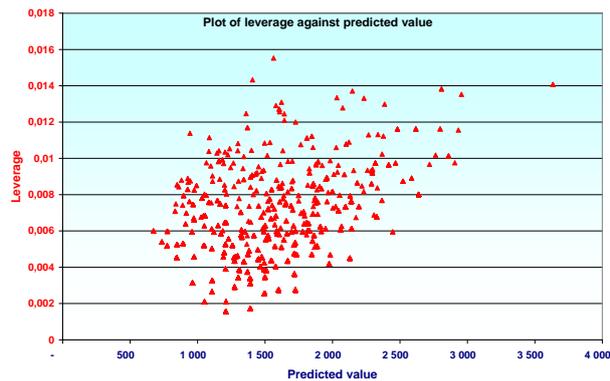
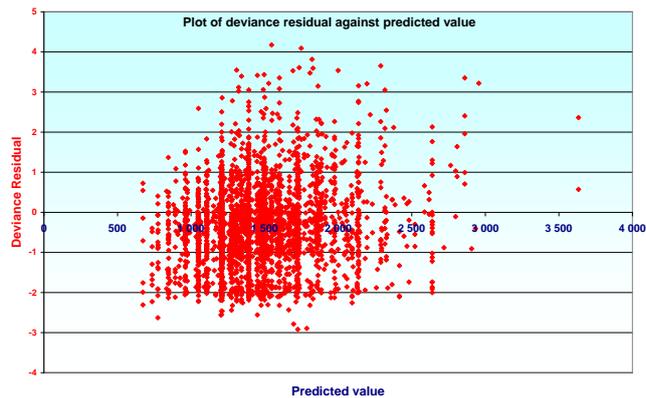
| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -51547.831 | | | |
| qualite_baleII | -51511.632 | 1 | 36.20 | <.0001 |
| T_Opt_multicond | -51501.955 | 1 | 9.68 | 0.0019 |
| T_Opt_PCC | -51492.366 | 1 | 9.59 | 0.0020 |
| T_USAGE_AGRI | -51489.599 | 1 | 2.77 | 0.0962 |
| T_Anc_peremis_0a5ans | -51460.422 | 1 | 29.18 | <.0001 |
| T_Anc_peremis_6et12an | -51453.097 | 1 | 7.32 | 0.0068 |
| T_Anc_peremis_13et16a | -51440.494 | 1 | 12.60 | 0.0004 |
| T_TMEC_10etPlus | -51411.238 | 1 | 29.27 | <.0001 |
| T_Fract_Mensuel | -51398.352 | 1 | 12.68 | 0.0004 |
| T_Classe_SRA_A_a_I | -51398.330 | 1 | 0.22 | 0.6378 |
| T_Classe_SRA_J | -51393.910 | 1 | 4.42 | 0.0325 |
| T_Classe_SRA_M | -51390.423 | 1 | 3.49 | 0.0619 |
| T_Classe_SRA_N_a_V | -51377.872 | 1 | 12.55 | 0.0004 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| qualite_baleII | 1 | 16.87 | <.0001 |
| T_Opt_multicond | 1 | 20.37 | <.0001 |
| T_Opt_PCC | 1 | 3.77 | 0.0521 |
| T_USAGE_AGRI | 1 | 11.17 | 0.0008 |
| T_Anc_peremis_0a5ans | 1 | 25.48 | <.0001 |
| T_Anc_peremis_6et12an | 1 | 7.14 | 0.0075 |
| T_Anc_peremis_13et16a | 1 | 8.07 | 0.0045 |
| T_TMEC_10etPlus | 1 | 31.01 | <.0001 |
| T_Fract_Mensuel | 1 | 12.58 | 0.0004 |
| T_Classe_SRA_A_a_I | 1 | 9.49 | 0.0021 |
| T_Classe_SRA_J | 1 | 11.83 | 0.0006 |
| T_Classe_SRA_M | 1 | 7.86 | 0.0051 |
| T_Classe_SRA_N_a_V | 1 | 12.55 | 0.0004 |

viii. Observation de la distribution des résidus en fonction des paramètres.

Le modèle converge et si nous analysons le graphe des résidus individuels en fonction des valeurs prédites, nous n'y constatons aucune structure particulière, mais des valeurs assez élevées de certains résidus, qui mettent en question la justesse du modèle.





Les deux derniers graphes imposent une application d'une clause pour retirer les points avec une distance de Cook supérieure à 0.035 car nous sommes en présence d'un fort effet de levier et d'une grande distance de Cook pour certaines valeurs prédites.

ix. Conclusions

Nous avons procédé de la même manière concernant les autres garanties et nous sommes tombés sur des résultats similaires aussi bien sur la modélisation de la fréquence que sur celle du coût moyen, sauf sur certaines garanties la RCC (Responsabilité Civile Corporelle) et l'AVA(avance), car, dans les 2 cas, sur la donnée Déviance normalisée/DF était trop proche de 0 et sur la donnée Scaled Pearson X^2 / DDL qui représente la sur-dispersion était bien au-delà de 2 voir proche de 3. Nous avons donc procédé à un réajustement pour que la prime pure modélisée se rapproche de la réalité des données empiriques. Ainsi nous pouvons maintenant estimer une réduction possible sur les différents critères de qualité du client bancassurance.

Un variable explicative « la qualité Bâle II » influence la sinistralité en fréquence et en coût moyen par garantie au même titre que le CRM et l'ancienneté de permis. Le fait d'enlever le CRM ou l'année de permis de la modélisation augmente considérablement l'effet de « la note Bâle II » sur la sinistralité.

Pour résumer les différents résultats à travers chaque garantie et chaque modélisation de la fréquence et du coût moyen, nous vous présentons un tableau reprenant les différentes variables retenues dans le modèle final:

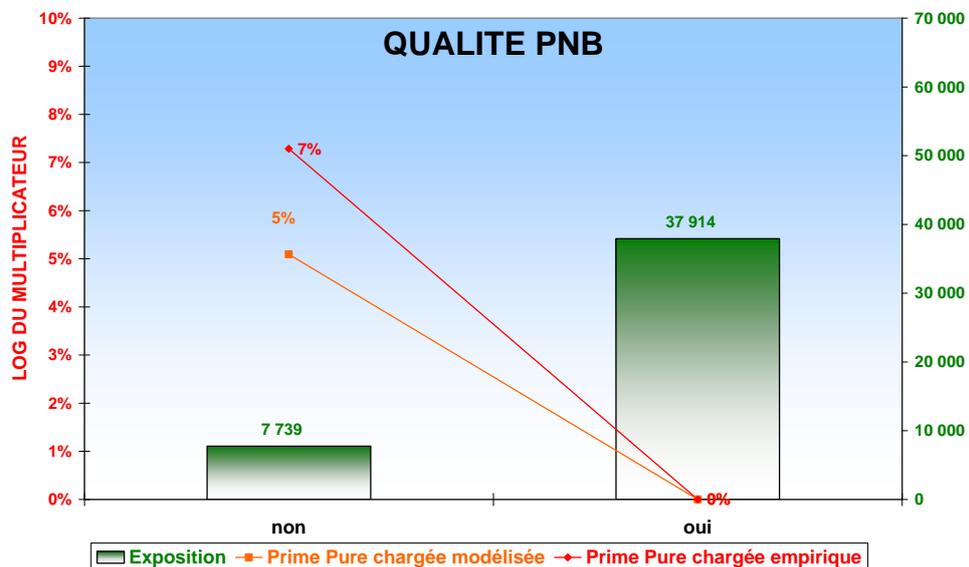
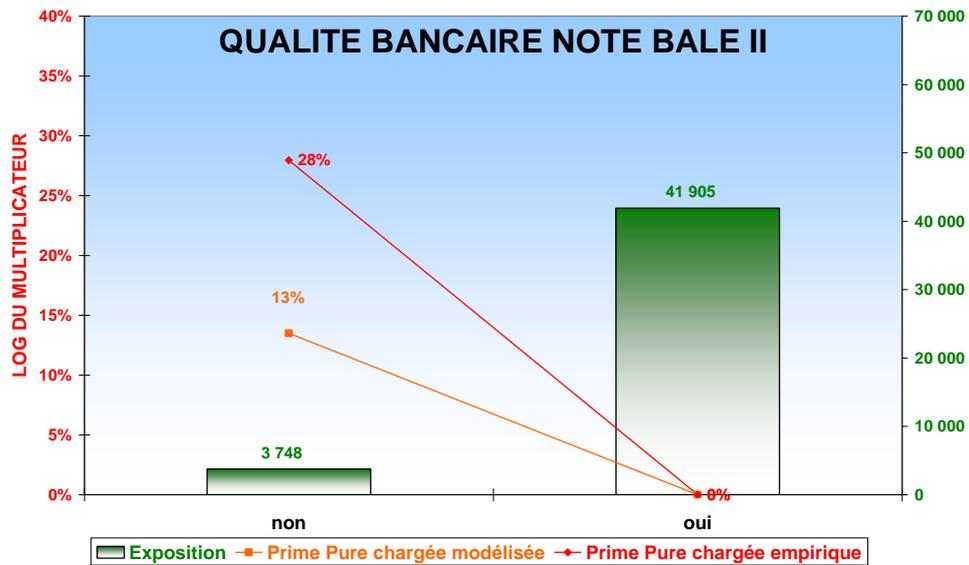
| | Freq RCM Hors IDA | Freq RCM IDA | Freq RCC Hors Avance | Freq RCC Avance | Freq Défense Recours Hors IDA | Freq Défense Recours IDA | Freq Incendie Vol | Freq Bris de glace | Freq Dommage Hors IDA | Freq Dommage IDA |
|--|-------------------|--------------|----------------------|-----------------|-------------------------------|--------------------------|-------------------|--------------------|-----------------------|------------------|
| Année | | | | | | | | | | |
| Formule | | | | | | | | | | |
| Pack Plus | | | | | | | | | | |
| Option Franchise | | | | | | | | | | |
| Option Contenu privé | | | | | | | ● | | | |
| Fractionnement | ● | | | | | | | ● | ● | |
| Garage fermé | | | | ● | | ● | ● | | | |
| Option Multiconducteur | | | | | | | | | | |
| Option Véhicule de remplacement | | | | | | | | | | |
| Option Protection Corporelle du conducteur | | | | | ● | | | | | |
| Option Sérénité | | | | | | | | | | |
| Option Tuning | | | | | | | | | | |
| Activité professionnelle | | | | | ● | ● | ● | ● | | |
| Tranche d'ancienneté du permis | | | | | | | | | | |
| Sexe | | ● | | | | | | | | |
| Tranche d'âge | | ● | | ● | | ● | ● | ● | | |
| Crédit Bonus | | ● | | | | | | | | |
| Tranche de Bonus Malus | | | | | | | | | | ● |
| Conducteur Novice Occasionnelle | | ● | | | | | | | | |
| Conduite Accompagnée | | | ● | | | | | | | |
| Tranche de Mise en circulation | ● | | ● | | ● | ● | ● | ● | ● | ● |
| Groupe SRA | | ● | ● | | | | | ● | | |
| Classe de prix SRA | ● | | | | | | ● | | | |
| Classe de réparation SRA | ● | | | | | | | ● | ● | |
| Type de véhicule | | | | | | | | | | |
| Type de carrosserie | | | | | ● | | ● | | ● | |
| Témoin emission CO2 | | | | ● | | | | | | |
| Energie | | | | | | | | ● | | ● |
| Catégorie de commune INSEE | | ● | ● | | ● | | | | | ● |
| Détention de PEL | | ● | | | | | | | | |
| Détention d'actions | | | | | | | | | | |
| Détention de Sécuricompte | | | | | | | | | | |
| Qualite PNB | | | | | | | | | | |
| Qualite Bale II | ● | | ● | | ● | ● | | | ● | |
| Qualite S/P Hors CNAT | | | | | | | | | | |
| Qualite FREQ Hors CNAT | | | | | | | | | | |

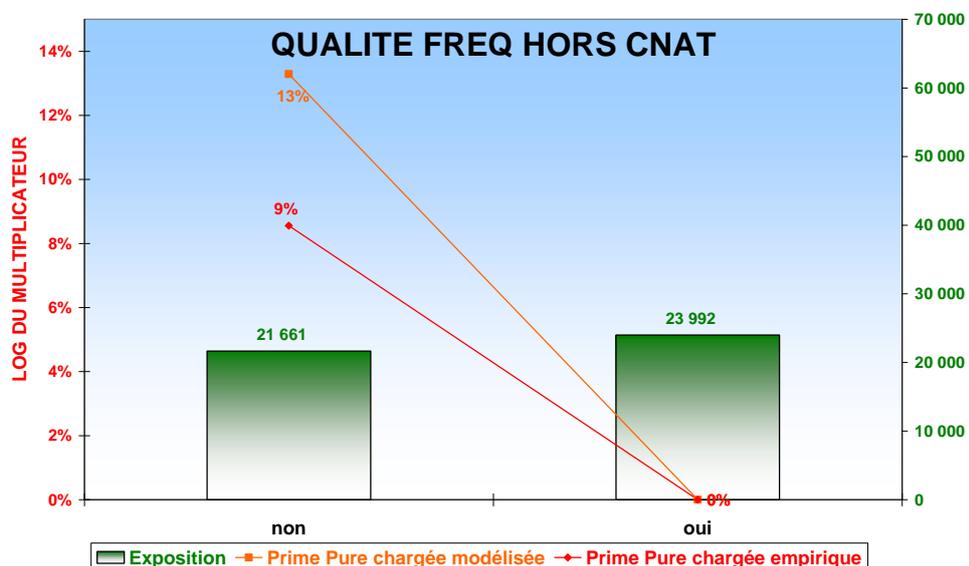
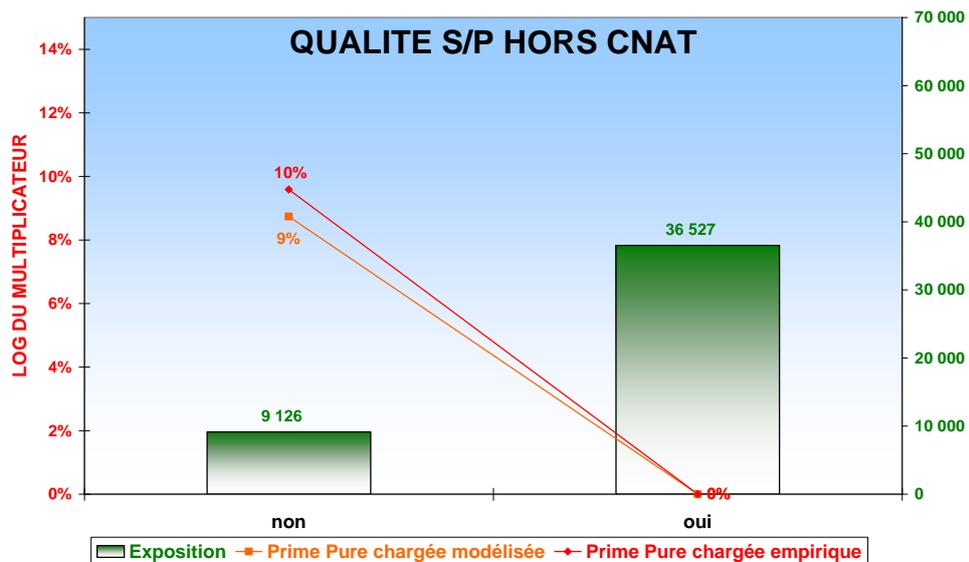
| | CM RCM Hors IDA | CM RCM IDA 1204€ | CM RCC Hors Avance | CM RCC Avance | CM Défense Recours Hors IDA | CM Défense Recours IDA | CM Incendie Vol | CM Bris de glace | CM Domage Hors IDA | CM Domage IDA |
|--|--------------------------|---------------------------|-----------------------------|---------------------|---|---------------------------------|-----------------------|---------------------------|--------------------------|---------------------|
| Année | | | | | | | | | | |
| Formule | | | ● | | | | | | | |
| Pack Plus | | | ● | | | | | | | |
| Option Franchise | | | | | | | | | | |
| Option Contenu privé | | | | | | | | | | |
| Fractionnement | | | | | | | ● | | ● | |
| Garage fermé | ● | | | | | | | | | |
| Option Multiconducteur | | | | | | | | | ● | |
| Option Véhicule de remplacement | | | ● | | | | | | | |
| Option Protection Corporelle du conducteur | | | | | | | | | ● | |
| Option Sérénité | | | | | | | | | | |
| Option Tuning | | | | | | | | | | |
| Activité professionnelle | | | | | | | | | ● | |
| Tranche d'ancienneté du permis | | | | ● | ● | | | ● | ● | |
| Sexe | | | | | | | | | | |
| Tranche d'âge | ● | | ● | | | | | | | |
| Crédit Bonus | | | | | | | | | | |
| Tranche de Bonus Malus | | | | | | | | | | ● |
| Conducteur Novice Occasionnelle | | | | | | | | | | |
| Conduite Accompagnée | | | | | | | | | | |
| Tranche de Mise en circulation | | | | | | ● | ● | ● | ● | |
| Groupe SRA | ● | | ● | | | | | | | |
| Classe de prix SRA | ● | | ● | | ● | | ● | | ● | |
| Classe de réparation SRA | ● | | | | | | | ● | | ● |
| Type de véhicule | | | ● | | | | ● | ● | | ● |
| Type de carrosserie | ● | | ● | | | | | ● | | |
| Témoin émission CO2 | | | | | | | | ● | | |
| Energie | ● | | ● | | | | | | | |
| Catégorie de commune INSEE | | | | | | | | | | |
| Détention de PEL | ● | | | | | | | | | |
| Détention d'actions | | | | | | | | | | |
| Détention de Sécuricompte | ● | | | ● | ● | | ● | | | ● |
| Qualite PNB | ● | | | | | | ● | | | |
| Qualite Bale II | | | ● | | | | | | ● | |
| Qualite S/P Hors CNAT | | | | | | | | | | |
| Qualite FREQ Hors CNAT | | | | | | | | | | |

14. Présentation des facteurs représentant la qualité du client en bancassurance

Les MLG donnent le vrai effet d'un facteur, tous les autres facteurs étant égaux.

Sur ces graphiques, les primes pures chargées sont exprimées en logarithme du multiplicateur (i.e. en variation par rapport à la moyenne du portefeuille).





15. Conclusion de la troisième partie

Il est évident qu'une réduction tarifaire sur cette typologie de client bancassurance est tout à fait appropriée. Mais l'utilisation de certaines données bancaires, notamment la qualité Bâle II, n'est pas légitime aux yeux de la CNIL. Par contre, dans le cadre du CRM, i.e. de la gestion de la relation client, elle peut faire l'objet de critère de ciblage. Plus particulièrement, le critère « qualité du risque bancaire » du client issu, bien sûr, de la note Bâle II imposée à toutes les banques et calculée tous les mois au niveau du groupe de risque, plus précisément, au niveau du partenaire commerciale du marché des particuliers, peut être utilisée aisément. Ce dernier critère traduit ce que nous constatons depuis des années « un bon risque bancaire est un bon risque assurantiel ».

Chapitre 5 Présentation de l'estimation des marges futures aussi bien en assurance qu'au niveau bancaire dans le cadre de la mise en marché du « contrat Fidelio »

1. Objectif

Il s'agira ici d'estimer au mieux les marges qui sont susceptibles d'être engrangées suite à la mise en marché du « contrat Fidelio », en dissociant celles engendrées par l'Assureur de celles engendrées par le banquier.

2. Hypothèses

a. Au niveau Assurance

Pour pouvoir évaluer les marges techniques, sur les 5 années suivant l'année de mise en marché du « contrat Fidelio », nous nous devons de faire différentes hypothèses, i.e. :

- contribution en % à l'enveloppe de ristourne par l'assurance pour les contrats souscrits dans le cadre du « contrat Fidelio »:
 - Auto : 25%
 - Mrh : 25%

| | sur la Primes émises HT | |
|--|-------------------------|---------|
| Enveloppe annuelle en montant accordé sur le contrat Auto | 1% | 202 884 |
| Enveloppe annuelle en montant accordé sur le contrat Mrh | 1% | 127 876 |
| Enveloppe annuelle en montant accordé sur le contrat auto dans le cadre du "contrat Fidelio" | | 25 927 |

Chaque année Pacifica tolère 1% de la prime émise Hors Taxe pour constituer une enveloppe de ristournes pour une famille de produit donnée. L'enveloppe utile à la commercialisation du « contrat Fidelio » représente 12.8% de l'enveloppe globale et 1.2% de la prime émise Hors Taxe issue de la souscription du « contrat Fidelio »

- Propositions d'un avantage tarifaire sur le produit d'appel Auto si le ou les contrats sont souscrits dans le cadre du « contrat Fidelio »:
 - 10%
- Taux d'actualisation :
 - Auto : 3%
 - Mrh : 6%
- Taux de marge technique sur les contrats :
 - Auto : 5.5%
 - Mrh : 5.9%

| | | |
|--|-----------|-------------|
| PNB Assurance Auto en € ou en % de la cotise HT par contrat | 19 | 5,5% |
| PNB Assurance Mrh en € ou en % de la cotise HT par contrat | 10 | 5,9% |

- Taux de résiliation auto ou Mrh pour les clients en portefeuille équipés d'une auto et d'une Mrh:
 - Auto : 13.3%
 - Mrh : 10.3%

| | |
|---|--------------|
| Taux de résiliation Auto pour les clients en portefeuille équipés Auto + Mrh | 13,3% |
| Taux de résiliation Mrh pour les clients en portefeuille équipés Auto + Mrh | 10,3% |

- Taux de résiliation sur affaire nouvelle :
 - Auto : 7.2%
 - Mrh : 3.2%

| | |
|--|-------------|
| Taux de résiliation sur affaire nouvelle Auto | 7,2% |
| Taux de résiliation sur affaire nouvelle Mrh | 3,2% |

- Durée de vie des clients mono détenteurs d'un contrat :
 - Auto : 10.3
 - Mrh : 11.5
 - Auto & Mrh : 16.7
- Taux d'affaire nouvelle des mono détenteurs d'un contrat :
 - Auto : 5.3% \Rightarrow 7.8% dans le cadre de la commercialisation du « contrat Fidelio »
 - Mrh : 2.5% \Rightarrow 5% dans le cadre de la commercialisation du « contrat Fidelio »

| | | |
|--|-------------|-------------|
| Taux d'affaire nouvelle contrat Auto sachant que le partenaire a une de Mrh | 5,3% | 7,8% |
| Taux d'affaire nouvelle contrat Mrh sachant que le partenaire a une Auto | 2,5% | 5,0% |

- Prime moyenne Hors Taxe suivant la qualité des clients au niveau « Bâle II » et mono détenteurs d'un contrat :
 - Auto : 379
 - Mrh : 197

| classe balell | Det auto seul | Det mrh seul | Nb ct Auto | Mt ct Auto | Nb ct Mrh | Mt ct Mrh | Cotise moyenne | |
|---------------|---------------|---------------|---------------|-------------------|---------------|-------------------|----------------|------------|
| | | | | | | | auto | mrh |
| Très bon | 10 373 | 22 933 | 35 465 | 13 111 334 | 41 876 | 8 456 697 | 370 | 202 |
| Bon | 3 513 | 9 425 | 12 456 | 4 988 286 | 15 811 | 3 039 614 | 400 | 192 |
| Moyen bon | 1 524 | 4 248 | 5 560 | 2 188 767 | 7 246 | 1 291 337 | 394 | 178 |
| | 15 410 | 36 606 | 53 481 | 20 288 387 | 64 933 | 12 787 648 | 379 | 197 |

- Du tableau ci-dessus, on en déduit le potentiel de clients équipables avec une note Bâle II (très bon, bon et moyen bon) et mono détenteurs d'un contrat :
 - Auto et pouvant être équipés en Mrh: 15 410
 - Mrh et pouvant être équipés en Auto : 36 606

b. Au niveau Bancaire

Pour pouvoir évaluer les marges techniques, sur les 5 années suivant l'année de mise en marché du « contrat Fidelio », nous nous devons de faire différentes hypothèses, i.e. :

- contribution en % à l'enveloppe de ristourne par le banquier pour les contrats souscrits:
 - Auto : 75%
 - Mrh : 75%

- Taux de marge technique sur les contrats :
 - Auto : 16.4%
 - Mrh : 17.7%

| | | |
|---|----|-------|
| PNB Bancaire Auto en € ou en % de la cotise HT par contrat | 58 | 16,4% |
| PNB Bancaire Mrh en € ou en % de la cotise HT par contrat | 31 | 17,7% |

- Taux d'inflation : 2%
- Estimation du coût horaire d'un commercial en banque sur le marché des particuliers :

D'après une enquête réalisée par Michael page, nous vous présentons la grille de salaire d'un commercial en banque sur le marché des particuliers :

| | | |
|--|----------------------------------|------------------|
| Pour la grille de salaire d'un commercial en banque (commercial particulier) | | |
| Rémunération | | |
| | Salaire fixe annuel brut en euro | Variable en euro |
| Débutant | | |
| (1 à 3 ans d'expérience) | Min. 17 000 | 0 à 2 000 |
| | Moy. 24 000 | 0 à 4 000 |
| | Max. 28 000 | 0 à 8 000 |
| Confirmé | | |
| (3 ans d'expérience et plus) | Min. 25 000 | 0 à 4 000 |
| | Moy. 30 000 | 0 à 6 000 |
| | Max. 34 000 | 0 à 8 000 |
| Le variable ne comprend ni la participation, ni l'intéressement. | | |
| source http://www.michaelpage.fr/controll [...] | | |

Nous en déduisons un coût horaire brut d'un conseiller de 13.6 € pour un salaire brut annuel hors intéressement et participation de 24 563 €

| | | |
|--|---|--------------------|
| Temps passé pour traité l'évènement | 30 | mn |
| coût brut d'un conseiller | 6,8 | euros/ heure |
| Nombre d'heures par jour | | 7 |
| Salaire Net mensuel | Nombre de jours ouvrés par mois d'après le tableau ci-dessous | |
| 1 450 | 228/12= | 19 |
| | | en euros en francs |
| Salaire Brut mensuel | coût horaire brut d'un conseiller | 13,6 89,4 |
| 1 813 | | |
| Salaire Brut annuel sur 14 mois | | |
| 24 563 | | |
| Prenons par exemple l'année 2012 pour calculer le nombre de jours travaillés dans l'année | | |
| Nombre de jours dans l'année 2012 : | | 366 |
| Nombre de jours ouvrés (pour information) : | | 253 |
| Nombre de jours de week-end : | - | 105 |
| Nombre de jours fériés : | - | 11 |
| Nombre de jours fériés tombant le week-end | + | 3 |
| Nombre de de jours de congés payés | - | 25 |
| Nombre total de jours ouvrés effectifs dans l'année 2012 | = | 228 |

- Estimation de la charge d'exploitation annuelle pour une mise en place d'une cible récurrente mensuelle : l'évènement « Multi équipement des assurés mono détenteurs en assurance Auto et Mrh » à l'aide du « contrat Fidelio » :

| | Stock | |
|-----------------------------------|--------|--------|
| | Auto | Mrh |
| Cible récurrente mensuelle | 2 000 | 4 000 |
| Nombre d'évènement par ETP | 2 | 4 |
| Taux de traitement des évènements | 30% | 30% |
| Nombre d'évènements traités | 600 | 1 200 |
| Taux de concrétisation | 8,0% | 9,5% |
| Nombre de concrétisation mensuel | 48 | 114 |
| Charge d'exploitation mensuelle | 4 088 | 8 177 |
| Nombre de concrétisation annuel | 576 | 1 368 |
| Charge d'exploitation annuelle | 49 060 | 98 120 |

3. Incertitude sur l'estimation des marges futures

a. Le taux d'inflation pris en compte dans le calcul des charges d'exploitation.

D'après le contexte économique pour les sociétés et les différentes crises énergétiques qui se succèdent, une inflation à 2% semble être la bonne approche et conforme aux prévisions réalisées fin 2011 par différents organismes. Les extraits ci-dessous me permettent de conforter mon postulat avec un plus haut proche de 4% atteint en juillet 2008 et un plus bas proche de -1% atteint en juillet 2009.

PREVISIONS INFLATION MOYENNE 2012 (réalisées fin 2011 par différents organismes)

| Organisme | 2011 | 2012 | 2013 |
|------------------|------|------|------|
| Société Générale | 2.1% | 1.5% | 1.6% |
| NATIXIS | 2.2% | 2.0% | 1.5% |
| BNP PARIBAS | 2.1% | 1.9% | 1.8% |
| Crédit Agricole | 2.2% | 1.6% | 1.8% |
| OCDE | 2.1% | 1.4% | 1.1% |
| FMI | 2.1% | 1.7% | 1.8% |

La plupart des organismes analysant la conjoncture économique prévoient une baisse de l'inflation en 2012. Pour l'Insee, la baisse mi-2012 devrait être assez forte. L'inflation diminuerait nettement, à 1,4 % en juin 2012 du fait du très fort recul des prix de l'énergie et de la baisse des produits alimentaires. L'effet hausse de la TVA ne serait que de 0,1% à Juin.

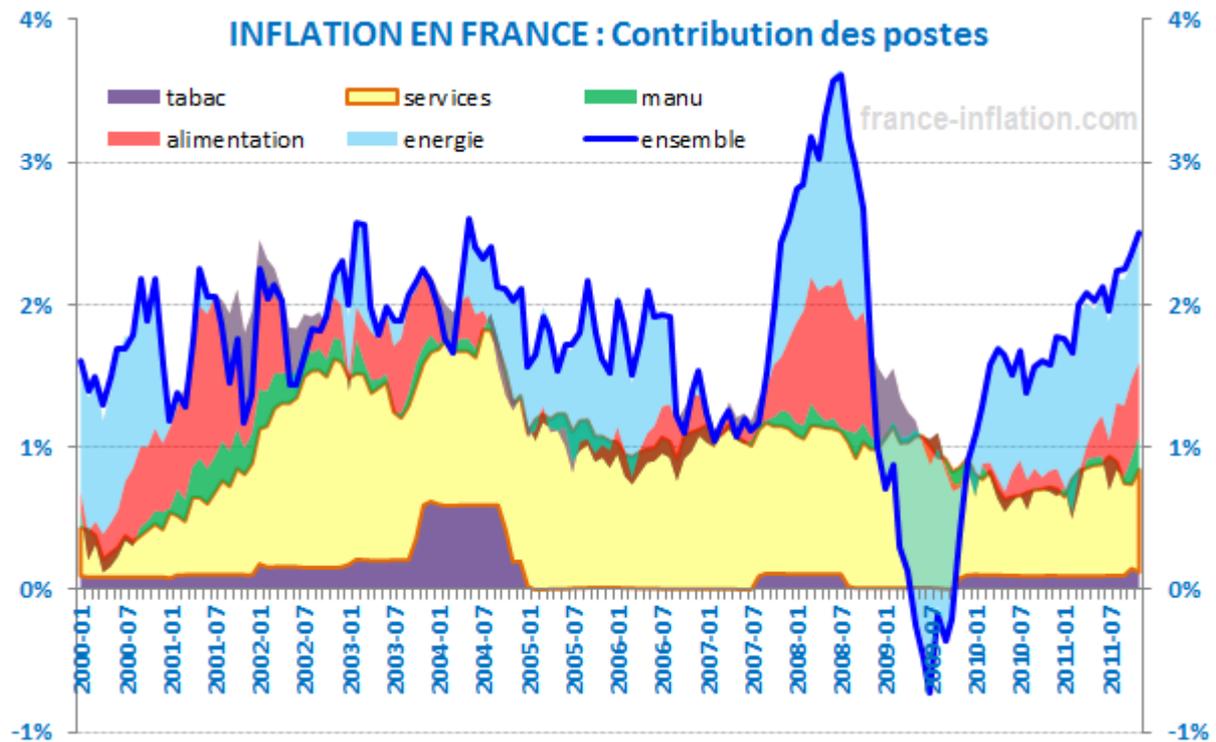
Pour faire une prévision de l'inflation en 2012, les deux facteurs clefs sont l'évolution des prix de l'énergie et l'évolution des prix alimentaires. On peut voir l'effet prépondérant de ces deux postes dans la décomposition de l'inflation depuis 2000 par grand contributeur qui suit.

Il est à noter que les prévisions de baisse d'inflation projetées pour 2012 anticipent une baisse

significative de ces deux postes. Le passé récent nous a cependant appris que l'évolution du prix du pétrole était très difficile à anticiper aussi on ne peut qu'être prudent sur ces prévisions.

Si l'instabilité politique s'amplifie dans les pays producteurs, le prix du pétrole peut toujours s'envoler et l'inflation suivrait.

INFLATION PAR GRANDS CONTRIBUTEURS



Dans le calcul de la marge future, le taux d'inflation intervient pour estimer l'évolution des salaires annuellement des conseillers. Il est évident qu'en pleine période de crise avec une problématique de charge d'exploitation à maîtriser et une période de déflation, une évolution des revenus au-delà de l'inflation me semble inappropriée, sauf à un retour à une inflation galopante causée par une sortie de l'Euro, par exemple. La probabilité de cet événement n'est pas nulle mais elle reste faible à cet instant.

b. Le taux d'actualisation pris en compte dans le calcul des primes Hors Taxes



Les produits les plus touchés seront encore l'assurance habitation avec une hausse prévue de l'ordre de 6%, l'assurance santé avec une augmentation de 4% et l'assurance automobile qui devrait augmenter un peu plus que l'inflation 2011 (autour de 3%).

Pourquoi une nouvelle hausse en 2012 alors même que les assurés ont vu leurs cotisations augmenter fortement l'année dernière ? Cette hausse est-elle justifiée ?

Christine Lagarde avait demandé un audit à la Direction Générale de la Concurrence, de la Consommation et de la Répression des Fraudes en décembre 2010, afin de savoir si l'augmentation des tarifs était justifiée.

Le rapport a démontré que la hausse des tarifs correspondait bien à une réelle augmentation de leurs coûts et non pas dans le but de gonfler leur marge.

En assurance habitation, ils ont perdu des milliards d'euros à cause des catastrophes naturelles (tempête Xynthia, inondations dans le Var...). Ils n'ont pas pu augmenter leur tarif de manière trop importante l'année dernière et ont donc opté pour une hausse progressive sur plusieurs années.

Pour l'assurance santé, le vieillissement de la population française, l'amélioration des technologies médicales ou encore les déremboursements de la Sécu ont considérablement augmenté les dépenses des assureurs.

Quant à l'assurance auto, les cotisations devraient suivre un peu plus que l'inflation.

Dans la lignée de 2011, les assureurs se voient donc obligés d'augmenter encore les cotisations en 2012 voir pendant 3 à 5 ans pour retourner à l'équilibre. Le but est, en fait, de lisser la hausse sur plusieurs années car les ménages subissent depuis 2009 une augmentation significative de leurs charges récurrentes de la vie courante.

⇒ Nous tablons donc pour une hausse constante de 3%/par an en auto et une hausse constante de 6%/an en habitation.

Pour conclure sur ces deux premiers points, qui sont pour moi les indicateurs pouvant remettre en cause l'estimation des marges futures, dans le cadre de la commercialisation du « contrat Fidelio », nous pensons que les deux effets peuvent à minima se neutraliser voire augmenter les marges puisque, en règle générale, les salaires augmentent moins vite que les charges de la vie courante, notamment, ceux liés à l'Assurance.

c. Les différents taux d'affaires nouvelles et de résiliations, sur les monodétenteurs de contrat Auto ou Mrh, sont estimés à partir des données historisées suivantes :

| | estimation | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 |
|---|---------------|-------|--------|--------|--------|--------|--------|
| taux d'affaires nouvelles Auto sachant monodétenteur Mrh en N-1 | 5,34% | 5,63% | 6,47% | 7,16% | 6,21% | 5,34% | 6,72% |
| taux d'affaires nouvelles Mrh sachant monodétenteur en Auto N-1 | 2,49% | 2,51% | 2,75% | 3,34% | 2,85% | 2,49% | 3,50% |
| taux de résiliations Auto N sachant mono détenteur Auto N-1 | 13,33% | 4,07% | 12,12% | 11,44% | 13,33% | 12,86% | 12,59% |
| taux de résiliations Mrh N sachant mono détenteur Mrh N-1 | 10,34% | 2,75% | 9,31% | 9,49% | 10,20% | 10,34% | 10,06% |

Au niveau des taux d'affaires nouvelles, nous les avons estimés en prenant le minimum des taux sur les 6 années. A l'inverse, concernant les taux de résiliations, nous les avons estimés en prenant le maximum des taux sur les 6 années.

d. Les taux de résiliations Auto et Mrh sur affaires nouvelles sont estimés à partir des données historisées suivantes :

| | 2004 | 2005 | 2006 | 2007 | 2008 | Estimation |
|--|-------|-------|-------|-------|-------|--------------|
| taux de résiliations Auto sur affaires nouvelles | 6,31% | 5,39% | 6,45% | 7,23% | 6,63% | 7,23% |
| taux de résiliations Mrh sur affaires nouvelles | 2,88% | 2,63% | 2,84% | 3,22% | 2,60% | 3,22% |

Au niveau des taux de résiliations, nous les avons estimés en prenant le maximum des taux sur les 6 années.

Nous pouvons clairement affirmer que les deux derniers points ne permettent pas de remettre en cause de manière significative l'estimation des marges futures liées à la commercialisation du « contrat Fidelio ».

e. Les taux des marges techniques des contrats Auto et Mrh :

Nous nous sommes appuyés sur un historique basé sur les 5 dernières années pour estimer ces taux. Force est de constater qu'ils ne varient pas de manière significative, ce qui peut s'expliquer par une qualité du risque assurantiel et une politique de réassurance maîtrisée.

4. Estimation des marges futures

a. Taux d'affaire nouvelle des partenaires mono détenteurs à 5.3% en Auto et en Mrh

| Variable d'ajustement du modèle : le taux d'affaire nouvelle 5,3% | Stock début d'année | Souscription s nets de résiliations sur affaires nouvelles | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année |
|---|---------------------|--|---|--|-------------------|---|--|---------------------------------|---|--|---|---|--|-------------------|---|--|-------------------|
| Année | 0 | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | | |
| Nombre de souscription de contrat auto ds le cadre du contrat fidelio | 0 | 764 | 662 | 726 | 1 388 | 1 203 | 690 | 1 893 | 1 640 | 656 | 2 236 | 1 990 | 623 | 2 613 | 2 265 | 592 | 2 857 |
| Nombre de souscription de contrat mih ds le cadre du contrat fidelio | 0 | 883 | 792 | 862 | 1 654 | 1 483 | 841 | 2 324 | 2 084 | 821 | 2 905 | 2 604 | 801 | 3 405 | 3 053 | 782 | 3 835 |
| enveloppe de ristournes utilisées auto | | 28 974 | 26 619 | 27 538 | 54 157 | 51 272 | 26 173 | 77 445 | 74 120 | 24 876 | 98 996 | 95 313 | 23 643 | 118 956 | 114 985 | 22 471 | 137 457 |
| enveloppe de ristournes utilisées Mrh | | | | | | | | | | | | | | | | | |
| PART CAISSE | | 21 731 | 19 964 | 20 654 | 40 618 | 38 454 | 19 630 | 58 084 | 55 590 | 18 657 | 74 247 | 71 485 | 17 732 | 89 217 | 86 239 | 16 854 | 103 093 |
| PART PACIFICA | | 7 244 | 6 655 | 6 885 | 13 539 | 12 818 | 6 543 | 19 361 | 18 530 | 6 219 | 24 749 | 23 828 | 5 911 | 29 739 | 28 746 | 5 618 | 34 364 |
| Cotises AUTO ht peçues dans le cadre du contrat fidelio | | 260 769 | 232 787 | 247 844 | 480 631 | 436 694 | 235 560 | 671 254 | 612 031 | 223 885 | 835 916 | 764 749 | 212 789 | 977 538 | 896 483 | 202 242 | 1 098 725 |
| Cotises MRH ht peçues dans le cadre du contrat fidelio | | 156 565 | 148 795 | 152 787 | 301 583 | 295 328 | 149 100 | 444 428 | 439 887 | 145 502 | 585 389 | 582 752 | 141 991 | 724 743 | 724 196 | 138 565 | 862 761 |
| MARGE Assurance auto | | 14 262 | 12 731 | 13 555 | 26 286 | 23 828 | 12 883 | 36 711 | 33 472 | 12 244 | 45 717 | 41 825 | 11 638 | 53 462 | 49 029 | 11 061 | 60 090 |
| MARGE Assurance mih | | 9 258 | 8 798 | 9 034 | 17 832 | 17 463 | 8 816 | 26 279 | 26 010 | 8 603 | 34 614 | 34 458 | 8 396 | 42 854 | 42 821 | 8 193 | 51 015 |
| Marge Auto & Mrh Assurance | | 23 519 | 21 529 | 22 589 | 44 118 | 41 291 | 21 699 | 62 990 | 59 483 | 20 848 | 80 331 | 76 283 | 20 033 | 96 316 | 91 851 | 19 254 | 111 105 |
| GAIN ASSURANCE PACIFICA | | 16 276 | 14 875 | 15 704 | 30 579 | 28 473 | 15 156 | 43 629 | 40 953 | 14 629 | 55 582 | 52 454 | 14 123 | 66 577 | 63 104 | 13 636 | 76 740 |
| MARGE bancaire auto | | 42 785 | 38 194 | 40 664 | 78 858 | 71 485 | 38 649 | 110 134 | 100 417 | 36 733 | 137 150 | 125 474 | 34 913 | 160 386 | 147 088 | 33 182 | 180 270 |
| MARGE bancaire mih | | 27 773 | 26 395 | 27 103 | 53 497 | 52 388 | 26 449 | 78 837 | 78 031 | 25 810 | 103 842 | 103 374 | 25 188 | 128 561 | 128 464 | 24 580 | 153 044 |
| Marge Auto & Mrh bancaire | | 70 558 | 64 588 | 67 767 | 132 355 | 123 873 | 65 098 | 188 971 | 178 448 | 62 544 | 240 992 | 228 848 | 60 100 | 288 948 | 275 552 | 57 762 | 333 314 |
| GAIN CAISSE REGIONALE NDF | | 48 827 | 44 624 | 47 113 | 91 738 | 85 419 | 45 468 | 130 887 | 122 858 | 43 887 | 166 745 | 157 363 | 42 368 | 199 731 | 189 313 | 40 909 | 230 221 |
| Charge d'exploitation (coût annuel salariale de l'opération avec une inflation à | 147 180 | | | | 150 124 | | | 153 127 | | | 156 189 | | | 159 313 | | | 765 933 |
| Projection des concrétisations sur 12 mois AUTO | 576 | avec | 30% | de taux de traitements | 8% | de taux de concrétisation sur la base de | 2 000 | cibles récurrentes mensuelles p | | | marge engrangée sur 5 ans par pacifica | | | | | | 273 107 |
| Projection des concrétisations sur 12 mois MRH | 1 368 | avec | 30% | de taux de traitements | 10% | de taux de concrétisation sur la base de | 4 000 | cibles récurrentes mensuelles p | | | marge engrangée sur 5 ans par crca ndf | | | | | | 619 322 |
| | | | | | | | | | | | marge engrangée sur 5 ans par crca ndf nette des charges d'exploitation | | | | | | 53 389 |

b. Taux d'affaire nouvelle des partenaires mon détenteurs à 7.8% en Auto et en Mrh

| Variable d'ajustement du modèle : le taux d'affaire nouvelle 7,8% | Stock début d'année | Souscription s nets de résiliations sur affaires nouvelles | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année | Stock début d'année net des résiliations de l'année | Souscription s nets de résiliations sur affaires nouvelles | Stock fin d'année |
|---|---------------------|--|---|--|-------------------|---|--|---------------------------------|---|--|---|---|--|-------------------|---|--|-------------------|
| Année | 0 | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | | |
| Nombre de souscription de contrat auto ds le cadre du contrat fidelio | 0 | 1 121 | 972 | 1 040 | 2 011 | 1 743 | 964 | 2 707 | 2 346 | 894 | 3 240 | 2 808 | 829 | 3 637 | 3 152 | 769 | 3 921 |
| Nombre de souscription de contrat mih ds le cadre du contrat fidelio | 0 | 1 769 | 1 586 | 1 684 | 3 270 | 2 931 | 1 602 | 4 534 | 4 065 | 1 525 | 5 589 | 5 011 | 725 | 5 736 | 5 143 | 1 416 | 6 559 |
| enveloppe de ristournes utilisées auto | | 42 533 | 39 075 | 39 439 | 78 513 | 74 304 | 36 569 | 110 873 | 106 011 | 33 908 | 139 919 | 134 493 | 31 441 | 165 935 | 160 025 | 29 154 | 189 179 |
| enveloppe de ristournes utilisées Mrh | | | | | | | | | | | | | | | | | |
| PART CAISSE | | 31 900 | 29 306 | 29 579 | 58 885 | 55 728 | 27 427 | 83 155 | 79 508 | 25 431 | 104 939 | 100 870 | 23 581 | 124 451 | 120 019 | 21 865 | 141 884 |
| PART PACIFICA | | 10 633 | 9 769 | 9 860 | 19 628 | 18 576 | 9 142 | 27 718 | 26 503 | 8 477 | 34 980 | 33 623 | 7 860 | 41 484 | 40 006 | 7 288 | 47 295 |
| Cotises AUTO ht peçues dans le cadre du contrat fidelio | | 382 798 | 341 721 | 354 947 | 696 668 | 631 417 | 329 122 | 960 538 | 875 359 | 305 176 | 1 180 535 | 1 079 117 | 282 972 | 1 362 089 | 1 247 632 | 262 384 | 1 510 016 |
| Cotises MRH ht peçues dans le cadre du contrat fidelio | | 313 542 | 297 982 | 298 390 | 596 372 | 583 791 | 283 970 | 867 761 | 858 054 | 270 247 | 1 128 301 | 1 121 366 | 128 425 | 1 249 791 | 1 219 804 | 250 981 | 1 470 785 |
| MARGE Assurance auto | | 20 935 | 18 689 | 19 412 | 38 101 | 34 533 | 18 000 | 52 532 | 47 874 | 16 690 | 64 564 | 59 018 | 15 476 | 74 493 | 68 234 | 14 350 | 82 584 |
| MARGE Assurance mih | | 18 540 | 17 620 | 17 644 | 35 263 | 34 519 | 16 791 | 51 310 | 50 736 | 15 980 | 66 716 | 66 306 | 7 594 | 73 900 | 72 127 | 14 840 | 86 967 |
| Marge Auto & Mrh Assurance | | 39 475 | 36 309 | 37 056 | 73 364 | 69 052 | 34 791 | 103 843 | 98 610 | 32 670 | 131 280 | 125 324 | 23 070 | 148 393 | 140 360 | 29 190 | 169 551 |
| GAIN ASSURANCE PACIFICA | | 28 842 | 26 540 | 27 196 | 53 736 | 50 476 | 25 649 | 76 125 | 72 108 | 24 193 | 96 300 | 91 700 | 15 209 | 106 909 | 100 354 | 21 902 | 122 256 |
| MARGE bancaire auto | | 62 806 | 56 067 | 58 237 | 114 304 | 103 598 | 54 000 | 157 597 | 143 622 | 50 071 | 193 692 | 177 053 | 46 428 | 223 480 | 204 701 | 43 050 | 247 751 |
| MARGE bancaire mih | | 55 619 | 52 859 | 52 931 | 105 790 | 103 558 | 50 373 | 153 931 | 152 209 | 47 939 | 200 148 | 198 918 | 22 781 | 221 699 | 216 380 | 44 521 | 260 901 |
| Marge Auto & Mrh bancaire | | 118 425 | 108 926 | 111 168 | 220 093 | 207 156 | 104 373 | 311 528 | 295 831 | 98 010 | 393 840 | 375 971 | 69 209 | 445 179 | 421 081 | 87 571 | 508 652 |
| GAIN CAISSE REGIONALE NDF | | 86 525 | 79 619 | 81 589 | 161 208 | 151 428 | 76 946 | 228 374 | 216 323 | 72 578 | 288 901 | 275 101 | 45 628 | 329 728 | 301 062 | 65 706 | 366 768 |
| Charge d'exploitation (coût annuel salariale de l'opération avec une inflation à | 147 180 | | | | 150 124 | | | 153 127 | | | 156 189 | | | 159 313 | | | 765 933 |
| Projection des concrétisations sur 12 mois AUTO | 576 | avec | 30% | de taux de traitements | 8% | de taux de concrétisation sur la base de | 2 000 | cibles récurrentes mensuelles p | | | marge engrangée sur 5 ans par pacifica | | | | | | 455 327 |
| Projection des concrétisations sur 12 mois MRH | 1 368 | avec | 30% | de taux de traitements | 10% | de taux de concrétisation sur la base de | 4 000 | cibles récurrentes mensuelles p | | | marge engrangée sur 5 ans par crca ndf | | | | | | 1 365 980 |
| | | | | | | | | | | | marge engrangée sur 5 ans par crca ndf nette des charges d'exploitation | | | | | | 600 047 |

5. Conclusions

La mise en marché du « contrat Fidelio » dans le cadre d'une politique de multi équipement semble être une opération financière rentable sur 5 ans. Nous avons déduit de la marge l'enveloppe de ristournes et la charge d'exploitation annuelle. Donc, avec un taux d'affaire nouvelle des partenaires mono détenteurs à 5,3% en Auto et 7,8% en Mrh, les marges à 5 ans dépassent les six cent mille euros pour la Caisse régionale Nord De France et plus de quatre cent cinquante mille euros pour Pacifica.

SYNTHESE DES RESULTATS

Nous avons démontré l'intérêt de commercialiser l'offre « Contrat Fidelio », en présentant des cartes projetant le taux d'équipement et de pénétration en Assurance auto et/ou habitation des clients bancassurance à la CRCA NDF et en s'appuyant sur la conception d'un schéma de rétention en fonction de la détention d'un contrat Iard et, enfin, en modélisant la durée de vie d'un client bancassurance pouvant répondre à une problématique d'attrition.

Nous avons ensuite qualifié nos clients bancaires notamment ceux non assurés en Assurance Iard à l'aide d'un score. Nous sommes donc en mesure d'avoir une approche globale du client à travers la qualité de ce dernier au niveau de la banque et de l'assurance en termes de risque et de marge.

La qualité du client bancassurance peut être reprise aisément comme critère de tarif compte tenu des résultats de la modélisation de la tarification automobile. Mais nous sommes confrontés à la problématique de la CNIL qui ne nous permet pas d'utiliser certaines données notamment la qualité Bâle II. Nous proposons cependant, dans le cadre du CRM, i.e. de la gestion de la relation client, de reprendre cette dernière comme critère de ciblage. Plus précisément, le critère « qualité Bâle II du client » issu, bien sûr, de la note Bâle II imposée à toutes les banques et calculée tous les mois au niveau du groupe de risque, i.e., au niveau du partenaire commercial du marché des particuliers, peut être utilisée aisément. Ce dernier critère traduit ce que nous constatons depuis des années « un bon risque bancaire est un bon risque assurantiel.

Enfin, la mise en marché du « contrat Fidelio » est une opération financièrement rentable pour l'assureur Pacifica avec plus de 450 K€ de rentabilité sur 5 ans et plus de 600 K€ de rentabilité sur 5 ans pour la CRCA NDF.

CONCLUSION

La première partie nous permet de démontrer l'intérêt de la commercialisation du « contrat Fidelio ». La deuxième partie nous permet d'avoir une bonne approche globale du client bancassurance et nous donne la possibilité d'intervenir en amont du possible départ du client, à l'aide d'un modèle de durée de vie, en lui proposant le « contrat Fidelio ». La troisième partie nous donne la qualité du client bancaire mais non assuré à l'aide d'un score. La quatrième partie permet de nous éclairer sur l'aspect réduction tarifaire sur le produit d'appel automobile tout en respectant l'équilibre des résultats techniques. Enfin, la cinquième et dernière partie nous permet de conclure que la commercialisation du « contrat Fidelio » est une opération rentable.

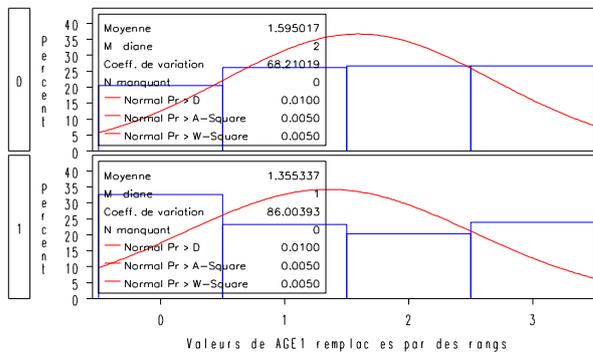
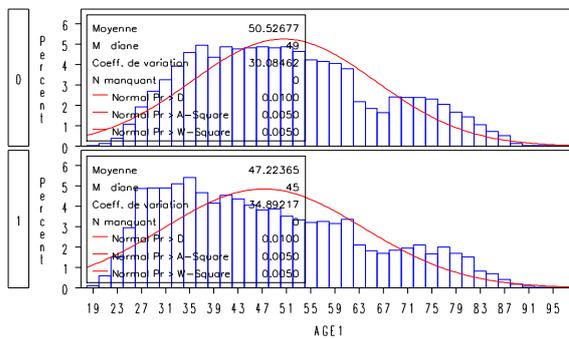
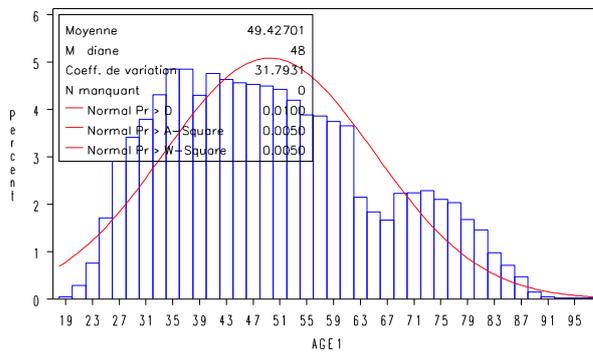
Il est évident que les différentes modélisations que l'on trouve dans ce mémoire, peuvent être améliorées, notamment le score, le modèle de durée de vie et la modélisation de la tarification automobile, mais elles constituent une bonne entrée en matière pour la suite des travaux au sein du crédit agricole et plus particulièrement avec Pacifica.

ANNEXE I

1. Statistiques univariées

a. Age de la personne la plus âgée dans le foyer

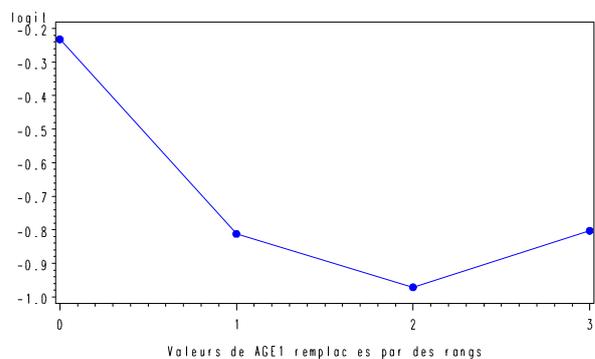
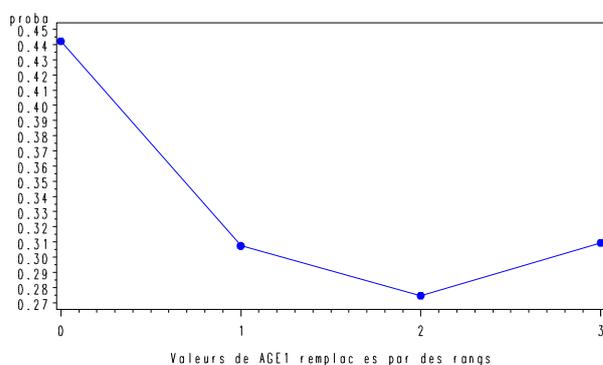
La variable **Age de la personne la plus âgée dans le foyer** a une distribution plutôt régulière, dont l'allure est éloignée de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.



Discrétisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : entre 18 et 36 ans (l'occurrence de référence),
- 1 : entre 37 et 47 ans,
- 2 : entre 48 et 59 ans,
- 3 : supérieure à 60 ans.

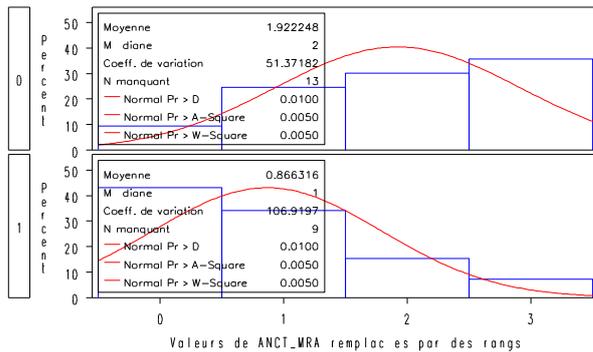
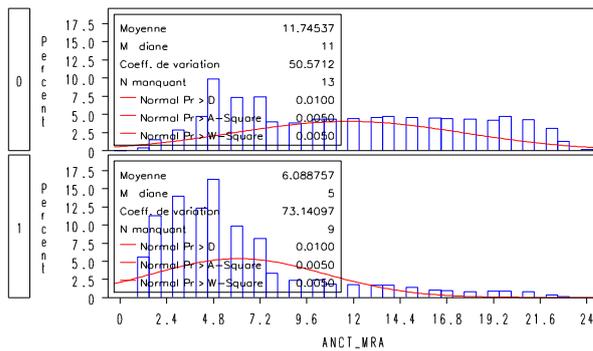
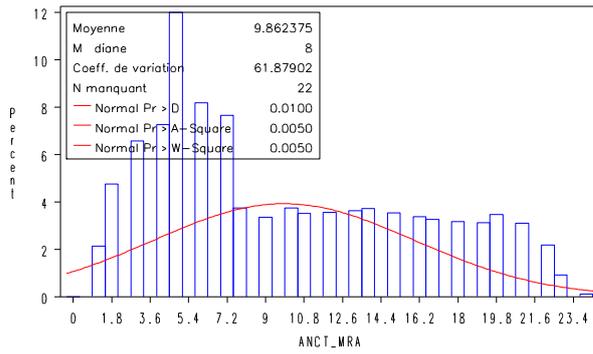
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur de l'âge de la personne la plus âgée dans le foyer. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



⇒ Ces deux graphes suggèrent un regroupement des modalités 1 et 3. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant l'âge de la personne la plus âgée dans le foyer sur les trois premières tranches.

b. Ancienneté de la relation du client au niveau de l'Assurance Iard

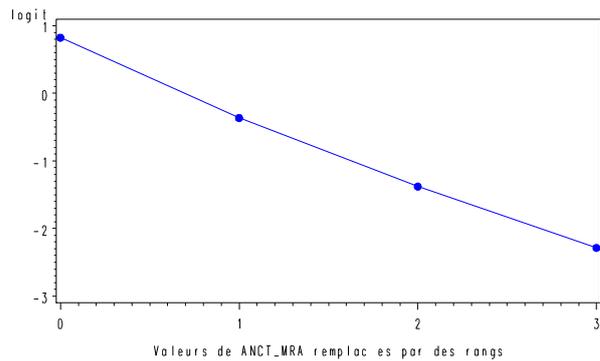
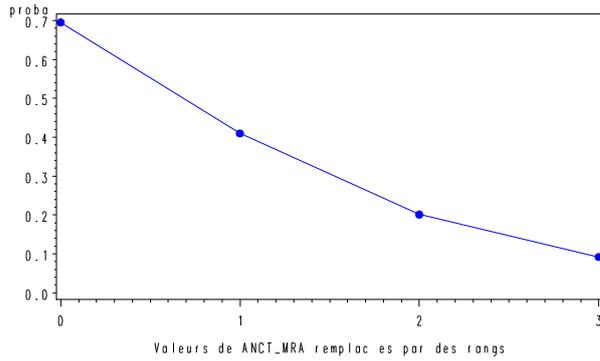
La variable **Ancienneté de la relation du client au niveau de l'Assurance Iard** a une distribution plutôt régulière, dont l'allure est éloignée de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.



Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : entre 0 et 4 ans (l'occurrence de référence),
- 1 : entre 5 et 7 ans,
- 2 : entre 8 et 14 ans,
- 3 : entre 15 et 24 ans.

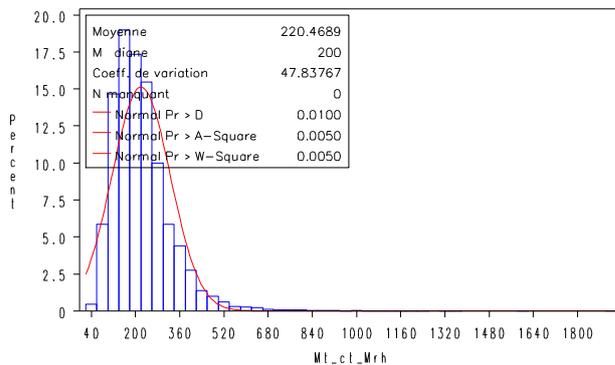
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur de l'ancienneté de la relation du client au niveau de l'Assurance Iard. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

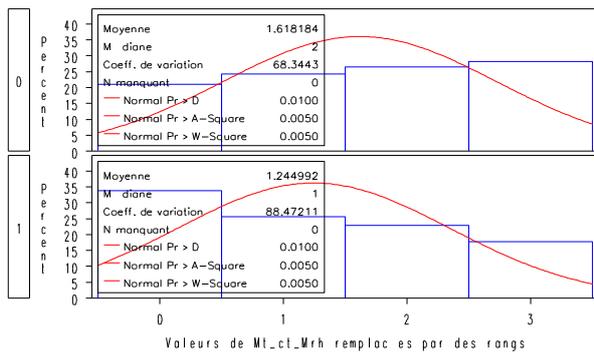
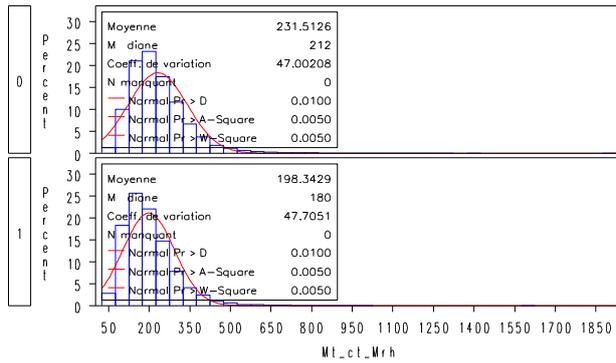


⇒ Ces deux graphes ne suggèrent aucun regroupement de ces modalités. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant l'ancienneté de sa relation au niveau de l'Assurance Iard.

c. Montant de la cotisation Multirisque habitation

La variable **Montant de la cotisation Multirisque habitation** a une distribution régulière, dont l'allure est proche de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

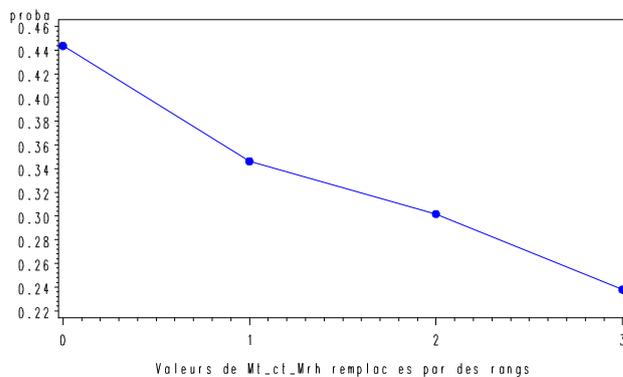


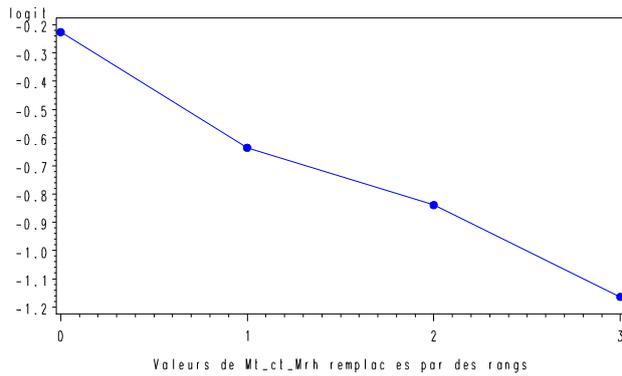


Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : entre 35 € et 150 € (l'occurrence de référence),
- 1 : entre 151 € et 200 €,
- 2 : entre 201 € et 268 €,
- 3 : supérieure ou égale à 269 €.

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du montant de la cotisation Multirisque habitation. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

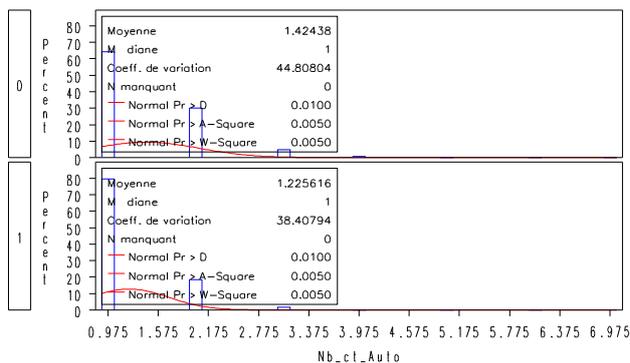
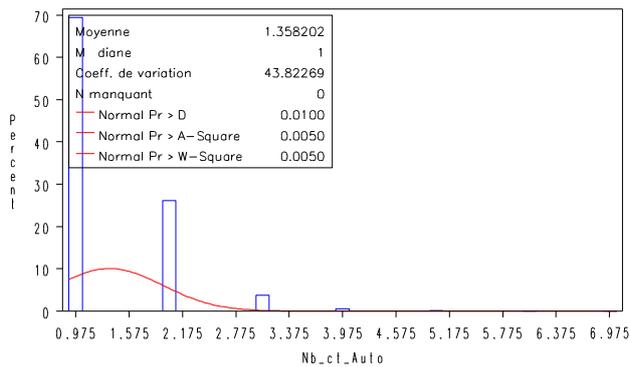


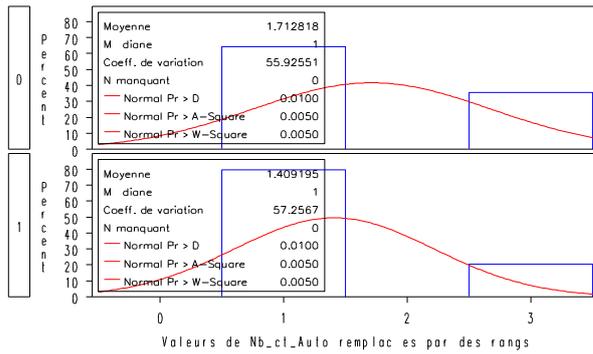


⇒ Ces deux graphes montrent des valeurs proches pour les occurrences 1 et 2 du Montant de la cotisation Mrh Hors Taxe, ce qui suggère un regroupement de ces deux modalités. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant le montant de la cotisation Mrh Hors Taxe.

d. Nombre de contrats Auto

La variable **Nombre de contrats Auto** a une distribution régulière. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

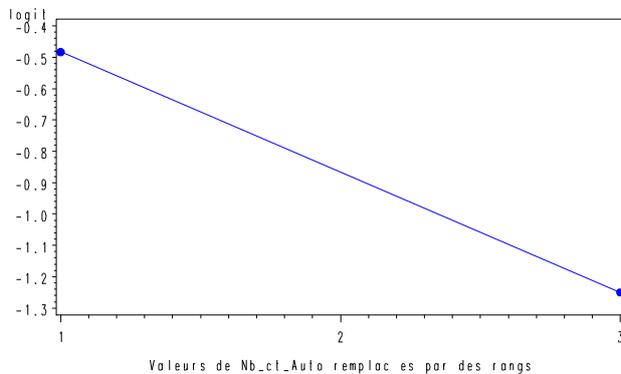
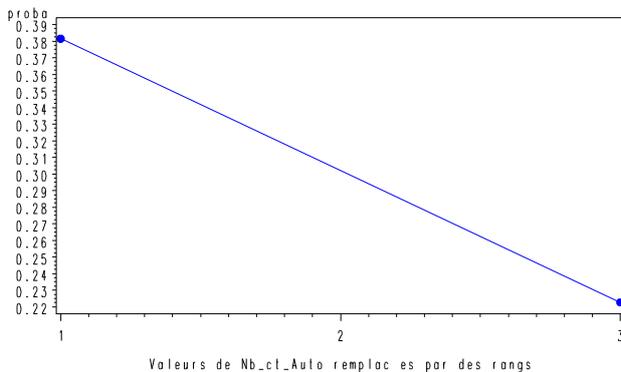




Discrétisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 :
- 1 : 1 (l'occurrence de référence),
- 2 :
- 3 : supérieure ou égale à 2.

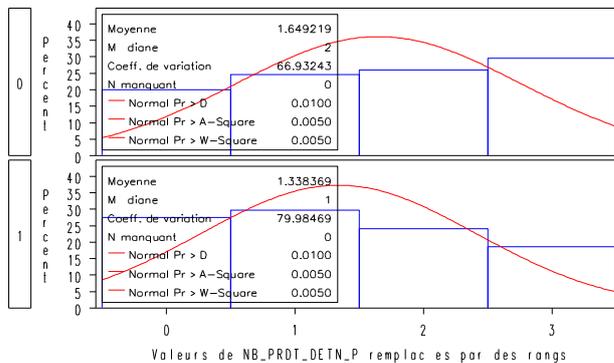
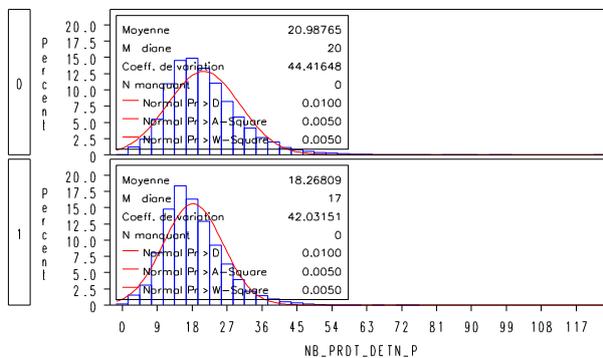
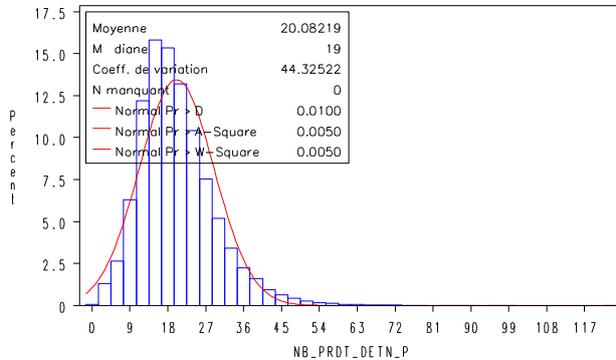
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du nombre de contrats Auto. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



⇒ Ces deux graphes montrent que seules deux occurrences sont à retenir 1 et 3 correspondant respectivement au fait d'avoir un contrat Auto ou plus d'un contrat Auto.

e. Nombre de produits détenus

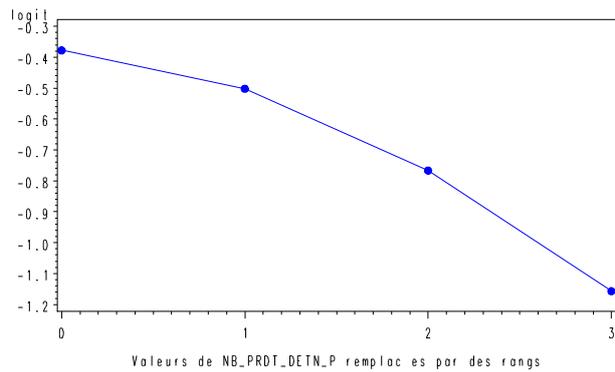
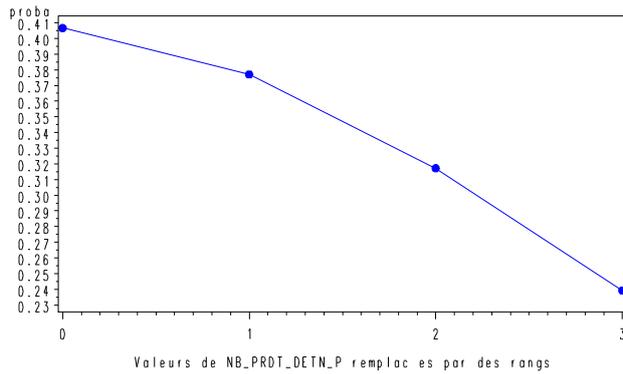
La variable **Nombre de produits détenus** a une distribution régulière, dont l'allure est proche de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.



Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : entre 0 et 13
- 1 : entre 14 et 18 (l'occurrence de référence),
- 2 : entre 19 et 24
- 3 : supérieure ou égale à 25.

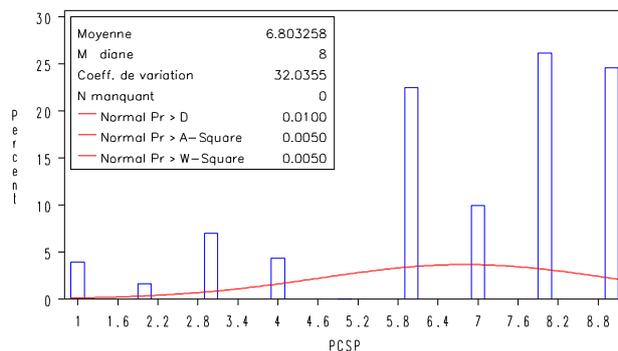
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du nombre de produits détenus. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

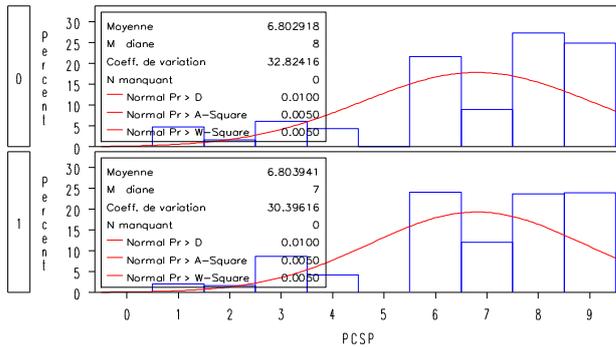
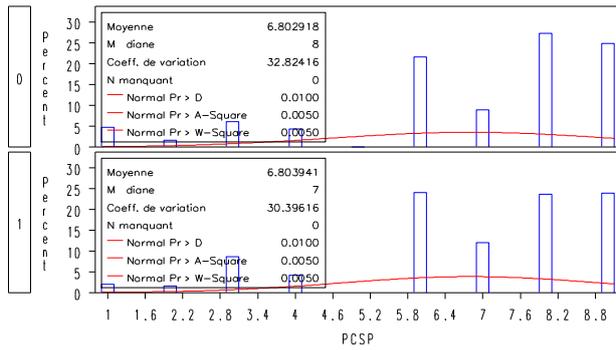


⇒ Ces deux graphes ne montrent pas de valeurs proches, ce qui ne suggère pas de regroupement des modalités. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant le nombre de produits détenus.

f. PCSP Catégorie Sociale Professionnelle de la personne la plus âgée du foyer

La variable **PCSP Catégorie Sociale Professionnelle de la personne la plus âgée du foyer** a une distribution irrégulière, ce qui n'a rien d'étonnant puisqu'il s'agit d'une variable qualitative. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

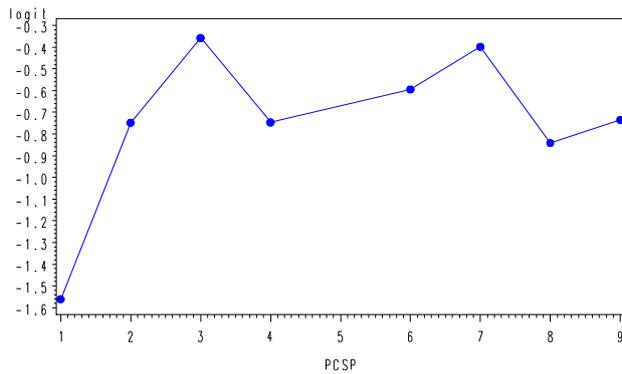
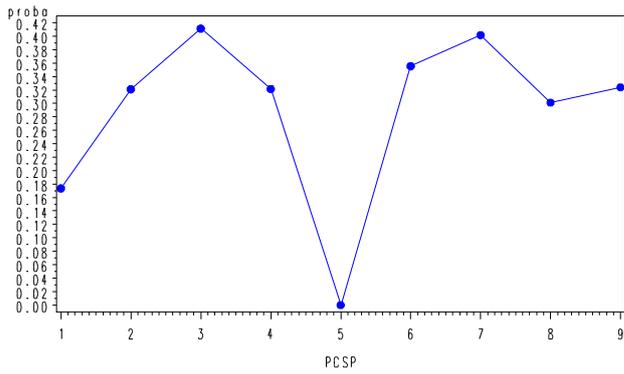




Discretisation de la variable à l'aide de la procédure proc rank en 10 occurrences :

- 0 : Non renseignée,
- 1 : Agriculteurs exploitants,
- 2 : Artisans commerçants, chefs d'entreprise,
- 3 : Autres personnes sans activités professionnelles,
- 4 : Cadres professions intellectuelles supérieures,
- 5 : Inconnue,
- 6 : Ouvriers,
- 7 : Professions intermédiaires,
- 8 : Retraités (l'occurrence de référence),
- 9 : Employés.

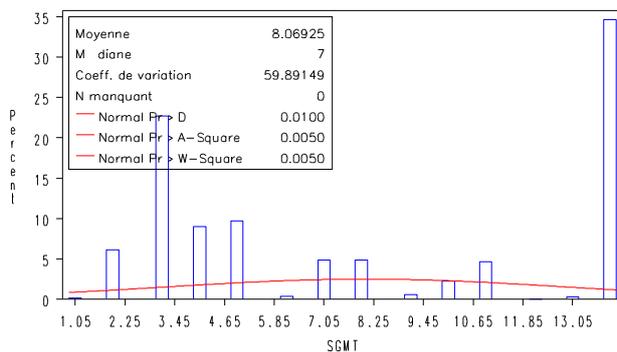
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur de la PCS. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

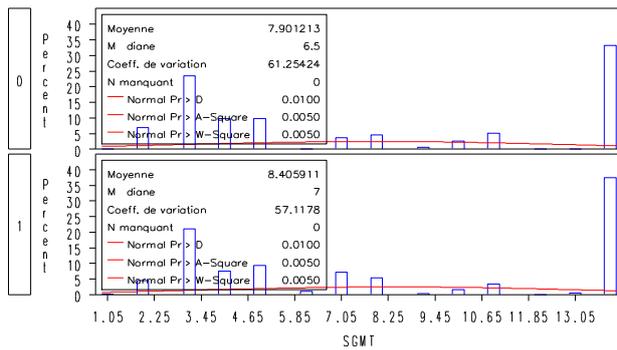


⇒ Ces deux graphes montrent des valeurs proches pour les valeurs 2, 4 et 9 de la PCS et pour les valeurs 3 et 7 ce qui suggère deux regroupements de ces différentes modalités.

g. Segmentation

La variable **Segmentation** a une distribution irrégulière, ce qui n'a rien d'étonnant puisqu'il s'agit d'une variable qualitative. On constate que certaines modalités de cette variable sont liées à la qualité et la valeur du client en Assurance Iard.

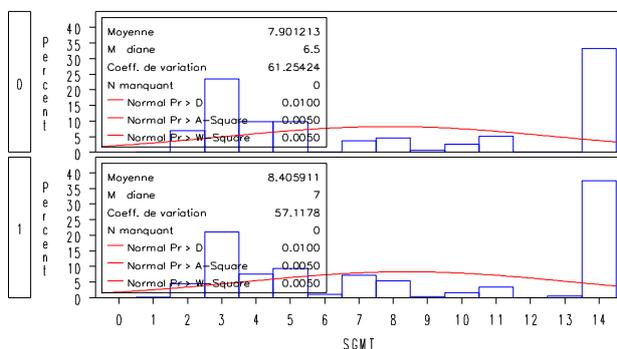


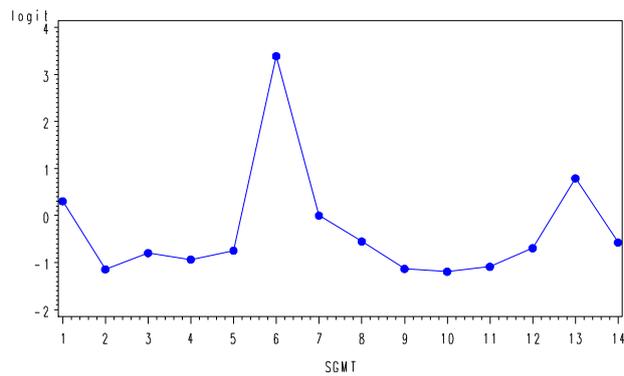
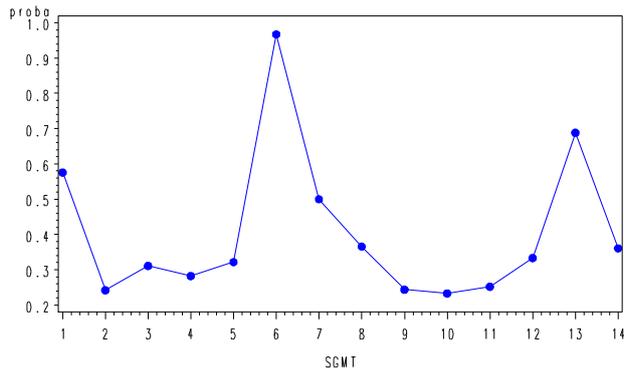


Discretisation de la variable à l'aide de la procédure proc rank en 15 occurrences :

- 0 : Non renseignée,
- 1 : Inconnue,
- 2 : Client Intermédiaire aisé,
- 3 : Client Intermédiaire dynamique,
- 4 : Client Intermédiaires épargnant,
- 5 : Client Intermédiaires mature,
- 6 : Nouveau client,
- 7 : Grand public futur Client Intermédiaire,
- 8 : Grand public modeste,
- 9 : Haut de gamme fortuné,
- 10 : Haut de gamme patrimonial,
- 11 : Haut de gamme traditionnel,
- 12 : Inactif,
- 13 : Jeune,
- 14 : Grand public cœur du grand public (l'occurrence de référence).

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur de la segmentation. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

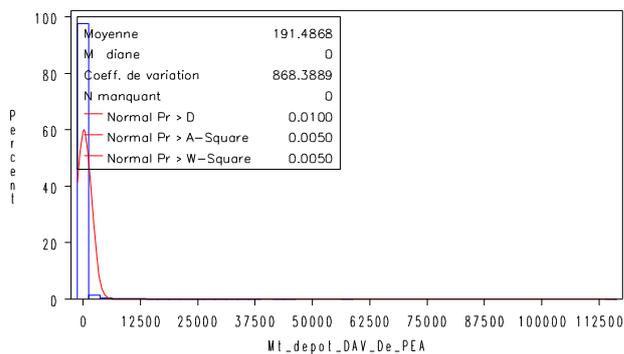


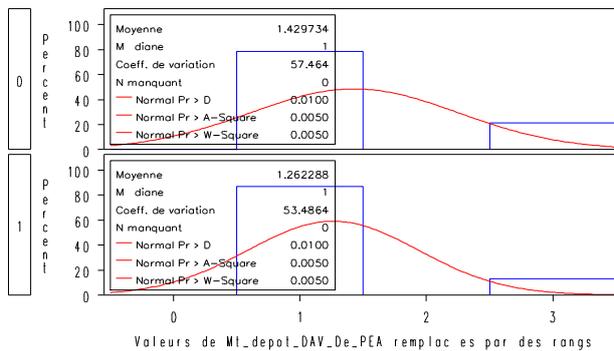
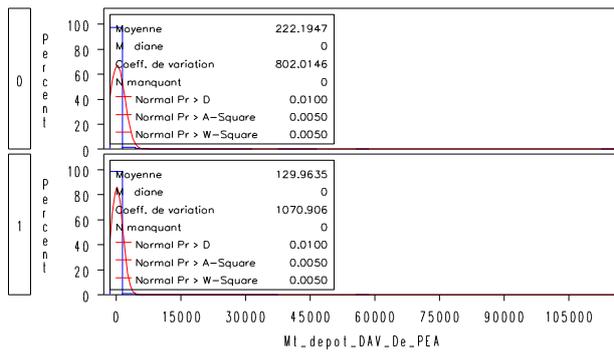


⇒ Ces deux graphes montrent des valeurs proches pour les valeurs 3, 4 et 5 et pour 2, 9, 10 et 11 de la segmentation, ce qui suggère un regroupement de ces différentes modalités.

h. Montant du dépôt sur le Dav de PEA

La variable **Montant du dépôt sur le Dav de PEA** a une distribution régulière, dont l'allure n'est pas proche de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

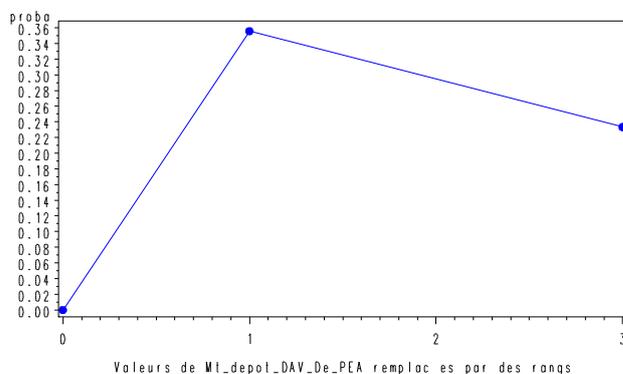


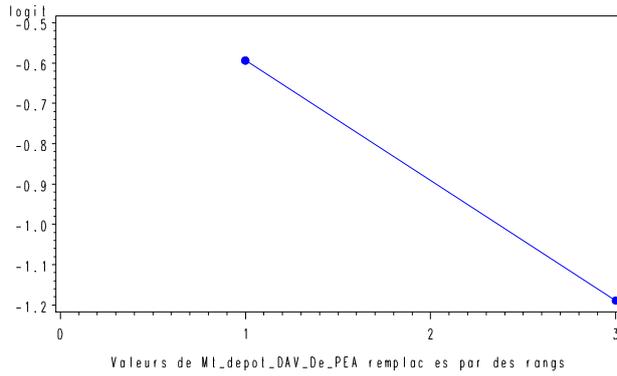


Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 :
- 1 : 0 € (l'occurrence de référence),
- 2 :
- 3 : supérieure à 0 €.

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du Montant du dépôt sur le Dav de PEA. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

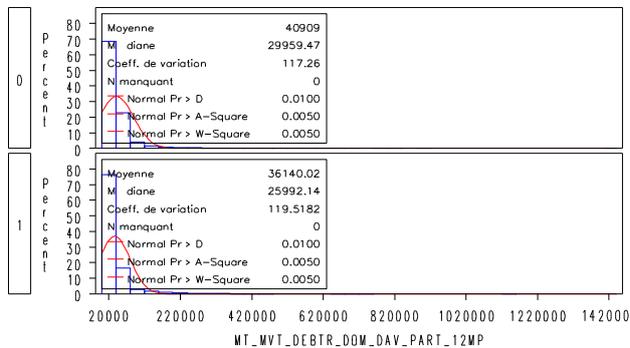
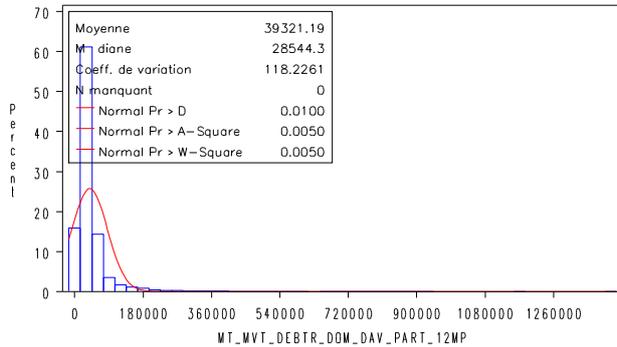


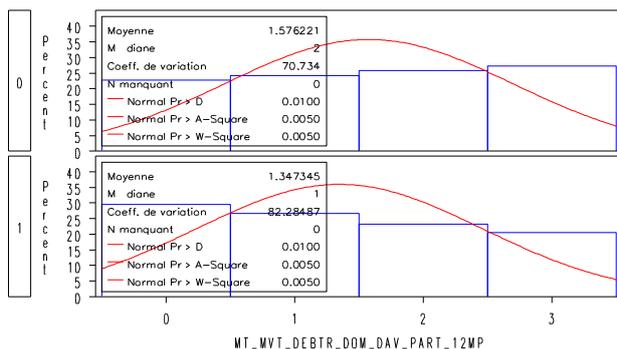


⇒ Ces deux graphes montrent que seules deux occurrences sont à retenir 3 et 1 correspondant respectivement au fait d'avoir un DAV de PEA ou pas.

i. Montant des mouvements débiteurs domiciliés sur le Dav des particuliers sur 12 mois glissants

La variable **Montant des mouvements débiteurs domiciliés sur le Dav des particuliers sur 12 mois glissants** a une distribution régulière, dont l'allure est proche de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

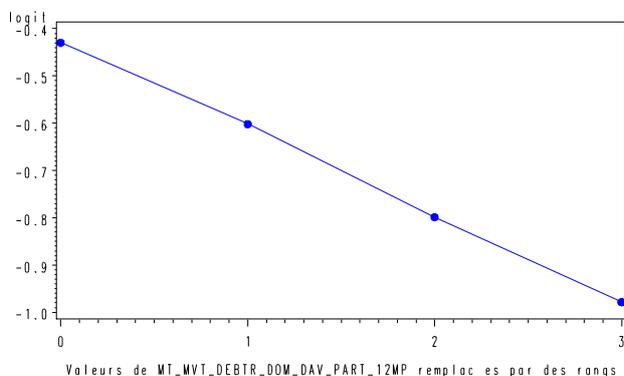
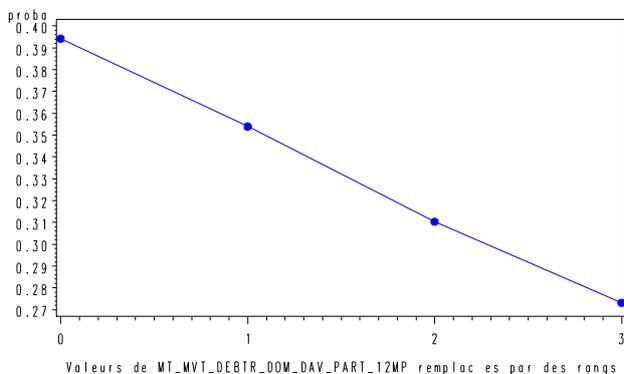




Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : entre 0 € et 18 691 € (l'occurrence de référence),
- 1 : entre 18 692 € et 28 542 €,
- 2 : entre 28 543 € et 42 979 €,
- 3 : supérieure ou égale à 42 980 €.

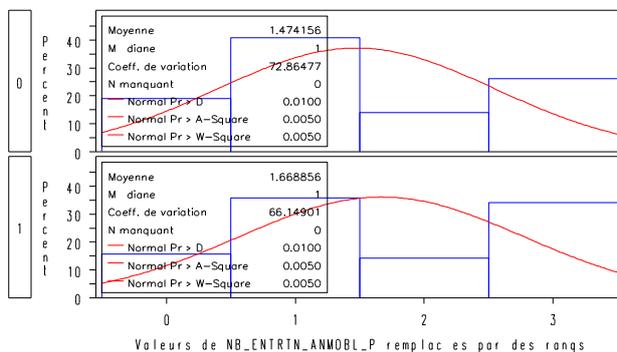
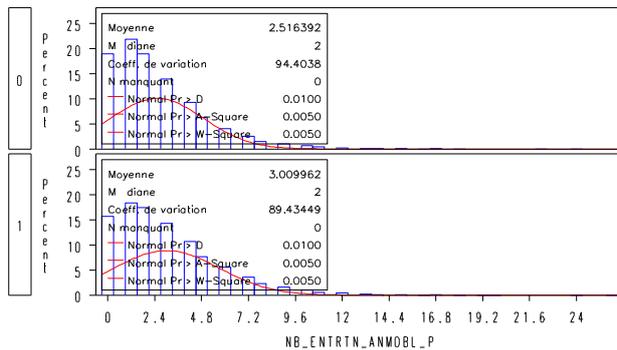
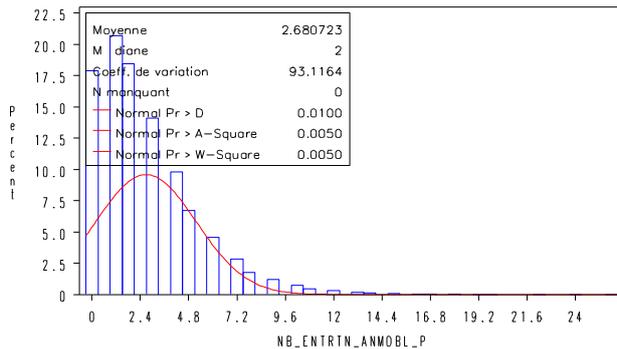
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du montant des mouvements débiteurs domiciliés sur le Dav des particuliers sur 12 mois glissants. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



⇒ Ces deux graphes ne montrent pas de valeurs proches, ce qui ne suggère pas de regroupement des modalités. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant le montant des mouvements débiteurs domiciliés sur le Dav des particuliers sur 12 mois glissants.

j. Nombre d'entretiens sur une année mobile

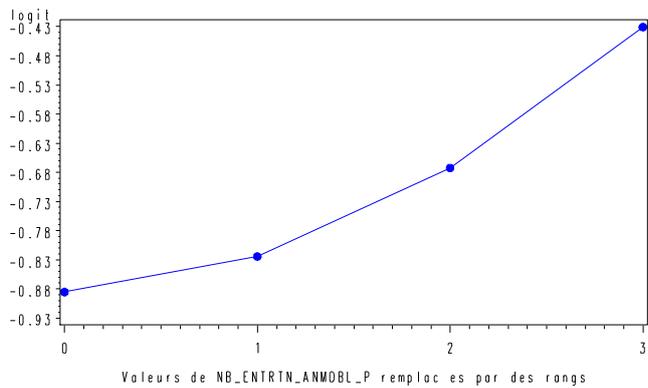
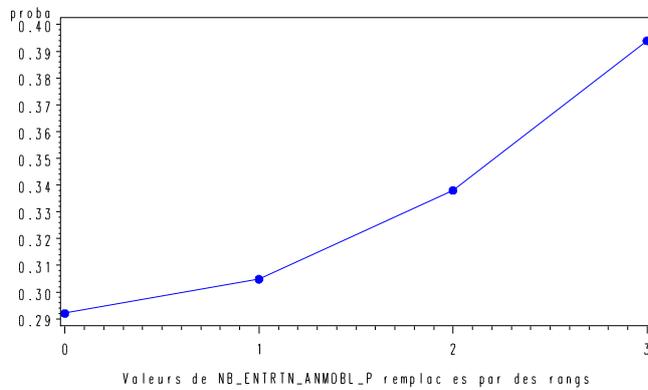
La variable **Nombre d'entretiens sur une année mobile** a une distribution irrégulière, dont l'allure est proche de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.



Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : 0,
- 1 : entre 1 et 2 (l'occurrence de référence),
- 2 : 3,
- 3 : supérieure ou égale à 4.

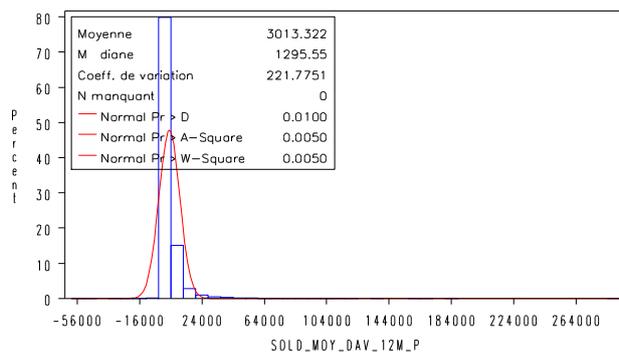
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du Nombre d'entretiens sur une année mobile. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

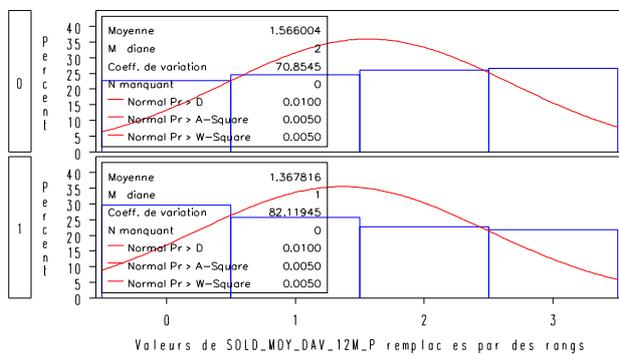
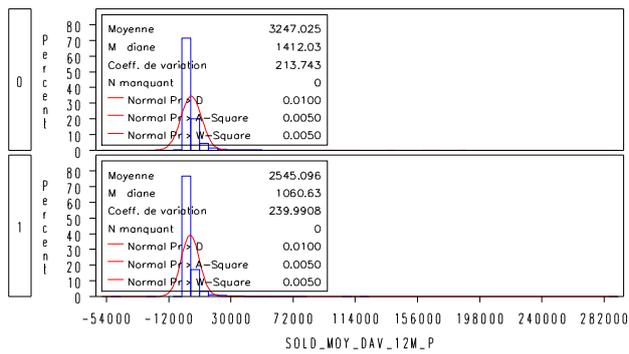


⇒ Ces deux graphes ne montrent pas de valeurs proches, ce qui ne suggère pas de regroupement des modalités. La qualité et la valeur du client en Assurance Iard est fonction croissante suivant le Nombre d'entretiens sur une année mobile.

k. Solde moyen du Dav sur 12 mois glissants

La variable **Solde moyen du Dav sur 12 mois glissants** a une distribution régulière, dont l'allure est proche de celle de loi normale. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

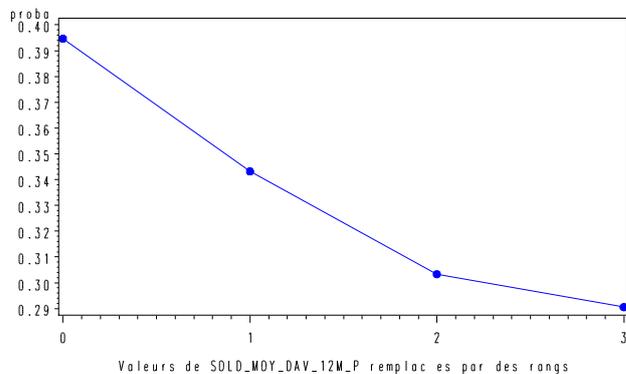


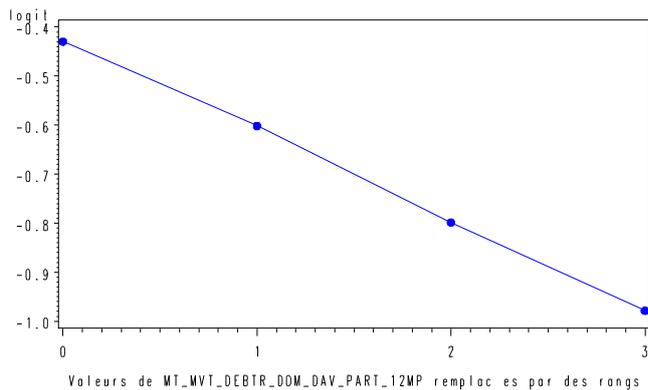


Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : inférieur ou égal 384 € (l'occurrence de référence),
- 1 : entre 385 € et 1 295 €,
- 2 : entre 1 296 € et 3 194 €,
- 3 : supérieure ou égale à 3 195 €.

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du Solde moyen du Dav sur 12 mois glissants. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

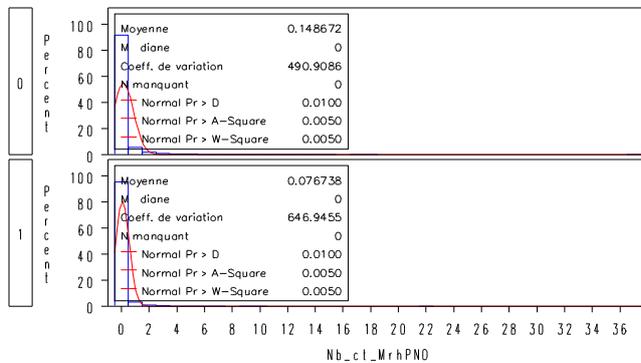
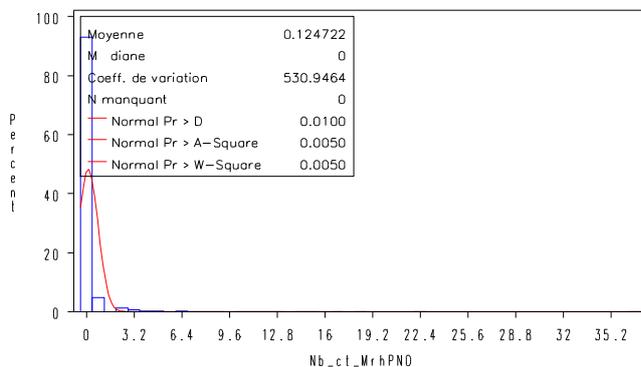


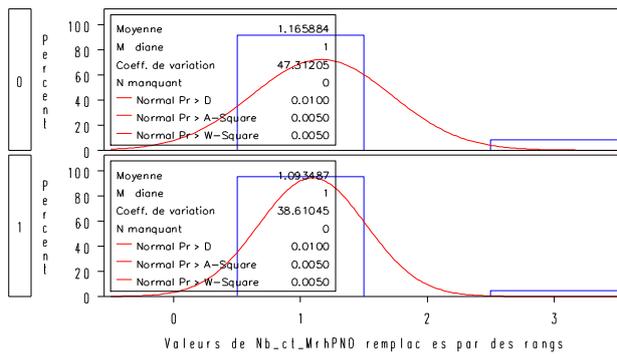


⇒ Ces deux graphes ne montrent pas de valeurs proches, ce qui ne suggère pas de regroupement des modalités. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant le Solde moyen du Dav sur 12 mois glissants.

1. Nombre de contrats Mrh PNO(Propriétaire Non Occupant)

La variable **Nombre de contrats Mrh PNO(Propriétaire Non Occupant)** a une distribution régulière. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

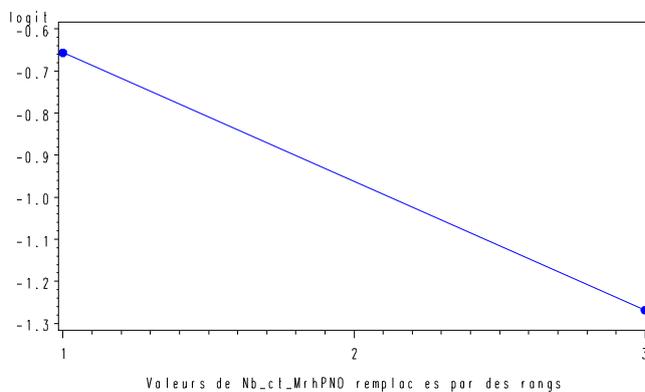
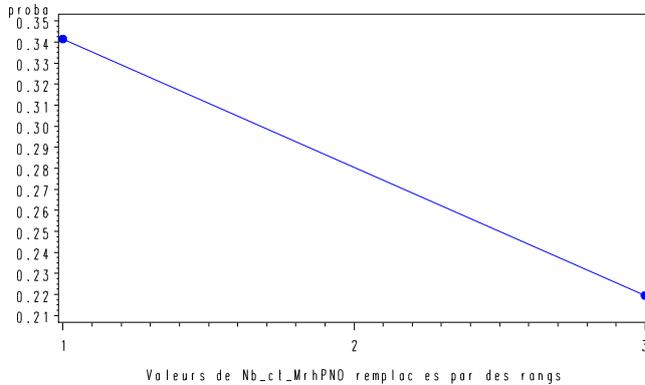




Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 :
- 1 : 0 (l'occurrence de référence),
- 2 :
- 3 : supérieure ou égale à 1.

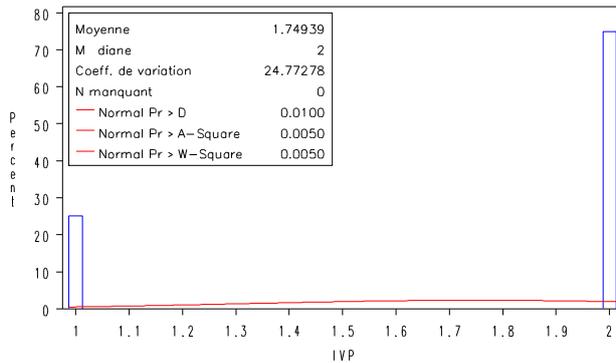
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du nombre de contrats Mrh PNO. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



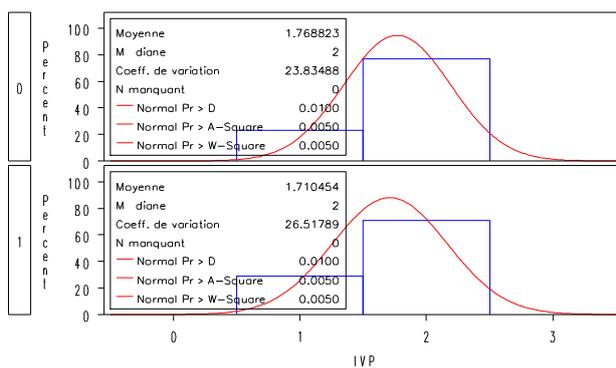
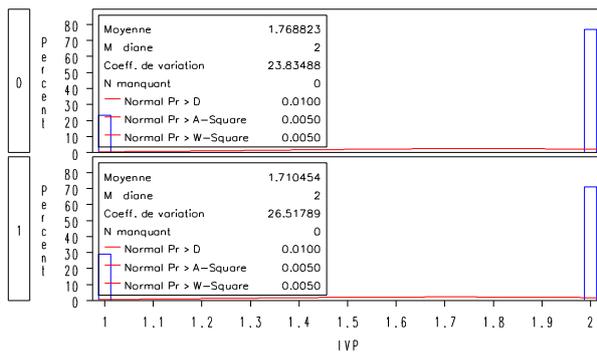
⇒ Ces deux graphes montrent que seules deux occurrences sont à retenir 1 et 3 correspondants respectivement au fait de ne pas avoir de contrat Mrh PNO ou d'en avoir au moins un.

m. Nombre de contrats IVP

La variable **Nombre de contrats IVP** a une distribution régulière. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.



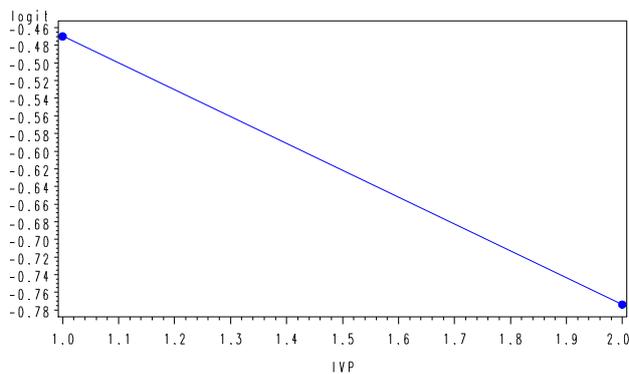
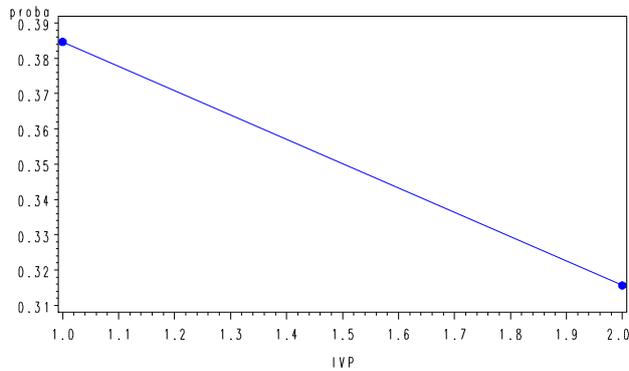
/



Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 :
- 1 : supérieure ou égale à 1,
- 2 : 0 (l'occurrence de référence).
- 3 :

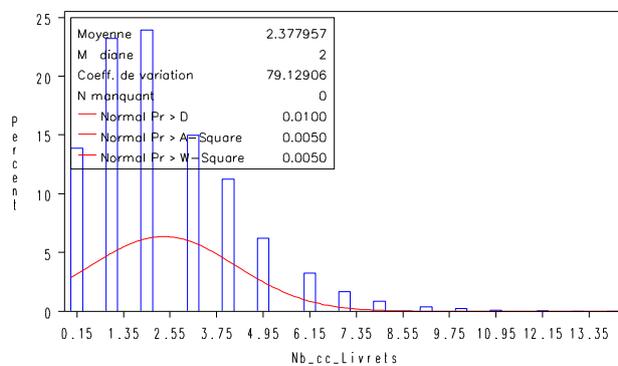
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du nombre de contrats IVP. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

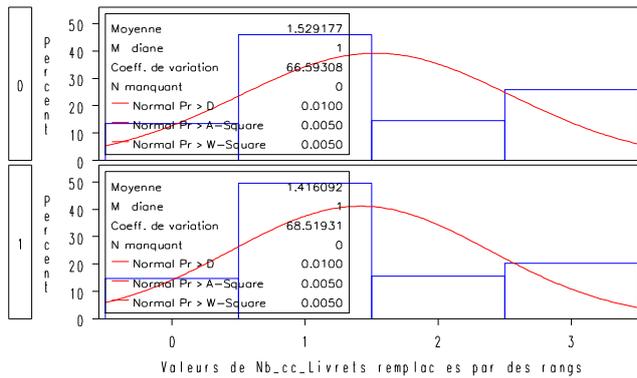
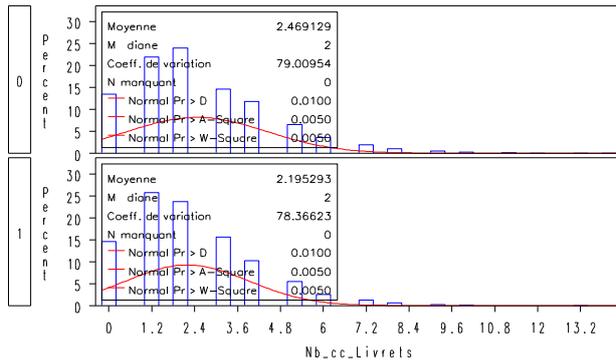


⇒ Ces deux graphes montrent que seules deux occurrences sont à retenir 1 et 2 correspondants respectivement au fait d'avoir au moins un contrat IVP ou de ne pas en avoir.

n. Nombre de livrets

La variable **Nombre de livrets** a une distribution irrégulière. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

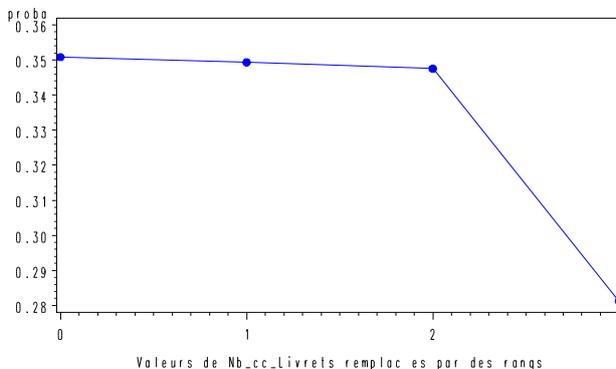


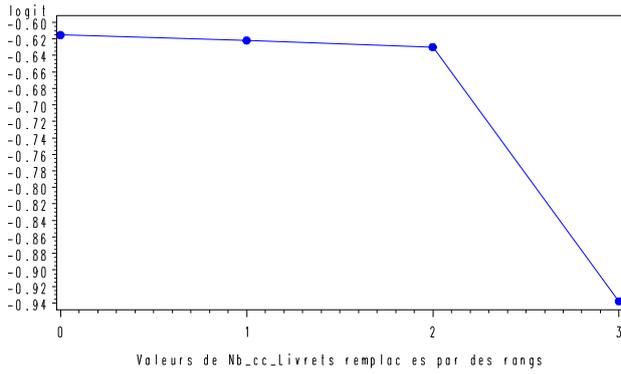


Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 : 0,
- 1 : 1 et 2 (l'occurrence de référence),
- 2 : 3,
- 3 : ≥ 4 .

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du nombre de livrets. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.

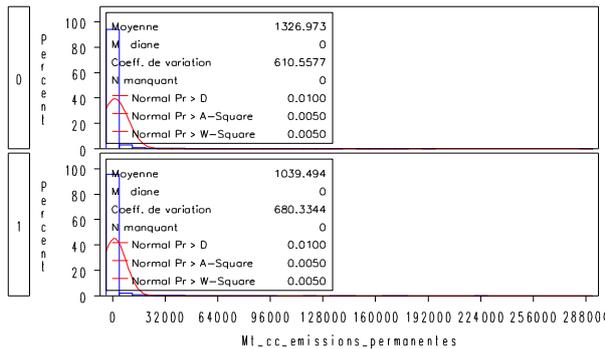
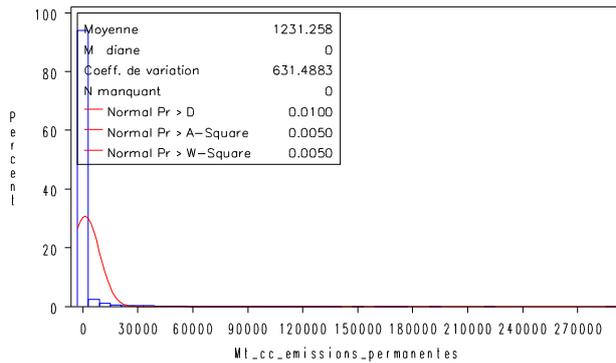


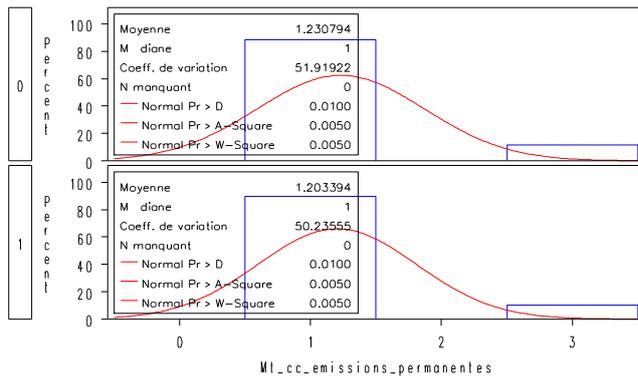


⇒ Ces deux graphes montrent des valeurs proches pour les occurrences 0, 1 et 2 de la variable nombre de livrets, ce qui suggère un regroupement de ces différentes modalités.

o. Montant des émissions permanentes

La variable **Montant des émissions permanentes** a une distribution régulière. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.

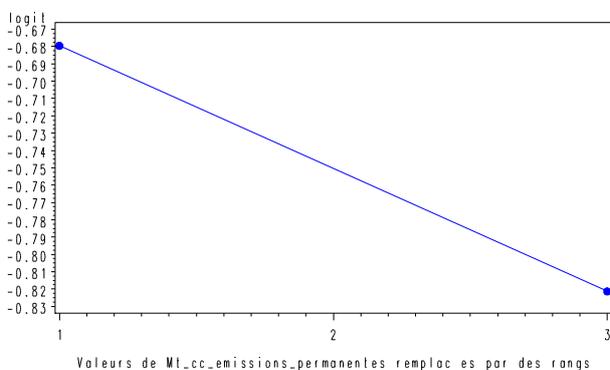
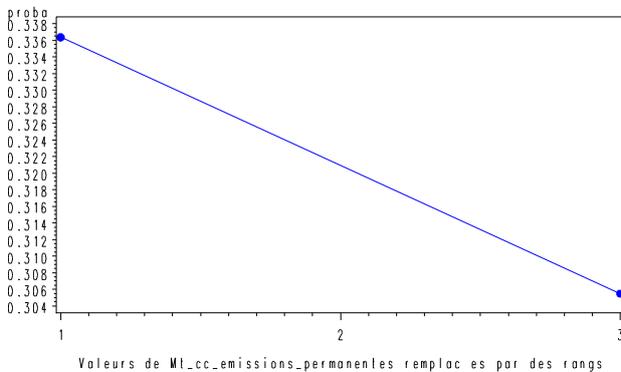




Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 :
- 1 : 0 (l'occurrence de référence),
- 2 :
- 3 : ≥ 1 .

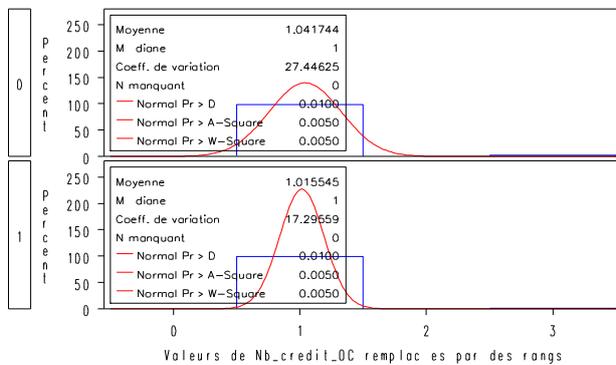
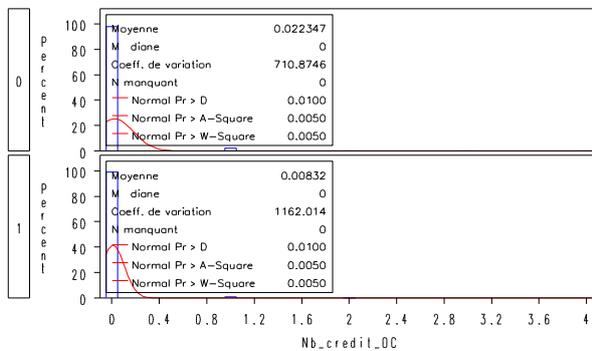
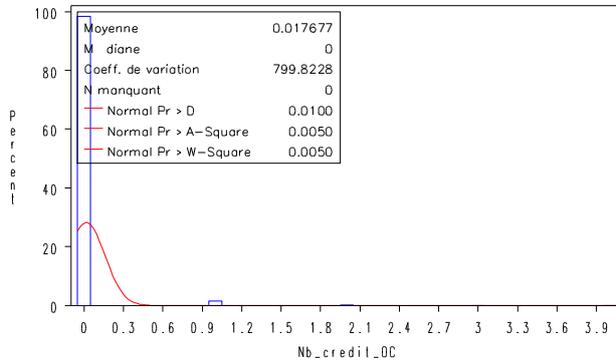
Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du montant des émissions permanentes. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



⇒ Ces deux graphes montrent que seules deux occurrences sont à retenir 1 et 3 correspondant respectivement au fait de ne pas avoir d'émissions permanentes ou d'en avoir.

p. Nombre de crédits OC(ouverture d'un crédit suite à un découvert autorisé pour un besoin temporaire le temps d'une transaction financière)

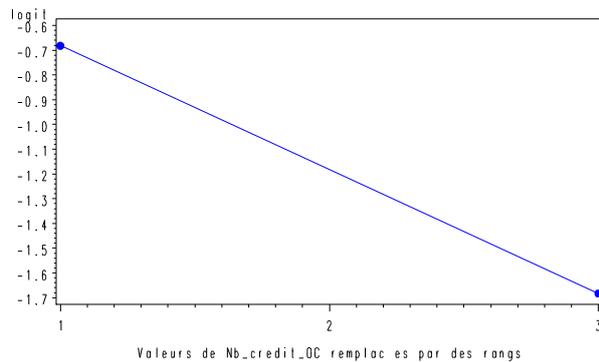
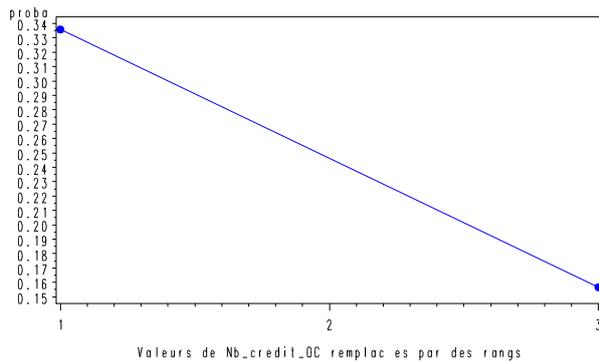
La variable **Nombre de crédits OC** a une distribution régulière. On constate que cette variable est liée à la qualité et la valeur du client en Assurance Iard.



Discretisation de la variable à l'aide de la procédure proc rank en 4 occurrences :

- 0 :
- 1 : 0 (l'occurrence de référence),
- 2 :
- 3 : supérieur ou égale à 1.

Présentation graphique de la probabilité d'être un client à fort potentiel de marge en Assurance Iard $P(Y=1)$ ou le logit $\ln P(Y=1) / P(Y=0)$ en fonction de la valeur du nombre de crédits OC. En vue d'une modélisation par régression logistique « logit », l'observation du graphique logit est plus appropriée, puisque les variables explicatives seront utilisées pour approcher le logit.



⇒ Ces deux graphes montrent des valeurs proches pour les occurrences 2 et 3 du nombre de connexions via à la BAM, ce qui suggère un regroupement de ces deux modalités. La qualité et la valeur du client en Assurance Iard est fonction décroissante suivant le nombre de connexions via à la BAM.

2. Sélection de variables après leur classification

La procédure VARCLUS résume son processus de classification sous la forme d'un tableau montrant l'affinement progressif de la classification, avec une proportion de la variance expliquée en augmentation régulière et importante, de 15.83% à 37.21%. On voit que, hormis le saut de 1 à 3 classes, l'évolution de la variance expliquée ne montre pas de point d'inflexion susceptible de suggérer un nombre de classes inférieur à 37. Qui plus est, on constate que c'est seulement à partir de 37 classes que le R2 minimum pour une variable, signalant quand il est bas une variable mal classée, fait un bond de 0.0273 à 0.1691. Tous ces éléments incitent à conserver le nombre de 37 classes, sans augmenter le seuil de seconde valeur propre de la procédure VARCLUS afin de diminuer le nombre de classes.

No cluster meets the criterion for splitting.

| Nombre de classifications | Variation totale expliquée par classifications | Proportion de variation expliquée par classifications | Proportion minimum expliquée par une classification | Valeur propre seconde maximum dans une classification | R carré minimum pour une variable | 1-R**2 Ratio maximum pour une variable |
|---------------------------|--|---|---|---|-----------------------------------|--|
| 1 | 20.583449 | 0.1583 | 0.1583 | 13.626600 | 0.0000 | |
| 2 | 32.035054 | 0.2464 | 0.2253 | 5.383095 | 0.0000 | 1.0000 |
| 3 | 36.597472 | 0.2815 | 0.2544 | 4.278662 | 0.0001 | 0.9999 |
| 4 | 40.720273 | 0.3132 | 0.2588 | 4.056911 | 0.0001 | 0.9999 |
| 5 | 44.787666 | 0.3445 | 0.2616 | 3.859453 | 0.0001 | 0.9999 |
| 6 | 48.430981 | 0.3725 | 0.2804 | 3.309359 | 0.0001 | 0.9999 |
| 7 | 51.481739 | 0.3960 | 0.2804 | 3.305712 | 0.0001 | 0.9999 |
| 8 | 54.572717 | 0.4198 | 0.2616 | 3.284187 | 0.0001 | 1.0000 |
| 9 | 57.758668 | 0.4443 | 0.2546 | 3.227877 | 0.0003 | 0.9998 |
| 10 | 60.984724 | 0.4691 | 0.2546 | 2.662129 | 0.0003 | 0.9998 |
| 11 | 63.208454 | 0.4862 | 0.2877 | 2.262273 | 0.0003 | 0.9998 |
| 12 | 65.477258 | 0.5037 | 0.2879 | 2.214946 | 0.0003 | 0.9998 |
| 13 | 67.638217 | 0.5203 | 0.2587 | 2.135024 | 0.0003 | 0.9998 |
| 14 | 69.510791 | 0.5347 | 0.2587 | 2.038068 | 0.0003 | 0.9998 |
| 15 | 71.418609 | 0.5494 | 0.2653 | 2.010146 | 0.0003 | 0.9998 |
| 16 | 73.423647 | 0.5648 | 0.2653 | 2.003994 | 0.0003 | 0.9998 |
| 17 | 75.422579 | 0.5802 | 0.2653 | 2.001290 | 0.0003 | 0.9998 |
| 18 | 77.422200 | 0.5956 | 0.2653 | 1.982136 | 0.0075 | 0.9943 |
| 19 | 79.366818 | 0.6105 | 0.2653 | 1.973826 | 0.0075 | 0.9943 |
| 20 | 81.339504 | 0.6257 | 0.2653 | 1.957706 | 0.0075 | 0.9943 |
| 21 | 83.290782 | 0.6407 | 0.2653 | 1.928864 | 0.0075 | 0.9943 |
| 22 | 85.212986 | 0.6555 | 0.2653 | 1.760830 | 0.0075 | 0.9943 |
| 23 | 86.973803 | 0.6690 | 0.2653 | 1.743415 | 0.0075 | 0.9943 |
| 24 | 88.715398 | 0.6824 | 0.3171 | 1.670785 | 0.0076 | 0.9942 |
| 25 | 90.359932 | 0.6951 | 0.3171 | 1.663227 | 0.0076 | 0.9942 |
| 26 | 91.973164 | 0.7075 | 0.3599 | 1.652598 | 0.0240 | 0.9802 |
| 27 | 93.625761 | 0.7202 | 0.3599 | 1.593014 | 0.0240 | 0.9802 |
| 28 | 95.037841 | 0.7311 | 0.3599 | 1.421874 | 0.0240 | 0.9802 |
| 29 | 96.406413 | 0.7416 | 0.3599 | 1.351344 | 0.0240 | 0.9802 |
| 30 | 97.582252 | 0.7506 | 0.3599 | 1.331455 | 0.0240 | 0.9804 |

| Nombre de classifications | Variation totale expliquée par classifications | Proportion de variation expliquée par classifications | Proportion minimum expliquée par une classification | Valeur propre seconde maximum dans une classification | R carré minimum pour une variable | 1-R**2 Ratio maximum pour une variable |
|---------------------------|--|---|---|---|-----------------------------------|--|
| 31 | 98.665663 | 0.7590 | 0.3441 | 1.230352 | 0.0240 | 0.9814 |
| 32 | 99.806126 | 0.7677 | 0.3908 | 1.126814 | 0.0253 | 0.9788 |
| 33 | 100.949267 | 0.7765 | 0.3004 | 1.069973 | 0.0253 | 0.9788 |
| 34 | 101.972668 | 0.7844 | 0.3004 | 1.069973 | 0.0253 | 0.9778 |
| 35 | 103.038991 | 0.7926 | 0.3721 | 1.069973 | 0.0273 | 0.9787 |
| 36 | 103.930670 | 0.7995 | 0.3721 | 1.004864 | 0.0273 | 0.9787 |
| 37 | 104.917154 | 0.8071 | 0.3721 | 0.965550 | 0.1691 | 0.8844 |

Le tableau suivant indique la répartition des variables analysées suivant les 37 classes finalement obtenues (« cluster 1 » à « cluster 37 »).

La colonne « R2 with own cluster » (« R carré avec propre classification ») indique le coefficient de corrélation entre la variable et sa classe. La colonne « R2 with next closest » (« R carré avec le plus proche ») indique le coefficient de corrélation entre la variable et la classe (autre que la sienne) la plus proche. La colonne suivante est le rapport entre 1-R2 (own cluster) et 1-R2 (next closest), qui sera d'autant plus petit que la variable sera bien classée et pourrait même être >1 si la variable était particulièrement mal classée.

| 37 Clusters | | R carré avec | | 1-R**2 Ratio |
|----------------|--------------------------------|-----------------------|----------------|--------------|
| Classification | Variable | Propre classification | Le plus proche | |
| Cluster 1 | Mt_cc_Epargne_Logement | 0.7355 | 0.3537 | 0.4092 |
| | Mt_cc_Epargne_Total | 0.9308 | 0.6116 | 0.1782 |
| | Mt_collecte_cam | 0.9354 | 0.6110 | 0.1661 |
| | SOLDE_COLLECTE_P | 0.8125 | 0.5254 | 0.3951 |
| | SOLDE_EP_BIL_P | 0.9126 | 0.5907 | 0.2135 |
| | Nb_cc_Epargne_Logement | 0.6782 | 0.3648 | 0.5066 |
| | NB_EVT_COMNTR_ANMOBL_P | 0.3308 | 0.1890 | 0.8251 |
| Cluster 2 | NB_EVT_SPECIF_ANMOBL_P | 0.2741 | 0.1792 | 0.8844 |
| | MT_MOY_M_PAINT_CHQ_P | 0.4197 | 0.1583 | 0.6894 |
| | MT_MOY_PAINT_M_CARTE_P | 0.4275 | 0.1545 | 0.6772 |
| | MT_MVT_CRDTR_DOM_DAV_PART_12MP | 0.6827 | 0.1692 | 0.3819 |
| | MT_MVT_CRED_DAV_PART_12M_P | 0.8387 | 0.2168 | 0.2060 |
| | MT_MVT_DEB_DAV_PART_12M_P | 0.8426 | 0.2184 | 0.2014 |

| 37 Clusters | | R carré avec | | 1-R**2 Ratio |
|----------------|--------------------------------|--------------------------|-------------------|-----------------|
| Classification | Variable | Propre classification | Le plus proche | |
| | MT_MVT_DEBTR_DOM_DAV_PART_12MP | 0.8257 | 0.1780 | 0.2121 |
| | NB_MVT_CRED_DAV_PART_12M_P | 0.5099 | 0.2016 | 0.6138 |
| | NB_MVT_DEB_DAV_PART_12M_P | 0.6044 | 0.2283 | 0.5127 |
| | NB_PRDT_DET_N_P | 0.4110 | 0.3133 | 0.8577 |
| Cluster 3 | Mt_ca_comptes_titres_pea_na | 0.8618 | 0.1688 | 0.1663 |
| | Mt_depot_DAV_De_PEA | 0.8560 | 0.2460 | 0.1910 |
| | Nb_ca_comptes_titres_pea_na | 0.9014 | 0.1753 | 0.1196 |
| | Nb_depot_DAV_De_PEA | 0.9022 | 0.2613 | 0.1323 |
| Cluster 4 | Mt_ADI_annee | 0.8952 | 0.2350 | 0.1370 |
| | Mt_ADI_hab | 0.8399 | 0.2141 | 0.2037 |
| | Mt_ADI_Total | 0.8952 | 0.2350 | 0.1370 |
| | Mt_credit_hab | 0.8254 | 0.2185 | 0.2235 |
| | Mt_Credit_Total | 0.8610 | 0.2488 | 0.1851 |
| | Nb_ADI_hab | 0.8155 | 0.1970 | 0.2298 |
| | Nb_ADI_Total | 0.8566 | 0.2536 | 0.1921 |
| | Nb_credit_hab | 0.8035 | 0.2035 | 0.2467 |
| | Nb_Credit_Total | 0.7073 | 0.2015 | 0.3665 |
| Cluster 5 | Mt_ADI_equipement | 0.9914 | 0.0400 | 0.0089 |
| | Mt_credit_equipement | 0.9914 | 0.0386 | 0.0089 |
| | Nb_ADI_equipement | 0.9914 | 0.0400 | 0.0089 |
| | Nb_credit_equipement | 0.9914 | 0.0386 | 0.0089 |
| Cluster 6 | Mt_ADI_conso_autres | 0.9639 | 0.0575 | 0.0383 |
| | Mt_credit_conso_autres | 0.9639 | 0.0549 | 0.0382 |
| | Nb_ADI_conso_autres | 0.9639 | 0.0575 | 0.0383 |
| | Nb_credit_conso_autres | 0.9639 | 0.0549 | 0.0382 |
| Cluster 7 | Mt_ADI_conso_auto | 0.9767 | 0.1280 | 0.0268 |
| | Mt_credit_conso_auto | 0.9767 | 0.1190 | 0.0265 |
| | Nb_ADI_conso_auto | 0.9767 | 0.1280 | 0.0268 |
| | Nb_credit_conso_auto | 0.9767 | 0.1190 | 0.0265 |

| 37 Clusters | | R carré avec | | 1-R**2 Ratio |
|----------------|--------------------------|--------------------------|-------------------|-----------------|
| Classification | Variable | Propre classification | Le plus proche | |
| Cluster 8 | Mt_dav | 0.9359 | 0.2131 | 0.0815 |
| | Mt_dav_clientele | 0.9359 | 0.2131 | 0.0814 |
| | Mt_dav_oc_privé | 0.9320 | 0.2106 | 0.0861 |
| | SOLD_FIN_MOIS_DAV_P | 0.9065 | 0.2273 | 0.1210 |
| | SOLD_MOY_DAV_12M_P | 0.7085 | 0.3545 | 0.4516 |
| | SOLD_MOY_DAV_PART_12M_P | 0.7217 | 0.3441 | 0.4242 |
| Cluster 9 | Mt_CT_Predica | 0.9085 | 0.2969 | 0.1302 |
| | Mt_CT_predica_euro | 0.8433 | 0.1322 | 0.1806 |
| | Mt_CT_Total | 0.9084 | 0.2968 | 0.1302 |
| | Nb_CT_Predica | 0.9353 | 0.2947 | 0.0917 |
| | Nb_CT_predica_euro | 0.8318 | 0.0815 | 0.1831 |
| | Nb_CT_Total | 0.9353 | 0.2946 | 0.0917 |
| Cluster 10 | Mt_ADI_tresorerie | 0.8162 | 0.0014 | 0.1840 |
| | Mt_credit_tresorerie | 0.8162 | 0.0041 | 0.1845 |
| | Nb_ADI_tresorerie | 0.8162 | 0.0014 | 0.1840 |
| | Nb_credit_tresorerie | 0.8162 | 0.0041 | 0.1845 |
| Cluster 11 | Nb_dav | 0.8509 | 0.2170 | 0.1904 |
| | Nb_dav_clientele | 0.9525 | 0.2022 | 0.0595 |
| | Nb_dav_oc_privé | 0.9510 | 0.2028 | 0.0615 |
| | NB_ENF_CLI | 0.4860 | 0.1955 | 0.6389 |
| Cluster 12 | Mt_contrat_OPEN | 0.9424 | 0.0988 | 0.0639 |
| | NB_CARTE_OPEN_P | 0.6814 | 0.0554 | 0.3373 |
| | Nb_contrat_OPEN | 0.9424 | 0.0988 | 0.0639 |
| Cluster 13 | MT_MOY_RETRAIT_M_CARTE_P | 0.1691 | 0.0347 | 0.8607 |
| | NB_CARTE_DI_P | 0.4593 | 0.0913 | 0.5950 |
| | NB_CARTE_DS_CSCA_P | 0.6711 | 0.0474 | 0.3453 |
| | NB_CARTE_PAIEMENT_P | 0.4586 | 0.2453 | 0.7174 |
| | NB_CONTRAT_BAM_P | 0.3243 | 0.0735 | 0.7293 |
| | NB_CSCA_P | 0.7071 | 0.0594 | 0.3114 |

| 37 Clusters | | R carré avec | | 1-R**2 Ratio |
|----------------|---------------------------------|--------------------------|-------------------|-----------------|
| Classification | Variable | Propre classification | Le plus proche | |
| Cluster 14 | Mt_ca_comptes_titres_pea_abonne | 0.9988 | 0.0686 | 0.0013 |
| | Nb_ca_comptes_titres_pea_abonne | 0.9988 | 0.0687 | 0.0013 |
| Cluster 15 | Mt_ca_comptes_titres | 0.8596 | 0.3574 | 0.2185 |
| | Mt_ca_comptes_titres_non_abonne | 0.8738 | 0.1793 | 0.1538 |
| | Nb_ca_comptes_titres | 0.9225 | 0.3162 | 0.1133 |
| | Nb_ca_comptes_titres_non_abonne | 0.8747 | 0.1312 | 0.1442 |
| Cluster 16 | Mt_AssuranceIARD_Total | 0.7607 | 0.2059 | 0.3013 |
| | Mt_ct_Auto | 0.8213 | 0.1525 | 0.2108 |
| | Nb_ct_Auto | 0.6598 | 0.1051 | 0.3802 |
| Cluster 17 | Mt_CT_Produits_Patrimoniaux | 1.0000 | 0.0005 | 0.0000 |
| | Nb_CT_Produits_Patrimoniaux | 1.0000 | 0.0005 | 0.0000 |
| Cluster 18 | Mt_credit_engagement_signature | 1.0000 | 0.0005 | 0.0000 |
| | Nb_credit_engagement_signature | 1.0000 | 0.0005 | 0.0000 |
| Cluster 19 | Mt_credit_OC | 1.0000 | 0.0186 | 0.0000 |
| | Nb_credit_OC | 1.0000 | 0.0186 | 0.0000 |
| Cluster 20 | Mt_depot_Collecte_Monetaire | 0.9955 | 0.0066 | 0.0046 |
| | Nb_depot_Collecte_Monetaire | 0.9955 | 0.0068 | 0.0046 |
| Cluster 21 | Mt_CT_predica_capitalisation | 1.0000 | 0.0144 | 0.0000 |
| | Nb_CT_predica_capitalisation | 1.0000 | 0.0144 | 0.0000 |
| Cluster 22 | Mt_ca_comptes_titres_abonnes | 0.9985 | 0.0059 | 0.0015 |
| | Nb_ca_comptes_titres_abonnes | 0.9985 | 0.0061 | 0.0015 |
| Cluster 23 | Mt_dav_oc_prof | 0.9479 | 0.0378 | 0.0541 |
| | Nb_dav_oc_prof | 0.9479 | 0.0373 | 0.0541 |
| Cluster 24 | Mt_dav_successions | 0.8773 | 0.0020 | 0.1229 |
| | Nb_dav_successions | 0.8773 | 0.0033 | 0.1231 |
| Cluster 25 | MT_MOY_M_REMISE_CHQ_P | 0.7671 | 0.2078 | 0.2940 |
| | NB_CHQ_12M_P | 0.9129 | 0.2431 | 0.1151 |
| | NB_REMISE_12M_P | 0.9215 | 0.2220 | 0.1009 |
| Cluster 26 | NB_CONNEXION_BAM_6M_P | 0.9808 | 0.1410 | 0.0224 |

| 37 Clusters | | R carré avec | | 1-R**2 Ratio |
|----------------|-----------------------------|--------------------------|-------------------|-----------------|
| Classification | Variable | Propre classification | Le plus proche | |
| | NB_CONNEXION_WEB_6M_P | 0.9808 | 0.1410 | 0.0224 |
| Cluster 27 | Mt_cc_émissions_permanentes | 0.9881 | 0.0801 | 0.0129 |
| | Nb_cc_émissions_permanentes | 0.9881 | 0.0773 | 0.0128 |
| Cluster 28 | Mt_CT_predica_UC | 1.0000 | 0.2141 | 0.0000 |
| | Nb_CT_predica_UC | 1.0000 | 0.2141 | 0.0000 |
| Cluster 29 | AGE1 | 0.6608 | 0.1509 | 0.3995 |
| | ANCT | 0.7642 | 0.1429 | 0.2751 |
| | ANCT_MRA | 0.4234 | 0.0531 | 0.6089 |
| Cluster 30 | Mt_cc_Livrets | 0.7099 | 0.5201 | 0.6045 |
| | SOLDE_LIVRETS_P | 0.6945 | 0.5254 | 0.6437 |
| | Nb_cc_Epargne_Total | 0.8955 | 0.5540 | 0.2342 |
| | Nb_cc_Livrets | 0.7335 | 0.2385 | 0.3500 |
| | Nb_collecte_autres | 0.8982 | 0.5881 | 0.2471 |
| | Nb_collecte_cam | 0.8982 | 0.5881 | 0.2471 |
| Cluster 31 | Nb_AssuranceIARD_Total | 0.6972 | 0.3473 | 0.4640 |
| | Nb_ct_GAV | 0.4172 | 0.0307 | 0.6013 |
| | Nb_ct_PJ | 0.5164 | 0.0174 | 0.4922 |
| Cluster 32 | NB_PERS_CLT | 0.8781 | 0.3447 | 0.1861 |
| | NB_PERS_MAJEUR | 0.8781 | 0.1601 | 0.1452 |
| Cluster 33 | Mt_ct_Mrh | 0.5385 | 0.1254 | 0.5277 |
| | NB_CARTE_DD_P | 0.2583 | 0.0726 | 0.7998 |
| | Nb_ct_Mrh | 0.4154 | 0.0176 | 0.5950 |
| | Nb_ct_MrhPNO | 0.2761 | 0.0349 | 0.7501 |
| Cluster 34 | NB_CARTE_RETRAIT_P | 0.5511 | 0.0132 | 0.4549 |
| | Nb_ct_2ROUES | 0.5511 | 0.0250 | 0.4604 |
| Cluster 35 | NB_CONNEXION_IMB_6M_P | 0.5464 | 0.0052 | 0.4560 |
| | NB_CONNEXION_WAP_6M_P | 0.5464 | 0.0069 | 0.4567 |
| Cluster 36 | NB_ENTRTN_ANMOBL_P | 1.0000 | 0.0908 | 0.0000 |
| Cluster 37 | NB_CONNEXION_VDTX_6M_P | 1.0000 | 0.0067 | 0.0000 |

On récupère en sortie du fichier OUTSTAT les enregistrements de type GROUP, qui nous fournit, pour chaque valeur du nombre _NCL_ de classes de variables retenues, la classe à laquelle appartient chacune des variables analysées. On retrouve dans les deux dernières colonnes respectivement le numéro de la classe d'appartenance de la variable « N° classe » ainsi que la sélection définie par une croix « sélection avec la classification ».

3. Présentation des résultats de la modélisation du score sur l'échantillon d'apprentissage

Il s'agira de modéliser la valeur de la variable à expliquer qualite_SP_hs_CNAT0=1. La procédure de sélection STEPWISE pas à pas peut commencer. Avant tout propos, il faut signaler que l'ordre dans lequel on entre les variables dans le modèle, c'est-à-dire de la plus corrélée à la moins corrélée avec la variable cible, est important.

Note: No (additional) effects met the 0.01 significance level for entry into the model.

Step 0. Intercept entered:

| État de convergence du modèle | |
|---|--|
| Convergence criterion (GCONV=1E-8) satisfied. | |

| | | |
|-----------------|---|-----------|
| -2 Log L | = | 23209.724 |
|-----------------|---|-----------|

| Test du Khi 2 résiduel | | |
|------------------------|----|------------|
| Khi 2 | DF | Pr > Khi 2 |
| 4276.9484 | 11 | <.0001 |

La déviance par rapport au modèle « saturé », qui est le modèle idéal, est égale à $-2 \ln$ (vraisemblance du modèle ajusté), et vaut 23 209.724 dans le cas réduit à la constante. Cette déviance est analogue pour la régression logistique de la somme des carrés résiduels pour la régression linéaire, et constitue un des principaux indicateurs de la qualité d'ajustement du modèle aux données observées.

L'algorithme de sélection ajoute au modèle les variables qui font le plus augmenter la vraisemblance (dont le maximum est 1, pour le modèle saturé, c'est-à-dire le plus baisser la déviance (dont le minimum est 0, pour le modèle saturé). Sous hypothèse H_0 de la nullité du coefficient d'une nouvelle variable ajoutée, la baisse de la déviance suit une loi du χ^2 à 1 degré de liberté (cette loi est à $p-1$ degrés de liberté dans le cas de l'ajout d'une variable qualitative à p modalités), et on n'ajoute donc cette variable que si la baisse de la déviance est supérieure au seuil critique 6.63 à 1% du χ^2 à 1 degré de liberté.

Step 1. Effect ANCT_MRA entered:

| État de convergence du modèle | |
|---|--|
| Convergence criterion (GCONV=1E-8) satisfied. | |

| Statistiques d'ajustement du modèle | | |
|-------------------------------------|-----------------------------------|---------------------------------------|
| Critère | Coordonnée à l'origine uniquement | Coordonnée à l'origine et covariables |
| AIC | 23211.724 | 19116.866 |

| Statistiques d'ajustement du modèle | | |
|-------------------------------------|-----------------------------------|---------------------------------------|
| Critère | Coordonnée à l'origine uniquement | Coordonnée à l'origine et covariables |
| SC | 23219.534 | 19148.105 |
| -2 Log L | 23209.724 | 19108.866 |

| | | | |
|----------|--------|-----------------------|--------|
| R-Square | 0.2016 | Max-rescaled R-Square | 0.2799 |
|----------|--------|-----------------------|--------|

| Test de l'hypothèse nulle globale : BETA=0 | | | |
|--|-----------|----|------------|
| Test | Khi 2 | DF | Pr > Khi 2 |
| Likelihood Ratio | 4100.8587 | 3 | <.0001 |
| Score | 3931.9193 | 3 | <.0001 |
| Wald | 3238.9145 | 3 | <.0001 |

| Test du Khi 2 résiduel | | |
|------------------------|----|------------|
| Khi 2 | DF | Pr > Khi 2 |
| 450.8869 | 8 | <.0001 |

Note: No effects for the model in Step 1 are removed.

La première variable entrée dans le modèle, la plus discriminante, est l'ancienneté de l'entrée en relation du client au niveau Bancaire. Cela n'est guère surprenant s'agissant de la prédiction de la présence d'un client à fort potentiel de marge en assurance. La déviance est passée de 23209.724 à 19108.866, soit une baisse de 4100.8587, très supérieure au seuil de 6.63 correspondant au χ^2 à 1 degré de liberté au seuil de probabilité de 1%.

Le tableau de « statistiques d'ajustement du modèle » fournit aussi les critères AIC et SC (ou BIC), respectivement critère d'Akaike et critère de Schwartz (Bayesian Information Criterion), qui sont dérivés de la déviance, à laquelle ils ajoutent un terme lié au nombre de degrés de liberté du modèle, donc au nombre de variables, afin de pénaliser les modèles complexes : en minimisant le AIC et le SC, on poursuit les deux objectifs de minimisation de la déviance et de parcimonie du modèle.

Par ailleurs, le R2 ajusté vaut 0.2799, ce qui n'est pas encore très élevé (il varie entre 0 et 1) et justifie de chercher à ajouter d'autres variables.

La totalité des étapes a suite en annexe

Step 2. Effect Nb_ct_Auto entered:

| État de convergence du modèle |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

| Statistiques d'ajustement du modèle | | |
|-------------------------------------|-----------------------------------|---------------------------------------|
| Critère | Coordonnée à l'origine uniquement | Coordonnée à l'origine et covariables |
| AIC | 23211.724 | 18895.245 |
| SC | 23219.534 | 18934.295 |
| -2 Log L | 23209.724 | 18885.245 |

| | | | |
|----------|--------|-----------------------|--------|
| R-Square | 0.2114 | Max-rescaled R-Square | 0.2934 |
|----------|--------|-----------------------|--------|

| Test de l'hypothèse nulle globale : BETA=0 | | | |
|--|-----------|----|------------|
| Test | Khi 2 | DF | Pr > Khi 2 |
| Likelihood Ratio | 4324.4791 | 4 | <.0001 |
| Score | 4097.3983 | 4 | <.0001 |
| Wald | 3328.5365 | 4 | <.0001 |

| Test du Khi 2 résiduel | | |
|------------------------|----|------------|
| Khi 2 | DF | Pr > Khi 2 |
| 236.1963 | 7 | <.0001 |

Note: No effects for the model in Step 2 are removed.

Avec la deuxième variable entrée dans le modèle, le nombre de contrats d'assurance auto, la déviance est passée de 19108.866 à 18885.245, soit une baisse de 223.621, très supérieure au seuil de 6.63 correspondant au χ^2 à 1 degré de liberté. Par ailleurs, le R2 ajusté a augmenté et vaut maintenant 0.2934.

Step 3. Effect Mt_ct_Mrh entered:

| État de convergence du modèle |
|---|
| Convergence criterion (GCONV=1E-8) satisfied. |

| Statistiques d'ajustement du modèle | | |
|-------------------------------------|-----------------------------------|---------------------------------------|
| Critère | Coordonnée à l'origine uniquement | Coordonnée à l'origine et covariables |
| AIC | 23211.724 | 18742.368 |
| SC | 23219.534 | 18804.847 |
| -2 Log L | 23209.724 | 18726.368 |

| | | | |
|-----------------|--------|------------------------------|--------|
| R-Square | 0.2182 | Max-rescaled R-Square | 0.3029 |
|-----------------|--------|------------------------------|--------|

| Test de l'hypothèse nulle globale : BETA=0 | | | |
|---|--------------|-----------|----------------------|
| Test | Khi 2 | DF | Pr > Khi 2 |
| Likelihood Ratio | 4483.3565 | 7 | <.0001 |
| Score | 4220.6975 | 7 | <.0001 |
| Wald | 3399.1386 | 7 | <.0001 |

| Test du Khi 2 résiduel | | |
|-------------------------------|-----------|----------------------|
| Khi 2 | DF | Pr > Khi 2 |
| 80.3741 | 4 | <.0001 |

Note: No effects for the model in Step 3 are removed.

Avec la troisième variable entrée dans le modèle, le nombre de contrats d'assurance auto, la déviance est passée de 18885.245 à 18726.368, soit une baisse de 158.877, très supérieure au seuil de 6.63 correspondant au χ^2 à 1 degré de liberté. Par ailleurs, le R2 ajusté a augmenté et vaut maintenant 0.3029. Nous passons les itérations suivantes en sachant que la démarche à chaque fois est la même.

Avec 7 variables, le R2 ajusté a atteint la valeur 0.3077

Après la septième itération, plus aucune variable n'est suffisamment significative pour entrer dans le modèle, et le processus de sélection pas à pas s'interrompt. Un tableau rappelle l'ordre d'entrée des variables, avec une avant dernière colonne indiquant que la variable en termes de probabilité du χ^2 est alors significative.

| Récapitulatif sur la sélection séquentielle | | | | | | | | |
|--|-------------------|-----------------|------------|--------------------|-----------------------|----------------------|----------------------|--|
| Étape | Effet | | D F | Nombre dans | Khi 2 du score | Khi 2 de Wald | Pr > Khi 2 | Libellé de variable |
| | Saisi | Supprimé | | | | | | |
| 1 | ANCT_MRA | | 3 | 1 | 3931.9193 | | <.0001 | Valeurs de ANCT_MRA remplacées par des rangs |
| 2 | Nb_ct_Auto | | 1 | 2 | 219.2914 | | <.0001 | Valeurs de Nb_ct_Auto remplacées par des rangs |
| 3 | Mt_ct_Mrh | | 3 | 3 | 157.4769 | | <.0001 | Valeurs de Mt_ct_Mrh remplacées par des rangs |
| 4 | SGMT | | 3 | 4 | 64.0364 | | <.0001 | |

| Récapitulatif sur la sélection séquentielle | | | | | | | | |
|---|----------------------|----------|-----|-------------|----------------|---------------|------------|---|
| Étape | Effet | | D F | Nombre dans | Khi 2 du score | Khi 2 de Wald | Pr > Khi 2 | Libellé de variable |
| | Saisi | Supprimé | | | | | | |
| 5 | NB_CONNEXION_BAM_6M_ | | 1 | 5 | 24.6502 | | <.0001 | Valeurs de NB_CONNEXION_BAM_6M_P remplacées par des rangs |
| 6 | NB_PRDT_DET_N_P | | 1 | 6 | 16.1766 | | <.0001 | Valeurs de NB_PRDT_DET_N_P remplacées par des rangs |
| 7 | Nb_cc_Livrets | | 2 | 7 | 20.6652 | | <.0001 | Valeurs de Nb_cc_Livrets remplacées par des rangs |
| 8 | Nb_credit_OC | | 1 | 8 | 10.8115 | | 0.0010 | Valeurs de Nb_credit_OC remplacées par des rangs |
| 9 | IVP | | 1 | 9 | 8.5914 | | 0.0034 | |
| 10 | Mt_cc_émissions_perm | | 1 | 10 | 8.6568 | | 0.0033 | Valeurs de Mt_cc_émissions_permanentes remplacées par des rangs |

ANNEXE II

1. Caractéristiques du preneur d'assurance

a. Variable Marché

Il s'agit d'une variable qualitative à 3 niveaux décrivant le marché d'appartenance de l'assuré. Dans notre portefeuille,

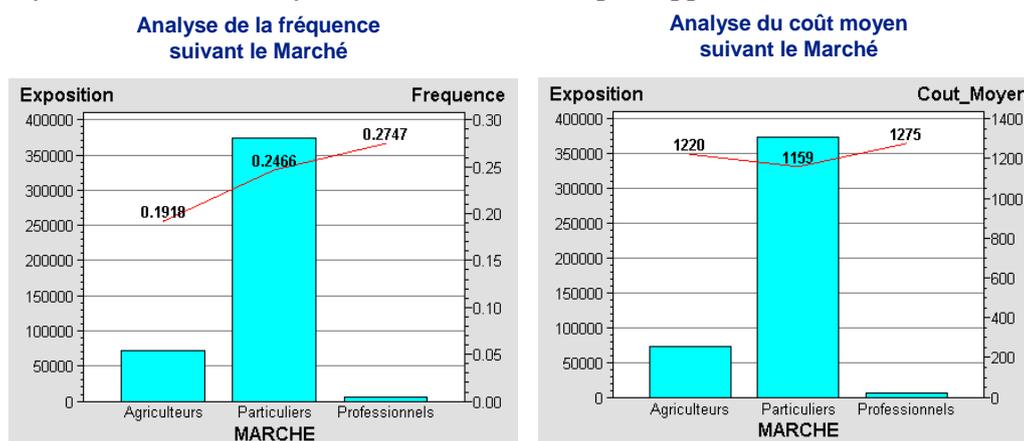
- **373 714** Marché des **Particuliers** (soit **83 %**) (**niveau de référence**),
- **72 268** Marché des **Agriculteurs** (soit **16 %**),
- **5 636** Marché des **Professionnels** (soit **1 %**).

Afin d'apprécier l'influence de la variable Marché sur la sinistralité, nous vous présentons, dans les graphes suivants, la fréquence des sinistres observée ainsi que le coût moyen suivant le marché d'appartenance de l'assuré.

On constate une sursinistralité en termes de fréquence et un coût moyen élevé pour le marché des professionnels.

Par contre, concernant le marché des agriculteurs, nous pouvons voir que la fréquence est faible mais que le coût moyen est élevé.

A contrario, le marché des particuliers enregistre une fréquence supérieure à la moyenne et un coût moyen relativement faible par rapport aux autres marchés.



b. Variable Usage du véhicule

Il s'agit d'une variable qualitative à 3 niveaux décrivant l'usage du véhicule de l'assuré. Dans notre portefeuille,

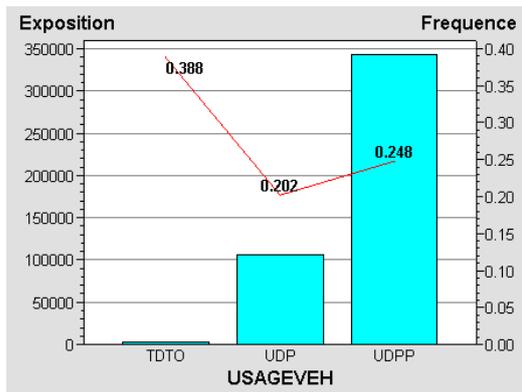
- **3 120** Usage **TDTO** Tous Déplacements Tournées (soit **1 %**),
- **105 457** Usage **UDP** Usage Déplacements Privé (soit **23 %**),
- **343 040** Usage **UDPP** Usage Déplacements Privé et Professionnel (soit **76 %**) (**niveau de référence**).

Afin d'apprécier l'influence de la variable Usage du véhicule sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant l'usage du véhicule de l'assuré.

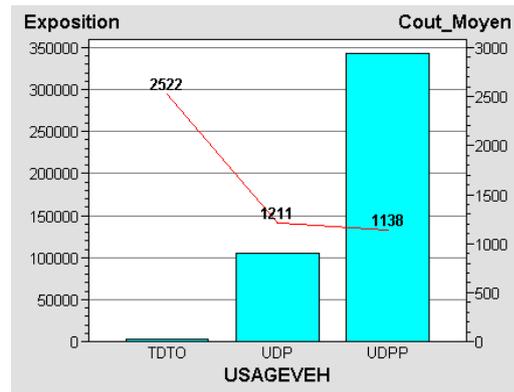
On constate une sursinistralité en termes de fréquence et de coût moyen pour l'usage TDTO et une fréquence faible et un coût moyen élevé par rapport à la moyenne pour l'usage UDP. Ce dernier est représenté principalement par des retraités, ce qui confortera les résultats vus au niveau de la variable âge.

L'usage UDPP enregistre une fréquence supérieure à la moyenne et un coût moyen relativement faible par rapport aux autres occurrences de la variable étudiée.

**Analyse de la fréquence
suivant l'usage du véhicule**



**Analyse du coût moyen
suivant l'usage du véhicule**



c. Variable Activité

Il s'agit d'une variable qualitative à 7 niveaux décrivant l'activité du véhicule de l'assuré. Dans notre portefeuille,

- **172 819** Activité **Salarié du privé** (soit **38 %**) (**niveau de référence**),
- **3 218** Activité **Artisan Commerçant** (soit **1 %**),
- **94 285** Activité **Fonctionnaire** (soit **21 %**),
- **68 797** Activité **Agriculteur** (soit **15 %**),
- **63 528** Activité **Retraité** (soit **14 %**),
- **45 849** Activité **Salarié CASA** (soit **10 %**),
- **3 120** Activité **Tournées** (soit **1 %**).

Afin d'apprécier l'influence de l'activité sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant l'activité de l'assuré.

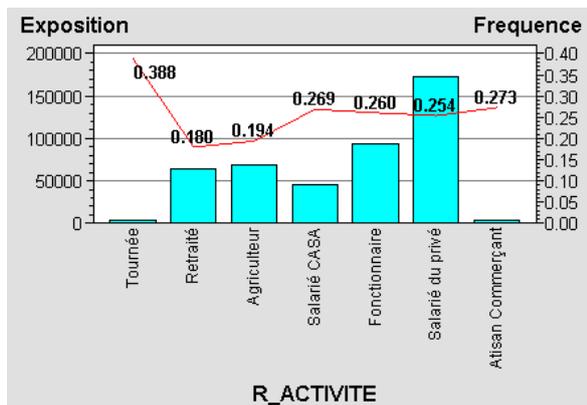
On constate une sursinistralité aussi bien en termes de fréquence que de coût moyen pour les activités Artisan Commerçant et Tournées.

La fréquence pour les activités Retraité et Agriculteur est relativement basse et leurs coûts moyens sont dans la moyenne.

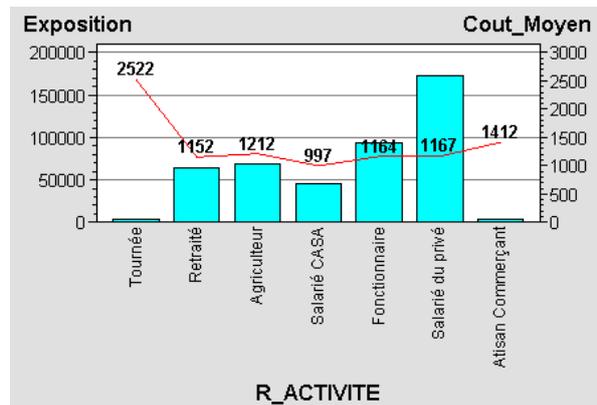
Concernant l'activité salarié CASA, Il est important de noter la valeur basse du coût moyen et une fréquence relativement élevée.

Forte similitude pour les activités Salarié du privé et Fonctionnaire aussi bien au niveau de la fréquence que du coût moyen, i.e. un coût moyen dans la moyenne et une fréquence nettement supérieure à la moyenne.

**Analyse de la fréquence
suivant l'activité**



**Analyse du coût moyen
suivant l'activité**



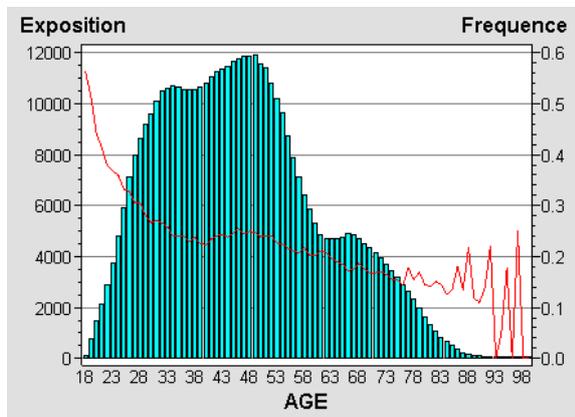
d. Variable Age

Il s'agit d'une variable quantitative à valeurs entières donnant l'âge du souscripteur(en années révolues). Plus que l'âge en tant que tel, c'est l'expérience de conduite qu'on espère récupérer par le biais de cette variable. Une variable équivalente (tant la corrélation entre les deux est grande) et souvent mieux acceptée par la clientèle est l'ancienneté du permis de conduire.

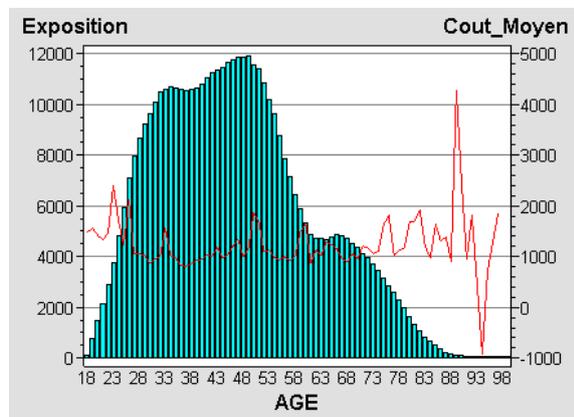
Les polices reprennent souvent la notion de conducteur habituel du véhicule, et les caractéristiques personnelles déterminant le montant de la prime sont alors celles du conducteur habituel mentionné aux conditions particulières, et pas celle du preneur d'assurance. Il est bon de rappeler que les caractéristiques personnelles (telles que l'âge) se rapportent souvent au preneur d'assurance, qui n'est pas nécessairement le conducteur habituel.

Examinons à présent la structure des âges du conducteur principal dans notre portefeuille. Celle-ci est décrite dans les graphes suivants.

**Analyse de la fréquence
suivant l'âge**



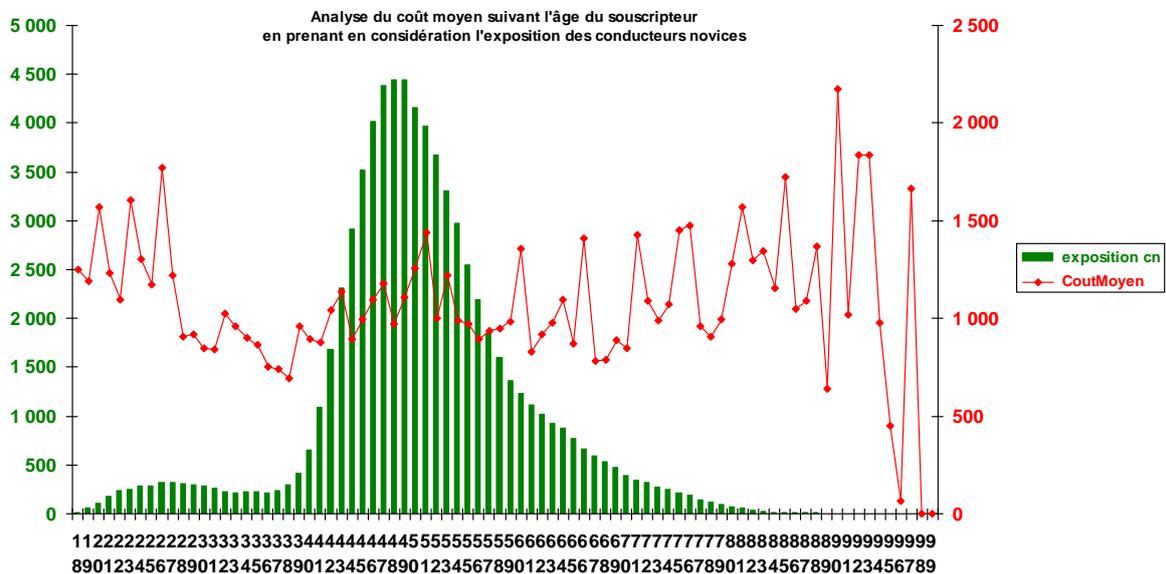
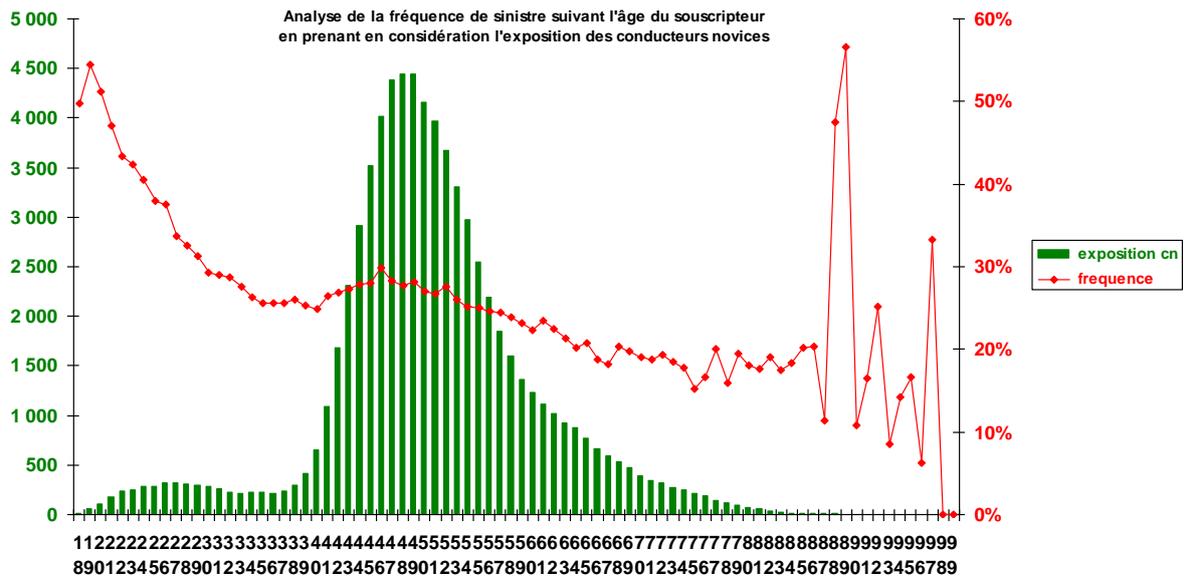
**Analyse du coût moyen
suivant l'âge**



La figure décrivant la fréquence suivant l'âge, montre clairement une sursinistralité des jeunes conducteurs à plus de 30 % et avec l'âge la fréquence diminue sauf pour la tranche 41 à 54 ans qui reste sensiblement égale à la tranche 34 à 40 ans, certainement dû à la présence de conducteur novice occasionnel (cf. graphes suivants) et qui annihile la baisse. Légère hausse pour les personnes plus âgées.

Nous constatons un coût moyen élevé pour les jeunes conducteurs par rapport à la moyenne (1 163 euros). Puis la diminution s'amorce avec l'âge pour remonter ensuite, à partir de la tranche 71 à 75 ans et doubler par rapport à la moyenne pour la tranche plus de 75 ans. Il faut noter toutefois le même effet constaté pour la fréquence pour les tranches 34 à 40 ans et 41 à 54 ans.

La variable Ages influence la fréquence ainsi que le coût moyen mais elle pourrait ainsi être causée par une autre variable fortement corrélée avec cette dernière telle que l'ancienneté de permis.

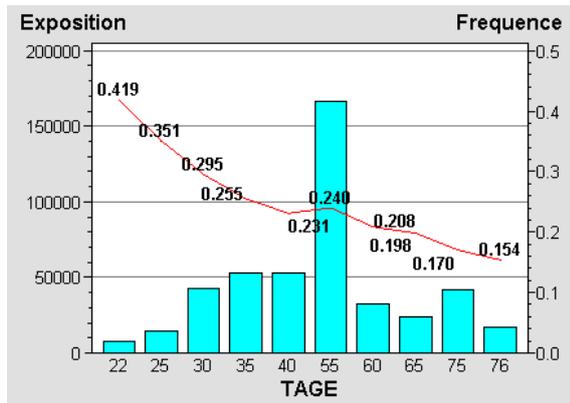


Après analyse de la sinistralité, nous proposons les tranches d'âge suivantes :

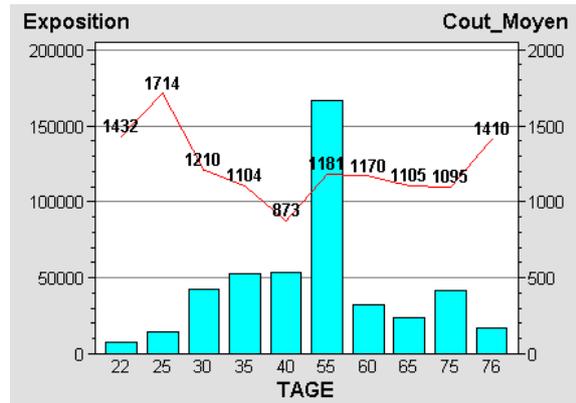
- ❖ **7 399 < 22 ans (soit 2 %)**
- ❖ **14 483 entre 23 et 25 ans (soit 3 %)**
- ❖ **42 547 entre 26 et 30 ans (soit 9 %)**
- ❖ **52 507 entre 31 et 35 ans (soit 12 %)**
- ❖ **53 104 entre 36 et 40 ans (soit 12 %)**
- ❖ **166 540 entre 41 et 55 ans (soit 37 %) (niveau de référence),**
- ❖ **32 599 entre 56 et 60 ans (soit 5 %)**
- ❖ **23 752 entre 61 et 65 ans (soit 7 %)**
- ❖ **41 669 entre 66 et 75 ans (soit 9 %)**
- ❖ **17 017 plus de 75 ans (soit 4 %)**

Elles peuvent être modifiées ou regroupées lors de la modélisation de la fréquence et du coût moyen. Nous présentons dans les graphes suivants la sinistralité suivant la tranche d'âge.

**Analyse de la fréquence
suivant la tranche d'âge d'appartenance**



**Analyse du coût moyen
suivant tranche d'âge d'appartenance**



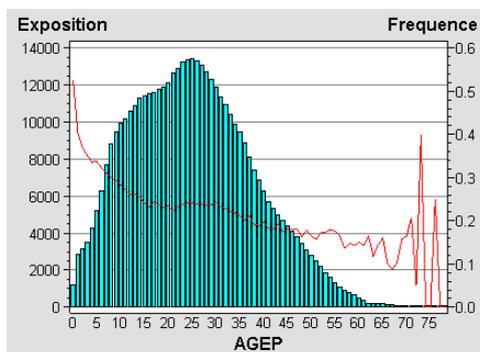
e. Variable Ancienneté du permis

Il s'agit d'une variable quantitative à valeurs entières donnant l'ancienneté du permis du souscripteur(en années révolues). C'est l'expérience de conduite qu'on espère récupérer par le biais de cette variable.

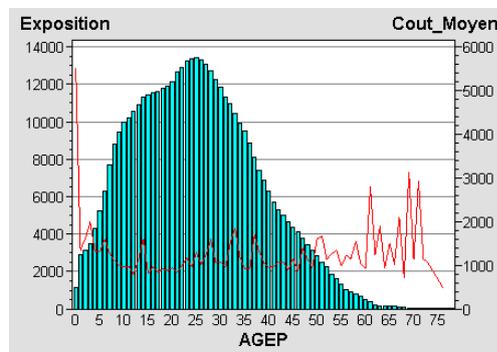
Il est bon de rappeler que les caractéristiques personnelles (telles que l'ancienneté du permis) se rapportent souvent au preneur d'assurance, qui n'est pas nécessairement le conducteur habituel.

Examinons à présent la structure de l'ancienneté du permis du conducteur principal dans notre portefeuille. Celle-ci est décrite dans les graphes suivants.

**Analyse de la fréquence
suivant l'ancienneté du permis**



**Analyse du coût moyen
suivant l'ancienneté du permis**

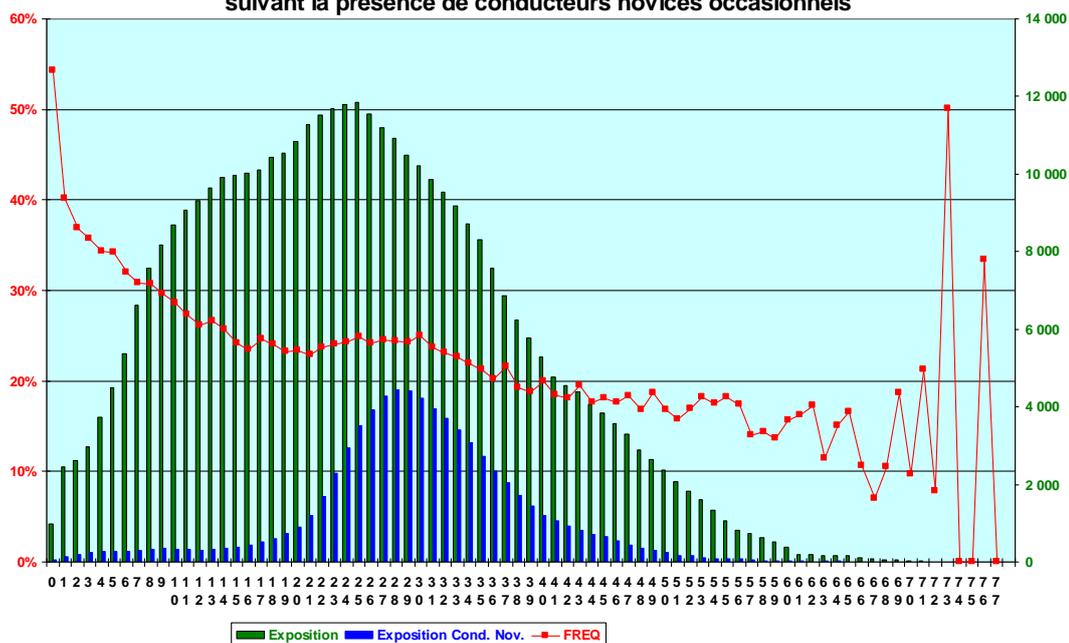


La figure décrivant la fréquence suivant l'ancienneté du permis, montre clairement une sursinistralité (plus de 30 %) des conducteurs avec peu d'expérience de conduite. La fréquence diminue avec l'ancienneté du permis sauf pour la tranche 22 à 37 ans qui est sensible à la présence de conducteurs novices occasionnels et qui annihile la baisse. Légère hausse pour une ancienneté supérieure à 50 ans.

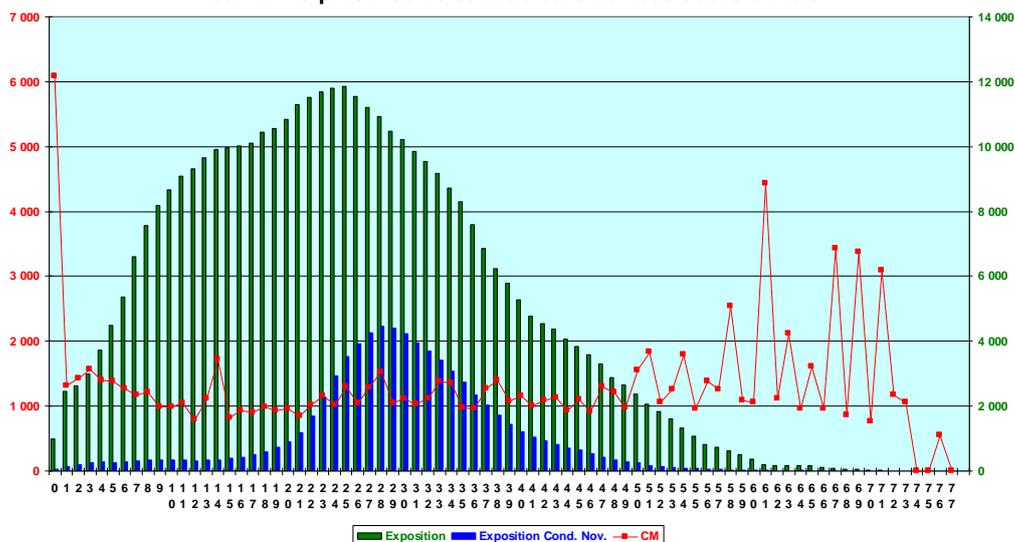
Nous constatons un coût moyen élevé par rapport à la moyenne (1 163 euros) pour les conducteurs avec peu d'expérience. Puis la diminution s'amorce avec l'ancienneté pour remonter ensuite, à partir de 22 ans mais l'effet de la présence des conducteurs novices occasionnels est un peu moins prononcé que pour la fréquence. Par contre la hausse du coût moyen est significative pour une ancienneté supérieure à 46 ans.

La variable ancienneté influence donc la fréquence mais aussi le coût moyen dans une moindre mesure. L'analyse conforte donc notre position concernant la corrélation entre cette dernière et la variable âge.

Analyse de la fréquence de sinistre suivant l'ancienneté du permis du souscripteur et suivant la présence de conducteurs novices occasionnels



Analyse du coût moyen suivant l'ancienneté du permis du souscripteur et suivant la présence de conducteurs novices occasionnels

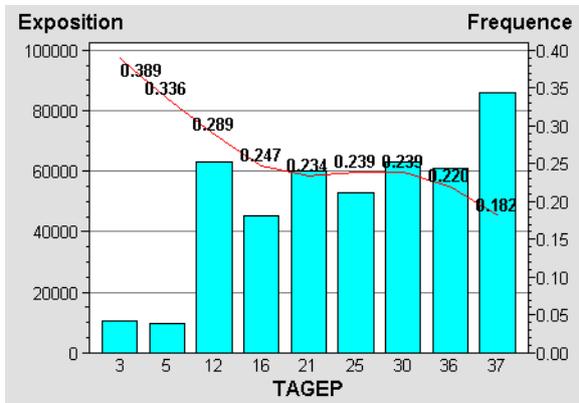


Après analyse de la sinistralité, nous proposons les tranches d'ancienneté de permis suivantes :

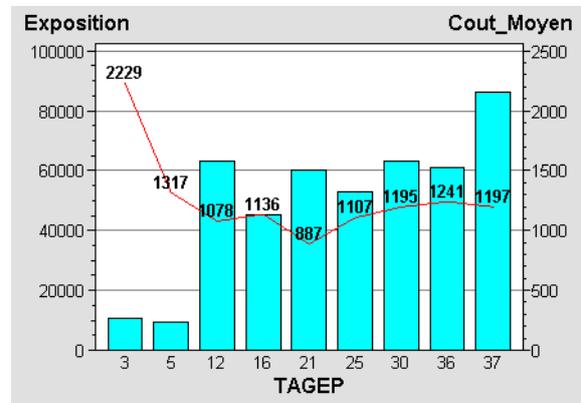
- ❖ 10 678 < 4 ans (soit 2 %),
- ❖ 9 521 entre 4 et 5 ans (soit 2 %),
- ❖ 62 965 entre 6 et 12 ans (soit 14 %),
- ❖ 45 140 entre 13 et 16 ans (soit 10 %),
- ❖ 60 019 entre 17 et 21 ans (soit 13 %),
- ❖ 52 965 entre 22 et 25 ans (soit 12 %),
- ❖ 63 245 entre 26 et 30 ans (soit 14 %),
- ❖ 60 949 entre 31 et 36 ans (soit 13 %),
- ❖ 86 136 entre 37 et + (soit 19 %) (niveau de référence).

Elles peuvent être modifiées ou regroupées lors de la modélisation de la fréquence et du coût moyen. Nous présentons dans les graphes suivants la sinistralité suivant les tranches d'ancienneté de permis suivantes.

**Analyse de la fréquence
suivant la tranche d'ancienneté du permis**



**Analyse du coût moyen
suivant tranche d'ancienneté du permis**



f. Variable Zone

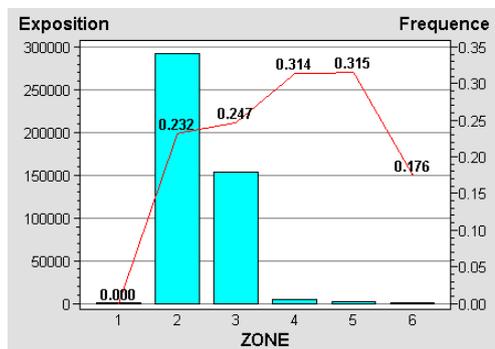
Il s'agit d'une variable qualitative à 5 niveaux décrivant la zone de tarification i.e. où est situé le garage du véhicule de l'assuré. Dans notre portefeuille,

- 292 417 zone 2 (soit 65 %) (**niveau de référence**),
- 152 953 zone 3 (soit 34 %),
- 4 516 zone 4 (soit 0,96 %),
- 1 663 zone 5 (soit 0,04 %),
- 68 zone 6 (soit 0 %).

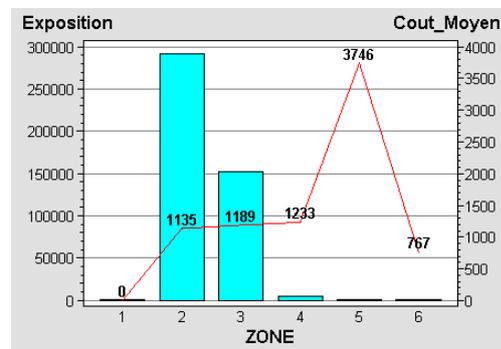
Afin d'apprécier l'influence de la zone de tarification sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la zone de tarification.

On constate une croissance de la sinistralité en termes de fréquence et de coût moyen suivant la zone de tarification.

**Analyse de la fréquence
suivant la zone**



**Analyse du coût moyen
suivant la zone**



g. Variable Agglomération

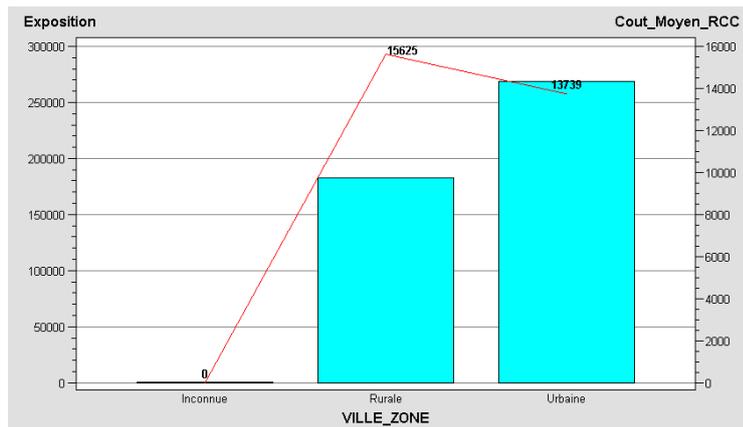
Cette variable donne une idée de l'environnement dans lequel vit le preneur d'assurance.

$AGGLOM = \begin{cases} 1, & \text{si l'agglomération est rurale } 182\ 865, \text{ (soit } 40 \%), \\ 0, & \text{si l'agglomération est urbaine } 268\ 724, \text{ (soit } 60 \%) \text{ (niveau de référence).} \end{cases}$

Si la commune est fortement urbanisée, la plupart des voiries seront situées en agglomération, et la vitesse des véhicules sera limitée. De ce fait, la fréquence des sinistres devrait être élevée, du fait de la densité importante du trafic, mais les conséquences des sinistres devraient être faibles, du fait de la vitesse réduite. Au contraire, dans des communes rurales, la plupart des voiries seront situées hors agglomération, les sinistres seront plus rares, du fait de la faible densité du trafic, mais leurs conséquences devraient être plus graves, du fait de la vitesse plus élevée des véhicules.

Le portefeuille est constitué, de 182 865 polices dont les souscripteurs résident dans une agglomération rurale et 268 724 polices dans une agglomération urbaine, représentant respectivement 40,5 % et 59,5 % (28 polices pour lesquelles l'agglomération est inconnue).

On peut noter que la fréquence est plus élevée en milieu urbain qu'en milieu rural par contre le coût moyen est plus élevé en milieu rural qu'en milieu urbain. Le dernier résultat est plus significatif en RC Corporelle, ce que vous pourrez vérifier dans le deuxième graphe qui suit.



L'INSEE nous propose, aujourd'hui, une définition des communes un peu plus précise. C'est pourquoi nous pensons qu'il est intéressant d'analyser la sinistralité suivant cette déclinaison en 4 catégories de communes.

Définition de :

- | Une **Commune monopolarisée**, ou commune périurbaine monopolarisée, est une commune appartenant à la **couronne périurbaine*** d'une **aire urbaine****.
- | Une **Commune multipolarisée** : communes rurales et unités urbaines situées hors des aires urbaines, dont au moins 40% de la population résidente ayant un emploi travaille dans plusieurs aires urbaines, sans atteindre ce seuil avec une seule d'entre elles, et qui forment avec elles un ensemble d'un seul tenant.
- | L'**Espace à dominante rurale** comprend à la fois des petites **unités urbaines***** et des communes rurales.
- | Un **pôle urbain** est une unité urbaine offrant au moins 5 000 emplois et qui n'est pas située dans la couronne périurbaine d'un autre pôle urbain.

Pour information, l'**Espace à dominante urbaine** est constitué des pôles urbains, des couronnes périurbaines et des communes multipolarisées.

Pacifica a pris en compte cette variable dans la modélisation et a obtenu des résultats probants. Elle devrait les présenter à sa direction générale.

- * recouvre l'ensemble des communes de l'aire urbaine à l'exclusion de son pôle urbain.
- ** ensemble de communes, d'un seul tenant et sans enclave, constitué par un pôle urbain, et par des communes rurales ou unités urbaines (couronne périurbaine) dont au moins 40% de la population résidente ayant un emploi travaille dans le pôle ou dans des communes attirées par celui-ci.
- *** est une commune ou un ensemble de communes qui comporte sur son territoire une zone bâtie d'au moins 2 000 habitants où aucune habitation n'est séparée de la plus proche de plus de 200 mètres. En outre, chaque commune concernée possède plus de la moitié de sa population dans cette zone bâtie.

Il s'agit d'une variable qualitative à 6 niveaux décrivant la Catégorie de la commune où résident nos assurés. Dans notre portefeuille,

- **101 717** Catégorie **Commune monopolarisée** (soit **23 %**),
- **75 346** Catégorie **Commune multipolarisée** (soit **17 %**),
- **6** Catégorie **DOM** (soit - %),
- **85 662** Catégorie **Espace à dominante rurale** (soit **19 %**),
- **28** Catégorie **Inconnue** (soit - %),
- **188 858** Catégorie **Pôle urbain** (soit **42 %**) (**niveau de référence**).

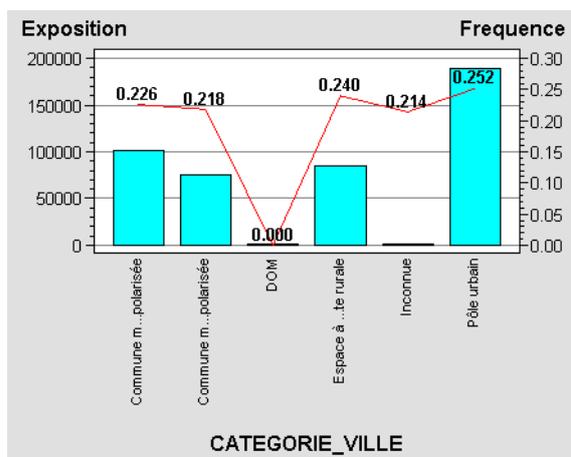
Afin d'apprécier l'influence de l'activité sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la Catégorie de la commune.

La fréquence associée à la catégorie de commune Pôle urbain est élevée ainsi que son coût moyen.

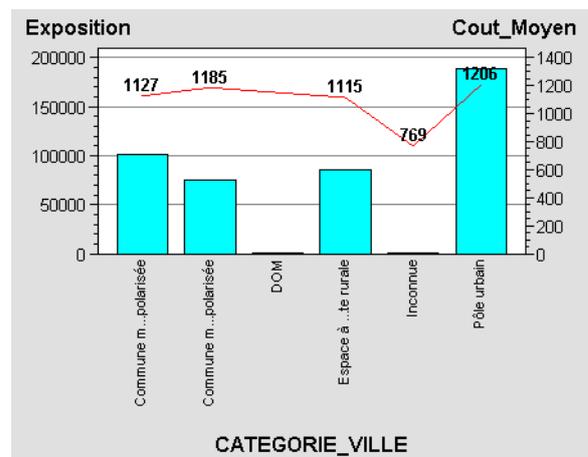
Concernant la catégorie de commune Espace à dominante rurale nous visualisons une fréquence supérieure à la moyenne et un coût moyen dans la moyenne.

Les Catégories de communes monopolarisées et multipolarisées sont assez proches en terme de sinistralité avec une fréquence et un coût moyen dans la moyenne.

**Analyse de la fréquence
suivant la catégorie de commune**



**Analyse du coût moyen
suivant la catégorie de commune**



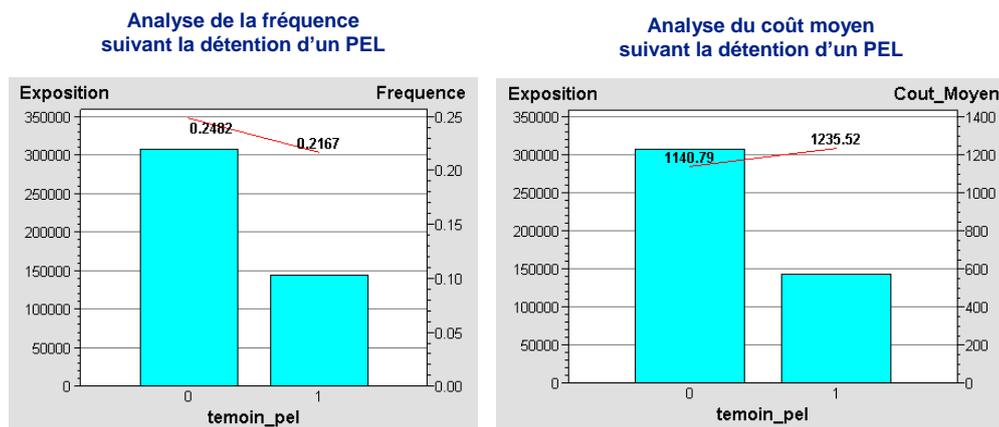
h. Variable PEL (détention d'un PEL)

Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose d'un PEL.

$$\text{PEL} = \begin{cases} 1, & \text{si l'assuré dispose d'un PEL } 143\ 846 \text{ (soit } 32\%), \\ 0, & \text{si l'assuré ne dispose pas d'un PEL } 307\ 771, \text{ (soit } 68\%) \text{ (niveau de référence).} \end{cases}$$

Afin d'apprécier l'influence de la variable **PEL** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **PEL**.

On constate que le fait que l'assuré soit équipé d'un **PEL** entraîne une baisse de la fréquence et une hausse du coût moyen.



i. Variable Actions (détention d'Actions)

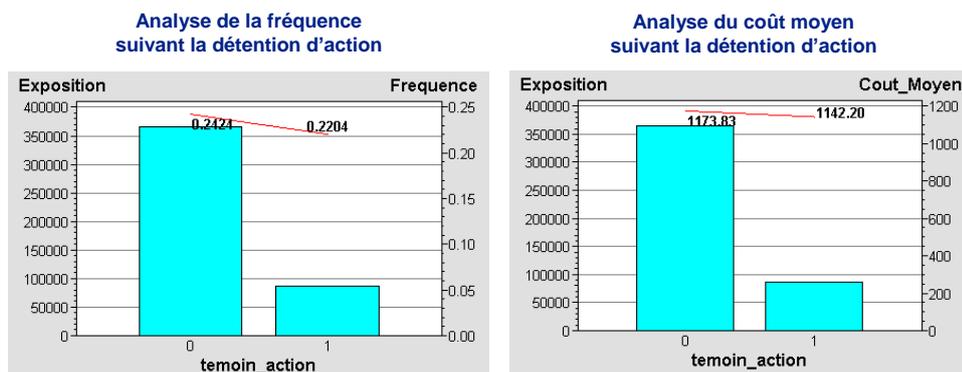
Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose d'Actions.

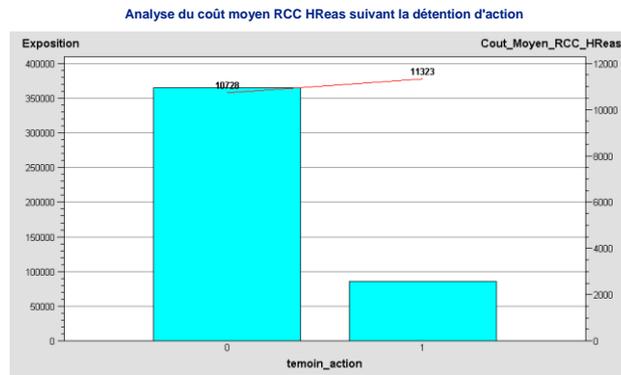
$$\text{ACTION} = \begin{cases} 1, & \text{si l'assuré dispose d'Actions } 86\ 325, \text{ (soit } 19\%), \\ 0, & \text{si l'assuré ne dispose pas d'Actions } 365\ 293, \text{ (soit } 81\%) \text{ (niveau de référence).} \end{cases}$$

Afin d'apprécier l'influence de la variable **Actions** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **Actions**.

On constate clairement que le fait que l'assuré soit équipé d'**Actions** entraîne une baisse de la sinistralité en termes de fréquence, mais en terme de coût moyen il est difficile de donner une tendance.

Par contre, si nous regardons, dans le deuxième graphe qui suit, le coût moyen en RC Corporelle Hors Réassurance, il s'avère que, le fait de détenir des actions entraîne une sursinistralité.





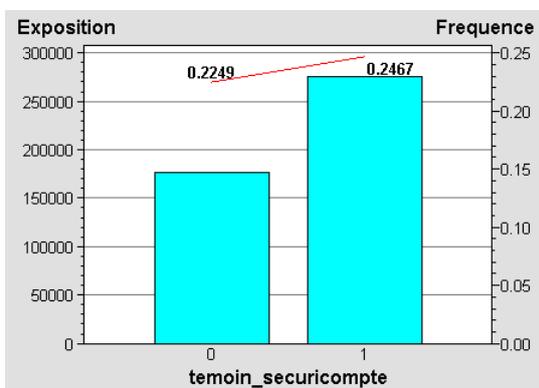
j. Variable SECURICOMPTE (détention d'un Sécuricompte)

Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose d'un Sécuricompte. **SECURICOMPTE** = 1, si l'assuré dispose d'un Sécuricompte 275 194, (soit 61 %) (niveau de référence),
 0, si l'assuré ne dispose pas d'un Sécuricompte 176 423, (soit 39 %).

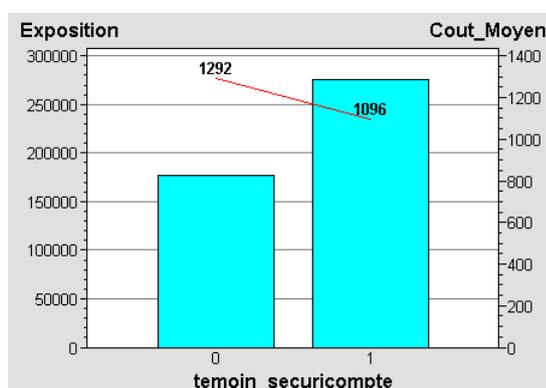
Afin d'apprécier l'influence de la variable **Sécuricompte** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **Sécuricompte**.

On constate clairement que le fait que l'assuré soit équipé d'un **Sécuricompte** entraîne une hausse de la fréquence et une baisse du coût moyen.

Analyse de la fréquence suivant la détention d'un temoin_securicompte



Analyse du coût moyen suivant la détention d'un temoin_securicompte



2. Caractéristiques du véhicule

a. Variable Groupe SRA

Cette variable a pour finalité d'aider les assureurs à tarifier la garantie Responsabilité Civile. C'est un indicateur qui est le reflet de la « dangerosité » intrinsèques des véhicules. Pour le calcul du groupe, il est pris en compte les paramètres suivants :

- la puissance réelle du véhicule.
- les masses du véhicule (Masse à vide et PTAC).
- la vitesse maximale du véhicule.
- un indicateur estimant la sécurité globale du véhicule. Ce paramètre dépend de la conception et des équipements de sécurité du véhicule.

Dans notre portefeuille,

- **60 752** véhicules (soit **13** %) ont un groupe SRA inférieur ou égal à **26**,
- **79 419** véhicules (soit **18** %) ont un groupe SRA égal à **27**,

- **90 285** véhicules (soit **20 %**) ont un groupe SRA égal à **28**(niveau de référence),

- **67 639** véhicules (soit **15 %**) ont un groupe SRA égal à **29**,

- **47 007** véhicules (soit **10 %**) ont un groupe SRA égal à **30**,

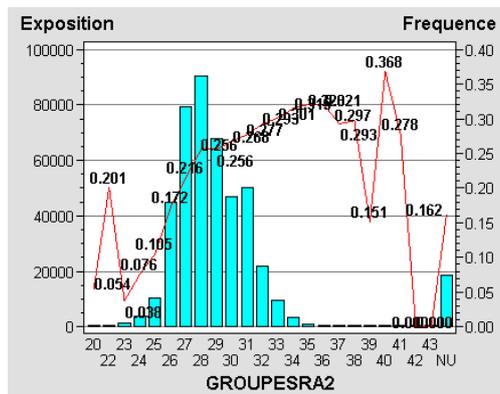
- **50 350** véhicules (soit **11 %**) ont un groupe SRA égal à **31**,

- **37 453** véhicules (soit **8 %**) ont un groupe SRA supérieur ou égal à **32**.

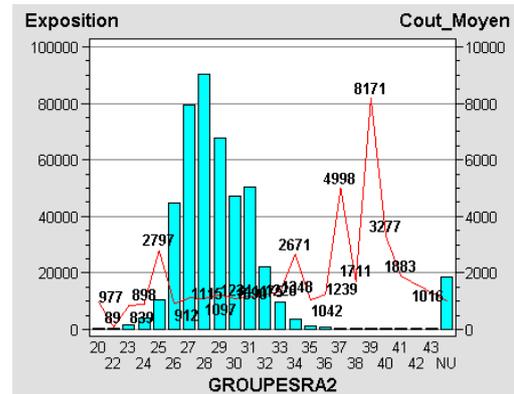
Afin d'apprécier l'influence du groupe SRA du véhicule sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre en fonction du groupe SRA du véhicule.

On constate que la fréquence, comme le coût, croît suivant le groupe SRA du véhicule.

**Analyse de la fréquence
suivant le GROUPE SRA**



**Analyse du coût moyen
suivant le GROUPE SRA**



b. Variable Classe de prix SRA

La classe de prix est un indicateur en cas de perte totale du véhicule ou de vol du véhicule. C'est la valeur à neuf du véhicule lors de sa mise sur le marché qui permet de calculer la valeur de cette Classe de prix. Pour le calcul de la Classe de prix, on définit un classement des véhicules en fonction des tranches de prix. Les valeurs de ces tranches sont réactualisées chaque année en fonction de l'évolution de l'indice INSEE du prix des véhicules neufs.

Dans notre portefeuille,

- **45 602** véhicules (soit **10 %**) ont une Classe de prix SRA inférieure ou égale à **C** ($\leq 8\ 700$ €),

- **39 408** véhicules (soit **9 %**) ont une Classe de prix SRA égale à **D** ($\leq 9\ 700$ €),

- **43 716** véhicules (soit **10 %**) ont une Classe de prix SRA égale à **E** ($\leq 10\ 800$ €),

- **52 777** véhicules (soit **12 %**) ont une Classe de prix SRA égale à **F** ($\leq 12\ 000$ €)(niveau de référence),

- **52 456** véhicules (soit **12 %**) ont une Classe de prix SRA égale à **G** ($\leq 13\ 400$ €),

- **47 884** véhicules (soit **9 %**) ont une Classe de prix SRA égale à **H** ($\leq 15\ 000$ €),

- **42 544** véhicules (soit **9 %**) ont une Classe de prix SRA égale à **I** ($\leq 16\ 800$ €),

- **34 267** véhicules (soit **8 %**) ont une Classe de prix SRA égale à **J** ($\leq 18\ 900$ €),

- **29 618** véhicules (soit **7 %**) ont une Classe de prix SRA égale à **K** ($\leq 21\ 200$ €),

- **18 405** véhicules (soit **4 %**) ont une Classe de prix SRA égale à **L** ($\leq 23\ 800$ €),

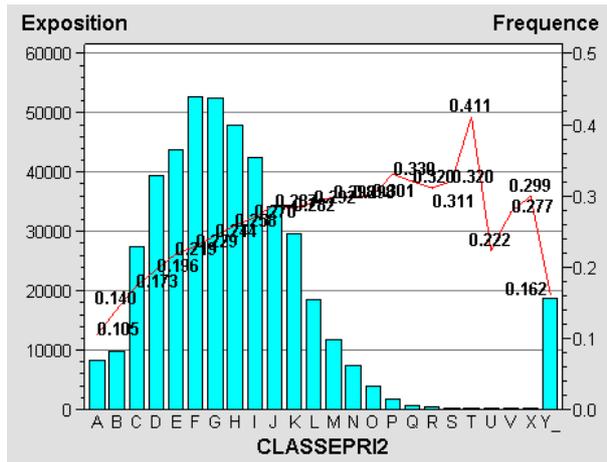
- **11 667** véhicules (soit **3 %**) ont une Classe de prix SRA égale à **M** ($\leq 26\ 700$ €),

- **14 563** véhicules (soit **3 %**) ont une Classe de prix SRA supérieure ou égale à **N** ($> 26\ 700$ €).

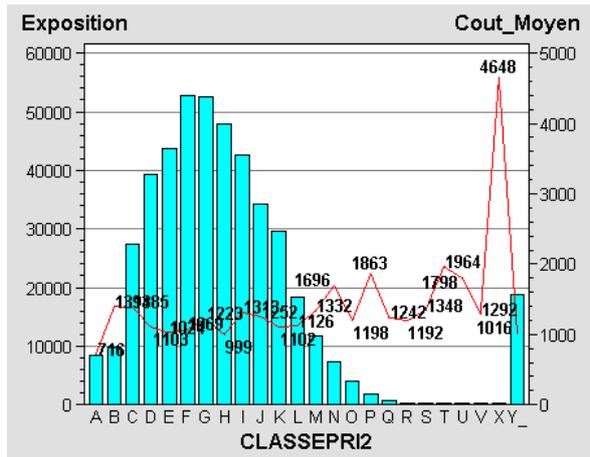
Afin d'apprécier l'influence de la Classe de prix SRA du véhicule sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre en fonction de Classe de prix SRA du véhicule.

On constate que la fréquence, comme le coût, croît suivant la Classe de prix SRA du véhicule. Il est à noter toutefois que la croissance du coût moyen est moins significative que la fréquence. Si nous prenons en compte uniquement l'évolution du coût moyen dans le cadre de la garantie Vol, Dommage, BDg et Incendie, la tendance à la hausse est plus nette.

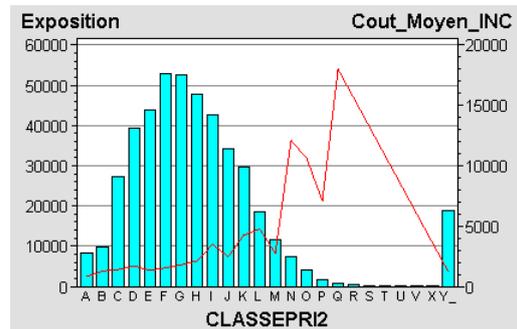
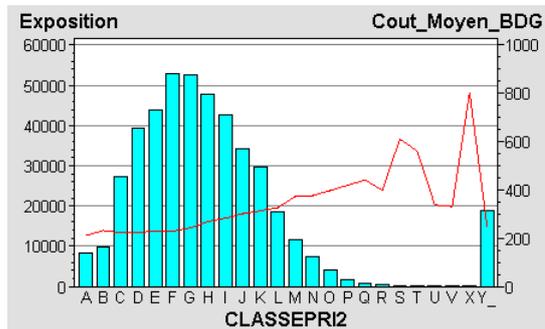
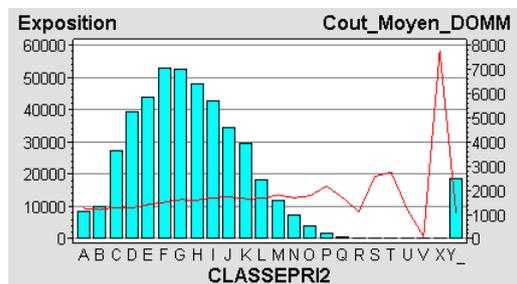
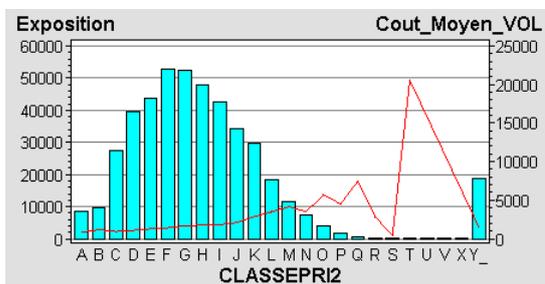
Analyse de la fréquence suivant la Classe de prix SRA



Analyse du coût moyen suivant le Classe de prix SRA



Analyse du Coût moyen Vol, Dommage, BDG et Incendie suivant la Classe de prix SRA



c. Variable Classe de réparation SRA

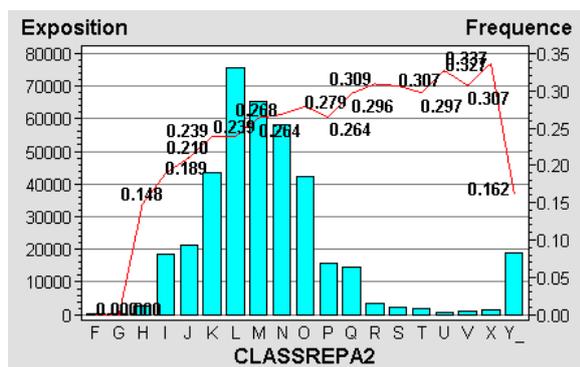
La classe de réparation est un indicateur estimant le coût de la réparation d'un véhicule. Pour le calcul de la Classe de réparation, il est fonction du coût pondéré de remplacement du panier de pièces, du coût d'un choc avant à 15 km/h et du coût d'un choc arrière à 15 km/h et on définit un classement des véhicules en fonction des tranches de coût de réparation. Les valeurs de ces tranches sont réactualisées chaque année en fonction de l'évolution de l'indice du coût de la réparation. Dans notre portefeuille,

- **21 393** véhicules (soit **5 %**) ont une Classe de réparation SRA inférieure ou égale à **I** ($\leq 1\,438$ €),
- **21 426** véhicules (soit **5 %**) ont une Classe de réparation SRA égale à **J** ($\leq 1\,580$ €),
- **43 405** véhicules (soit **10 %**) ont une Classe de réparation SRA égale à **K** ($\leq 1\,729$ €),
- **75 705** véhicules (soit **18 %**) ont une Classe de réparation SRA égale à **L** ($\leq 1\,884$ €)
(niveau de référence),
- **65 200** véhicules (soit **14 %**) ont une Classe de réparation SRA égale à **M** ($\leq 2\,046$ €),
- **58 156** véhicules (soit **13 %**) ont une Classe de réparation SRA égale à **N** ($\leq 2\,216$ €),
- **42 167** véhicules (soit **9 %**) ont une Classe de réparation SRA égale à **O** ($\leq 2\,393$ €),
- **15 910** véhicules (soit **4 %**) ont une Classe de réparation SRA égale à **P** ($\leq 2\,578$ €),
- **25 420** véhicules (soit **7 %**) ont une Classe de réparation SRA supérieure ou égale à **Q** ($> 2\,578$ €),
- **18 712** véhicules (soit **4 %**) ont une Classe de réparation SRA **Inconnue**.

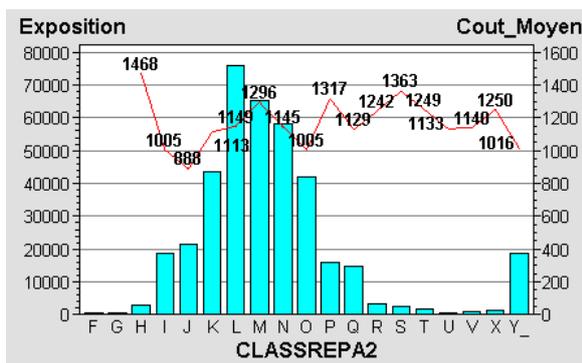
Afin d'apprécier l'influence de la Classe de réparation SRA du véhicule sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre en fonction de Classe de réparation SRA du véhicule.

On constate que la fréquence, comme le coût, croît suivant la Classe de réparation SRA du véhicule. Il est à noter toutefois que la croissance du coût moyen est moins significative que la fréquence. Si nous prenons en compte uniquement l'évolution du coût moyen dans le cadre de la garantie Vol, Dommage, Bdg et Incendie, la tendance à la hausse est plus nette.

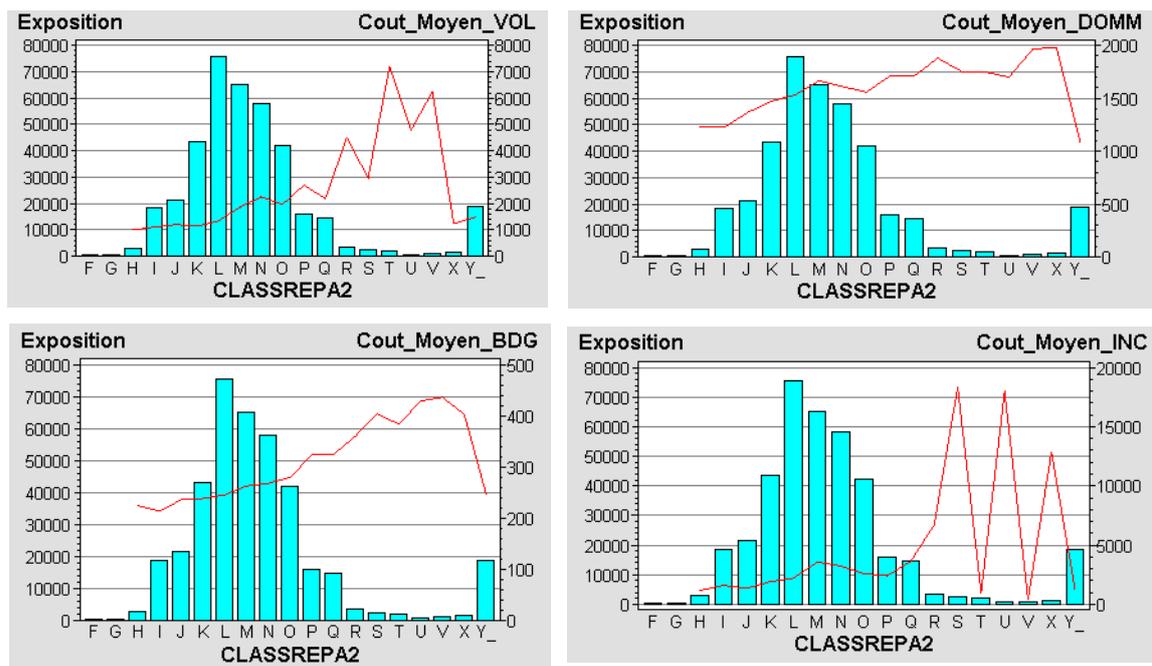
**Analyse de la fréquence
suivant la Classe de réparation SRA**



**Analyse du coût moyen
suivant le Classe de réparation SRA**



Analyse du Coût moyen Vol, Dommage, BDG et Incendie suivant la Classe de réparation SRA



d. Variable Type de carburant

Cette variable binaire prenant 2 modalités donne le type de carburant du véhicule assuré, i.e.

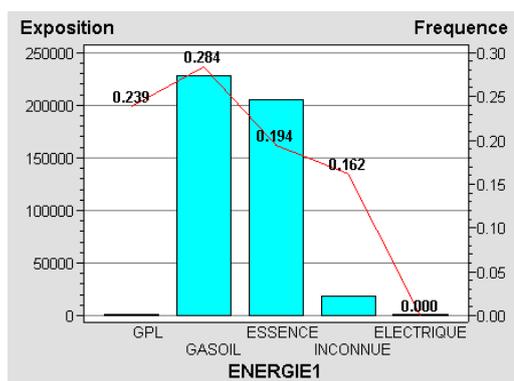
CARB= {
 véhicule **GPL** , **168**, (soit 0 %),
 véhicule **diesel** , **227 931**, (soit 50 %) (**niveau de référence**),
 véhicule **essence** , **204 806**, (soit 45 %),
 véhicule **inconnu** , **18 712**, (soit 45 %),
 véhicule **électrique** , **1**, (soit 0 %).

Cette variable n'influence peut-être pas directement la sinistralité, mais on peut raisonnablement s'attendre à ce qu'elle soit fortement corrélée avec le kilométrage annuel (l'acquisition d'un véhicule diesel en France se justifiant souvent par un usage plus fréquent, et pour de plus longues distances puisque le prix d'achat est plus élevé, de même que la taxe de circulation, mais le prix du litre du carburant est moins élevé).

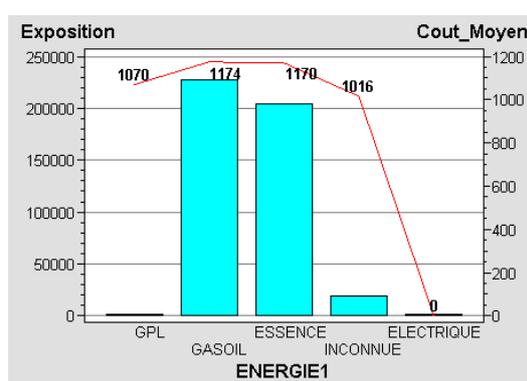
Dans notre portefeuille, 35 709 concernent des véhicules essence (soit 37 %), 60270 véhicules diesel (soit 62 %). On peut voir, au niveau des graphes suivants, que la fréquence des sinistres est plus élevée pour les véhicules diesel, mais que le coût moyen pour ces mêmes véhicules est dans la moyenne. La fréquence plus élevée pour les véhicules diesel pourrait s'expliquer par une exposition au risque(en kilomètres) plus importante.

Nous devons prendre en considération cette variable dans la tarification car elle n'est pas intégrée dans le calcul de la nouvelle classification du groupe SRA alors que dans le groupe GTA elle l'était.

Analyse de la fréquence suivant l'énergie



Analyse du coût moyen suivant l'énergie



e. Variable Type de carrosserie

La variable qualitative Type de carrosserie est défini par 10 modalités. Dans notre portefeuille,

- 65 véhicules de type **4X4** (soit 0 %),
- 350 865 véhicules de type **Berline** (soit 78 %) (niveau de référence),
- 22 412 véhicules de type **Break** (soit 5 %),
- 342 véhicules de type **Commerciale** (soit 0 %),
- 1 749 véhicules de type **Cabriolet** (soit 0 %),
- 3 854 véhicules de type **Coupe** (soit 1 %),
- 18 916 véhicules de type **Monospace** (soit 4 %),
- 34 675 véhicules de type **Utilitaire** (soit 8 %),
- 26 véhicules de type **Divers** (soit 0 %),
- 18 712 véhicules de type **Non Renseigné** (soit 5 %) : mauvaise saisie du code auto ou saisie de code provisoire ou non renseigné contrat Monceau.

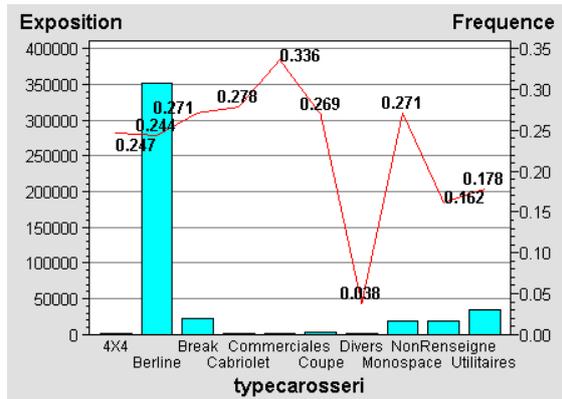
Afin d'apprécier l'influence du Type de carrosserie sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant le Type de carrosserie.

On constate que la fréquence, comme le coût moyen, est élevé pour la carrosserie Cabriolet et Coupe. Les breaks, les commerciales et les Monospaces ont une fréquence nettement au-dessus de la moyenne, à contrario, le coût moyen est bien en deçà de sa valeur moyenne sauf pour les commerciales.

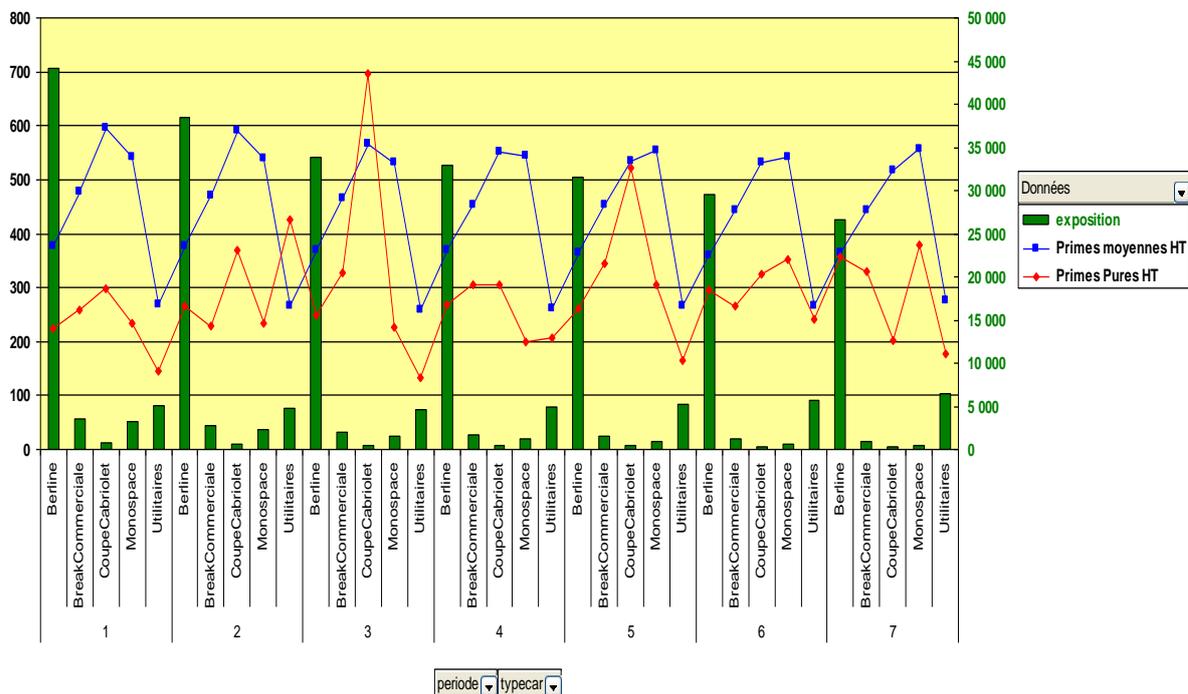
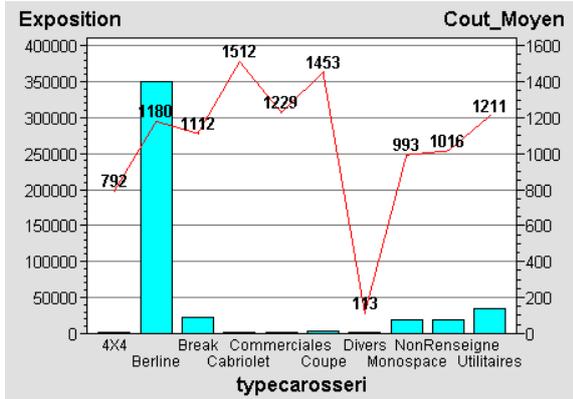
Toutefois le réseau nous a notifié l'année dernière que les monospaces étaient sur tarifés par rapport à nos principaux concurrents. C'est pourquoi nous avons comparé la prime moyenne HT avec la prime pure HT suivant le type de carrosserie. La deuxième diapositive qui suit confirme les dires du réseau.

De plus, nous voyons une très nette amélioration de la fréquence depuis 2003 pour les monospaces.

Analyse de la fréquence suivant le type de carrosserie



Analyse du coût moyen suivant le type de carrosserie



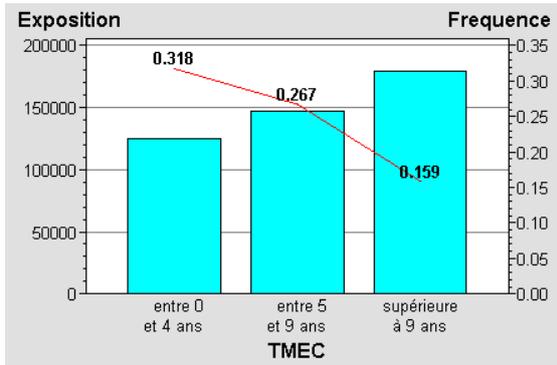
f. Variable Tranche de mise en circulation du véhicule

La variable qualitative TMEC(Tranche de mise en circulation du véhicule) est définie par 3 modalités. Dans notre portefeuille,

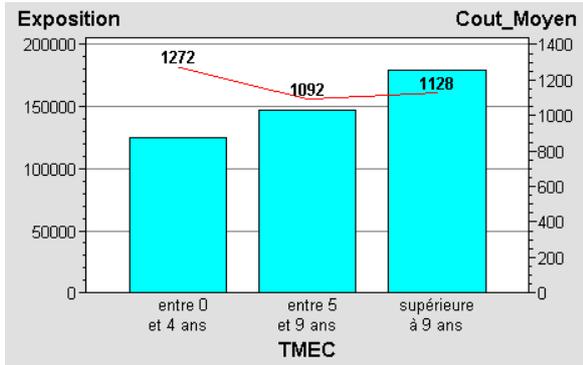
- **15 336** véhicules de TMEC 1 i.e. entre 0 et 4 ans(soit **27 %**),
- **17 702** véhicules de TMEC 2 i.e. entre 5 et 9 ans(soit **31 %**),
- **24 890** véhicules de TMEC 3 i.e. 10 ans et plus(soit **42 %**) (**niveau de référence**).

Afin d'apprécier l'influence de la TMEC sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la TMEC. On constate que la fréquence est décroissante en fonction de l'ancienneté du véhicule. Il est à noter toutefois que la dégradation du coût moyen est moins significative que la fréquence. Si nous prenons en compte uniquement l'évolution du coût moyen dans le cadre des garanties Vol, Dommages, Incendie, Bdg, **DREC** et CNAT, nous retrouvons cette tendance baissière suivant l'ancienneté du véhicule.

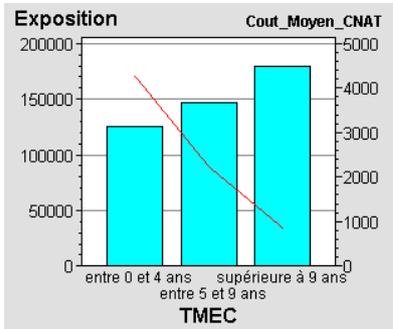
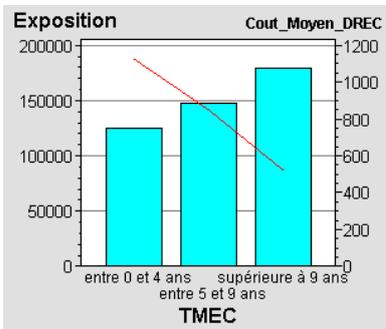
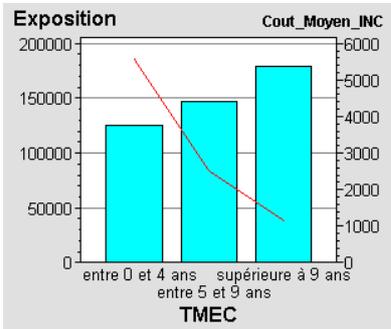
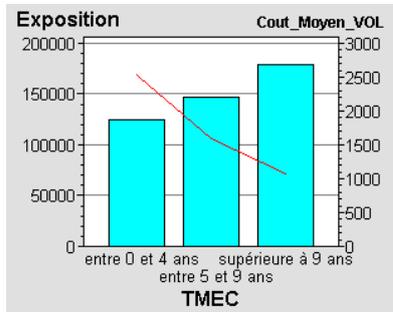
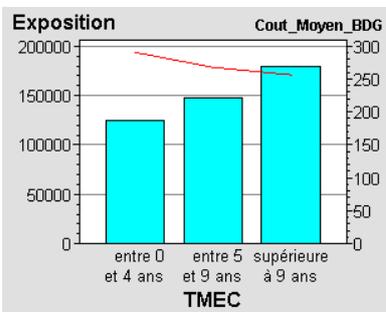
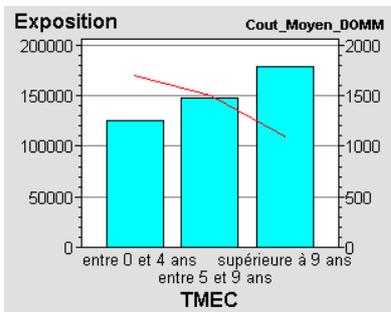
Analyse de la fréquence suivant la tranche de mise en circulation



Analyse du coût moyen suivant la tranche de mise en circulation



Analyse du Coût moyen Domm, BDG, Vol, Incendie, DREC et CNAT suivant la tranche de mise en circulation



3. Caractéristiques de la police

a. Variable nombre de conducteurs

Il s'agit d'une variable quantitative à valeurs entières donnant le nombre de conducteurs déclarés sur un même contrat. Nous cherchons à montrer la relation logique suivante :

Au plus il y a de conducteurs déclarés sur un même véhicule, au plus il y a de « chance » qu'un des conducteurs ait un accident.

L'occurrence à 0, de la variable nombre de conducteurs, provient du fait que certains vieux contrats sont encore présents dans la base et les différents conducteurs ne sont pas renseignés.

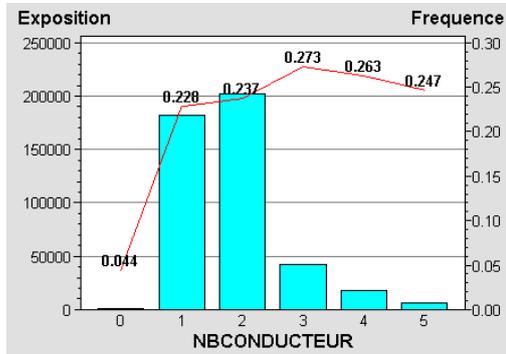
Le portefeuille est composé d'au plus 5 conducteurs déclarés. 40 % des contrats en portefeuille sont composés d'un conducteur déclaré et 45% de deux conducteurs déclarés. Ainsi les 15 % restants sont des contrats ayant plus de 2 conducteurs déclarés.

Nous constatons, en termes de fréquence, une sursinistralité pour les contrats détenant plus de deux conducteurs déclarés. Le coût moyen est une fonction croissante suivant le nombre de

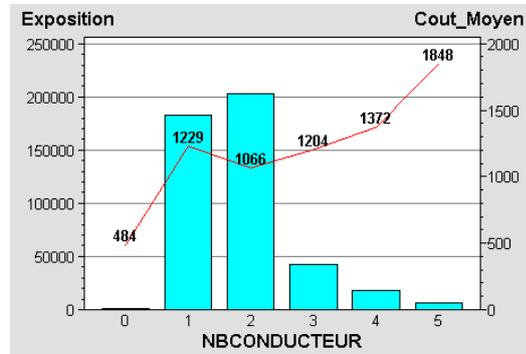
conducteurs, sauf, pour une valeur de l'occurrence à 1 puisqu'il correspond en grande partie aux assurés célibataires.

La variable conducteur novice occasionnel peut être corrélée avec la variable nombre de conducteurs déclarés par contrat.

Analyse de la fréquence suivant le nombre de conducteurs



Analyse du coût moyen suivant le nombre de conducteurs



b. Variable Fractionnement

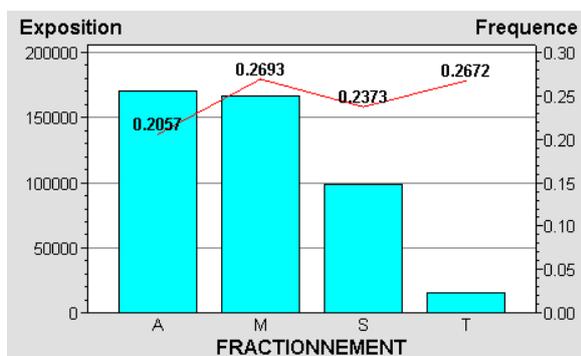
Il s'agit d'une variable qualitative à quatre niveaux spécifiant la fréquence de paiement de la prime. Dans notre portefeuille,

- **170 551** Fractionnement **Annuel** (soit **38 %**) (**niveau de référence**),
- **166 696** Fractionnement **Mensuel** (soit **37 %**),
- **99 065** Fractionnement **Semestriel** (soit **22 %**),
- **15 305** Fractionnement **Trimestriel** (soit **3 %**).

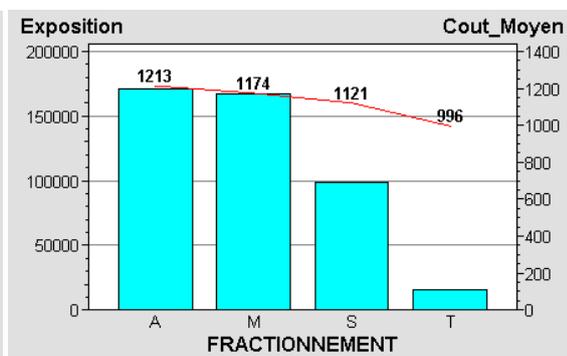
Encore une fois, on voit mal comment cette variable pourrait influencer directement la sinistralité. L'idée est ici qu'un fractionnement de la prime peut traduire une position sociale fragilisée, donc un véhicule moins bien entretenu. Il faut toutefois se méfier des paiements mensuels, car certaines compagnies proposent à leurs clients de verser la prime chaque mois (par domiciliation et sans frais supplémentaires) à titre gratuit. De plus, beaucoup souhaitent, aujourd'hui, mensualiser les impôts, les taxes, l'assurance, ...

Les graphes suivants illustrent l'influence du fractionnement sur la sinistralité des assurés. Les assurés payant leur prime une fois par an ont une fréquence de sinistres inférieure à celle de ceux ayant opté pour le fractionnement. En comparant les coûts moyens par rapport à la moyenne (1 163 €), nous constatons une sursinistralité pour les assurés payant leur prime une fois par an et mensuellement.

Analyse de la fréquence suivant le fractionnement de la prime



Analyse du coût moyen suivant le fractionnement de la prime



c. Variable conduite accompagnée

Cette variable qualitative binaire nous renseigne sur la présence dans la police, d'un conducteur principal CCA, ie ayant eu le permis dans le cadre de la conduite accompagnée, i.e.

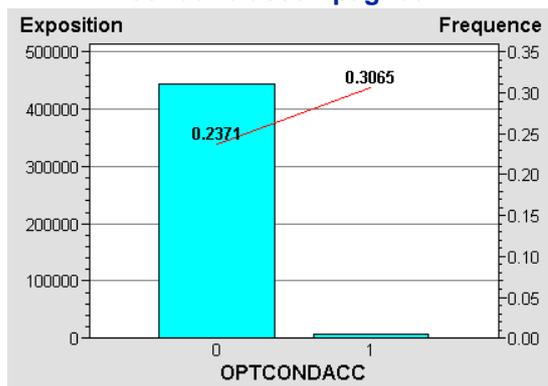
$$\text{CONDACCO} = \begin{cases} 0, & \text{si pas de présence d'un conducteur CCA } \mathbf{444\ 935} \text{ (soit } 98,5\%) \\ & \text{(niveau de référence),} \\ 1, & \text{si présence d'un conducteur CCA } \mathbf{6\ 682} \text{ (soit } 1,5\%). \end{cases}$$

Afin d'apprécier l'influence de la variable CONDACCO sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la variable CONDACCO.

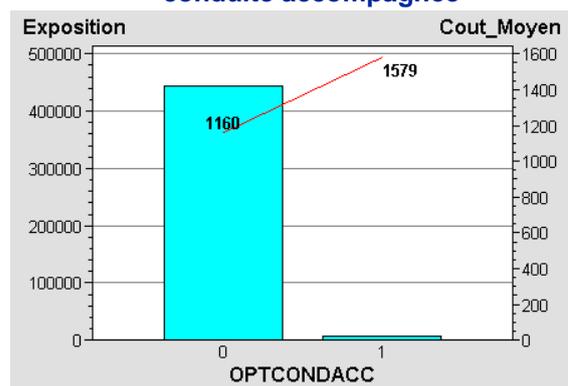
On peut noter, au niveau des graphes suivants, aussi bien, en termes de fréquence que de coût moyen, une sursinistralité significative pour les contrats ayant des conducteurs principaux CCA.

Pacifica et la FFSA constate la même sinistralité.

Analyse de la fréquence suivant l'option permis obtenu dans le cadre de la conduite accompagnée



Analyse du coût moyen suivant l'option permis obtenu dans le cadre de la conduite accompagnée



d. Variable Formule

Il s'agit d'une variable qualitative à trois niveaux décrivant le degré de couverture de l'assuré. Dans notre portefeuille,

- **271 334** Formule **Maxi** (soit **60 %**) (**niveau de référence**),
- **123 749** Formule **Médian** (soit **27,5 %**),
- **56 534** Formule **Mini** (soit **12,5 %**).

On peut ici s'attendre à deux types de comportement :

- soit la théorie de l'aléa moral s'applique, et on observe plus de sinistre de la part des individus les plus couverts. De la même manière, en vertu de la théorie du signal, les assurés ayant souscrit uniquement la RC l'on fait en connaissance de cause, se sachant moins risqués. La modalité Mini devrait donc apparaître comme un facteur diminuant la sinistralité.
- soit la souscription de garanties plus étendues traduit une aversion au risque plus forte de la part des individus concernés. Ceci se traduirait alors par un mode de conduite particulièrement prudent, et donc donnerait lieu à moins de sinistres. Dans ce cas, la modalité Mini devient un facteur aggravant.

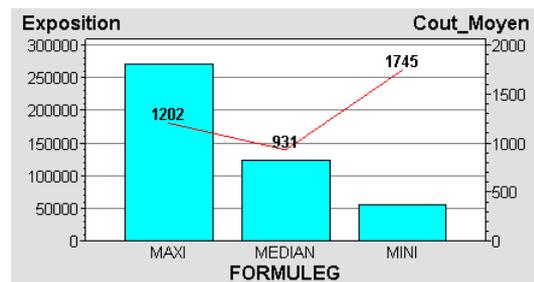
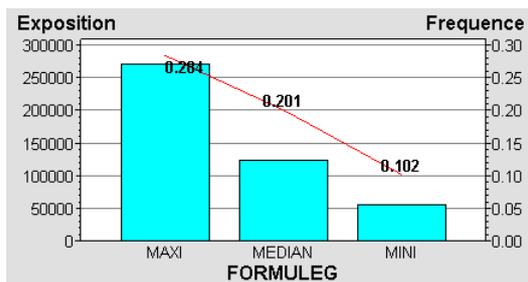
Les actuaires, par essence pragmatiques, n'essaient pas de déterminer quel mécanisme prévaut et se bornent à constater sur base des données quel est l'effet des différentes modalités de la formule.

Afin d'apprécier l'influence du choix de la formule sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la formule.

On constate que la fréquence est croissante en fonction du niveau de garantie et que le coût moyen des formules Mini et Maxi est élevé, par contre celui de la formule Médian est relativement bas.

Nous pouvons donc affirmer que la formule Médian contribue fortement à la bonne tenue du portefeuille en termes de sinistralité.

Analyse de la fréquence et du coût moyen suivant la formule souscrite



e. Variable Pack +

Il s'agit d'une variable qualitative binaire indiquant si l'assuré a souscrit l'option Pack +. Dans notre portefeuille,

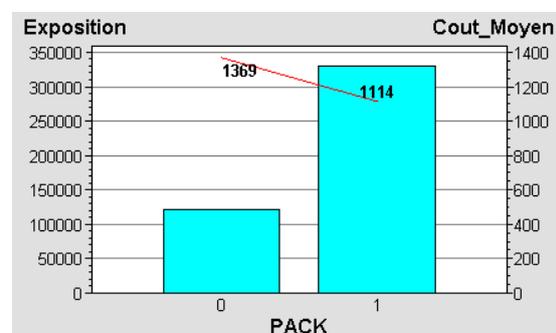
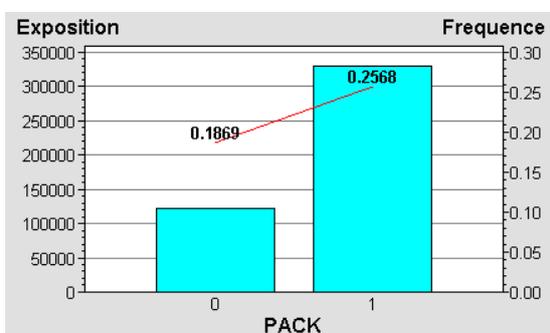
Pack+= $\begin{cases} 1, & \text{si l'assuré a souscrit à l'option Pack+ } 122\ 043 \text{ (soit 27\%),} \\ 0, & \text{si l'assuré n'a pas souscrit à l'option Pack+ } 329\ 574, \text{ (soit 73 \% (niveau de référence).} \end{cases}$

Afin d'apprécier l'influence de la variable **Pack +** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la variable **Pack +**.

On constate que la fréquence est nettement au-dessus de la moyenne alors que le coût moyen est en dessous de la valeur moyenne.

Ce résultat nous montre que la théorie de l'aléa moral s'applique, car on observe plus de sinistres de la part des individus les plus couverts en termes de garanties

Analyse de la fréquence et du coût moyen suivant l'option Pack+ souscrite



f. Variable Option Franchise

Il s'agit d'une variable qualitative à cinq niveaux décrivant le niveau de franchise de l'assuré. Dans notre portefeuille,

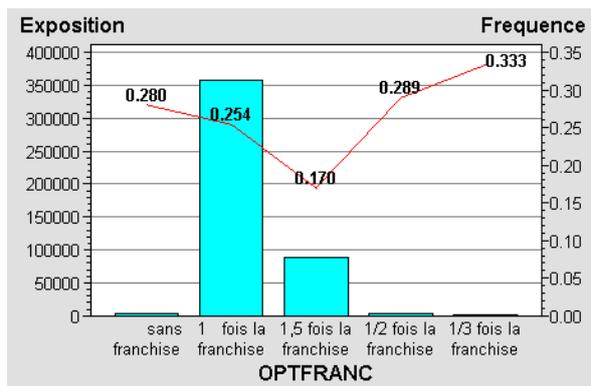
- **2693** niveau de franchise 0 : **pas de franchise de base** (soit **0,6 %**),
- **210** niveau de franchise 1 : **1/3 fois la franchise de base** (soit **0,05 %**),
- **3 922** niveau de franchise 2 : **1/2 fois la franchise de base** (soit **0,85 %**),
- **356 829** niveau de franchise 3 : **1 fois la franchise de base** (soit **79 %**) (**niveau de référence**),
- **87 963** niveau de franchise 4 : **1,5 fois la franchise de base** (soit **19,5 %**).

Afin d'apprécier l'influence du choix du niveau de franchise sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant le niveau de franchise.

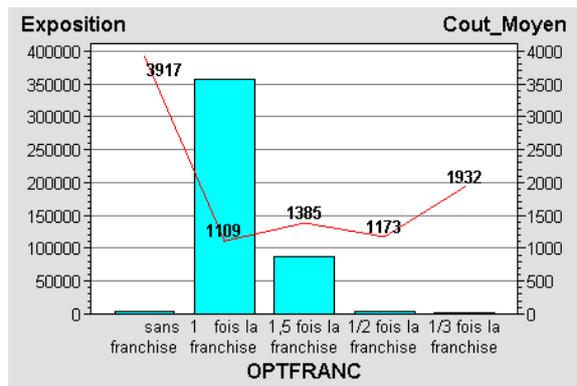
On constate que la fréquence est décroissante en fonction du niveau de franchise. Nous pouvons remarquer que le coût moyen du niveau de franchise 0 est élevé et dans une moindre mesure celui du niveau 1 l'est aussi. Le niveau de franchise 3 est inférieur à la valeur moyenne.

A noter que, pour les agris, le niveau de franchise 0 est au prix du niveau de franchise 3.

Analyse de la fréquence suivant le niveau de la franchise



Analyse du coût moyen suivant le niveau de la franchise



g. Variable Crédit Bonus

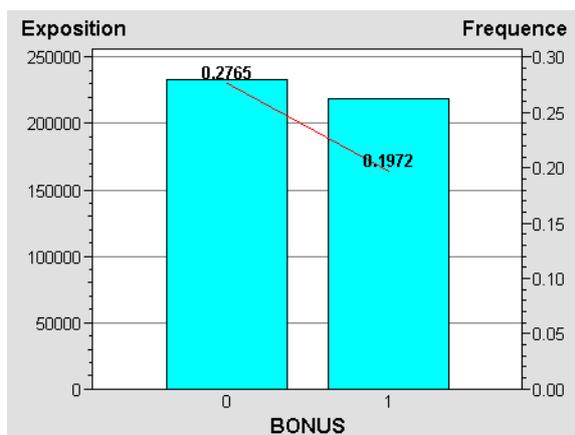
Il s'agit d'une variable qualitative binaire indiquant si l'assuré est éligible au crédit bonus, i.e. ayant un CRM à 50 et sans sinistre depuis plus de 3 ans. Dans notre portefeuille,

CREDITBONUS = $\begin{cases} 1, & \text{si l'assuré est éligible au crédit bonus } \mathbf{218\ 337}, \text{ (soit } \mathbf{52\ %}), \\ 0, & \text{si l'assuré n'est pas éligible au crédit bonus } \mathbf{233\ 280}, \text{ (soit } \mathbf{48\ %}) \\ \end{cases}$ (**niveau de référence**).

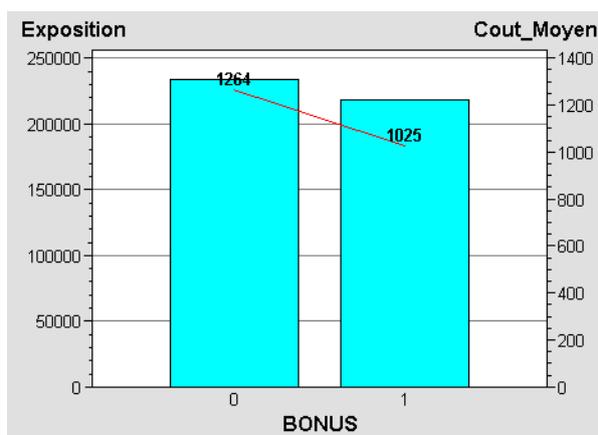
Afin d'apprécier l'influence de la variable crédit bonus sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable crédit bonus.

On constate clairement que le fait que l'assuré soit éligible au crédit bonus entraîne une baisse significative de la sinistralité aussi bien en termes de fréquence qu'en terme de coût moyen.

Analyse de la fréquence suivant le Crédit bonus



Analyse du coût moyen suivant le Crédit bonus



h. Variable Tranche de CRM (Coefficient Réduction Majoration)

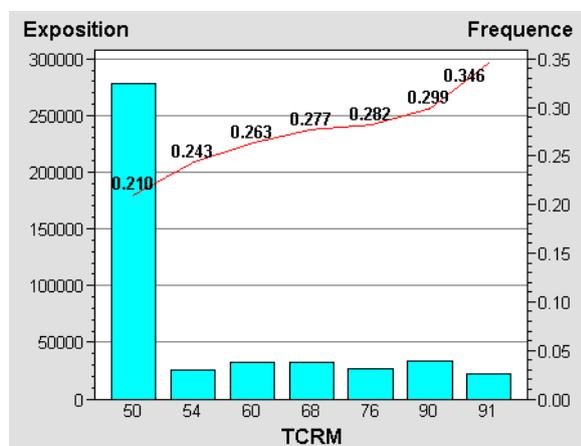
Il s'agit d'une variable qualitative à 7 niveaux décrivant la Tranche de CRM i.e. l'historique de la sinistralité de l'assuré. Dans notre portefeuille,

- 278 537 CRM 50 (soit 62 %) (niveau de référence),
- 25 163 CRM] 50 ; 54] (soit 6 %),
- 32 280 CRM] 54 ; 60] (soit 7 %),
- 32 673 CRM] 60 ; 68] (soit 7 %),
- 27 041 CRM] 68 ; 76] (soit 6 %).
- 33 371 CRM] 76 ; 90] (soit 7 %).
- 22 551 CRM 91 & + (soit 5 %).

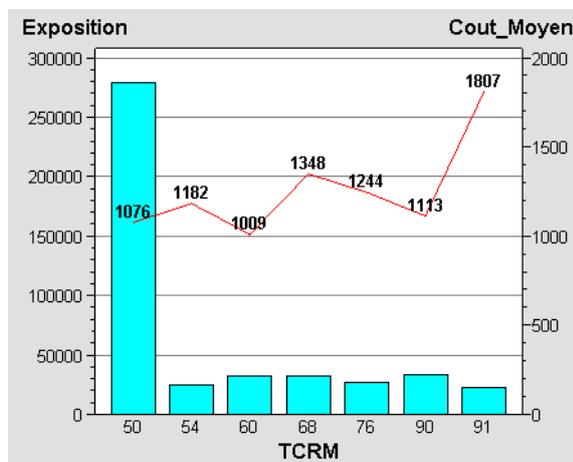
Afin d'apprécier l'influence du CRM sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen par sinistre suivant la Tranche de CRM.

On constate une croissance de la sinistralité en termes de fréquence mais en terme de coût moyen aucune tendance ne ressort vraiment suivant la Tranche de CRM. Mais une sursinistralité significative apparaît pour la Tranche de CRM] 60 ; 68] et pour la Tranche de CRM 91 & +.

Analyse de la fréquence suivant la tranche de CRM



Analyse du coût moyen suivant la tranche de CRM



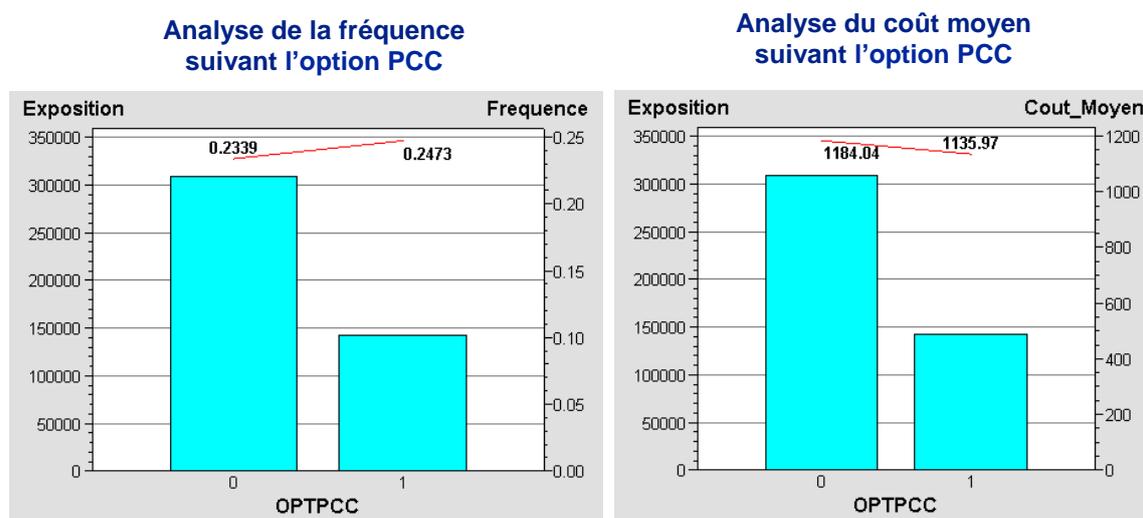
i. Variable OPTPCC (Option Protection Corporelle du Conducteur)

Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose de l'Option PCC.

$$\text{OPTPCC} = \begin{cases} 1, & \text{si l'assuré dispose de l'Option PCC 218 337, (soit 52 \%),} \\ 0, & \text{si l'assuré ne dispose pas de l'Option PCC 233 280, (soit 48 \%) } \\ & \text{(niveau de référence).} \end{cases}$$

Afin d'apprécier l'influence de la variable **OPTPCC** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **OPTPCC**.

On constate clairement que le fait que l'assuré soit équipé de l'option **OPTPCC** entraîne une hausse de la sinistralité en termes de fréquence et une baisse du coût moyen.



j. Variable OPTMULTI (Option Multiconducteur)

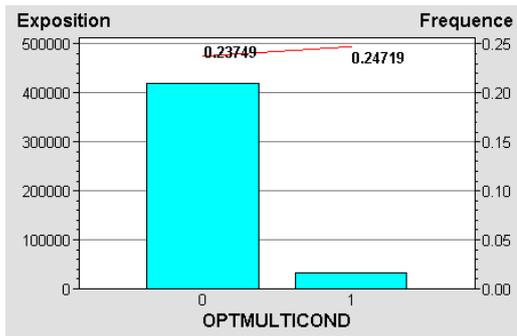
Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose de l'Option **Multiconducteur**.

$$\text{OPTMULTI} = \begin{cases} 1, & \text{si l'assuré dispose de l'Option Multiconducteur 31 798 (soit 7 \%),} \\ 0, & \text{si l'assuré ne dispose pas de l'Option Multiconducteur 419 820 (soit 93 \%) } \\ & \text{(niveau de référence).} \end{cases}$$

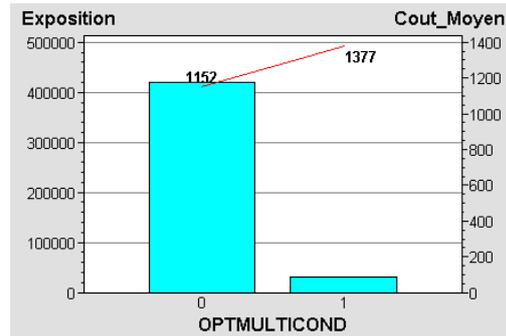
Afin d'apprécier l'influence de la variable **OPTMULTI** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **OPTMULTI**.

On constate clairement que le fait que l'assuré soit équipé de l'option **OPTMULTI** entraîne une hausse de la sinistralité en termes de fréquence ainsi qu'en terme de coût moyen.

**Analyse de la fréquence
suivant l'option Multiconducteur**



**Analyse du coût moyen
suivant l'option Multiconducteur**



k. Variable OPTCTPRIVE (Option Contenu Privé)

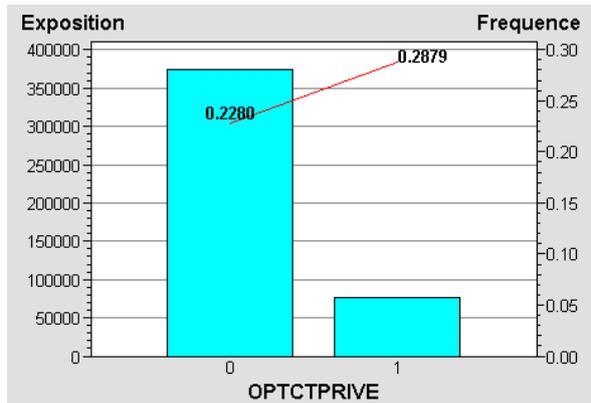
Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose de l'Option **Contenu Privé**.

OPTCTPRIVE = $\begin{cases} 1, & \text{si l'assuré dispose de l'Option Contenu Privé } 76\,983 \text{ (soit } 17\% \text{),} \\ 0, & \text{si l'assuré ne dispose pas de l'Option Contenu Privé } 374\,634 \text{ (soit } 83\% \text{) (niveau de référence).} \end{cases}$

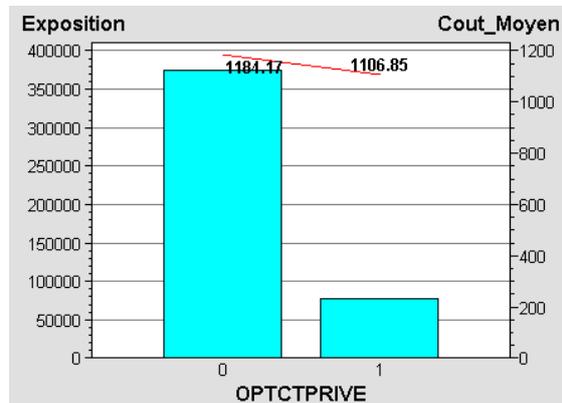
Afin d'apprécier l'influence de la variable **OPTCTPRIVE** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **OPTCTPRIVE**.

On constate clairement que le fait que l'assuré soit équipé de l'option **OPTCTPRIVE** entraîne une hausse de la sinistralité en termes de fréquence mais en terme de coût moyen nous sommes inférieurs à la moyenne.

**Analyse de la fréquence
suivant l'option Contenu privé**



**Analyse du coût moyen
suivant l'option Contenu privé**



l. Variable GARFERME (Garage Fermé)

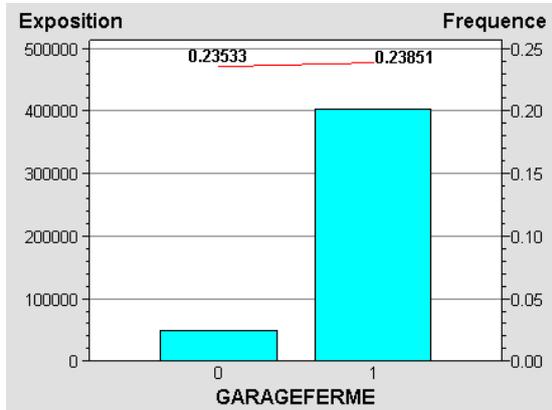
Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose d'un Garage Fermé.

GARFERME = $\begin{cases} 1, & \text{si l'assuré dispose d'un Garage Fermé } 403\,481 \text{ (soit } 89\% \text{) (niveau de référence),} \\ 0, & \text{si l'assuré ne dispose pas d'un Garage Fermé } 48\,136 \text{ (soit } 11\% \text{).} \end{cases}$

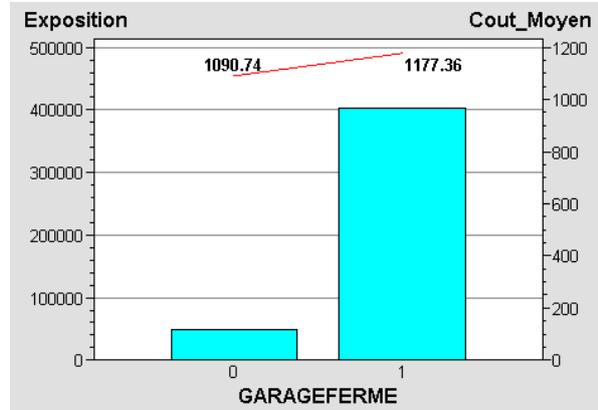
Afin d'apprécier l'influence de la variable **GARFERME** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **GARFERME**.

Il est très difficile de donner une orientation sur la sinistralité, toutes garanties confondues, de cette variable au niveau fréquence par contre au niveau coût moyen si l'assuré dispose d'un garage fermé alors il enregistre une valeur supérieure à la moyenne.

**Analyse de la fréquence
suivant la présence d'un garage**

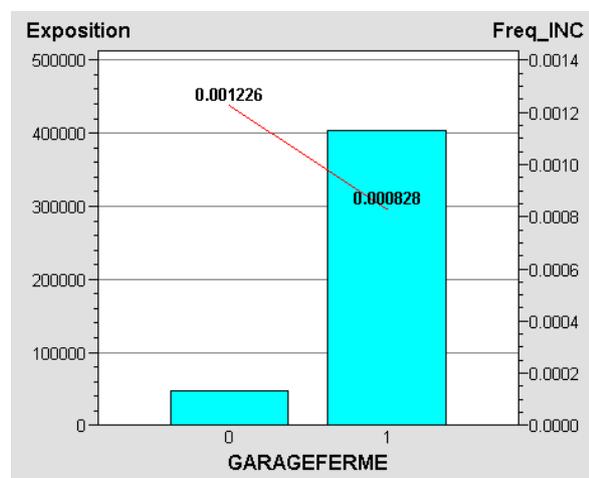
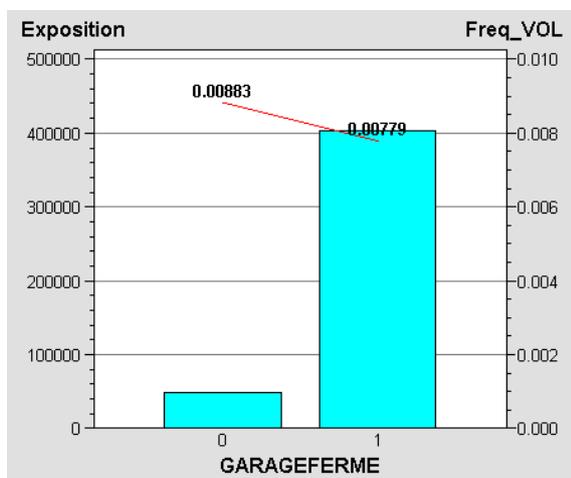


**Analyse du coût moyen
suivant la présence d'un garage**



Nous allons maintenant nous attacher à regarder uniquement les garanties Vol et Incendie. Nous constatons clairement que, le fait que l'assuré dispose d'un Garage Fermé entraîne une baisse de la sinistralité en termes de fréquence.

Analyse de la fréquence Vol et Incendie suivant la présence d'un garage



m. Variable OPTVEHRPLT (Option Contenu Privé)

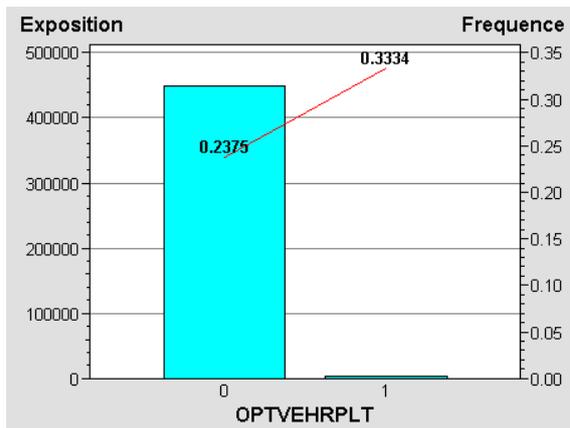
Il s'agit d'une variable qualitative binaire indiquant si l'assuré dispose de l'Option **Véhicule de Remplacement**.

OPTVEHRPLT = $\begin{cases} 1, & \text{si l'assuré dispose de l'Option Véhicule de Remplacement 2 934} \\ & \text{(soit 1 \%),} \\ 0, & \text{si l'assuré ne dispose pas de l'Option Véhicule de Remplacement} \\ & \text{448 684 (soit 99 \%) (niveau de référence).} \end{cases}$

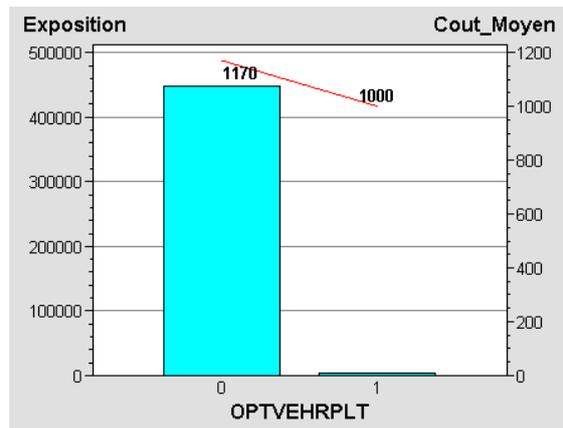
Afin d'apprécier l'influence de la variable **OPTVEHRPLT** sur la sinistralité, on peut voir dans les graphes suivants la fréquence des sinistres observée et le coût moyen suivant la variable **OPTVEHRPLT**.

On constate clairement que le fait que l'assuré soit équipé de l'option **OPTVEHRPLT** entraîne une hausse de la sinistralité en termes de fréquence et à contrario une baisse de la sinistralité en termes de coût moyen.

Analyse de la fréquence suivant l'option Véhicule de remplacement



Analyse du coût moyen suivant l'option Véhicule de remplacement



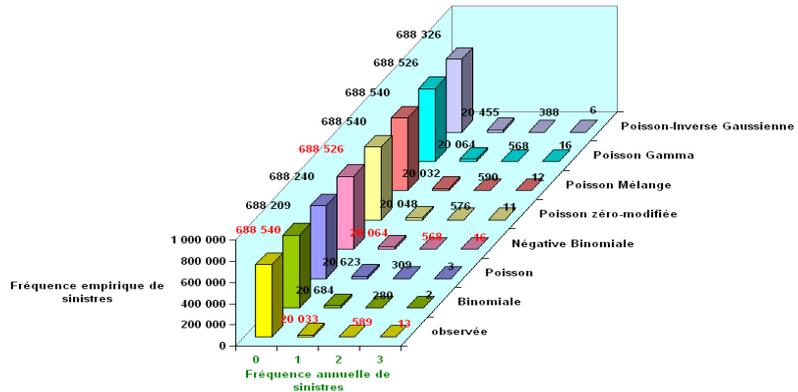
4. Analyse de la distribution du coût moyen d'un sinistre suivant la garantie :

Nous allons présenter les différents modèles de comptage concernant les garanties suivantes : RCM(Responsabilité Civile Matériel), RCC(Responsabilité Civile Corporelle), DREC(Défense Recours), BDG(Bris De Glaces), DOMM(Dommage tous accidents), INC(Incendie) et VOL(Vol ou tentative de Vol).

a. concernant la garantie RCM

| Modèle de comptage concernant la Garantie RCM | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | |
|---|----------|---|------------|------------|--------------|----------|--------------|-----------|--------------------|----------|-----------------------|-----------|-----------------|-------------|---------------|----------|----------------------------|----------|
| | | | $p^n/n!$ | $p^n/n!$ | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | |
| | | | 0,00599 | 0,00300 | 1 | | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | |
| | | | α | α | α | α | λ | λ | p | r | λ | λ | λ_1 | λ_2 | p | r | μ | β |
| | 5 | 10 | 5 | 10 | | 0,0300 | 0,9725 | 1,0601 | | 0,0575 | 0,00292 | 0,06226 | 0,97251 | 1,06010 | 0,02996 | 0,00836 | | |
| classes de fréquence annuelle | observée | fréquence relative empirique f _j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux |
| 0 | 688 540 | 97,09% | 688 178,18 | 688 209,26 | 0,19 | 0,16 | 688 240,22 | 0,13 | 688 526,35 | 0,00 | 688 540,00 | 0,00 | 688 540,20 | 0,00 | 688 526,35 | 0,00 | 688 325,66 | 0,07 |
| 1 | 20 033 | 2,82% | 20 745,17 | 20 683,75 | 24,45 | 20,47 | 20 622,70 | 16,86 | 20 064,15 | 0,05 | 20 047,76 | 0,01 | 20 032,20 | 0,00 | 20 064,15 | 0,05 | 20 455,06 | 8,71 |
| 2 | 589 | 0,08% | 250,15 | 279,74 | 459,02 | 341,90 | 306,97 | 253,79 | 568,11 | 0,77 | 576,04 | 0,29 | 590,19 | 0,00 | 588,11 | 0,77 | 387,99 | 104,15 |
| 3 | 13 | 0,00% | 1,51 | 2,24 | 87,57 | 51,62 | 3,09 | 31,85 | 16,39 | 0,70 | 11,19 | 0,29 | 12,41 | 0,03 | 16,39 | 0,70 | 6,29 | 7,15 |
| nb occurrence | 4 | | 709 175 | 709 175 | 571,23 | 414,16 | 709 175 | 302,63 | 709 175 | 1,52 | 709 175 | 0,59 | 709 175 | 0,03 | 709 175 | 1,52 | 709 175 | 120 |
| μ | 0,0300 | moeyenne | | | | | | | | | | | | | | | | |
| σ^2 | 0,0317 | moment d'ordre 2 | | | | | | | | | | | | | | | | |
| σ | 0,0308 | variance | | | | | | | | | | | | | | | | |
| cv | 5,86 | coefficient de variati | | | | | | | | | | | | | | | | |
| μ_3 | 0,0354 | moment d'ordre 3 | | | | | | | | | | | | | | | | |
| μ_4 | 0,0430 | moment d'ordre 4 | | | | | | | | | | | | | | | | |
| μ_3 | 0,0326 | moment centré d'ordre 3 | | | | | | | | | | | | | | | | |
| μ_4 | 0,0389 | moment centré d'ordre 4 | | | | | | | | | | | | | | | | |
| γ_1 | 6,02 | coefficient d'asymétrie skewness | | | | | | | | | | | | | | | | |
| γ_2 | 37,96 | coefficient d'aplatissement | | | | | | | | | | | | | | | | |
| $P[N=0]$ | 97,09% | | | | | | | | | | | | | | | | | |

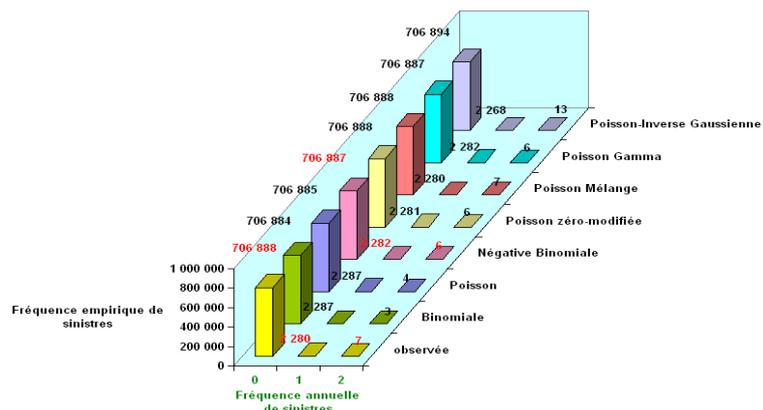
Diagramme en bâtons sur données MRACA concernant la garantie RCM Automobile



b. concernant la garantie RCC

| Modèle de comptage concernant la Garantie RCC | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | |
|---|----------|---|------------|------------|--------------|----------|--------------|-----------|--------------------|----------|-----------------------|-----------|-----------------|-------------|---------------|----------|----------------------------|----------|
| | | | $p^n/n!$ | $p^n/n!$ | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | |
| | | | 0,00065 | 0,00032 | 1 | | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | |
| | | | α | α | α | α | λ | λ | p | r | λ | λ | λ_1 | λ_2 | p | r | μ | β |
| | 5 | 10 | 5 | 10 | | 0,0032 | 0,9981 | 1,6736 | | 0,00507 | -0,0108 | 0,0039 | 0,9981 | 1,6736 | 0,00323 | 0,00836 | | |
| classes de fréquence annuelle | observée | fréquence relative empirique f _j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux |
| 0 | 706 888 | 99,68% | 706 883,97 | 706 884,34 | 0,00 | 0,00 | 706 884,71 | 0,00 | 706 886,91 | 0,00 | 706 888,00 | 0,00 | 706 888,00 | 0,00 | 706 886,91 | 0,00 | 706 884,18 | 0,00 |
| 1 | 2 280 | 0,32% | 2 288,07 | 2 287,33 | 0,03 | 0,02 | 2 286,59 | 0,02 | 2 282,19 | 0,00 | 2 281,21 | 0,00 | 2 280,00 | 0,00 | 2 282,19 | 0,00 | 2 287,75 | 0,07 |
| 2 | 7 | 0,00% | 2,96 | 3,33 | 5,49 | 4,03 | 3,70 | 2,94 | 5,90 | 0,21 | 5,79 | 0,25 | 7,00 | 0,00 | 5,90 | 0,21 | 13,07 | 2,82 |
| nb occurrence | 3 | | 709 175 | 709 175 | 5,52 | 4,06 | 709 175 | 2,96 | 709 175 | 0,21 | 709 175 | 0 | 709 175 | 0,00 | 709 175 | 0,21 | 709 175 | 2,88 |
| μ | 0,0032 | moeyenne | | | | | | | | | | | | | | | | |
| σ^2 | 0,0033 | moment d'ordre 2 | | | | | | | | | | | | | | | | |
| σ | 0,0032 | variance | | | | | | | | | | | | | | | | |
| cv | 17,61 | coefficient de variati | | | | | | | | | | | | | | | | |
| μ_3 | 0,0033 | moment d'ordre 3 | | | | | | | | | | | | | | | | |
| μ_4 | 0,0034 | moment d'ordre 4 | | | | | | | | | | | | | | | | |
| μ_3 | 0,0033 | moment centré d'ordre 3 | | | | | | | | | | | | | | | | |
| μ_4 | 0,0033 | moment centré d'ordre 4 | | | | | | | | | | | | | | | | |
| γ_1 | 17,66 | coefficient d'asymétrie skewness | | | | | | | | | | | | | | | | |
| γ_2 | 313,48 | coefficient d'aplatissement | | | | | | | | | | | | | | | | |
| $P[N=0]$ | 99,68% | | | | | | | | | | | | | | | | | |

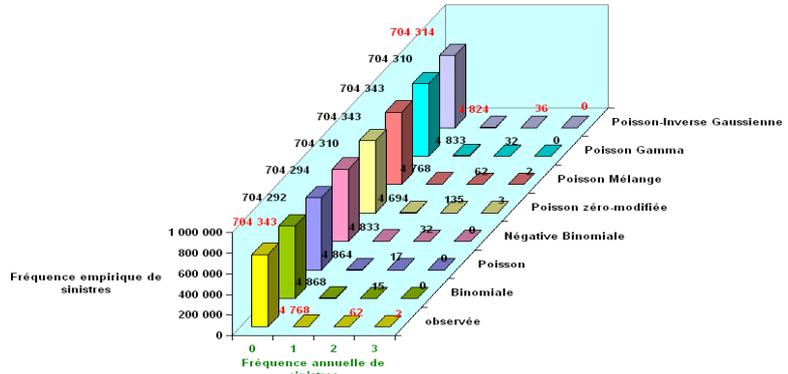
Diagramme en bâtons sur données MRACA concernant la garantie RCC Automobile



c. concernant la garantie DREC

| Modèle de comptage concernant la Garantie DREC | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | | | | | |
|--|----------|---|---------------------|---------------------|--------------|----------|--------------|--------------|--------------------|--------------|-----------------------|--------------|-----------------|--------------|---------------|--------------|----------------------------|--------------|------------|-----------|------|--|
| | | | p ^m /n/α | p ^m /n/α | paramètre(s) | | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | | paramètre(s) | | | | |
| | | | 0,00138 | 0,00069 | 1 | | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | | 2 | | | |
| | | | α | α | α | α | λ = | | β | r | λ = | | λ1 | λ2 | p = | r | μ | β | | | | |
| | | | 5 | 10 | 5 | 10 | λ = | | 0,0089 | 0,9935 | 1,0601 | λ = | | 0,0575 | 0,00541 | 0,10285 | 0,99363 | 1,08010 | 0,00691 | 0,00836 | | |
| classes de fréquence annuelle | observée | fréquence relative empirique f _j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | | |
| 0 | 704 343 | 99,32% | 704 290,51 | 704 292,19 | 0,00 | 0,00 | 704 293,88 | 0,00 | 704 309,65 | 0,00 | 704 343,00 | 0,00 | 704 343,05 | 0,00 | 704 309,65 | 0,00 | 704 309,65 | 0,00 | 704 314,03 | 0,00 | | |
| 1 | 4 768 | 0,67% | 4 870,99 | 4 867,64 | 2,18 | 2,04 | 4 864,29 | 1,91 | 4 832,91 | 0,87 | 4 694,49 | 1,15 | 4 767,80 | 0,00 | 4 832,91 | 0,87 | 4 824,28 | 0,66 | 4 824,28 | 0,66 | | |
| 2 | 62 | 0,01% | 13,48 | 15,14 | 174,73 | 145,05 | 16,80 | 121,64 | 32,22 | 27,52 | 134,89 | 39,39 | 62,30 | 0,00 | 32,22 | 27,52 | 36,35 | 18,11 | 36,35 | 18,11 | | |
| 3 | 2 | 0,00% | 0,02 | 0,03 | 210,61 | 139,39 | 0,04 | 99,47 | 0,21 | 14,89 | 2,62 | 0,15 | 1,85 | 0,01 | 0,21 | 14,89 | 0,34 | 8,10 | 0,34 | 8,10 | | |
| | | | 709 175 | 100,00% | 709 175 | 387,53 | 286,48 | 709 175 | 223,02 | 709 175 | 43,28 | 709 175 | 40,89 | 709 175 | 0,01 | 709 175 | 43,28 | 709 175 | 27 | 27 | | |
| nb occurrence | 4 | | P-value = | | 0% | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | #NOMBREI | P-value = | 0% | P-value = | 0% | P-value = | 0% | |
| ^n = | 0,0069 | moeyne | K2 = | | 5,99 | 5,99 | K2 = | 5,99 | K2 = | 3,84 | K2 = | 5,99 | K2 = | 5,99 | #NOMBREI | K2 = | 3,84 | K2 = | 3,84 | K2 = | 3,84 | |
| ^m2 = | 0,0071 | moment d'ord | | | | | | | | | | | | | | | | | | | | |
| ^σ2 = | 0,0071 | variance | | | | | | | | | | | | | | | | | | | | |
| cv = | 12,16 | coefficient de | | | | | | | | | | | | | | | | | | | | |
| ^m3 = | 0,0075 | moment d'ordre 3 | | | | | | | | | | | | | | | | | | | | |
| ^m4 = | 0,0084 | moment d'ordre 4 | | | | | | | | | | | | | | | | | | | | |
| ^μ3 = | 0,0074 | moment centré d'ordre 3 | | | | | | | | | | | | | | | | | | | | |
| ^μ4 = | 0,0081 | moment centré d'ordre 4 | | | | | | | | | | | | | | | | | | | | |
| ^γ1 = | 12,42 | coefficient d'asymétrie skewness | | | | | | | | | | | | | | | | | | | | |
| ^γ2 = | 160,85 | coefficient d'aplatissement | | | | | | | | | | | | | | | | | | | | |
| P(N=0) = | 99,32% | | | | | | | | | | | | | | | | | | | | | |

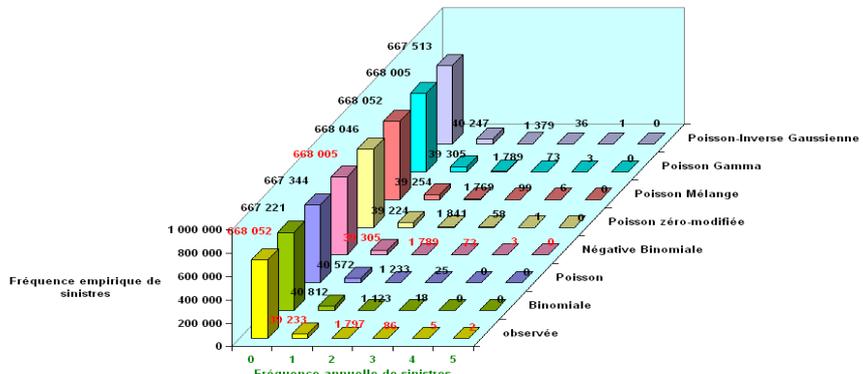
Diagramme en bâtons sur données MRACA concernant la garantie DREC Automobile



d. concernant la garantie BDG

| Modèle de comptage concernant la Garantie BDG | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | | | | | |
|---|----------|---|---------------------|---------------------|--------------|-----------|--------------|--------------|--------------------|--------------|-----------------------|--------------|-----------------|--------------|---------------|--------------|----------------------------|--------------|--------------|----------|--|--|
| | | | p ^m /n/α | p ^m /n/α | paramètre(s) | | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | paramètre(s) | | | |
| | | | 0,012159199 | 0,0060796 | 1 | | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | | 2 | | | |
| | | | α | α | α | α | λ = | | β | r | λ = | | λ1 | λ2 | p = | r | μ | β | | | | |
| | | | 5 | 10 | 5 | 10 | λ = | | 0,0808 | 0,9678 | 1,8290 | λ = | | 0,09395 | 0,04982 | 0,26492 | 0,96783 | 1,82898 | 0,08080 | 0,00836 | | |
| classes de fréquence annuelle | observée | fréquence relative empirique f _j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | | |
| 0 | 668 052 | 94,20% | 667 095,82 | 667 220,63 | 1,37 | 1,04 | 667 344,45 | 0,75 | 668 004,51 | 0,00 | 668 052,00 | 0,00 | 668 046,22 | 0,00 | 668 004,51 | 0,00 | 667 512,57 | 0,44 | 667 512,57 | 0,44 | | |
| 1 | 39 233 | 5,53% | 41 055,96 | 40 812,47 | 80,94 | 61,13 | 40 571,87 | 44,18 | 39 305,47 | 0,13 | 39 223,52 | 0,00 | 39 254,48 | 0,01 | 39 305,47 | 0,13 | 40 247,20 | 25,58 | 40 247,20 | 25,58 | | |
| 2 | 1 797 | 0,25% | 1 010,70 | 1 123,39 | 611,71 | 403,92 | 1 233,30 | 257,64 | 1 788,62 | 0,04 | 1 840,52 | 1,03 | 1 768,89 | 0,45 | 1 768,62 | 0,04 | 1 378,72 | 126,90 | 1 378,72 | 126,90 | | |
| 3 | 86 | 0,01% | 12,44 | 18,32 | 434,94 | 249,95 | 24,99 | 148,91 | 73,44 | 2,15 | 57,58 | 14,03 | 99,23 | 1,76 | 73,44 | 2,15 | 35,72 | 70,79 | 35,72 | 70,79 | | |
| 4 | 5 | 0,00% | 0,08 | 0,20 | 316,60 | 117,65 | 0,38 | 56,19 | 2,85 | 1,62 | 1,35 | 9,86 | 5,86 | 0,13 | 2,85 | 1,62 | 0,78 | 22,85 | 0,78 | 22,85 | | |
| 5 | 2 | 0,00% | 0,00 | 0,00 | 21 217,78 | 2 760,12 | 0,00 | 853,24 | 0,11 | 32,13 | 0,03 | 151,32 | 0,32 | 8,92 | 0,11 | 32,13 | 0,02 | 249,01 | 0,02 | 249,01 | | |
| | | | 709 175 | 100,00% | 709 175 | 22 863,35 | 3 593,80 | 709 175 | 1 360,92 | 709 175 | 36,07 | 709 175 | 176,24 | 709 175 | 11,27 | 709 175 | 36,07 | 709 175 | 495,35 | 495,35 | | |
| nb occurrence | 6 | | P-value = | | 0% | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | P-value = | 0% | | |
| ^n = | 0,0608 | moeyne | K2 = | | 9,49 | 9,49 | K2 = | 9,49 | K2 = | 7,81 | K2 = | 9,49 | K2 = | 9,49 | K2 = | 7,81 | K2 = | 7,81 | K2 = | 7,81 | | |
| ^m2 = | 0,0667 | moment d'ordre 2 | | | | | | | | | | | | | | | | | | | | |
| ^σ2 = | 0,0630 | variance | | | | | | | | | | | | | | | | | | | | |
| cv = | 4,13 | coefficient de variatio | | | | | | | | | | | | | | | | | | | | |
| ^m3 = | 0,0797 | moment d'ordre 3 | | | | | | | | | | | | | | | | | | | | |
| ^m4 = | 0,1093 | moment d'ordre 4 | | | | | | | | | | | | | | | | | | | | |
| ^μ3 = | 0,0679 | moment centré d'ordr | | | | | | | | | | | | | | | | | | | | |
| ^μ4 = | 0,0913 | moment centré d'ordre 4 | | | | | | | | | | | | | | | | | | | | |
| ^γ1 = | 1,29 | coefficient d'asymétrie skewn | | | | | | | | | | | | | | | | | | | | |
| ^γ2 = | 19,98 | coefficient d'aplatissement | | | | | | | | | | | | | | | | | | | | |
| P(N=0) = | 94,20% | | | | | | | | | | | | | | | | | | | | | |

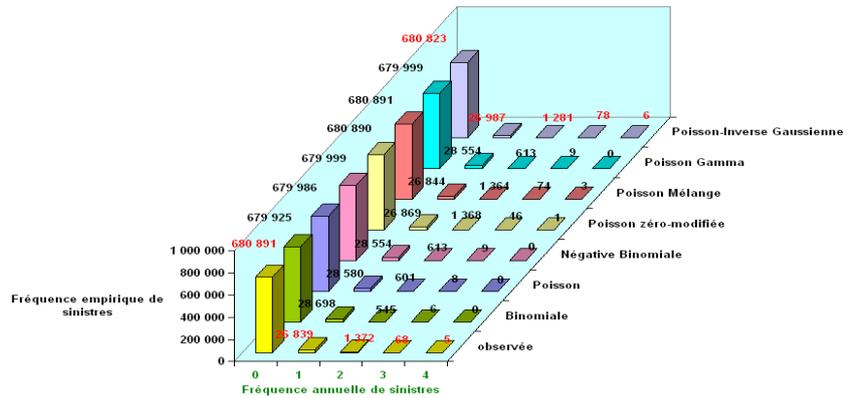
Diagramme en bâtons sur données MRACA concernant la garantie BDG Automobile



e. concernant la garantie DOMM

| Modèle de comptage concernant la Garantie DOMM | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | |
|--|----------|----------------------------------|------------|------------|--------------|-------------|--------------|-------------|--------------------|----------|-----------------------|-----------|-----------------|-------------|---------------|----------|----------------------------|----------|
| | | | $p^n/n!$ | $p^n/n!$ | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | |
| | | | α | α | α | α | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | |
| | | | 5 | 10 | 5 | 10 | λ | λ | p | r | λ | λ | λ_1 | λ_2 | p | r | μ | β |
| | | | | | | $\lambda =$ | | $\lambda =$ | | | | | $\lambda =$ | | | | | |
| | | | | | | 0,0420 | 0,9991 | 44,51675734 | | 0,10179 | 0,02163 | 0,17877 | 0,99906 | 44,51676 | 0,04203 | 0,96215 | | |
| | | | | | | | | | | | $p =$ | 0,87126 | β | 1059,15293 | | | | |
| classes de fréquence annuelle | observée | fréquence relative empirique f_j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux |
| 0 | 680 891 | 96,01% | 679 864,93 | 679 925,49 | 1,55 | 1,37 | 679 985,72 | 1,21 | 679 999,20 | 1,17 | 680 891,00 | 0,00 | 680 889,76 | 0,00 | 679 999,20 | 1,17 | 680 823,11 | 0,01 |
| 1 | 26 839 | 3,78% | 26 817,32 | 26 698,25 | 135,81 | 120,45 | 26 580,16 | 106,07 | 28 553,77 | 102,98 | 26 866,87 | 0,03 | 26 844,11 | 0,00 | 26 553,77 | 102,98 | 26 987,22 | 0,81 |
| 2 | 1 372 | 0,19% | 488,59 | 545,08 | 1 597,28 | 1 254,48 | 600,62 | 990,69 | 612,97 | 939,91 | 1 367,52 | 0,01 | 1 364,02 | 0,05 | 612,97 | 939,91 | 1 280,78 | 6,50 |
| 3 | 68 | 0,01% | 4,14 | 6,14 | 984,52 | 623,83 | 8,41 | 421,92 | 8,97 | 388,74 | 46,40 | 10,05 | 73,75 | 0,45 | 8,97 | 388,74 | 77,87 | 1,25 |
| 4 | 5 | 0,00% | 0,02 | 0,05 | 1 411,59 | 538,93 | 0,09 | 270,46 | 0,10 | 236,69 | 1,21 | 11,95 | 3,37 | 0,78 | 10,10 | 236,69 | 6,02 | 0,17 |
| | 709 175 | 100,00% | 709 175 | 709 175 | 4 130,75 | 2 539,06 | 709 175 | 1 790,35 | 709 175 | 1 669,49 | 709 175 | 22,05 | 709 175 | 1,28 | 709 175 | 1 669,49 | 709 175 | 8,74 |
| nb occurrence | 5 | | | | | | | | | | | | | | | | | |
| n | 0,0420 | moyenne | | | | | | | | | | | | | | | | |
| m_2 | 0,0466 | moment d'ord | | | | | | | | | | | | | | | | |
| σ_2 | 0,0448 | variance | | | | | | | | | | | | | | | | |
| cv | 5,04 | coefficient de | | | | | | | | | | | | | | | | |
| m_3 | 0,0564 | moment d'ord | | | | | | | | | | | | | | | | |
| m_4 | 0,0784 | moment d'ord | | | | | | | | | | | | | | | | |
| μ_3 | 0,0506 | moment centré d'ordre 3 | | | | | | | | | | | | | | | | |
| μ_4 | 0,0694 | moment centré d'ordre 4 | | | | | | | | | | | | | | | | |
| γ_1 | 5,34 | coefficient d'asymétrie skewness | | | | | | | | | | | | | | | | |
| γ_2 | 31,58 | coefficient d'aplatissement | | | | | | | | | | | | | | | | |
| $P[N=0]$ | 96,01% | | | | | | | | | | | | | | | | | |

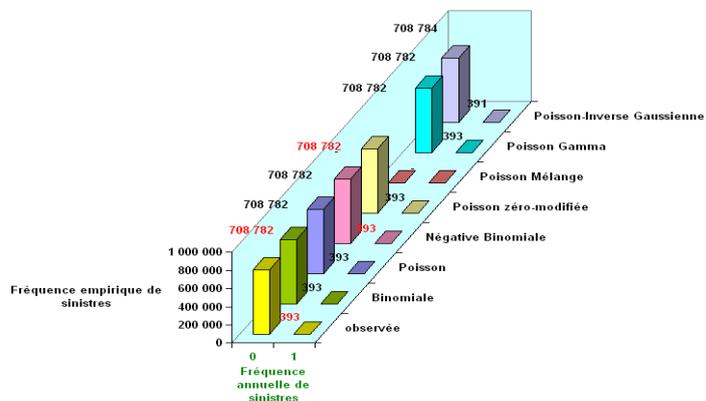
Diagramme en bâtons sur données MRACA concernant la garantie DOMM Automobile



f. concernant la garantie INC

| Modèle de comptage concernant la Garantie INC | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | |
|---|----------|----------------------------------|------------|------------|--------------|----------|--------------|-----------|--------------------|----------|-----------------------|-----------|-----------------|-------------|---------------|----------|----------------------------|----------|
| | | | $p^n/n!$ | $p^n/n!$ | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | |
| | | | α | α | α | α | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | |
| | | | 5 | 10 | 5 | 10 | λ | λ | p | r | λ | λ | λ_1 | λ_2 | p | r | μ | β |
| | | | | | | 0,0006 | 0,9999 | 7,6817 | | - | 0,00615 | 0,00000 | 0,00000 | 0,99993 | 7,68170 | 0,00055 | 0,00836 | |
| | | | | | | | | | | $p =$ | #N/D! | #N/D! | β | 13861,75690 | | | | |
| classes de fréquence annuelle | observée | fréquence relative empirique f_j | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux |
| 0 | 708 782 | 99,94% | 708 782,09 | 708 782,10 | 0,00 | 0,00 | 708 782,11 | 0,00 | 708 782,12 | 0,00 | 708 782,00 | - | #N/D! | #N/D! | 708 782,12 | 0,00 | 708 783,74 | 0,00 |
| 1 | 393 | 0,06% | 392,91 | 392,90 | 0,00 | 0,00 | 392,89 | 0,00 | 392,88 | 0,00 | 393,00 | 0,00 | #N/D! | #N/D! | 392,89 | 0,00 | 391,26 | 0,01 |
| | 709 175 | 100,00% | 709 175 | 709 175 | 0,00 | 0,00 | 709 175 | 0,00 | 709 175 | 0,00 | 709 175 | 0,00 | #N/D! | #N/D! | 709 175 | 0,00 | 709 175 | 0,01 |
| nb occurrence | 2 | | | | | | | | | | | | | | | | | |
| n | 0,0006 | moyenne | | | | | | | | | | | | | | | | |
| m_2 | 0,0006 | moment d'ord | | | | | | | | | | | | | | | | |
| σ_2 | 0,0006 | variance | | | | | | | | | | | | | | | | |
| cv | 42,47 | coefficient de variation | | | | | | | | | | | | | | | | |
| m_3 | 0,0006 | moment d'ordre 3 | | | | | | | | | | | | | | | | |
| m_4 | 0,0006 | moment d'ordre 4 | | | | | | | | | | | | | | | | |
| μ_3 | 0,0006 | moment centré d'ordre 3 | | | | | | | | | | | | | | | | |
| μ_4 | 0,0006 | moment centré d'ordre 4 | | | | | | | | | | | | | | | | |
| γ_1 | 42,44 | coefficient d'asymétrie skewness | | | | | | | | | | | | | | | | |
| γ_2 | 1 799,52 | coefficient d'aplatissement | | | | | | | | | | | | | | | | |
| $P[N=0]$ | 99,94% | | | | | | | | | | | | | | | | | |

Diagramme en bâtons sur données MRACA concernant la garantie INC Automobile

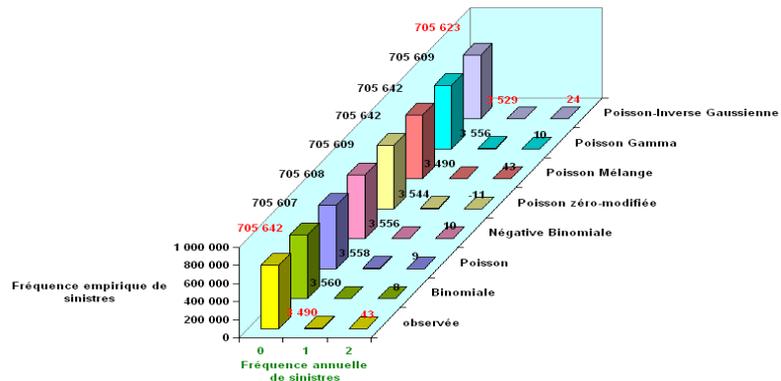


g. concernant la garantie VOL

| Modèle de comptage concernant la Garantie VOL | | | Binomiale | | | | Poisson | | Négative Binomiale | | Poisson zéro-modifiée | | Poisson Mélange | | Poisson Gamma | | Poisson-Inverse Gaussienne | | | |
|---|----------|---------------------------------|---------------------|------------|---------------------|----------|--------------|----------|--------------------|----------|-----------------------|----------|--------------------|----------|--------------------|----------|----------------------------|----------|----------|--|
| | | | p ⁿ /n!α | | p ⁿ /n!α | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | paramètre(s) | | | |
| | | | 0,001008496 | | 0,000504248 | | 1 | | 2 | | 1 | | 3 | | 2 | | 2 | | | |
| | | | α | α | α | α | λ | λ | λ | λ | λ1 | λ2 | p | p | μ | β | μ | β | | |
| 5 | 10 | 5 | 10 | λ = | 0,0050 | 0,9993 | 7,6817 | λ = | - 0,00615 | -0,01598 | 0,00960 | 0,99934 | 7,68170 | 0,00504 | 0,00836 | | | | | |
| | | | | | | | | | | | | | | p = | 0,17819 | β | 1523,39778 | | | |
| classes de fréquence annuelle | observée | fréquence relative empirique fj | nsin_théo | nsin_théo | khi-deux | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | nsin_théo | khi-deux | | |
| 0 | 705 642 | 99,50% | 705 606,21 | 705 607,10 | 0,00 | 0,00 | 705 608,00 | 0,00 | 705 609,17 | 0,00 | 705 642,00 | - | 705 642,00 | 0,00 | 705 609,17 | 0,00 | 705 622,74 | 0,00 | | |
| 1 | 3 490 | 0,49% | 3 561,60 | 3 559,80 | 1,44 | 1,37 | 3 558,01 | 1,30 | 3 555,69 | 1,21 | 3 543,88 | 0,82 | 3 490,00 | 0,00 | 3 555,69 | 1,21 | 3 528,73 | 0,42 | | |
| 2 | 43 | 0,01% | 7,20 | 8,09 | 178,07 | 150,57 | 8,99 | 128,76 | 10,15 | 106,38 | - 10,88 | - 266,84 | 43,00 | 0,00 | 10,15 | 106,38 | 23,53 | 16,11 | | |
| 709 175 | | | 100,00% | 709 175 | 709 175 | 179,51 | 151,94 | 709 175 | 130,06 | 709 175 | 107,59 | 709 175 | - 266,02 | 709 175 | 0,00 | 709 175 | 107,59 | 709 175 | 16,53 | |
| nb occurrence = 3 | | | P-value = 0% | | 0% | | P-value = 0% | | #NOMBRE! | | P-value = #NOMBRE! | | P-value = #NOMBRE! | | P-value = #NOMBRE! | | P-value = #NOMBRE! | | #NOMBRE! | |
| ^n = 0,0050 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^m2 = 0,0052 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^σ2 = 0,0051 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| cv = 14,22 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^m3 = 0,0054 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^m4 = 0,0059 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^μ3 = 0,0053 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^μ4 = 0,0058 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^γ1 = 14,47 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| ^γ2 = 216,04 | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |
| P(N=0) = 99,50% | | | K2 = 3,84 | | 3,84 | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = 3,84 | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | K2 = #NOMBRE! | | #NOMBRE! | |

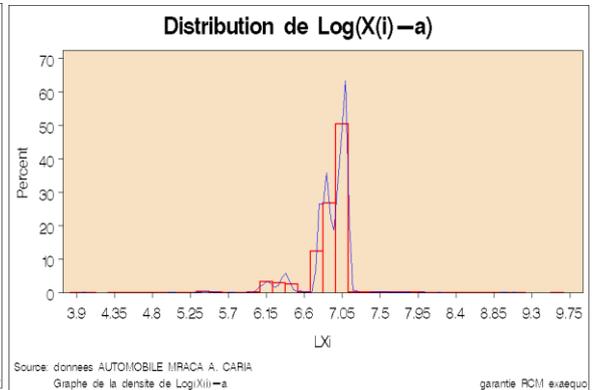
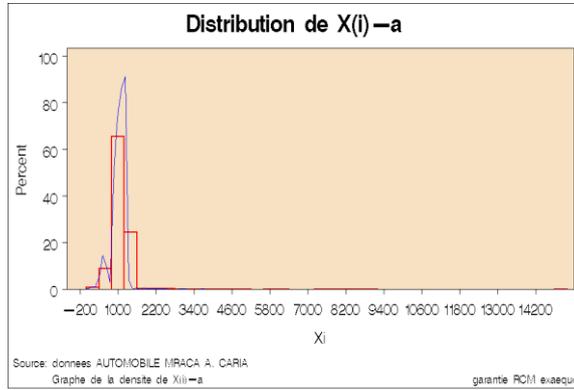
| | |
|-----------------|--------|
| nb occurrence = | 3 |
| ^n = | 0,0050 |
| ^m2 = | 0,0052 |
| ^σ2 = | 0,0051 |
| cv = | 14,22 |
| ^m3 = | 0,0054 |
| ^m4 = | 0,0059 |
| ^μ3 = | 0,0053 |
| ^μ4 = | 0,0058 |
| ^γ1 = | 14,47 |
| ^γ2 = | 216,04 |
| P(N=0) = | 99,50% |

Diagramme en bâtons sur données MRACA concernant la garantie VOL Automobile

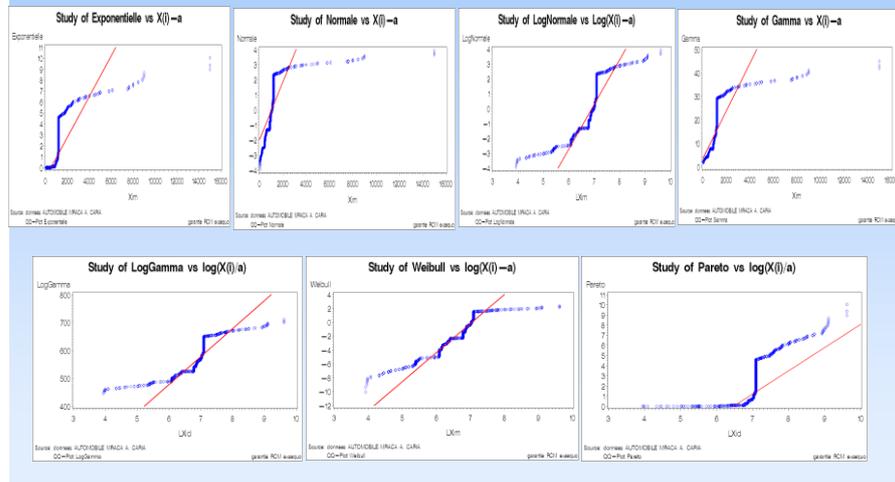


5. Analyse de la distribution du coût moyen d'un sinistre suivant la garantie :

a. X(i) Montant des sinistres RCM (Responsabilité Civile Matériel) exaequo et a la franchise égale à 0



Q-Q Plots RCM(Responsabilité Civile Matériel) exaequo

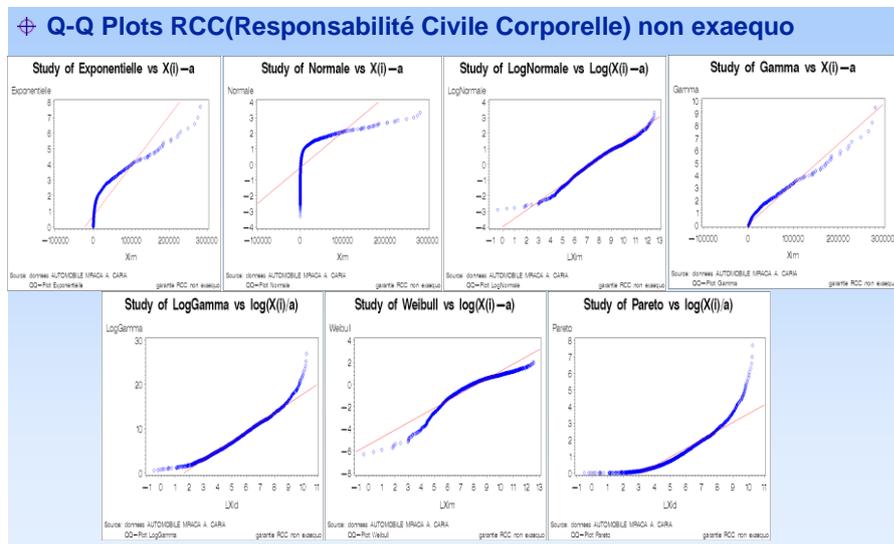
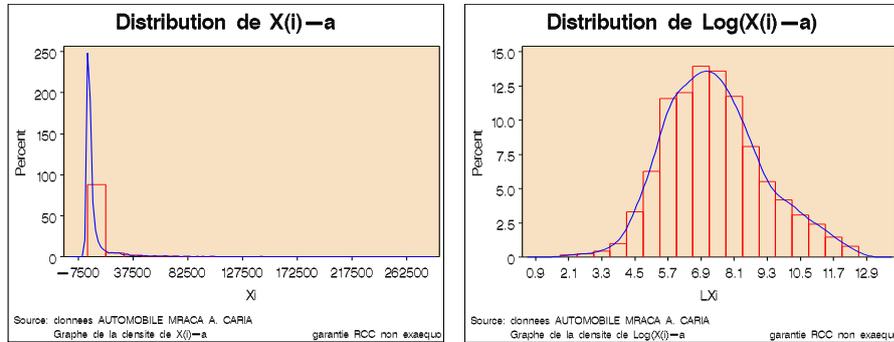


↓ Résultats statistiques RCM(Responsabilité Civile Matériel) exaequo

| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----------|--------------------|---|--------------------|--------------------|--------------------|--------------------|---------------------|
| Paramètre 1 | | | | | | | | |
| Estimation du paramètre 1 | | 0,0009 | 1 170,3840 | 7,0195 | 11,1475 | 205,7287 | 4,2523 | 0,1425 |
| Paramètre 2 | | | | | | | | |
| Estimation du paramètre 2 | | | 310,6574 | 0,4022 | 0,0095 | 29,3082 | 1751,6853 | |
| risque relatif paramètre1 | | 0,0076 | 0,0020 | 0,0004 | 0,0106 | 0,0108 | 0,0063 | 0,0076 |
| risque relatif paramètre2 | | | 3,3520 | 0,0108 | 0,0109 | 0,0108 | 0,0019 | |
| risque relatif moy | | 0,0076 | 0,0020 | 0,0032 | 0,0023 | 0,000003 | 0,0020 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 |
| moy | 1 171 | 1 171 | 1 171 | 1 213 | 1 265 | 1 171 | 1 594 | infini |
| var | 96 514 | 1 369 822 | 96 508 | 298 075 | 122 882 | 467 975 | 178 854 | infini |
| sigma | 311 | 1 170 | 311 | 508 | 351 | 684 | 423 | infini |
| Minimum | 6 | | | | | | | |
| Maximum | 10 000 | | | | | | | |
| moy_borne_inf | | 1 154 | 1 167 | 1 206 | 1 166 | 1 265 | 1 598 | - |
| moy_borne_sup | | 1 189 | 1 176 | 1 221 | 1 177 | 1 265 | 1 601 | - |
| CV | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0,07 |
| G1 | 302 | 2 | 0 | 1 | 1 | 0 | 0 | infini |
| G2 | 11 | 6 | 0 | 3 | 1 | 0 | 0 | 0,00 |
| Dp | 0,23 | 0,52 | 1,00 | 0,33 | 0,29 | 1,00 | 0,27 | 16 123,15 |
| G | | 0,50 | 2,00 | 0,47 | 0,17 | | 0,75 | -1,40 |
| Q25 | 1 204 | 337 | 981 | 853 | 919 | 799 | 1 307 | 9 |
| mediane | 1 217 | 811 | 1 170 | 1 118 | 1 138 | 1 106 | 1 607 | 130 |
| Q75 | 1 237 | 1 623 | 1 380 | 1 467 | 1 384 | 1 546 | 1 892 | 16 833 |
| Q90 | 1 269 | 2 695 | 1 569 | 1 872 | 1 636 | 2 108 | 2 131 | 10 459 237 |
| Q95 | 1 270 | 3 506 | 1 681 | 2 167 | 1 800 | 2 549 | 2 267 | 1 357 000 656 |
| Q99 | 1 270 | 5 390 | 1 893 | 2 850 | 2 136 | 3 670 | 2 509 | 109 395 640 000 000 |
| Dn | | 0,53 | 0,37 | 0,43 | 0,41 | 0,43 | 0,77 | 0,58 |
| W2n | | 1 318,57 | 773,16 | 890,42 | 846,14 | 918,03 | 2 791,40 | 1 636,54 |
| A2n | | 6 116,06 | 3 635,43 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,31 | 7 482,10 |
| P_KS | | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 69,29 | 48,39 | 55,89 | 54,17 | 56,59 | 100,57 | 76,42 |
| W2n_T | | 1 318,58 | 776,10 | 890,48 | 846,19 | 918,08 | 2 795,66 | 1 636,64 |
| A2n_T | | 6 116,27 | 3 635,59 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,55 | 7 482,10 |
| P_value_KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0,00000000 | 0 | 0,00000 | 0 | 0 |
| Mh2_theo | | 670 647 | 147 229 | 244 002 | 176 438 | 302 037 | 330 630 | 4 857 713 |
| la distribution vérifie ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | OUI | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 8,13873E-05 | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

b. X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) non exaequo et a la franchise égale à 0

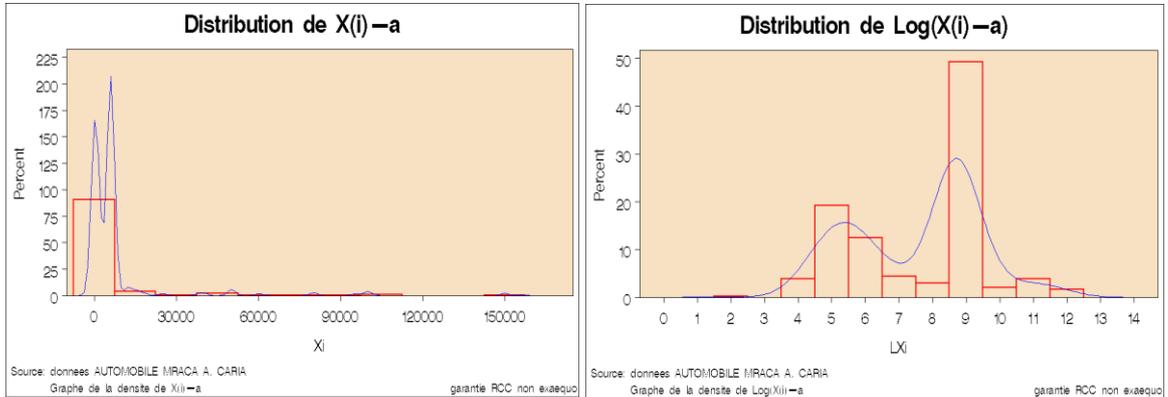


↓ Résultats statistiques RCC(Responsabilité Civile Corporelle) non exaequo

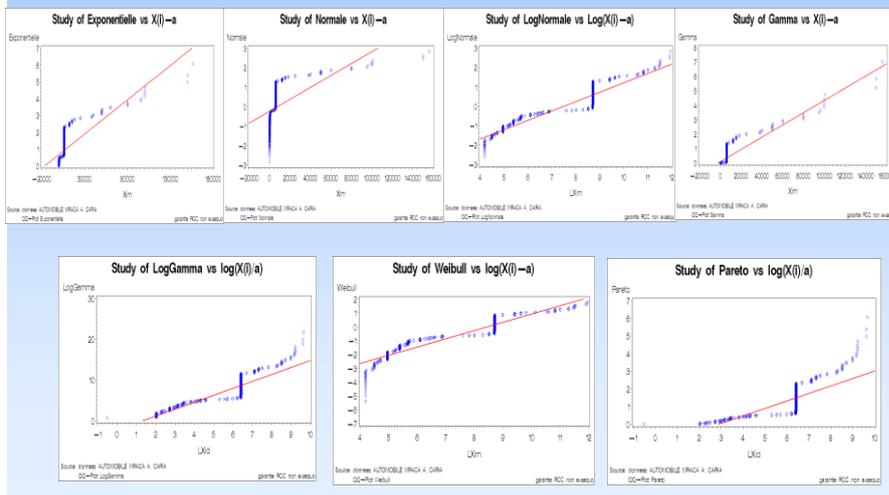
| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|---------------|---------------|---------------|---|-------------|-----------------------|-------------|---------------------|
| Paramètre 1 | | θ | μ | μ | v | v | τ | α |
| Estimation du paramètre 1 | | 0,0001 | 11 439 4727 | 7,3994 | 0,3463 | 15,9787 | 0,5008 | 0,1351 |
| Paramètre 2 | | σ | σ | σ | β | β | α | α |
| Estimation du paramètre 2 | | 37 973 7950 | 1 8421 | 1 8421 | 0,00003 | 2 1595 | 4271,0173 | 0,0218 |
| risque relatif paramètre1 | | 0,0218 | 0,0723 | 0,0054 | 0,0246 | 0,0305 | 0,0180 | 0,0218 |
| risque relatif paramètre2 | | 1 169,3936 | 0,0308 | 0,0308 | 0,0444 | 0,0310 | 0,0461 | 0,0461 |
| risque relatif moy | | 0,0218 | 0,0723 | 0,0659 | 0,0370 | 0,000002 | 0,0470 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | 0,6448 |
| Constante K2 utile pour Weibull | | | | | | | | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 | 2 109 |
| moy | 11 440 | 11 440 | 11 440 | 8 921 | 11 440 | 20 687 | 8 517 | infini |
| var | 1 442 693 169 | 130 661 535 | 1 442 009 104 | 2 288 968 905 | 377 911 303 | 1 210 000 000 000 000 | 363 908 388 | infini |
| sigma | 37 983 | 11 439 | 37 974 | 47 643 | 19 440 | 1 100 000 000 | 18 998 | infini |
| Minimum | 4 | | | | | | | |
| Maximum | 400 000 | | | | | | | |
| moy borne inf | | 10 952 | 9 620 | 7 770 | 10 611 | 20 687 | 7 732 | 0,85 |
| moy borne sup | | 11 929 | 13 061 | 10 073 | 12 270 | 20 687 | 9 302 | 0,84 |
| CV | 3 | 1 | 3 | 0 | 2 | 0 | 2 | 0,25 |
| G1 KURTOSIS | 42 | 2 | 0 | 170 | 3 | 1 | 0 | infini |
| G2 SKEWNESS | 6 | 6 | 0 | 840 448 | 17 | 0 | 0 | 0,00 |
| Dp | 0,89 | 0,52 | 1,00 | 0,84 | 0,73 | 1,00 | 0,78 | 29 715,62 |
| G | 406 | 0,50 | 2,00 | 1,24 | 0,71 | 0,14 | 1,00 | -1,37 |
| Q25 | 406 | 3 391 | -14 173 | 472 | 437 | 439 | 355 | 8 |
| mediane | 1 359 | 7 929 | 11 439 | 1 635 | 3 458 | 1 402 | 2 055 | 169 |
| Q75 | 5 039 | 15 658 | 37 062 | 5 664 | 13 622 | 5 167 | 8 199 | 26 502 |
| Q90 | 19 392 | 26 340 | 60 105 | 17 329 | 33 089 | 18 932 | 22 562 | 25 093 972 |
| Q95 | 59 095 | 34 270 | 73 951 | 33 640 | 49 923 | 43 647 | 36 190 | 4 234 913 435 |
| Q99 | 236 697 | 52 881 | 99 790 | 118 751 | 92 928 | 235 938 | 90 118 | 629 200 610 000 000 |
| Dn | | 0,43 | 0,38 | 0,04 | 0,19 | 0,02 | 0,11 | 0,44 |
| W2n | | 178,11 | 110,85 | 1,40 | 30,22 | 0,14 | 9,17 | 116,63 |
| A2n | | 1 153,62 | 530,87 | 10,21 | 154,93 | 1,65 | 80,42 | 557,43 |
| P-JKS | | 0,04 | 0,02 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| P-CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P-AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 19,85 | 17,53 | 2,03 | 8,85 | 0,92 | 5,20 | 20,05 |
| W2n_T | | 178,12 | 112,06 | 1,40 | 30,23 | 0,14 | 9,21 | 116,69 |
| A2n_T | | 1 153,95 | 531,06 | 10,21 | 154,93 | 1,65 | 80,68 | 557,43 |
| P_valeur KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,25 | P<=0,010 | P<=0,001 |
| P_valeur CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,25 | P<=0,010 | P<=0,001 |
| P_valeur AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | 0,10-P<=0,150 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 6,91E-13 | 0 | 0,00285 | 0 | 0 |
| khi2 theo | | 4 333 | 48 774 | 72 | 921 | 22 | 403 | 3 928 |
| la distribution vérifie ou pas, la règle de Pareto ? | OUI | NON | OUI | OUI | NON | OUI | NON | OUI |
| A | | | | non adéquation du modèle Normal aux données | | | | |
| | | | | Prob. charge. négative | | | | |

Graphiquement et statistiquement la loi log normale semble être la plus appropriée, vient ensuite la loi log gamma et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

c. X(i) Montant des sinistres RCC (Responsabilité Civile Corporelle) exaquo et a la franchise égale à 0



Q-Q Plots RCC(Responsabilité Civile Corporelle) exaquo

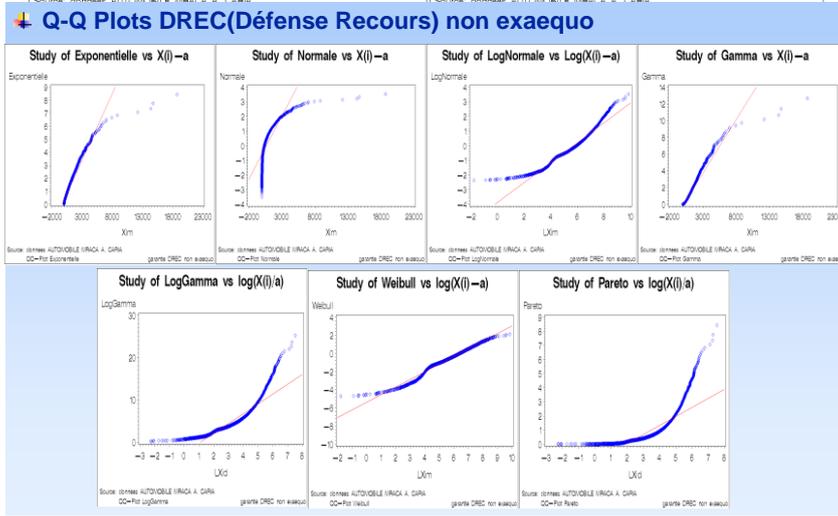
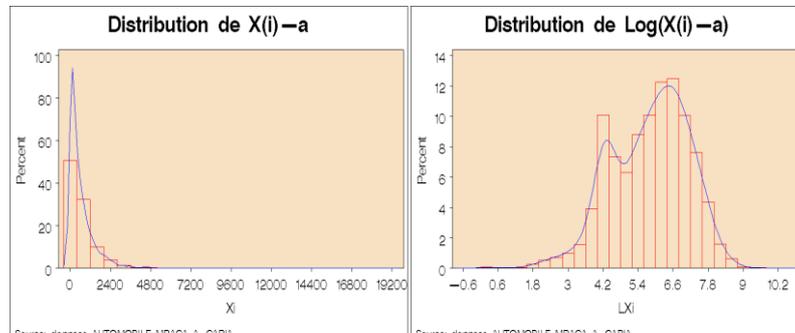


| Paramètre 1 | Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----|-----------|---------------|------------|------------|----------|----------|-----------|---------------------|
| Estimation du paramètre 1 | | | 0,0009 | 1 170,3940 | 7,0195 | 11,1475 | 205,7287 | 4,2523 | 0,1425 |
| Paramètre 2 | | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | | 310,6574 | 0,4022 | 0,0095 | 29,3082 | 1751,6853 | |
| risque relatif paramètre1 | | | 0,0076 | 0,0020 | 0,0004 | 0,0106 | 0,0108 | 0,0063 | 0,0076 |
| risque relatif paramètre2 | | | | 3,3520 | 0,0108 | 0,0109 | 0,0108 | 0,0019 | |
| risque relatif moy | | | 0,0076 | 0,0020 | 0,0032 | 0,0023 | 0,000003 | 0,0020 | infini |
| Constante K1 utile pour Weibull | | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | | 0,6449 | |
| a | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 | 17 179 |
| moy | | 1 171 | 1 171 | 1 171 | 1 213 | 1 171 | 1 265 | 1 594 | infini |
| var | | 96 514 | 1 369 822 | 96 508 | 258 075 | 122 882 | 467 975 | 178 854 | infini |
| sigma | | 311 | 1 170 | 311 | 508 | 351 | 684 | 423 | infini |
| Minimum | | 6 | | | | | | | |
| Maximum | | 10 000 | | | | | | | |
| moy_borne_inf | | | 1 154 | 1 167 | 1 206 | 1 166 | 1 265 | 1 588 | - |
| moy_borne_sup | | | 1 189 | 1 176 | 1 221 | 1 177 | 1 265 | 1 601 | |
| CV | | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0,07 |
| G1 | | 302 | 2 | 0 | 1 | 1 | 0 | 0 | infini |
| G2 | | 11 | 6 | 0 | 3 | 1 | 0 | 0 | 0,00 |
| Dp | | 0,23 | 0,52 | 1,00 | 0,33 | 0,29 | 1,00 | 0,27 | 16 123,15 |
| G | | | 0,50 | 2,00 | 0,47 | 0,17 | 0,75 | 0,75 | -1,40 |
| Q25 | | 1 204 | 337 | 961 | 853 | 919 | 799 | 1 307 | 8 |
| mediane | | 1 217 | 811 | 1 170 | 1 118 | 1 136 | 1 106 | 1 607 | 130 |
| Q75 | | 1 237 | 1 623 | 1 380 | 1 467 | 1 384 | 1 545 | 1 892 | 16 833 |
| Q90 | | 1 269 | 2 695 | 1 569 | 1 872 | 1 636 | 2 108 | 2 131 | 10 459 237 |
| Q95 | | 1 270 | 3 906 | 1 681 | 2 167 | 1 800 | 2 549 | 2 267 | 1 357 000 656 |
| Q99 | | 1 270 | 5 390 | 1 883 | 2 850 | 2 136 | 3 670 | 2 509 | 109 395 640 000 000 |
| Dn | | | 0,53 | 0,37 | 0,43 | 0,41 | 0,43 | 0,77 | 0,58 |
| W2n | | | 1 318,57 | 773,16 | 890,42 | 846,14 | 918,03 | 2 791,40 | 1 636,54 |
| A2n | | | 6 116,06 | 3 635,43 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,31 | 7 482,10 |
| P_KS | | | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | | 69,29 | 48,39 | 55,89 | 54,17 | 56,59 | 100,57 | 76,42 |
| W2n_T | | | 1 318,58 | 776,10 | 890,48 | 846,19 | 918,08 | 2 795,66 | 1 636,64 |
| A2n_T | | | 6 116,27 | 3 635,59 | 4 143,50 | 3 949,87 | 4 294,97 | 13 286,65 | 7 482,10 |
| P value KS | | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| kh2 theo | | | 670 647 | 147 229 | 244 002 | 176 438 | 302 037 | 330 630 | 4 857 713 |
| la distribution vérifiée ou pas, la règle de Pareto ? | | NON | NON | OUI | NON | NON | OUI | OUI | OUI |

non adéquation du modèle Normal aux données
Prob_charge_negative 8,13973E-05

Graphiquement et statistiquement la loi Weibull semble être la plus appropriée, vient ensuite la loi Normale et la loi gamma. On note toutefois des problèmes de sous-estimation ou de sur-estimation au niveau des queues de distribution.

d. X(i) Montant des sinistres DREC (Défense Recours) non exaequo et a la franchise égale à 0

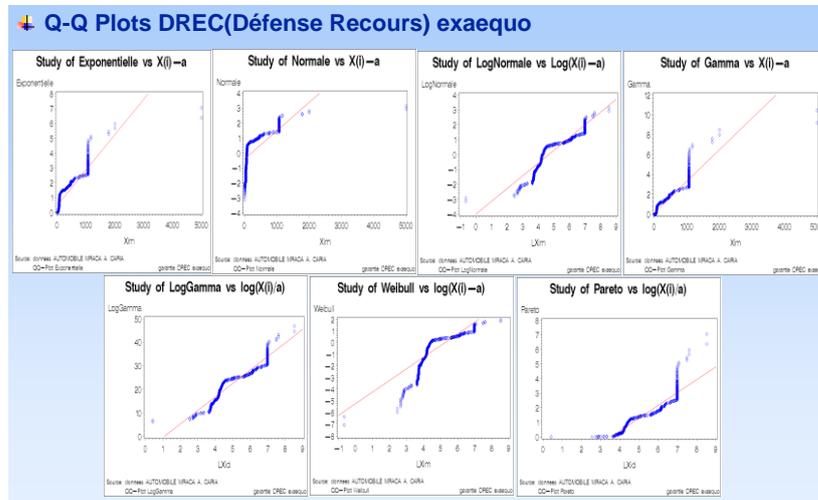
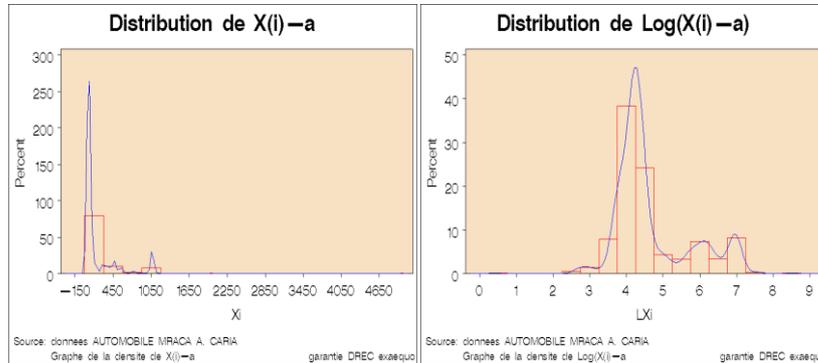


↓ Résultats statistiques DREC(Défense Recours) non exaequo

| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|--|-----------|---------------|---|-------------|----------|------------|-------------|-----------------|
| Paramètre 1 | | | | | | | | |
| Estimation du paramètre 1 | | 0,0015 | 674,3866 | 5,7105 | 0,7457 | 13,7041 | 0,8037 | 0,1751 |
| Paramètre 2 | | | | | | | | |
| Estimation du paramètre 2 | | | 1 101,4028 | 1,3667 | 0,0011 | 2,3988 | 687,8711 | |
| risque_relatif_parametre1 | | 0,0142 | 0,0232 | 0,0034 | 0,0172 | 0,0198 | 0,0117 | 0,0142 |
| risque_relatif_parametre2 | | | 22,1390 | 0,0201 | 0,0238 | 0,0202 | 0,0197 | |
| risque_relatif_moy | | 0,0142 | 0,0232 | 0,0275 | 0,0164 | 0,0000 | 0,0178 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,422784336 | |
| Constante K2 utile pour Weibull | | | | | | | 0,644934067 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 |
| moy | 675 | 675 | 675 | 790 | 675 | 1 616 | 665 | infini |
| var | 1 213 333 | 454 811 | 1 213 088 | 3 623 866 | 609 881 | 46 400 000 | 693 387 | infini |
| sigma | 1 102 | 674 | 1 101 | 1 904 | 781 | 215 407 | 833 | infini |
| Minimum | 0 | | | | | | | |
| Maximum | 34 587 | | | | | | | |
| moy_borne_inf | | 657 | 645 | 747 | 654 | 1 616 | 642 | 1 |
| moy_borne_sup | | 684 | 706 | 833 | 697 | 1 616 | 688 | 1 |
| CV | 2 | 1 | 2 | 0 | 1 | 0 | 1 | 0 |
| G1 KURTOSIS | 239 | 2 | 0 | 21 | 2 | 0 | 0 | infini |
| G2 SKEWNESS | 10 | 6 | 0 | 2 936 | 8 | 0 | 0 | 0 |
| Bp | 0,62 | 0,52 | 1,00 | 0,71 | 0,57 | 1,00 | 0,60 | 1,961,13 |
| G | | 0,50 | 2,00 | 0,98 | 0,56 | 0,15 | 0,98 | -1,54 |
| Q25 | 102 | 194 | -68 | 119 | 137 | 100 | 125 | 5 |
| mediane | 332 | 467 | 674 | 302 | 407 | 263 | 373 | 52 |
| Q75 | 874 | 936 | 1 417 | 769 | 930 | 792 | 893 | 2 742 |
| Q90 | 1 652 | 1 563 | 2 086 | 1 784 | 1 868 | 2 333 | 1 660 | 513 467 |
| Q95 | 2 325 | 2 020 | 2 486 | 2 951 | 2 244 | 4 731 | 2 302 | 26 887 349 |
| Q99 | 4 592 | 3 106 | 3 237 | 7 587 | 3 612 | 19 866 | 3 931 | 263 648 247 387 |
| Dn | 0,14 | 0,27 | 0,05 | 0,07 | 0,06 | 0,06 | 0,06 | 0,43 |
| W2n | 29,60 | 96,45 | 3,07 | 8,08 | 7,45 | 4,15 | 273,95 | |
| A2n | 169,54 | 610,25 | 19,91 | 47,01 | 63,99 | 29,71 | 1 311,79 | |
| P_KS | 0,02 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_CVM | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | 9,62 | 19,01 | 3,38 | 4,90 | 4,52 | 4,40 | 30,31 | |
| W2n_T | 29,60 | 96,13 | 3,07 | 8,08 | 7,46 | 4,16 | 274,01 | |
| A2n_T | 169,56 | 610,33 | 19,91 | 47,01 | 63,99 | 29,79 | 1 311,79 | |
| P_value_KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_value_CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,001 |
| P_value_AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| lnli2_theo | | 1 011 | 11 736 | 308 | 676 | 428 | 559 | 10 127 |
| la distribution vérifie ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 0,269867134 | | | | |

Graphiquement et statistiquement la loi Gamma semble être la plus appropriée. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

e. X(i) Montant des sinistres DREC (Défense Recours) exaequo et a la franchise égale à 0

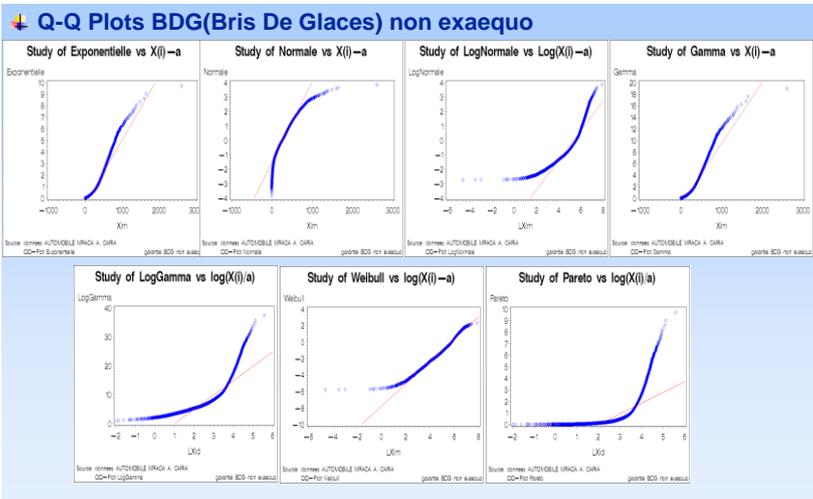
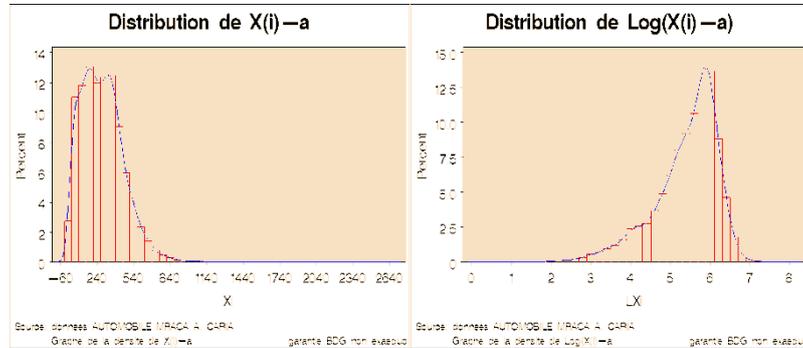


| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|--|--|--|--|--|--|--|--|--|
| Paramètre 1 | | θ | μ | μ | β | β | α | α |
| Estimation du paramètre 1 | | 0,0015 | 674,3986 | 571,05 | 0,7457 | 13,7041 | 0,037 | 0,1751 |
| Paramètre 2 | | σ | σ | σ | β | β | β | α |
| Estimation du paramètre 2 | | | 1 101,4028 | 1,3857 | 0,00111 | 2,3998 | 587,8711 | |
| risque relatif paramètre1 | | 0,0142 | 0,0232 | 0,0034 | 0,0172 | 0,0199 | 0,0117 | 0,0142 |
| risque relatif paramètre2 | | | 22,1390 | 0,0201 | 0,0238 | 0,0202 | 0,0187 | |
| risque relatif moy | | 0,0142 | 0,0232 | 0,0275 | 0,0164 | 0,0000 | 0,0178 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0,6449 | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 | 4 950 |
| moy | 675 | 675 | 675 | 790 | 675 | 1 616 | 665 | infini |
| var | 1 213 333 | 454 811 | 1 213 088 | 3 623 856 | 609 881 | 46 400 000 000 | 693 387 | infini |
| sigma | 1 102 | 674 | 1 101 | 1 904 | 781 | 215 407 | 833 | infini |
| Minimum | 0 | | | | | | | |
| Maximum | 34 587 | | | | | | | |
| moy borne inf | | 657 | 645 | 747 | 654 | 1 616 | 642 | 1 |
| moy borne sup | | 694 | 706 | 833 | 697 | 1 616 | 688 | 1 |
| CV | 2 | 1 | 2 | 0 | 1 | 0 | 1 | 0 |
| G1 KURTOSIS | 239 | 2 | 0 | 21 | 2 | 1 | 0 | infini |
| G2 SKEWNESS | 10 | 6 | 0 | 2 936 | 8 | 0 | 0 | 0 |
| Dp | 0,62 | 0,52 | 1,00 | 0,71 | 0,57 | 1,00 | 0,60 | 1 961,13 |
| G | | 0,50 | 2,00 | 0,98 | 0,56 | 0,15 | 0,98 | -1,54 |
| Q25 | 102 | 194 | -68 | 119 | 137 | 100 | 125 | 5 |
| mediane | 332 | 467 | 674 | 302 | 407 | 263 | 373 | 52 |
| Q75 | 874 | 935 | 1 417 | 769 | 930 | 782 | 883 | 2 742 |
| Q90 | 1 852 | 1 953 | 2 986 | 1 784 | 1 868 | 2 533 | 1 660 | 513 467 |
| Q95 | 2 325 | 2 020 | 2 486 | 2 951 | 2 244 | 4 731 | 2 302 | 26 887 349 |
| Q99 | 4 592 | 3 106 | 3 237 | 7 587 | 3 612 | 19 866 | 3 931 | 263 648 247 387 |
| Dn | | 0,14 | 0,27 | 0,05 | 0,07 | 0,06 | 0,06 | 0,43 |
| W2n | | 29,60 | 95,45 | 3,07 | 8,08 | 7,45 | 4,15 | 273,95 |
| A2n | | 159,54 | 510,25 | 19,91 | 47,01 | 53,99 | 29,71 | 1 311,79 |
| P_KS | | 0,02 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 9,62 | 19,01 | 3,38 | 4,90 | 4,52 | 4,40 | 30,31 |
| W2n_T | | 29,60 | 96,13 | 3,07 | 8,08 | 7,46 | 4,16 | 274,01 |
| A2n_T | | 159,56 | 510,33 | 19,91 | 47,01 | 53,99 | 29,79 | 1 311,79 |
| P value KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| kh2_theo | | 1 011 | 11 736 | 308 | 676 | 428 | 559 | 10 127 |
| la distribution vérifie ou pas, la règle de Pareto ? | la distribution ne vérifie pas, la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution vérifie la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution vérifie la règle de Pareto | la distribution ne vérifie pas, la règle de Pareto | la distribution vérifie la règle de Pareto |

non adéquation du modèle Normal aux données
Prob_charge_negative 0,27

Graphiquement et statistiquement la loi normale semble être la plus appropriée, vient ensuite la loi Weibull et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

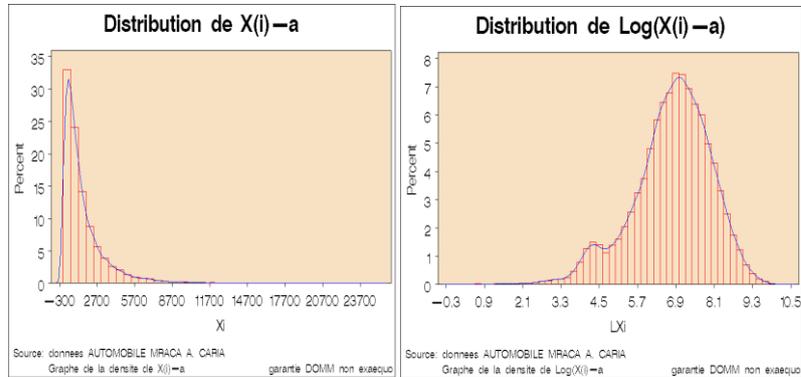
f. X(i) Montant des sinistres BDG (Bris De Glaces) non exæquo et a la franchise égale à 0



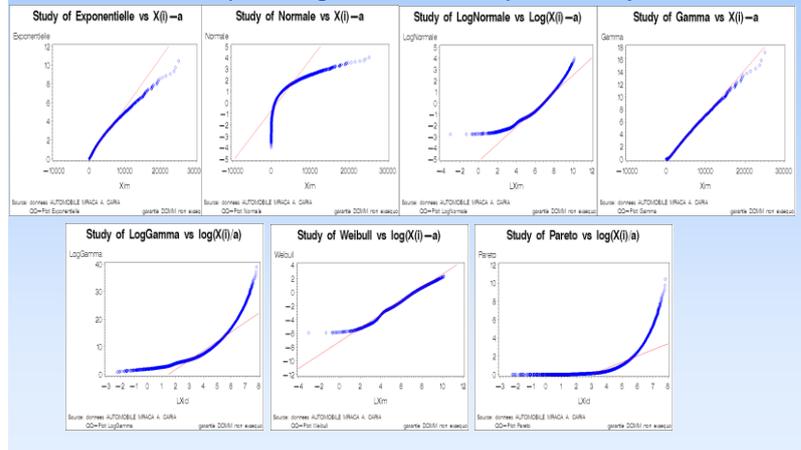
| ❑ Résultats statistiques BDG(Bris De Glaces) non exæquo | | | | | | | | |
|---|-----------|---------------|---|------------|----------|----------|----------|----------------|
| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
| Paramètre 1 | | θ | μ | μ | v | v | τ | α |
| Estimation du paramètre 1 | | 0,0036 | 274,454 | 5,4179 | 2,6840 | 56,1327 | 1,7845 | 0,1846 |
| Paramètre 2 | | β | σ | σ | β | β | α | α |
| Estimation du paramètre 2 | | 0,0049 | 160,6321 | 0,6867 | 0,00977 | 10,3607 | 309,1898 | 0,0049 |
| risque_relatif_paramètre1 | | 0,0049 | 1,1059 | 0,0069 | 0,0071 | 0,0069 | 0,0028 | 0,0028 |
| risque_relatif_paramètre2 | | 0,0049 | 0,0028 | 0,0037 | 0,0030 | 0,000012 | 0,0028 | infini |
| risque_relatif_moy | | | | | | | 0,4228 | |
| Constante K1 utile pour Weibull | | | | | | | 0,6449 | |
| Constante K2 utile pour Weibull | | | | | | | | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| moy | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 | 42 198 |
| var | 276 | 276 | 276 | 286 | 276 | 299 | 276 | 276 |
| sigma | 26 803 | 75 485 | 25 803 | 48 798 | 28 124 | 80 399 | 25 402 | infini |
| Minimum | 161 | 275 | 161 | 221 | 168 | 264 | 159 | infini |
| Maximum | 2 | | | | | | | |
| moy_borne_inf | 2 982 | | | | | | | |
| moy_borne_sup | | 273 | 274 | 284 | 274 | 299 | 275 | 0,77 |
| CV | | 278 | 277 | 286 | 277 | 299 | 270 | 0,77 |
| G1 KURTOSIS | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,13 |
| G2 SKEWNESS | 5 | 2 | 0 | 3 | 1 | 0 | 0 | infini |
| Dp | 1 | 6 | 0 | 16 | 2 | 0 | 0 | 0,00 |
| G | 0,37 | 0,52 | 1,00 | 0,44 | 0,39 | 1,00 | 0,38 | 1 224,49 |
| Q25 | 0,50 | 2,00 | 2,00 | 0,61 | 0,33 | 0,08 | 0,88 | -1,59 |
| mediane | 140 | 79 | 166 | 142 | 151 | 136 | 154 | 5 |
| Q75 | 270 | 150 | 275 | 225 | 241 | 218 | 252 | 43 |
| Q90 | 365 | 381 | 383 | 365 | 362 | 360 | 371 | 1 828 |
| Q95 | 470 | 633 | 481 | 543 | 500 | 580 | 493 | 261 737 |
| Q99 | 554 | 823 | 539 | 696 | 596 | 780 | 572 | 11 189 362 |
| Dn | 755 | 1 265 | 648 | 1 111 | 805 | 1 395 | 728 | 68 506 229 264 |
| W2n | | 0,19 | 0,06 | 0,11 | 0,07 | 0,12 | 0,05 | 0,50 |
| A2n | | 593,52 | 44,85 | 126,22 | 56,33 | 189,06 | 26,99 | 3 203,81 |
| P | | 3 284,00 | 380,35 | 663,42 | 285,29 | 891,54 | 159,73 | 14 923,50 |
| P_KS | | 0,01 | 0,00 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 0,75 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 38,46 | 11,91 | 22,70 | 15,06 | 25,62 | 9,65 | 102,60 |
| W2n_T | | 593,52 | 44,96 | 126,22 | 56,33 | 189,07 | 27,02 | 3 203,88 |
| A2n_T | | 3 284,05 | 380,35 | 663,42 | 285,29 | 891,54 | 159,89 | 14 923,50 |
| P value KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| kh2 theo | | 16 989 | 4 374 | 6 440 | 3 517 | 8 350 | 2 617 | 241 512 |
| la distribution vérifie ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | | | | | |

Graphiquement et statistiquement la loi gamma semble être la plus appropriée, vient ensuite la loi weibull. On note toutefois des problèmes de sous-estimation log au niveau des queues de distribution.

g. X(i) Montant des sinistres DOMM(Dommage tous accidents) non exaequo et à la franchise égale à 0



Q-Q Plots DOMM(Dommage tous accidents) non exaequo

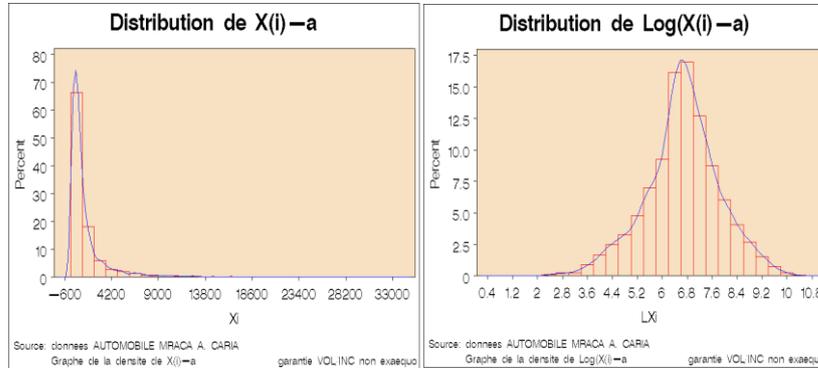


Résultats statistiques DOMM(Dommage tous accidents) non exaequo

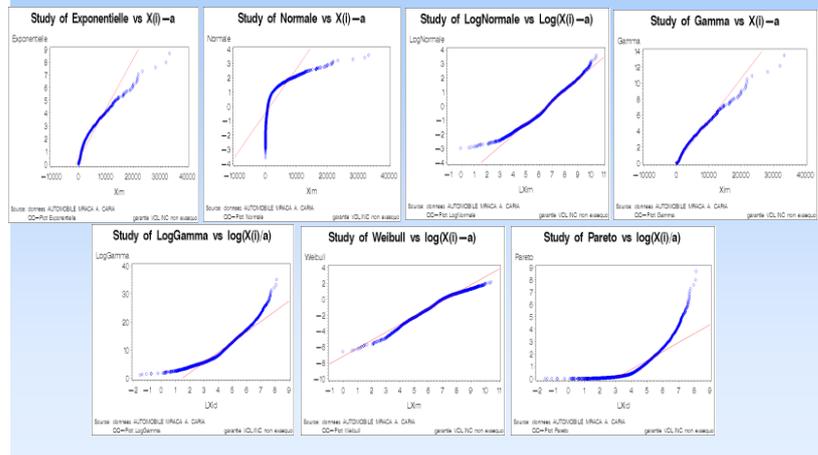
| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----------|--------------------|---|--------------------|--------------------|--------------------|--------------------|--------------------|
| Paramètre 1 | | 0 | μ | μ | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0006 | 1 646,8184 | 6 7826 | 0,9328 | 23 7268 | 0 9304 | 0 1497 |
| Paramètre 2 | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | 1 986,8324 | 1 2983 | 0 00057 | 3 4382 | 1638 6216 | |
| risque relatif paramètre1 | | 0 0060 | 0 0072 | 0 0011 | 0 0074 | 0 0084 | 0 0060 | 0 0069 |
| risque relatif paramètre2 | | | 16 8504 | 0 0085 | 0 0097 | 0 0085 | 0 0068 | |
| risque relatif moy | | 0 0060 | 0 0072 | 0 0101 | 0 0062 | 0 000003 | 0 0064 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0 4228 | |
| Constante K2 utile pour Weibull | | | | | | | 0 6449 | |
| λ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 | 29 115 |
| moy | 1 646 | 1 646 | 1 646 | 1 646 | 1 646 | 1 646 | 1 646 | infini |
| var | 3 946 950 | 2 712 340 | 3 946 708 | 14 885 257 | 2 907 792 | 538 080 000 | 3 121 862 | infini |
| sigma | 1 987 | 1 647 | 1 987 | 3 632 | 1 705 | 23 195 | 1 767 | infini |
| Minimum | | | | | | | | |
| Maximum | 32 000 | | | | | | | |
| moy_borne_inf | | 1 629 | 1 625 | 1 910 | 1 629 | 2 948 | 1 623 | 0 83 |
| moy_borne_sup | | 1 667 | 1 671 | 1 967 | 1 668 | 2 948 | 1 664 | 0 82 |
| CV | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 22 |
| G1 KURTOSIS | 17 | 2 | 0 | 14 | 2 | 0 | 0 | infini |
| G2 SKEWNESS | 3 | 6 | 0 | 860 | 6 | 0 | 0 | 0 00 |
| dp | 0 57 | 0 52 | 1 00 | 0 66 | 0 53 | 1 00 | 0 55 | 9 356 14 |
| g | 0 50 | 0 50 | 2 10 | 0 91 | 0 51 | 0 12 | 0 36 | -1 43 |
| Q25 | 461 | 474 | 307 | 378 | 440 | 330 | 416 | 7 |
| mediane | 1 013 | 1 142 | 1 647 | 882 | 1 109 | 802 | 1 071 | 103 |
| Q75 | 2 059 | 2 283 | 2 987 | 2 062 | 2 283 | 2 129 | 2 257 | 10 526 |
| Q90 | 3 828 | 3 792 | 4 193 | 4 426 | 3 899 | 5 525 | 4 796 950 | 4 796 950 |
| Q95 | 6 296 | 4 934 | 4 915 | 6 991 | 5 057 | 10 134 | 5 166 | 452 180 459 |
| Q99 | 9 883 | 7 584 | 6 269 | 16 481 | 7 856 | 34 026 | 8 201 | 23 009 772 000 000 |
| Dn | | 0 05 | 0 20 | 0 06 | 0 04 | 0 10 | 0 03 | 0 43 |
| W2n | | 21 12 | 389 89 | 33 31 | 14 75 | 86 54 | 10 80 | 1 822 38 |
| A2n | | 110 85 | 2 150 46 | 193 35 | 62 93 | 502 30 | 67 65 | 8 655 87 |
| P_KS | | 0 01 | 0 01 | 0 01 | 0 01 | 0 01 | 0 01 | 0 01 |
| P_CVM | | 0 46 | 0 13 | 0 46 | 0 46 | 0 46 | 0 46 | 0 46 |
| P_AD | | 2 49 | 0 75 | 2 49 | 2 49 | 2 49 | 2 49 | 2 49 |
| Dn_T | | 8 02 | 33 99 | 10 31 | 7 02 | 16 33 | 5 72 | 72 89 |
| W2n_T | | 21 12 | 391 06 | 33 31 | 14 75 | 86 54 | 10 81 | 1 822 44 |
| A2n_T | | 110 86 | 2 150 52 | 193 35 | 62 93 | 502 30 | 67 73 | 8 655 87 |
| P value KS | | P<=0 010 | P<=0 010 | P<=0 001 | P<=0 001 | P<=0 001 | P<=0 010 | P<=0 001 |
| P value CVM | | P<=0 005 | P<=0 005 | P<=0 001 | P<=0 001 | P<=0 001 | P<=0 010 | P<=0 001 |
| P value AD | | P<=0 005 | P<=0 005 | P<=0 001 | P<=0 001 | P<=0 001 | P<=0 010 | P<=0 001 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| kh2_theo | | 390 | 22 523 | 1 242 | 394 | 2 418 | 407 | 36 196 |
| la distribution vérifiée ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |
| A | | | non adéquation du modèle Normal aux données | | | | | |
| | | | Prob_charge_negative | 0 203409796 | | | | |

Graphiquement et statistiquement la loi weibull semble être la plus appropriée, vient ensuite la loi gamma et la loi exponentiel. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

h. X(i) Montant des sinistres VOL/INC(Incendie) non exæquo et a la franchise égale à 0



Q-Q Plots VOL / INC(Incendie) non exæquo



| Loi | Empirique | Exponentielle | Normale | LogNormale | Gamma | LogGamma | Weibull | Pareto |
|---|-----------|---------------|----------|------------|----------|----------|----------|-----------------|
| Paramètre 1 | | θ | μ | μ | ν | ν | τ | α |
| Estimation du paramètre 1 | | 0,0028 | 355,4086 | 5,7051 | 3,1344 | 70,9172 | 2,0277 | 0,1753 |
| Paramètre 2 | | | σ | σ | β | β | α | |
| Estimation du paramètre 2 | | | 183,4918 | 0,6435 | 0,00882 | 12,4305 | 401,1241 | |
| risque relatif paramètre1 | | 0,0136 | 0,0070 | 0,0015 | 0,0184 | 0,0192 | 0,0113 | 0,0136 |
| risque relatif paramètre2 | | | 3,5395 | 0,0193 | 0,0199 | 0,0193 | 0,0071 | |
| risque relatif moy | | 0,0136 | 0,0070 | 0,0096 | 0,0077 | 0,0000 | 0,4223 | infini |
| Constante K1 utile pour Weibull | | | | | | | 0,6449 | |
| Constante K2 utile pour Weibull | | | | | | | | |
| a | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| n | 5,375 | 5,375 | 5,375 | 5,375 | 5,375 | 5,375 | 5,375 | 5,375 |
| moy | 357 | 356 | 356 | 371 | 356 | 384 | 356 | infini |
| var | 33 675 | 126 315 | 33 669 | 70 050 | 40 299 | 106 188 | 33 669 | infini |
| sigma | 184 | 355 | 183 | 265 | 201 | 326 | 183 | infini |
| Minimum | 1 | | | | | | | |
| Maximum | 1 609 | | | | | | | |
| moy_borne_inf | | 347 | 352 | 364 | 351 | 384 | 351 | 1 |
| moy_borne_sup | | 366 | 361 | 377 | 362 | 384 | 361 | 1 |
| CV | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| G1 KURTOSIS | 1 | 2 | 0 | 3 | 1 | 0 | 0 | infini |
| G2 SKEWNESS | 1 | 6 | 0 | 13 | 2 | 0 | 0 | 0 |
| Dp | 0,35 | 0,52 | 1,00 | 0,42 | 0,38 | 1,00 | 0,36 | 1 944,18 |
| G | | 0,50 | 2,00 | 0,59 | 0,31 | 0,07 | 0,86 | -1,54 |
| Q25 | 229 | 102 | 232 | 195 | 208 | 188 | 217 | 5 |
| mediane | 346 | 246 | 355 | 300 | 318 | 292 | 335 | 52 |
| Q75 | 460 | 493 | 479 | 464 | 463 | 467 | 471 | 2 722 |
| Q90 | 597 | 818 | 591 | 685 | 625 | 727 | 605 | 507 128 |
| Q95 | 676 | 1 065 | 657 | 866 | 737 | 957 | 689 | 26 456 297 |
| Q99 | 856 | 1 637 | 782 | 1 342 | 979 | 1 633 | 852 | 257 178 763 772 |
| Dn | | 0,24 | 0,05 | 0,14 | 0,10 | 0,16 | 0,07 | 0,50 |
| W2n | | 105,48 | 3,60 | 25,89 | 11,81 | 33,63 | 4,74 | 424,36 |
| A2n | | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | -5 326,37 |
| P_KS | | 0,02 | 0,01 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 |
| P_CVM | | 0,46 | 0,13 | 0,46 | 0,46 | 0,46 | 0,46 | 0,46 |
| P_AD | | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 | 2,49 |
| Dn_T | | 17,51 | 3,57 | 10,55 | 7,56 | 11,83 | 5,11 | 37,08 |
| W2n_T | | 105,48 | 3,62 | 25,90 | 11,81 | 33,64 | 4,75 | 424,44 |
| A2n_T | | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | -5 326,37 |
| P value KS | | P<=0,010 | P<=0,010 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value CVM | | P<=0,005 | P<=0,005 | P<=0,001 | P<=0,001 | P<=0,001 | P<=0,010 | P<=0,001 |
| P value AD | | P>=0,25 | P>=0,25 | P>=0,25 | P>=0,25 | P>=0,25 | P>=0,25 | P>=0,25 |
| P_chi2 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| khi2 theo | | 3 032 | 414 | 1 173 | 673 | 1 433 | 435 | 39 236 |
| la distribution vérifiée ou pas, la règle de Pareto ? | NON | NON | OUI | NON | NON | OUI | NON | OUI |

non adéquation du modèle Normal aux données
Prob_charge_negative 0,03

Graphiquement et statistiquement la loi normale semble être la plus appropriée, vient ensuite la loi weibull et la loi gamma. On note toutefois des problèmes de sous-estimation au niveau des queues de distribution.

Sortie sas des différentes proc genmod par garantie :

- **nn : Responsabilité Civile Matérielle Hors forfait IDA**
 - **Modélisation de la fréquence de sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | grienn |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_TMEC_5a9 | 2 | 0 1 |
| T_Fract_Annuel | 2 | 0 1 |
| qualite_baleII | 2 | 0 1 |
| T_Classe_SRA_N_a_V | 2 | 0 1 |
| T_ville_Rurale | 2 | 0 1 |
| T_Classe_rep_SRA_I | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 75E3 | 4642.2247 | 0.0619 |
| Scaled Deviance | 75E3 | 4642.2247 | 0.0619 |
| Pearson Chi-Square | 75E3 | 125216.8308 | 1.6691 |
| Scaled Pearson X2 | 75E3 | 125216.8308 | 1.6691 |
| Log Likelihood | | -3077.9079 | |
| Full Log Likelihood | | -3080.6804 | |
| AIC (smaller is better) | | 6177.3609 | |
| AICC (smaller is better) | | 6177.3628 | |
| BIC (smaller is better) | | 6251.1658 | |

Algorithm converged.

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_TMEC_5a9 | 1 | . | . |
| T_Fract_Annuel | 1 | . | . |
| qualite_baleII | 1 | 20.27 | <.0001 |
| T_Classe_SRA_N_a_V | 1 | 13.20 | 0.0003 |
| T_ville_Rurale | 1 | 11.97 | 0.0005 |
| T_Classe_rep_SRA_I | 1 | . | . |

The GENMOD Procedure

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--|---------------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | 1 | -4.2028 | 0.2257 | -4.6448 -3.7603 | 346.84 | <.0001 |
| T_TMEC_5a9 | 0 1 | -0.3544 | 0.0915 | -0.5337 -0.1752 | 15.01 | 0.0001 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Fract_Annuel | 0 1 | 0.4690 | 0.1078 | 0.2577 0.6803 | 18.92 | <.0001 |
| T_Fract_Annuel | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| qualite_baleII | 0 1 | 0.5181 | 0.1095 | 0.3035 0.7326 | 22.40 | <.0001 |
| qualite_baleII | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_SRA_N_a_V | 0 1 | -0.5860 | 0.1499 | -0.8798 -0.2922 | 15.28 | <.0001 |
| T_Classe_SRA_N_a_V | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_ville_Rurale | 0 1 | 0.3041 | 0.0994 | 0.1092 0.4989 | 9.36 | 0.0022 |
| T_ville_Rurale | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| T_Classe_rep_SRA_I | 0 1 | -0.3902 | 0.1305 | -0.6460 -0.1343 | 8.94 | 0.0028 |
| T_Classe_rep_SRA_I | 1 | 0 | 0.0000 | 0.0000 0.0000 | . | . |
| Dispersion | 1 | 0.7803 | 0.7080 | 0.1318 4.6190 | . | . |

Five binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -6248.3253 | | | |
| T_TMEC_5a9 | -6236.7660 | 1 | 11.56 | 0.0007 |
| T_Fract_Annuel | -6208.7031 | 1 | 28.06 | <.0001 |
| qualite_baleII | -6187.2758 | 1 | 21.43 | <.0001 |
| T_Classe_SRA_N_a_V | -6173.5568 | 1 | 13.72 | 0.0002 |
| T_ville_Rurale | -6165.9964 | 1 | 7.56 | 0.0060 |
| T_Classe_rep_SRA_I | -6155.8157 | 1 | 10.18 | 0.0014 |

- Modélisation du coût moyen d'un sinistre

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Chann |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| Qualite_PNB | 2 | 0 1 |
| T_AGE_inf22ans | 2 | 0 1 |
| T_Classe_SRA_A_a_H | 2 | 0 1 |
| T_Classe_rep_SRA_O_a | 2 | 0 1 |
| T_Energie_GO | 2 | 0 1 |
| T_Garage_ferme | 2 | 0 1 |
| T_Groupe_SRA_00_a_27 | 2 | 0 1 |
| T_Groupe_SRA_29_a_32 | 2 | 0 1 |
| T_Groupe_SRA_33_a_43 | 2 | 0 1 |
| T_car_Utilitaire | 2 | 0 1 |
| temoin_pel | 2 | 0 1 |
| temoin_secuicompte | 2 | 0 1 |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | |
|--|-----|----------------|-------------|--|---------------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | 1 | 9.1925 | 0.4333 | 8.3432 10.0418 | 450.04 | <.0001 |
| Qualite_PNB | 0 1 | 0.3390 | 0.0919 | 0.1590 0.5190 | 13.62 | 0.0002 |
| Qualite_PNB | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_AGE_inf22ans | 0 1 | -0.5960 | 0.1898 | -0.9680 -0.2239 | 9.86 | 0.0017 |
| T_AGE_inf22ans | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Classe_SRA_A_a_H | 0 1 | -0.3271 | 0.1118 | -0.5463 -0.1079 | 8.56 | 0.0034 |
| T_Classe_SRA_A_a_H | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Classe_rep_SRA_O_a | 0 1 | 0.5332 | 0.2129 | 0.1159 0.9505 | 6.27 | 0.0123 |
| T_Classe_rep_SRA_O_a | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Energie_GO | 0 1 | -0.4155 | 0.0991 | -0.6096 -0.2213 | 17.58 | <.0001 |
| T_Energie_GO | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Garage_ferme | 0 1 | 0.2789 | 0.0923 | 0.0979 0.4598 | 9.13 | 0.0025 |
| T_Garage_ferme | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Groupe_SRA_00_a_27 | 0 1 | -0.5668 | 0.1479 | -0.8566 -0.2769 | 14.69 | 0.0001 |
| T_Groupe_SRA_00_a_27 | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Groupe_SRA_29_a_32 | 0 1 | -0.5456 | 0.1224 | -0.7656 -0.3057 | 19.86 | <.0001 |
| T_Groupe_SRA_29_a_32 | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_Groupe_SRA_33_a_43 | 0 1 | -0.9235 | 0.2067 | -1.3387 -0.5184 | 19.96 | <.0001 |
| T_Groupe_SRA_33_a_43 | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| T_car_Utilitaire | 0 1 | 0.3661 | 0.1649 | 0.0428 0.6893 | 4.93 | 0.0265 |
| T_car_Utilitaire | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| temoin_pel | 0 1 | -0.2948 | 0.0890 | -0.4693 -0.1203 | 10.97 | 0.0009 |
| temoin_pel | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| temoin_secuicompte | 0 1 | 0.2206 | 0.0854 | 0.0532 0.3881 | 6.67 | 0.0098 |
| temoin_secuicompte | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | | |
| Scale | 1 | 1.1111 | 0.0592 | 1.0010 1.2332 | | |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 544 | 571.8297 | 1.0512 |
| Scaled Deviance | 544 | 635.3570 | 1.1679 |
| Pearson Chi-Square | 544 | 944.2297 | 1.7257 |
| Scaled Pearson X2 | 544 | 1049.1288 | 1.9285 |
| Log Likelihood | | -4967.7678 | |
| Full Log Likelihood | | -4967.7678 | |
| AIC (smaller is better) | | 9763.5356 | |
| AICC (smaller is better) | | 9764.3105 | |
| BIC (smaller is better) | | 9824.0515 | |

The GENMOD Procedure

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -9893.6484 | | | |
| Qualite_PNB | -9854.7291 | 1 | 38.92 | <.0001 |
| T_AGE_inf22ans | -9802.3847 | 1 | 52.34 | <.0001 |
| T_Classe_SRA_A_a_H | -9802.1224 | 1 | 0.26 | 0.6085 |
| T_Classe_rep_SRA_O_a | -9800.2247 | 1 | 1.90 | 0.1683 |
| T_Energie_GO | -9788.2764 | 1 | 11.95 | 0.0005 |
| T_Garage_ferme | -9785.8999 | 1 | 2.38 | 0.1232 |
| T_Groupe_SRA_00_a_27 | -9781.3322 | 1 | 4.57 | 0.0326 |
| T_Groupe_SRA_29_a_32 | -9776.8244 | 1 | 4.50 | 0.0339 |
| T_Groupe_SRA_33_a_43 | -9755.9414 | 1 | 20.89 | <.0001 |
| T_car_Utilitaire | -9752.1575 | 1 | 3.78 | 0.0517 |
| temoin_pel | -9742.1034 | 1 | 10.05 | 0.0015 |
| temoin_secuicompte | -9735.3356 | 1 | 6.57 | 0.0104 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| Qualite_PNB | 1 | 13.98 | 0.0002 |
| T_AGE_inf22ans | 1 | 10.83 | 0.0010 |
| T_Classe_SRA_A_a_H | 1 | 8.70 | 0.0032 |
| T_Classe_rep_SRA_O_a | 1 | 5.77 | 0.0163 |
| T_Energie_GO | 1 | 16.70 | <.0001 |
| T_Garage_ferme | 1 | 9.28 | 0.0023 |
| T_Groupe_SRA_00_a_27 | 1 | 14.94 | 0.0001 |
| T_Groupe_SRA_29_a_32 | 1 | 18.65 | <.0001 |
| T_Groupe_SRA_33_a_43 | 1 | 20.93 | <.0001 |
| T_car_Utilitaire | 1 | 4.52 | 0.0336 |
| temoin_pel | 1 | 11.09 | 0.0009 |
| temoin_secuicompte | 1 | 6.57 | 0.0104 |

• ii : Responsabilité Civile Matérielle avec forfait IDA
 – Modélisation de la fréquence de sinistre

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | grisii |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_Credit_bonus | 2 | 0 1 |
| T_ville_Rurale | 2 | 0 1 |
| T_Groupe_SRA_00_a26 | 2 | 0 1 |
| T_AGE_inf25ans | 2 | 0 1 |
| temoin_pel | 2 | 0 1 |
| T_Sexe | 2 | 0 1 |
| T_Cond_novice_occas | 2 | 0 1 |
| qualite_baleII | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 76E3 | 14377.9346 | 0.1881 |
| Scaled Deviance | 76E3 | 14377.9346 | 0.1881 |
| Pearson Chi-Square | 76E3 | 96727.0381 | 1.2653 |
| Scaled Pearson X2 | 76E3 | 96727.0381 | 1.2653 |
| Log Likelihood | | -9771.3372 | |
| Full Log Likelihood | | -9810.6767 | |
| AIC (smaller is better) | | 19641.3533 | |
| AICC (smaller is better) | | 19641.3562 | |
| BIC (smaller is better) | | 19733.7980 | |

Algorithm converged.

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Credit_bonus | 1 | . | . |
| T_ville_Rurale | 1 | 49.26 | <.0001 |
| T_Groupe_SRA_00_a26 | 1 | . | . |
| T_AGE_inf25ans | 1 | . | . |
| temoin_pel | 1 | 10.96 | 0.0009 |
| T_Sexe | 1 | 6.00 | 0.0143 |
| T_Cond_novice_occas | 1 | . | . |
| qualite_baleII | 1 | 16.65 | <.0001 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | |
|--|-----|----------------|-------------|--|---------------|------------|--|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | Khi-2 de Wald | Pr > Khi-2 | |
| Intercept | 1 | -4.1624 | 0.2178 | -4.5893 -3.7355 | 365.15 | <.0001 | |
| T_Credit_bonus | 0 1 | 1.0655 | 0.1668 | 0.7385 1.3925 | 40.78 | <.0001 | |
| T_Credit_bonus | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| T_ville_Rurale | 0 1 | 0.3317 | 0.0483 | 0.2371 0.4263 | 47.22 | <.0001 | |
| T_ville_Rurale | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| T_Groupe_SRA_00_a26 | 0 1 | 0.6018 | 0.1033 | 0.3993 0.8044 | 33.93 | <.0001 | |
| T_Groupe_SRA_00_a26 | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| T_AGE_inf25ans | 0 1 | -0.6616 | 0.0759 | -0.8104 -0.5129 | 76.00 | <.0001 | |
| T_AGE_inf25ans | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| temoin_pel | 0 1 | 0.1486 | 0.0451 | 0.0601 0.2370 | 10.84 | 0.0010 | |
| temoin_pel | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| T_Sexe | 0 1 | -0.1118 | 0.0454 | -0.2008 -0.0229 | 6.07 | 0.0138 | |
| T_Sexe | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| T_Cond_novice_occas | 0 1 | -0.3764 | 0.0690 | -0.5097 -0.2430 | 30.59 | <.0001 | |
| T_Cond_novice_occas | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| qualite_baleII | 0 1 | 0.2386 | 0.0574 | 0.1271 0.3520 | 17.43 | <.0001 | |
| qualite_baleII | 1 0 | 0.0000 | 0.0000 | 0.0000 0.0000 | . | . | |
| Dispersion | 1 | 0.3141 | 0.1454 | 0.1268 0.7781 | | | |

ve binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -19845.294 | | | |
| T_Credit_bonus | -19783.294 | 1 | 62.00 | <.0001 |
| T_ville_Rurale | -19725.839 | 1 | 57.45 | <.0001 |
| T_Groupe_SRA_00_a26 | -19685.997 | 1 | 39.84 | <.0001 |
| T_AGE_inf25ans | -19605.698 | 1 | 80.31 | <.0001 |
| temoin_pel | -19591.178 | 1 | 14.51 | 0.0001 |
| T_Sexe | -19596.679 | 1 | 4.50 | 0.0339 |
| T_Cond_novice_occas | -19559.319 | 1 | 27.36 | <.0001 |
| qualite_baleII | -19542.674 | 1 | 16.65 | <.0001 |

- **rc : Responsabilité Civile Corporelle hors avance**
 - **Modélisation de la fréquence de sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | grierc |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_Cond_accompagne | 2 | 0 1 |
| T_Groupe_SRA_29 | 2 | 0 1 |
| T_ville_Rurale | 2 | 0 1 |
| qualite_baleII | 2 | 0 1 |
| T_Pack_plus | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 75E3 | 2187.6645 | 0.0294 |
| Scaled Deviance | 75E3 | 2187.6645 | 0.0294 |
| Pearson Chi-Square | 75E3 | 184506.0371 | 2.4757 |
| Scaled Pearson X2 | 75E3 | 184506.0371 | 2.4757 |
| Log Likelihood | | -1272.8322 | |
| Full Log Likelihood | | -1272.8322 | |
| AIC (smaller is better) | | 2559.6645 | |
| AICC (smaller is better) | | 2559.6660 | |
| BIC (smaller is better) | | 2624.1974 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | -4.3288 | 0.4570 | -5.2246 | -3.4331 | 89.72 | <.0001 |
| T_Cond_accompagne | 0 | 1 | -1.0607 | 0.4156 | -1.8752 | -0.2461 | 6.51 | 0.0107 |
| T_Cond_accompagne | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Groupe_SRA_29 | 0 | 1 | -0.5588 | 0.1736 | -0.8991 | -0.2185 | 10.36 | 0.0013 |
| T_Groupe_SRA_29 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_ville_Rurale | 0 | 1 | 0.3523 | 0.1666 | 0.0259 | 0.6788 | 4.47 | 0.0344 |
| T_ville_Rurale | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| qualite_baleII | 0 | 1 | 0.4124 | 0.1910 | 0.0380 | 0.7868 | 4.66 | 0.0309 |
| qualite_baleII | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | -0.3238 | 0.1546 | -0.6269 | -0.0207 | 4.38 | 0.0363 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Dispersion | | 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | | |

e binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -2573.7221 | | | |
| T_Cond_accompagne | -2568.2379 | 1 | 5.48 | 0.0192 |
| T_Groupe_SRA_29 | -2559.1185 | 1 | 9.12 | 0.0025 |
| T_ville_Rurale | -2553.9824 | 1 | 5.14 | 0.0234 |
| qualite_baleII | -2549.9102 | 1 | 4.07 | 0.0436 |
| T_TMEC_5a9 | -2545.6645 | 1 | 4.25 | 0.0393 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Cond_accompagne | 1 | . | . |
| T_Groupe_SRA_29 | 1 | 9.35 | 0.0022 |
| T_ville_Rurale | 1 | . | . |
| qualite_baleII | 1 | . | . |
| T_TMEC_5a9 | 1 | 4.25 | 0.0394 |

- **Modélisation du coût moyen d'un sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Charrc |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| qualite_baleII | 2 | 0 1 |
| T_AGE_inf22ans | 2 | 0 1 |
| T_AGE_56et60ans | 2 | 0 1 |
| T_AGE_61et65ans | 2 | 0 1 |
| T_AGE_plusde65ans | 2 | 0 1 |
| T_Classe_SRA_A_a_D | 2 | 0 1 |
| T_Classe_SRA_E | 2 | 0 1 |
| T_Classe_SRA_F | 2 | 0 1 |
| T_Energie_GO | 2 | 0 1 |
| T_Groupe_SRA_29 | 2 | 0 1 |
| T_Groupe_SRA_30 | 2 | 0 1 |
| T_Groupe_SRA_31 | 2 | 0 1 |
| T_Groupe_SRA_32 | 2 | 0 1 |
| T_Groupe_SRA_33 | 2 | 0 1 |
| T_Groupe_SRA_34 | 2 | 0 1 |
| T_Groupe_SRA_35 | 2 | 0 1 |
| T_Opt_veh_rplt | 2 | 0 1 |
| T_Pack_plus | 2 | 0 1 |
| T_Tarif_SW | 2 | 0 1 |
| T_formule_mini | 2 | 0 1 |
| T_veh_Citadine | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 173 | 237.2348 | 1.3713 |
| Scaled Deviance | 173 | 226.3720 | 1.3085 |
| Pearson Chi-Square | 173 | 239.0372 | 1.3817 |
| Scaled Pearson X2 | 173 | 228.0918 | 1.3184 |
| Log Likelihood | | -1885.2966 | |
| Full Log Likelihood | | -1885.2966 | |
| AIC (smaller is better) | | 3816.5932 | |
| AICC (smaller is better) | | 3823.0493 | |
| BIC (smaller is better) | | 3891.8722 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 30.8504 | 3.0040 | 24.9627 | 36.7381 | 105.47 | <.0001 |
| qualite_baleII | 0 | 1 | 0.5496 | 0.2152 | 0.1278 | 0.9713 | 6.52 | 0.0106 |
| qualite_baleII | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_inf22ans | 0 | 1 | 0.9750 | 0.3081 | 0.3711 | 1.5790 | 10.01 | 0.0016 |
| T_AGE_inf22ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_56et60ans | 0 | 1 | -0.8231 | 0.3186 | -1.4475 | -0.1987 | 6.67 | 0.0098 |
| T_AGE_56et60ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_61et65ans | 0 | 1 | -2.5436 | 0.3798 | -3.2879 | -1.7993 | 44.87 | <.0001 |
| T_AGE_61et65ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_plusde65ans | 0 | 1 | -0.7219 | 0.2785 | -1.2678 | -0.1760 | 6.72 | 0.0095 |
| T_AGE_plusde65ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_A_a_D | 0 | 1 | -1.5007 | 0.3997 | -2.2842 | -0.7173 | 14.09 | 0.0002 |
| T_Classe_SRA_A_a_D | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_E | 0 | 1 | -2.1571 | 0.4249 | -2.9898 | -1.3243 | 25.77 | <.0001 |
| T_Classe_SRA_E | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_F | 0 | 1 | -1.5983 | 0.3068 | -2.1997 | -0.9969 | 27.13 | <.0001 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | |
|--|-----|----------------|-------------|--|---------|---------------|--------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| T_Classe_SRA_F | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Energie_GO | 0 | 1 | -0.8100 | 0.2265 | -1.2539 | -0.3661 | 12.79 0.0003 |
| T_Energie_GO | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_29 | 0 | 1 | -1.1957 | 0.2678 | -1.7205 | -0.6709 | 19.94 <.0001 |
| T_Groupe_SRA_29 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_30 | 0 | 1 | -1.4018 | 0.3325 | -2.0535 | -0.7500 | 17.77 <.0001 |
| T_Groupe_SRA_30 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_31 | 0 | 1 | -1.7124 | 0.3387 | -2.3762 | -1.0486 | 25.56 <.0001 |
| T_Groupe_SRA_31 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_32 | 0 | 1 | -0.8837 | 0.3773 | -1.6232 | -0.1443 | 5.49 0.0192 |
| T_Groupe_SRA_32 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_33 | 0 | 1 | -2.2759 | 0.5253 | -3.3055 | -1.2463 | 18.77 <.0001 |
| T_Groupe_SRA_33 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_34 | 0 | 1 | -4.3763 | 1.0820 | -6.4970 | -2.2556 | 16.36 <.0001 |
| T_Groupe_SRA_34 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Groupe_SRA_35 | 0 | 1 | -1.3106 | 0.6107 | -2.5075 | -0.1138 | 4.61 0.0319 |
| T_Groupe_SRA_35 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Opt_veh_rplt | 0 | 1 | -2.0315 | 0.8519 | -3.7012 | -0.3618 | 5.69 0.0171 |
| T_Opt_veh_rplt | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Pack_plus | 0 | 1 | 0.4358 | 0.1713 | 0.1001 | 0.7715 | 6.47 0.0110 |
| T_Pack_plus | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_Tarif_SW | 0 | 1 | 1.3709 | 0.4899 | 0.4106 | 2.3311 | 7.83 0.0051 |
| T_Tarif_SW | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_formule_mini | 0 | 1 | -1.4640 | 0.2262 | -1.9074 | -1.0206 | 41.88 <.0001 |
| T_formule_mini | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| T_veh_Citadine | 0 | 1 | -0.6127 | 0.2008 | -1.0063 | -0.2191 | 9.31 0.0023 |
| T_veh_Citadine | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| Scale | | 1 | 0.9542 | 0.0847 | 0.8018 | 1.1356 | |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -3929.6907 | | | |
| qualite_baleII | -3921.9787 | 1 | 7.71 | 0.0055 |
| T_AGE_inf22ans | -3911.2605 | 1 | 10.72 | 0.0011 |
| T_AGE_56et60ans | -3902.5346 | 1 | 8.73 | 0.0031 |
| T_AGE_61et65ans | -3881.0986 | 1 | 21.44 | <.0001 |
| T_AGE_plusde65ans | -3877.6902 | 1 | 3.41 | 0.0649 |
| T_Classe_SRA_A_a_D | -3877.6729 | 1 | 0.02 | 0.8952 |
| T_Classe_SRA_E | -3876.5799 | 1 | 1.09 | 0.2958 |
| T_Classe_SRA_F | -3869.4963 | 1 | 7.08 | 0.0078 |
| T_Energie_GO | -3864.6961 | 1 | 4.80 | 0.0285 |
| T_Groupe_SRA_29 | -3864.6135 | 1 | 0.08 | 0.7738 |
| T_Groupe_SRA_30 | -3864.6134 | 1 | 0.00 | 0.9939 |
| T_Groupe_SRA_31 | -3864.6033 | 1 | 0.01 | 0.9200 |
| T_Groupe_SRA_32 | -3855.3709 | 1 | 9.23 | 0.0024 |
| T_Groupe_SRA_33 | -3844.1077 | 1 | 11.26 | 0.0008 |
| T_Groupe_SRA_34 | -3831.7194 | 1 | 12.39 | 0.0004 |
| T_Groupe_SRA_35 | -3830.5499 | 1 | 1.17 | 0.2795 |
| T_Opt_veh_rplt | -3826.1732 | 1 | 4.38 | 0.0364 |
| T_Pack_plus | -3822.0298 | 1 | 4.14 | 0.0418 |
| T_Tarif_SW | -3818.0382 | 1 | 3.99 | 0.0457 |
| T_formule_mini | -3779.6650 | 1 | 38.37 | <.0001 |
| T_veh_Citadine | -3770.5932 | 1 | 9.07 | 0.0026 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| qualite_baleII | 1 | 6.87 | 0.0088 |
| T_AGE_inf22ans | 1 | 8.03 | 0.0046 |
| T_AGE_56et60ans | 1 | 7.74 | 0.0054 |
| T_AGE_61et65ans | 1 | 53.69 | <.0001 |
| T_AGE_plusde65ans | 1 | 7.27 | 0.0070 |
| T_Classe_SRA_A_a_D | 1 | 13.47 | 0.0002 |
| T_Classe_SRA_E | 1 | 29.74 | <.0001 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Classe_SRA_F | 1 | 28.58 | <.0001 |
| T_Energie_GO | 1 | 11.40 | 0.0007 |
| T_Groupe_SRA_29 | 1 | 17.95 | <.0001 |
| T_Groupe_SRA_30 | 1 | 16.70 | <.0001 |
| T_Groupe_SRA_31 | 1 | 22.88 | <.0001 |
| T_Groupe_SRA_32 | 1 | 5.42 | 0.0199 |
| T_Groupe_SRA_33 | 1 | 23.69 | <.0001 |
| T_Groupe_SRA_34 | 1 | 37.96 | <.0001 |
| T_Groupe_SRA_35 | 1 | 5.43 | 0.0197 |
| T_Opt_veh_rplt | 1 | 7.12 | 0.0076 |
| T_Pack_plus | 1 | 6.22 | 0.0126 |
| T_Tarif_SW | 1 | 5.53 | 0.0187 |
| T_formule_mini | 1 | 45.10 | <.0001 |
| T_veh_Citadine | 1 | 9.07 | 0.0026 |

- **av : Responsabilité Civile Corporelle avec avance**
 - **Modélisation de la fréquence de sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | grieav |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_AGE_inf25ans | 2 | 0 1 |
| T_Garage_ferme | 2 | 0 1 |
| T_emission_CO2 | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 74E3 | 884.9559 | 0.0119 |
| Scaled Deviance | 74E3 | 884.9559 | 0.0119 |
| Pearson Chi-Square | 74E3 | 203770.4043 | 2.7385 |
| Scaled Pearson X2 | 74E3 | 203770.4043 | 2.7385 |
| Log Likelihood | | -503.4781 | |
| Full Log Likelihood | | -503.4781 | |
| AIC (smaller is better) | | 1016.9561 | |
| AICC (smaller is better) | | 1016.9569 | |
| BIC (smaller is better) | | 1063.0430 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|-----|----------------|-------------|--|---------|---------------|------------|--|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 | |
| | | | | | | | | |
| Intercept | 1 | -5.2548 | 0.4843 | -6.2039 | -4.3056 | 117.75 | <.0001 | |
| T_AGE_inf25ans | 0 1 | -1.0431 | 0.3688 | -1.7660 | -0.3202 | 8.00 | 0.0047 | |
| T_AGE_inf25ans | 1 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . | |
| T_Garage_ferme | 0 1 | 0.7338 | 0.2625 | 0.2193 | 1.2482 | 7.81 | 0.0052 | |
| T_Garage_ferme | 1 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . | |
| T_emission_CO2 | 0 1 | -0.9248 | 0.3806 | -1.6708 | -0.1788 | 5.90 | 0.0151 | |
| T_emission_CO2 | 1 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . | |
| Dispersion | 1 | 0.0000 | 0.0011 | . | . | . | . | |

nominal dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -1028.5227 | | | |
| T_AGE_inf25ans | -1019.0653 | 1 | 9.46 | 0.0021 |
| T_Garage_ferme | -1011.6745 | 1 | 7.39 | 0.0066 |
| T_emission_CO2 | -1006.9561 | 1 | 4.72 | 0.0298 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_AGE_inf25ans | 1 | 6.36 | 0.0117 |
| T_Garage_ferme | 1 | 7.45 | 0.0063 |
| T_emission_CO2 | 1 | . | . |

– **Modélisation du coût moyen d'un sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Charav |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_Anc_permis_4et5ans | 2 | 0 1 |
| temoin_securicompte_ | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|-----------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 55 | 53.2039 | 0.9673 |
| Scaled Deviance | 55 | 65.4444 | 1.1899 |
| Pearson Chi-Square | 55 | 40.1358 | 0.7297 |
| Scaled Pearson X2 | 55 | 49.3698 | 0.8976 |
| Log Likelihood | | -510.8266 | |
| Full Log Likelihood | | -510.8266 | |
| AIC (smaller is better) | | 1029.6533 | |
| AICC (smaller is better) | | 1030.4080 | |
| BIC (smaller is better) | | 1037.8950 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 6.6307 | 0.6376 | 5.3811 | 7.8803 | 108.16 | <.0001 |
| T_Anc_permis_4et5ans | 0 | 1 | 1.6180 | 0.6578 | 0.3287 | 2.9073 | 6.05 | 0.0139 |
| T_Anc_permis_4et5ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| temoin_securicompte_ | 0 | 1 | -0.8651 | 0.2424 | -1.3402 | -0.3901 | 12.74 | 0.0004 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|--------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| temoin_securicompte_ | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 1 | 1.2301 | 0.2047 | 0.8877 | 1.7045 | | |

parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -1035.1673 | | | |
| T_Anc_permis_4et5ans | -1032.7681 | 1 | 2.40 | 0.1214 |
| temoin_securicompte_ | -1021.6533 | 1 | 11.11 | 0.0009 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Anc_permis_4et5ans | 1 | 3.82 | 0.0508 |
| temoin_securicompte_ | 1 | 11.11 | 0.0009 |

- **ni : Défense recours Hors forfait Ida**
 - **Modélisation de la fréquence de sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | grieni |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_USAGE_AGRI | 2 | 0 1 |
| T_USAGE_FONC | 2 | 0 1 |
| T_USAGE_RETR | 2 | 0 1 |
| T_TMEC_0a4 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_ville_Rurale | 2 | 0 1 |
| T_Opt_PCC | 2 | 0 1 |
| qualite_baleII | 2 | 0 1 |
| T_car_Utilitaire | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 75E3 | 2671.0300 | 0.0356 |
| Scaled Deviance | 75E3 | 2671.0300 | 0.0356 |
| Pearson Chi-Square | 75E3 | 113612.8847 | 1.5145 |
| Scaled Pearson X2 | 75E3 | 113612.8847 | 1.5145 |
| Log Likelihood | | -1586.5710 | |
| Full Log Likelihood | | -1587.2641 | |
| AIC (smaller is better) | | 3196.5282 | |
| AICC (smaller is better) | | 3196.5317 | |
| BIC (smaller is better) | | 3298.0101 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|----------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | -10.8765 | 0.6227 | -12.0970 | -9.6560 | 305.07 | <.0001 |
| T_USAGE_AGRI | 0 | 1 | 0.8413 | 0.3056 | 0.2422 | 1.4403 | 7.58 | 0.0059 |
| T_USAGE_AGRI | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_FONC | 0 | 1 | 0.9032 | 0.2359 | 0.4409 | 1.3656 | 14.66 | 0.0001 |
| T_USAGE_FONC | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_RETR | 0 | 1 | 0.9041 | 0.2311 | 0.4511 | 1.3570 | 15.30 | <.0001 |
| T_USAGE_RETR | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_0a4 | 0 | 1 | 1.8867 | 0.2709 | 1.3558 | 2.4176 | 48.51 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | 0.5883 | 0.1479 | 0.2984 | 0.8783 | 15.82 | <.0001 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_ville_Rurale | 0 | 1 | 0.5062 | 0.1558 | 0.2008 | 0.8117 | 10.55 | 0.0012 |
| T_ville_Rurale | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Opt_PCC | 0 | 1 | -0.6468 | 0.1720 | -0.9840 | -0.3097 | 14.14 | 0.0002 |
| T_Opt_PCC | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| qualite_baleII | 0 | 1 | 0.4774 | 0.1561 | 0.1714 | 0.7833 | 9.35 | 0.0022 |
| qualite_baleII | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_car_Utilitaire | 0 | 1 | 0.7876 | 0.3447 | 0.1119 | 1.4632 | 5.22 | 0.0223 |
| T_car_Utilitaire | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Dispersion | | 1 | 0.0360 | 1.0399 | 0.0000 | 1.474E23 | | |

ve binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -3343.9649 | | | |
| T_USAGE_AGRI | -3333.5240 | 1 | 10.44 | 0.0012 |
| T_USAGE_FONC | -3315.0528 | 1 | 18.47 | <.0001 |
| T_USAGE_RETR | -3291.1411 | 1 | 23.91 | <.0001 |
| T_TMEC_0a4 | -3235.9935 | 1 | 55.15 | <.0001 |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| T_TMEC_5a9 | -3221.5455 | 1 | 14.45 | 0.0001 |
| T_ville_Rurale | -3206.1640 | 1 | 15.38 | <.0001 |
| T_Opt_PCC | -3187.5002 | 1 | 18.66 | <.0001 |
| qualite_baleII | -3179.7604 | 1 | 7.74 | 0.0054 |
| T_car_Utilitaire | -3173.1419 | 1 | 6.62 | 0.0101 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_USAGE_AGRI | 1 | 9.49 | 0.0021 |
| T_USAGE_FONC | 1 | . | . |
| T_USAGE_RETR | 1 | . | . |
| T_TMEC_0a4 | 1 | 79.27 | <.0001 |
| T_TMEC_5a9 | 1 | . | . |
| T_ville_Rurale | 1 | 11.46 | 0.0007 |
| T_Opt_PCC | 1 | 16.05 | <.0001 |
| qualite_baleII | 1 | 8.56 | 0.0034 |
| T_car_Utilitaire | 1 | 6.62 | 0.0101 |

- **Modélisation du coût moyen d'un sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Charni |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_Anc_permis_inf5ans | 2 | 0 1 |
| T_Anc_permis_13a21an | 2 | 0 1 |
| T_Anc_permis_22et25a | 2 | 0 1 |
| T_Anc_permis_plusde2 | 2 | 0 1 |
| T_Classe_SRA_A_a_C | 2 | 0 1 |
| T_Classe_SRA_D | 2 | 0 1 |
| T_Classe_SRA_E | 2 | 0 1 |
| T_Classe_SRA_F | 2 | 0 1 |
| T_Classe_SRA_G | 2 | 0 1 |
| T_Classe_SRA_I_a_V | 2 | 0 1 |
| temoin_securicompte_ | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 246 | 335.5118 | 1.3639 |
| Scaled Deviance | 246 | 301.6920 | 1.2264 |
| Pearson Chi-Square | 246 | 326.2568 | 1.3262 |
| Scaled Pearson X2 | 246 | 293.3699 | 1.1926 |
| Log Likelihood | | -1902.2343 | |
| Full Log Likelihood | | -1902.2343 | |
| AIC (smaller is better) | | 3830.4686 | |
| AICC (smaller is better) | | 3831.9604 | |
| BIC (smaller is better) | | 3876.6570 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 8.4070 | 1.3156 | 5.8284 | 10.9857 | 40.83 | <.0001 |
| T_Anc_permis_inf5ans | 0 | 1 | 0.5296 | 0.2213 | 0.0958 | 0.9634 | 5.73 | 0.0167 |
| T_Anc_permis_inf5ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Anc_permis_13a21an | 0 | 1 | 0.4352 | 0.2128 | 0.0181 | 0.8524 | 4.18 | 0.0409 |
| T_Anc_permis_13a21an | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Anc_permis_22et25a | 0 | 1 | 0.9136 | 0.2732 | 0.3782 | 1.4491 | 11.18 | 0.0008 |
| T_Anc_permis_22et25a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Anc_permis_plusde2 | 0 | 1 | 0.3953 | 0.1902 | 0.0225 | 0.7680 | 4.32 | 0.0377 |
| T_Anc_permis_plusde2 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_A_a_C | 0 | 1 | -0.7374 | 0.3252 | -1.3748 | -0.1000 | 5.14 | 0.0234 |
| T_Classe_SRA_A_a_C | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_D | 0 | 1 | -0.7527 | 0.3266 | -1.3928 | -0.1126 | 5.31 | 0.0212 |
| T_Classe_SRA_D | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_E | 0 | 1 | -0.9672 | 0.3210 | -1.5964 | -0.3381 | 9.08 | 0.0026 |
| T_Classe_SRA_E | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_F | 0 | 1 | -0.8220 | 0.2952 | -1.4005 | -0.2435 | 7.76 | 0.0054 |
| T_Classe_SRA_F | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_G | 0 | 1 | -0.7398 | 0.2919 | -1.3119 | -0.1677 | 6.42 | 0.0113 |
| T_Classe_SRA_G | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_I_a_V | 0 | 1 | -0.9742 | 0.2414 | -1.4474 | -0.5011 | 16.29 | <.0001 |
| T_Classe_SRA_I_a_V | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| temoin_secucompte_ | 0 | 1 | 0.4925 | 0.1450 | 0.2084 | 0.7767 | 11.54 | 0.0007 |
| temoin_secucompte_ | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 1 | 0.8992 | 0.0690 | 0.7736 | 1.0452 | | |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -3847.8754 | | | |
| T_Anc_permis_inf5ans | -3847.5022 | 1 | 0.37 | 0.5413 |
| T_Anc_permis_13a21an | -3847.3100 | 1 | 0.19 | 0.6610 |
| T_Anc_permis_22et25a | -3839.3663 | 1 | 7.94 | 0.0048 |
| T_Anc_permis_plusde2 | -3830.0592 | 1 | 9.31 | 0.0023 |
| T_Classe_SRA_A_a_C | -3829.6746 | 1 | 0.38 | 0.5352 |
| T_Classe_SRA_D | -3829.4755 | 1 | 0.20 | 0.6555 |
| T_Classe_SRA_E | -3829.4715 | 1 | 0.00 | 0.9493 |
| T_Classe_SRA_F | -3829.4490 | 1 | 0.02 | 0.8809 |
| T_Classe_SRA_G | -3829.2229 | 1 | 0.23 | 0.6344 |
| T_Classe_SRA_I_a_V | -3815.6743 | 1 | 13.55 | 0.0002 |
| temoin_securicompte_ | -3804.4686 | 1 | 11.21 | 0.0008 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Anc_permis_inf5ans | 1 | 5.61 | 0.0179 |
| T_Anc_permis_13a21an | 1 | 4.08 | 0.0434 |
| T_Anc_permis_22et25a | 1 | 9.72 | 0.0018 |
| T_Anc_permis_plusde2 | 1 | 4.38 | 0.0364 |
| T_Classe_SRA_A_a_C | 1 | 5.13 | 0.0235 |
| T_Classe_SRA_D | 1 | 5.30 | 0.0213 |
| T_Classe_SRA_E | 1 | 8.79 | 0.0030 |
| T_Classe_SRA_F | 1 | 7.31 | 0.0069 |
| T_Classe_SRA_G | 1 | 6.07 | 0.0137 |
| T_Classe_SRA_I_a_V | 1 | 13.03 | 0.0003 |
| temoin_securicompte_ | 1 | 11.21 | 0.0008 |

- ai : Défense recours avec forfait Ida
 - Modélisation de la fréquence de sinistre
- The GENMOD Procedure*

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | gtieai |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_AGE_inf30ans | 2 | 0 1 |
| T_Garage_ferme | 2 | 0 1 |
| T_TMEC_0a4 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_USAGE_RETR | 2 | 0 1 |
| T_USAGE_AGRI | 2 | 0 1 |
| qualite_baleII | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 75E3 | 2024.2662 | 0.0270 |
| Scaled Deviance | 75E3 | 2024.2662 | 0.0270 |
| Pearson Chi-Square | 75E3 | 112558.3554 | 1.5018 |
| Scaled Pearson X2 | 75E3 | 112558.3554 | 1.5018 |
| Log Likelihood | | -1620.9529 | |
| Full Log Likelihood | | -1624.4186 | |
| AIC (smaller is better) | | 3266.8373 | |
| AICC (smaller is better) | | 3266.8397 | |
| BIC (smaller is better) | | 3349.8592 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | -9.7113 | 0.6232 | -10.9328 | -8.4898 | 242.80 | <.0001 |
| T_AGE_inf30ans | 0 | 1 | -0.6518 | 0.1506 | -0.9471 | -0.3565 | 18.72 | <.0001 |
| T_AGE_inf30ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Garage_ferme | 0 | 1 | 0.4132 | 0.1354 | 0.1479 | 0.6786 | 9.32 | 0.0023 |
| T_Garage_ferme | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_0a4 | 0 | 1 | 2.2285 | 0.3272 | 1.5871 | 2.8699 | 46.38 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | 0.5311 | 0.1454 | 0.2462 | 0.8161 | 13.35 | 0.0003 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_RETR | 0 | 1 | 0.8160 | 0.2527 | 0.3207 | 1.3114 | 10.43 | 0.0012 |
| T_USAGE_RETR | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_AGRI | 0 | 1 | 1.6494 | 0.4188 | 0.8286 | 2.4701 | 15.51 | <.0001 |
| T_USAGE_AGRI | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| qualite_baleII | 0 | 1 | 0.4187 | 0.1615 | 0.1021 | 0.7353 | 6.72 | 0.0095 |
| qualite_baleII | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Dispersion | | 1 | 4.7290 | 2.4912 | 1.6841 | 13.2795 | | |

binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -3426.9659 | | | |
| T_AGE_inf30ans | -3403.5059 | 1 | 23.46 | <.0001 |
| T_Garage_ferme | -3364.8449 | 1 | 38.66 | <.0001 |
| T_TMEC_0a4 | -3306.7245 | 1 | 58.12 | <.0001 |
| T_TMEC_5a9 | -3281.7921 | 1 | 24.93 | <.0001 |
| T_USAGE_RETR | -3285.0370 | 1 | 0.00 | 1.0000 |
| T_USAGE_AGRI | -3248.1702 | 1 | 36.87 | <.0001 |
| qualite_baleII | -3241.9058 | 1 | 6.26 | 0.0123 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_AGE_inf30ans | 1 | 17.38 | <.0001 |
| T_Garage_ferme | 1 | 9.13 | 0.0025 |
| T_TMEC_0a4 | 1 | 88.38 | <.0001 |
| T_TMEC_5a9 | 1 | 14.18 | 0.0002 |
| T_USAGE_RETR | 1 | . | . |
| T_USAGE_AGRI | 1 | 26.13 | <.0001 |
| qualite_baleII | 1 | 6.26 | 0.0123 |

- **Modélisation du coût moyen d'un sinistre**
The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Charai |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_TMEC_0a4 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 268 | 279.3924 | 1.0425 |
| Scaled Deviance | 268 | 309.2528 | 1.1539 |
| Pearson Chi-Square | 268 | 238.3453 | 0.8893 |
| Scaled Pearson X2 | 268 | 263.8188 | 0.9844 |
| Log Likelihood | | -2067.4560 | |
| Full Log Likelihood | | -2067.4560 | |
| AIC (smaller is better) | | 4142.9120 | |
| AICC (smaller is better) | | 4143.0624 | |
| BIC (smaller is better) | | 4157.3205 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 8.1380 | 0.3039 | 7.5423 | 8.7337 | 716.91 | <.0001 |
| T_TMEC_0a4 | 0 | 1 | -1.1098 | 0.2831 | -1.6648 | -0.5548 | 15.36 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | -0.6124 | 0.1307 | -0.8687 | -0.3562 | 21.94 | <.0001 |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | |
|--|-----|----------------|-------------|--|--------|---------------|------------|
| Paramètre | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . |
| Scale | 1 | 1.1069 | 0.0845 | 0.9531 | 1.2855 | | |

eter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -4170.1506 | | | |
| T_TMEC_0a4 | -4157.7927 | 1 | 12.36 | 0.0004 |
| T_TMEC_5a9 | -4134.9120 | 1 | 22.88 | <.0001 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_TMEC_0a4 | 1 | 20.81 | <.0001 |
| T_TMEC_5a9 | 1 | 22.88 | <.0001 |

- iv : Incendie Vol
 - Modélisation de la fréquence de sinistre
- The GENMOD Procedure*

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | gtieiv |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_TMEC_0a4 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_Objt | 2 | 0 1 |
| T_AGE_inf30ans | 2 | 0 1 |
| T_AGE_31et35ans | 2 | 0 1 |
| T_Garage_ferme | 2 | 0 1 |
| T_car_Utilitaire | 2 | 0 1 |
| T_Classe_SRA_N_a_V | 2 | 0 1 |
| T_USAGE_FONC | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 19E3 | 2379.3249 | 0.1253 |
| Scaled Deviance | 19E3 | 2379.3249 | 0.1253 |
| Pearson Chi-Square | 19E3 | 39512.2410 | 2.0811 |
| Scaled Pearson X2 | 19E3 | 39512.2410 | 2.0811 |
| Log Likelihood | | -1525.5036 | |
| Full Log Likelihood | | -1527.5830 | |
| AIC (smaller is better) | | 3077.1661 | |
| AICC (smaller is better) | | 3077.1800 | |
| BIC (smaller is better) | | 3163.5379 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|----------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 1.6117 | 0.4644 | 0.7015 | 2.5219 | 12.05 | 0.0005 |
| T_TMEC_0a4 | 0 | 1 | -2.7957 | 0.1385 | -3.0671 | -2.5242 | 407.39 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | -0.9861 | 0.1312 | -1.2432 | -0.7290 | 56.50 | <.0001 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Objt | 0 | 1 | -1.4476 | 0.1958 | -1.8313 | -1.0638 | 54.65 | <.0001 |
| T_Objt | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_inf30ans | 0 | 1 | -0.3040 | 0.1360 | -0.5705 | -0.0375 | 5.00 | 0.0254 |
| T_AGE_inf30ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_31et35ans | 0 | 1 | -0.3751 | 0.1644 | -0.6973 | -0.0529 | 5.21 | 0.0225 |
| T_AGE_31et35ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Garage_ferme | 0 | 1 | 0.3066 | 0.1145 | 0.0822 | 0.5310 | 7.17 | 0.0074 |
| T_Garage_ferme | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_car_Utilitaire | 0 | 1 | 0.6466 | 0.2651 | 0.1270 | 1.1663 | 5.95 | 0.0147 |
| T_car_Utilitaire | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_N_a_V | 0 | 1 | -0.5227 | 0.2154 | -0.9449 | -0.1004 | 5.89 | 0.0153 |
| T_Classe_SRA_N_a_V | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_FONC | 0 | 1 | -0.4757 | 0.1311 | -0.7327 | -0.2186 | 13.16 | 0.0003 |
| T_USAGE_FONC | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Dispersion | | 1 | 0.0000 | 0.0000 | 0.0000 | 2.74E210 | | |

ve binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|--------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -3472.1431 | | | |
| T_TMEC_0a4 | -3191.2594 | 1 | 280.88 | <.0001 |
| T_TMEC_5a9 | -3127.1253 | 1 | 64.13 | <.0001 |
| T_Objt | -3093.7540 | 1 | 33.37 | <.0001 |
| T_AGE_inf30ans | -3088.8404 | 1 | 4.91 | 0.0266 |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| T_AGE_31et35ans | -3082.3392 | 1 | 6.50 | 0.0108 |
| T_Garage_ferme | -3076.2082 | 1 | 6.13 | 0.0133 |
| T_car_Utilitaire | -3067.7330 | 1 | 8.48 | 0.0036 |
| T_Classe_SRA_N_a_V | -3063.0139 | 1 | 4.72 | 0.0298 |
| T_USAGE_FONC | -3051.0072 | 1 | 12.01 | 0.0005 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_TMEC_0a4 | 1 | . | . |
| T_TMEC_5a9 | 1 | . | . |
| T_Objt | 1 | 37.79 | <.0001 |
| T_AGE_inf30ans | 1 | . | . |
| T_AGE_31et35ans | 1 | . | . |
| T_Garage_ferme | 1 | . | . |
| T_car_Utilitaire | 1 | 7.26 | 0.0071 |
| T_Classe_SRA_N_a_V | 1 | . | . |
| T_USAGE_FONC | 1 | . | . |

- **Modélisation du coût moyen d'un sinistre**

The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Chariv |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| Qualite_PNB | 2 | 0 1 |
| T_TMEC_0a4 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_Classe_SRA_L | 2 | 0 1 |
| T_Classe_SRA_M_a_V | 2 | 0 1 |
| T_veh_Microcitadine | 2 | 0 1 |
| temoin_securicompte_ | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 356 | 399.2061 | 1.1214 |
| Scaled Deviance | 356 | 417.9828 | 1.1741 |
| Pearson Chi-Square | 356 | 352.9729 | 0.9915 |
| Scaled Pearson X2 | 356 | 369.5749 | 1.0381 |
| Log Likelihood | | -3054.4764 | |
| Full Log Likelihood | | -3054.4764 | |
| AIC (smaller is better) | | 6126.9527 | |
| AICC (smaller is better) | | 6127.4612 | |
| BIC (smaller is better) | | 6162.0271 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 9.1409 | 0.3411 | 8.4723 | 9.8095 | 718.00 | <.0001 |
| Qualite_PNB | 0 | 1 | 0.3019 | 0.1170 | 0.0724 | 0.5313 | 6.65 | 0.0099 |
| Qualite_PNB | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_0a4 | 0 | 1 | -1.1315 | 0.1359 | -1.3978 | -0.8652 | 69.36 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | -0.4981 | 0.1273 | -0.7477 | -0.2486 | 15.31 | <.0001 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_L | 0 | 1 | -0.5032 | 0.2106 | -0.9160 | -0.0904 | 5.71 | 0.0169 |
| T_Classe_SRA_L | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_SRA_M_a_V | 0 | 1 | -0.6252 | 0.1761 | -0.9704 | -0.2801 | 12.60 | 0.0004 |
| T_Classe_SRA_M_a_V | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_veh_Microcitadine | 0 | 1 | 0.5517 | 0.1943 | 0.1708 | 0.9326 | 8.06 | 0.0045 |
| T_veh_Microcitadine | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| temoin_securicompte_ | 0 | 1 | -0.3418 | 0.1054 | -0.5485 | -0.1352 | 10.52 | 0.0012 |
| temoin_securicompte_ | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 1 | 1.0470 | 0.0686 | 0.9208 | 1.1905 | | |

parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -6263.9999 | | | |
| Qualite_PNB | -6258.2909 | 1 | 5.71 | 0.0169 |
| T_TMEC_0a4 | -6179.7852 | 1 | 78.51 | <.0001 |
| T_TMEC_5a9 | -6151.2726 | 1 | 28.51 | <.0001 |
| T_Classe_SRA_L | -6146.4578 | 1 | 4.81 | 0.0282 |
| T_Classe_SRA_M_a_V | -6125.6767 | 1 | 20.78 | <.0001 |
| T_veh_Microcitadine | -6119.2907 | 1 | 6.39 | 0.0115 |
| temoin_securicompte_ | -6108.9527 | 1 | 10.34 | 0.0013 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| Qualite_PNB | 1 | 6.87 | 0.0088 |
| T_TMEC_0a4 | 1 | 66.86 | <.0001 |
| T_TMEC_5a9 | 1 | 15.31 | <.0001 |
| T_Classe_SRA_L | 1 | 6.40 | 0.0114 |
| T_Classe_SRA_M_a_V | 1 | 13.73 | 0.0002 |
| T_veh_Microcitadine | 1 | 6.91 | 0.0086 |
| temoin_securicompte_ | 1 | 10.34 | 0.0013 |

- **bg : Bris de glace**
 - **Modélisation de la fréquence de sinistre**
The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | gtiebg |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_Energie_GO | 2 | 0 1 |
| T_USAGE_RETR | 2 | 0 1 |
| T_Classe_rep_SRA_K_a | 2 | 0 1 |
| T_Groupe_SRA_00_a26 | 2 | 0 1 |
| T_Groupe_SRA_27 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_TMEC_0a4 | 2 | 0 1 |
| T_Fract_Annuel | 2 | 0 1 |
| T_AGE_inf30ans | 2 | 0 1 |
| T_USAGE_AGRI | 2 | 0 1 |
| T_Classe_rep_SRA_N_a | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 93E3 | 28794.6443 | 0.3096 |
| Scaled Deviance | 93E3 | 28794.6443 | 0.3096 |
| Pearson Chi-Square | 93E3 | 127377.1238 | 1.3694 |
| Scaled Pearson X2 | 93E3 | 127377.1238 | 1.3694 |
| Log Likelihood | | -20957.3144 | |
| Full Log Likelihood | | -21190.2254 | |
| AIC (smaller is better) | | 42406.4508 | |
| AICC (smaller is better) | | 42406.4547 | |
| BIC (smaller is better) | | 42529.1796 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | -2.5939 | 0.1532 | -2.8941 | -2.2937 | 286.78 | <.0001 |
| T_Energie_GO | 0 | 1 | -0.3672 | 0.0352 | -0.4361 | -0.2983 | 109.09 | <.0001 |
| T_Energie_GO | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_RETR | 0 | 1 | 0.3545 | 0.0464 | 0.2636 | 0.4453 | 58.46 | <.0001 |
| T_USAGE_RETR | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_K_a | 0 | 1 | -0.1442 | 0.0317 | -0.2062 | -0.0822 | 20.75 | <.0001 |
| T_Classe_rep_SRA_K_a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Groupe_SRA_00_a26 | 0 | 1 | 0.4478 | 0.0817 | 0.2877 | 0.6079 | 30.06 | <.0001 |
| T_Groupe_SRA_00_a26 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Groupe_SRA_27 | 0 | 1 | 0.1842 | 0.0520 | 0.0823 | 0.2862 | 12.55 | 0.0004 |
| T_Groupe_SRA_27 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | -0.3928 | 0.0343 | -0.4601 | -0.3255 | 130.91 | <.0001 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_0a4 | 0 | 1 | -0.3801 | 0.0364 | -0.4515 | -0.3086 | 108.75 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Fract_Annuel | 0 | 1 | 0.1098 | 0.0321 | 0.0469 | 0.1726 | 11.71 | 0.0006 |
| T_Fract_Annuel | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_AGE_inf30ans | 0 | 1 | -0.2078 | 0.0352 | -0.2767 | -0.1388 | 34.86 | <.0001 |
| T_AGE_inf30ans | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_USAGE_AGRI | 0 | 1 | 0.1659 | 0.0508 | 0.0664 | 0.2654 | 10.68 | 0.0011 |
| T_USAGE_AGRI | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_N_a | 0 | 1 | -0.1074 | 0.0518 | -0.2089 | -0.0059 | 4.30 | 0.0381 |
| T_Classe_rep_SRA_N_a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Dispersion | | 1 | 0.3413 | 0.0651 | 0.2348 | 0.4960 | | |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|--------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -42943.818 | | | |
| T_Energie_GO | -42429.989 | 1 | 513.83 | <.0001 |
| T_USAGE_RETR | -42335.374 | 1 | 94.61 | <.0001 |
| T_Classe_rep_SRA_K_a | -42282.301 | 1 | 53.07 | <.0001 |
| T_Groupe_SRA_00_a26 | -42213.193 | 1 | 69.11 | <.0001 |
| T_Groupe_SRA_27 | -42170.041 | 1 | 43.15 | <.0001 |
| T_TMEC_5a9 | -42118.644 | 1 | 51.40 | <.0001 |
| T_TMEC_0a4 | -41993.944 | 1 | 124.70 | <.0001 |
| T_Fract_Annuel | -41966.276 | 1 | 27.67 | <.0001 |
| T_AGE_inf30ans | -41929.620 | 1 | 36.66 | <.0001 |
| T_USAGE_AGRI | -41918.833 | 1 | 10.79 | 0.0010 |
| T_Classe_rep_SRA_N_a | -41914.629 | 1 | 4.20 | 0.0403 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|--------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_Energie_GO | 1 | 113.81 | <.0001 |
| T_USAGE_RETR | 1 | 62.74 | <.0001 |
| T_Classe_rep_SRA_K_a | 1 | 20.56 | <.0001 |
| T_Groupe_SRA_00_a26 | 1 | 33.52 | <.0001 |
| T_Groupe_SRA_27 | 1 | 13.01 | 0.0003 |
| T_TMEC_5a9 | 1 | 132.66 | <.0001 |
| T_TMEC_0a4 | 1 | 109.23 | <.0001 |
| T_Fract_Annuel | 1 | 11.86 | 0.0006 |
| T_AGE_inf30ans | 1 | 33.71 | <.0001 |
| T_USAGE_AGRI | 1 | 11.05 | 0.0009 |
| T_Classe_rep_SRA_N_a | 1 | 4.20 | 0.0403 |

- Modélisation du coût moyen d'un sinistre
The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Charbg |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_emission_CO2 | 2 | 0 1 |
| T_car_Utilitaire | 2 | 0 1 |
| T_TMEC_0a4 | 2 | 0 1 |
| T_TMEC_5a9 | 2 | 0 1 |
| T_Fract_Annuel | 2 | 0 1 |
| T_Classe_rep_SRA_H | 2 | 0 1 |
| T_Classe_rep_SRA_I | 2 | 0 1 |
| T_Classe_rep_SRA_J | 2 | 0 1 |
| T_Classe_rep_SRA_K | 2 | 0 1 |
| T_Classe_rep_SRA_L | 2 | 0 1 |
| T_Classe_rep_SRA_M | 2 | 0 1 |
| T_Classe_rep_SRA_N_a | 2 | 0 1 |
| T_veh_Micro_ou_citad | 2 | 0 1 |
| T_veh_Familiale_ou_R | 2 | 0 1 |
| T_Anc_permis_13et16a | 2 | 0 1 |
| T_Anc_permis_17et21a | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 5424 | 1736.8269 | 0.3202 |
| Scaled Deviance | 5424 | 5714.1880 | 1.0535 |
| Pearson Chi-Square | 5424 | 1248.5077 | 0.2302 |
| Scaled Pearson X2 | 5424 | 4107.6103 | 0.7573 |
| Log Likelihood | | -35658.5802 | |
| Full Log Likelihood | | -35658.5802 | |
| AIC (smaller is better) | | 71353.1605 | |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|-----|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| AICC (smaller is better) | | 71353.2866 | |
| BIC (smaller is better) | | 71471.9914 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 7.4500 | 0.1730 | 7.1110 | 7.7890 | 1855.11 | <.0001 |
| T_emission_CO2 | 0 | 1 | -0.0789 | 0.0328 | -0.1432 | -0.0147 | 5.79 | 0.0161 |
| T_emission_CO2 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_car_Utilitaire | 0 | 1 | 0.1488 | 0.0322 | 0.0858 | 0.2119 | 21.42 | <.0001 |
| T_car_Utilitaire | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_0a4 | 0 | 1 | -0.1123 | 0.0231 | -0.1575 | -0.0670 | 23.66 | <.0001 |
| T_TMEC_0a4 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_TMEC_5a9 | 0 | 1 | -0.0541 | 0.0192 | -0.0918 | -0.0164 | 7.92 | 0.0049 |
| T_TMEC_5a9 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Fract_Annuel | 0 | 1 | 0.0516 | 0.0170 | 0.0182 | 0.0850 | 9.19 | 0.0024 |
| T_Fract_Annuel | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_H | 0 | 1 | -0.0774 | 0.0278 | -0.1319 | -0.0228 | 7.72 | 0.0055 |
| T_Classe_rep_SRA_H | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_I | 0 | 1 | -0.1184 | 0.0310 | -0.1791 | -0.0577 | 14.63 | 0.0001 |
| T_Classe_rep_SRA_I | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_J | 0 | 1 | -0.2008 | 0.0330 | -0.2654 | -0.1362 | 37.10 | <.0001 |
| T_Classe_rep_SRA_J | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_K | 0 | 1 | -0.2249 | 0.0345 | -0.2924 | -0.1574 | 42.62 | <.0001 |
| T_Classe_rep_SRA_K | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_L | 0 | 1 | -0.2792 | 0.0363 | -0.3503 | -0.2080 | 59.11 | <.0001 |
| T_Classe_rep_SRA_L | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_M | 0 | 1 | -0.3107 | 0.0414 | -0.3919 | -0.2296 | 56.30 | <.0001 |
| T_Classe_rep_SRA_M | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| T_Classe_rep_SRA_N_a | 0 | 1 | -0.4310 | 0.0430 | -0.5153 | -0.3467 | 100.42 | <.0001 |
| T_Classe_rep_SRA_N_a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_veh_Micro_ou_citad | 0 | 1 | 0.0558 | 0.0246 | 0.0075 | 0.1041 | 5.13 | 0.0236 |
| T_veh_Micro_ou_citad | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_veh_Familiale_ou_R | 0 | 1 | -0.0651 | 0.0241 | -0.1123 | -0.0179 | 7.31 | 0.0069 |
| T_veh_Familiale_ou_R | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Anc_permis_13et16a | 0 | 1 | -0.0631 | 0.0248 | -0.1117 | -0.0145 | 6.47 | 0.0110 |
| T_Anc_permis_13et16a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Anc_permis_17et21a | 0 | 1 | -0.0585 | 0.0227 | -0.1030 | -0.0140 | 6.64 | 0.0099 |
| T_Anc_permis_17et21a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 1 | 3.2900 | 0.0602 | 3.1742 | 3.4101 | | |

parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|--------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -71982.519 | | | |
| T_emission_CO2 | -71977.462 | 1 | 5.06 | 0.0245 |
| T_car_Utilitaire | -71937.786 | 1 | 39.68 | <.0001 |
| T_TMEC_0a4 | -71849.411 | 1 | 88.38 | <.0001 |
| T_TMEC_5a9 | -71819.244 | 1 | 30.17 | <.0001 |
| T_Fract_Annuel | -71806.451 | 1 | 12.79 | 0.0003 |
| T_Classe_rep_SRA_H | -71781.578 | 1 | 24.87 | <.0001 |
| T_Classe_rep_SRA_I | -71773.702 | 1 | 7.88 | 0.0050 |
| T_Classe_rep_SRA_J | -71772.331 | 1 | 1.37 | 0.2416 |
| T_Classe_rep_SRA_K | -71764.744 | 1 | 7.59 | 0.0059 |
| T_Classe_rep_SRA_L | -71729.213 | 1 | 35.53 | <.0001 |
| T_Classe_rep_SRA_M | -71667.126 | 1 | 62.09 | <.0001 |
| T_Classe_rep_SRA_N_a | -71341.176 | 1 | 325.95 | <.0001 |
| T_veh_Micro_ou_citad | -71335.743 | 1 | 5.43 | 0.0198 |
| T_veh_Familiale_ou_R | -71328.778 | 1 | 6.96 | 0.0083 |

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| T_Anc_permis_13et16a | -71323.891 | 1 | 4.89 | 0.0271 |
| T_Anc_permis_17et21a | -71317.160 | 1 | 6.73 | 0.0095 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|--------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_emission_CO2 | 1 | 5.86 | 0.0155 |
| T_car_Utilitaire | 1 | 20.64 | <.0001 |
| T_TMEC_0a4 | 1 | 23.65 | <.0001 |
| T_TMEC_5a9 | 1 | 7.90 | 0.0050 |
| T_Fract_Annuel | 1 | 9.11 | 0.0025 |
| T_Classe_rep_SRA_H | 1 | 7.78 | 0.0053 |
| T_Classe_rep_SRA_I | 1 | 14.77 | 0.0001 |
| T_Classe_rep_SRA_J | 1 | 37.32 | <.0001 |
| T_Classe_rep_SRA_K | 1 | 42.67 | <.0001 |
| T_Classe_rep_SRA_L | 1 | 59.06 | <.0001 |
| T_Classe_rep_SRA_M | 1 | 56.87 | <.0001 |
| T_Classe_rep_SRA_N_a | 1 | 101.13 | <.0001 |
| T_veh_Micro_ou_citad | 1 | 5.13 | 0.0235 |
| T_veh_Familiale_ou_R | 1 | 7.33 | 0.0068 |
| T_Anc_permis_13et16a | 1 | 6.56 | 0.0104 |
| T_Anc_permis_17et21a | 1 | 6.73 | 0.0095 |

- di : Dommage Ida
 - Modélisation de la fréquence de sinistre
- The GENMOD Procedure*

| Informations sur le modèle | |
|----------------------------|-------------------|
| Data Set | WORK.ECH_20_21 |
| Distribution | Negative Binomial |
| Link Function | Log |
| Dependent Variable | gtiedi |
| Scale Weight Variable | Nb_annees |
| Offset Variable | Lduree |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| T_TMEC_10etPlus | 2 | 0 1 |
| T_CRM_55et76 | 2 | 0 1 |
| T_CRM_77et90 | 2 | 0 1 |
| T_CRM_plusde90 | 2 | 0 1 |
| T_ville_Rurale | 2 | 0 1 |
| T_Energie_GO | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 45E3 | 8284.3060 | 0.1834 |
| Scaled Deviance | 45E3 | 8284.3060 | 0.1834 |
| Pearson Chi-Square | 45E3 | 58317.3844 | 1.2912 |
| Scaled Pearson X2 | 45E3 | 58317.3844 | 1.2912 |
| Log Likelihood | | -5788.5149 | |
| Full Log Likelihood | | -5818.0326 | |
| AIC (smaller is better) | | 11652.0651 | |
| AICC (smaller is better) | | 11652.0683 | |
| BIC (smaller is better) | | 11721.8111 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | -1.0888 | 0.1970 | -1.4750 | -0.7025 | 30.53 | <.0001 |
| T_TMEC_10etPlus | 0 | 1 | 0.3336 | 0.0852 | 0.1666 | 0.5006 | 15.32 | <.0001 |
| T_TMEC_10etPlus | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_CRM_55et76 | 0 | 1 | -0.5404 | 0.0669 | -0.6715 | -0.4094 | 65.33 | <.0001 |
| T_CRM_55et76 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_CRM_77et90 | 0 | 1 | -0.8265 | 0.0986 | -1.0198 | -0.6333 | 70.25 | <.0001 |
| T_CRM_77et90 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_CRM_plusde90 | 0 | 1 | -1.4706 | 0.1056 | -1.6776 | -1.2637 | 193.98 | <.0001 |
| T_CRM_plusde90 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_ville_Rurale | 0 | 1 | 0.2699 | 0.0626 | 0.1472 | 0.3926 | 18.60 | <.0001 |
| T_ville_Rurale | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Energie_GO | 0 | 1 | -0.1646 | 0.0659 | -0.2938 | -0.0354 | 6.23 | 0.0125 |
| T_Energie_GO | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Dispersion | | 1 | 0.5001 | 0.1948 | 0.2331 | 1.0730 | | |

binomial dispersion parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|--------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -11878.864 | | | |
| T_TMEC_10etPlus | -11833.455 | 1 | 45.41 | <.0001 |
| T_CRM_55et76 | -11817.339 | 1 | 16.12 | <.0001 |
| T_CRM_77et90 | -11765.699 | 1 | 51.64 | <.0001 |
| T_CRM_plusde90 | -11601.021 | 1 | 164.68 | <.0001 |
| T_ville_Rurale | -11583.414 | 1 | 17.61 | <.0001 |
| T_Energie_GO | -11577.030 | 1 | 6.38 | 0.0115 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| T_TMEC_10etPlus | 1 | 16.44 | <.0001 |
| T_CRM_55et76 | 1 | 60.93 | <.0001 |
| T_CRM_77et90 | 1 | . | . |
| T_CRM_plusde90 | 1 | . | . |
| T_ville_Rurale | 1 | 19.25 | <.0001 |
| T_Energie_GO | 1 | 6.38 | 0.0115 |

- Modélisation du coût moyen d'un sinistre
The GENMOD Procedure

| Informations sur le modèle | |
|----------------------------|-----------|
| Data Set | WORK.TEST |
| Distribution | Gamma |
| Link Function | Log |
| Dependent Variable | Chardi |

| Informations sur le niveau de classe | | |
|--------------------------------------|---------|---------|
| Classe | Niveaux | Valeurs |
| temoin_securicompte_ | 2 | 0 1 |
| T_veh_Familiale | 2 | 0 1 |
| T_veh_Routiere | 2 | 0 1 |
| T_Classe_rep_SRA_G_a | 2 | 0 1 |
| T_Classe_rep_SRA_L_a | 2 | 0 1 |
| T_CRM_51a90 | 2 | 0 1 |
| T_CRM_plusde90 | 2 | 0 1 |

| Critères d'évaluation de l'adéquation | | | |
|---------------------------------------|------|-------------|------------|
| Critère | DDL | Valeur | Valeur/DDL |
| Deviance | 1380 | 1628.9812 | 1.1804 |
| Scaled Deviance | 1380 | 1605.1368 | 1.1631 |
| Pearson Chi-Square | 1380 | 1466.7765 | 1.0629 |
| Scaled Pearson X2 | 1380 | 1445.3063 | 1.0473 |
| Log Likelihood | | -11496.7498 | |
| Full Log Likelihood | | -11496.7498 | |
| AIC (smaller is better) | | 23011.4995 | |
| AICC (smaller is better) | | 23011.6301 | |
| BIC (smaller is better) | | 23058.6201 | |

Algorithm converged.

| Paramètres estimés par l'analyse du maximum de vraisemblance | | | | | | | | |
|--|---|-----|----------------|-------------|--|---------|---------------|------------|
| Paramètre | | DDL | Valeur estimée | Erreur type | Intervalle de confiance de Wald à 95 % | | Khi-2 de Wald | Pr > Khi-2 |
| Intercept | | 1 | 7.4189 | 0.1925 | 7.0415 | 7.7962 | 1484.84 | <.0001 |
| temoin_securicompte_ | 0 | 1 | 0.1160 | 0.0549 | 0.0085 | 0.2235 | 4.47 | 0.0344 |
| temoin_securicompte_ | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_veh_Familiale | 0 | 1 | 0.2548 | 0.0882 | 0.0818 | 0.4277 | 8.34 | 0.0039 |
| T_veh_Familiale | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_veh_Routiere | 0 | 1 | 0.3385 | 0.1283 | 0.0870 | 0.5899 | 6.96 | 0.0083 |
| T_veh_Routiere | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_G_a | 0 | 1 | -0.1885 | 0.0755 | -0.3365 | -0.0405 | 6.23 | 0.0125 |
| T_Classe_rep_SRA_G_a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_Classe_rep_SRA_L_a | 0 | 1 | -0.3593 | 0.0952 | -0.5459 | -0.1727 | 14.24 | 0.0002 |
| T_Classe_rep_SRA_L_a | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_CRM_51a90 | 0 | 1 | -0.1684 | 0.0572 | -0.2806 | -0.0563 | 8.67 | 0.0032 |
| T_CRM_51a90 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| T_CRM_plusde90 | 0 | 1 | -0.3172 | 0.1010 | -0.5151 | -0.1193 | 9.87 | 0.0017 |
| T_CRM_plusde90 | 1 | 0 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | . | . |
| Scale | | 1 | 0.9854 | 0.0329 | 0.9230 | 1.0520 | | |

parameter was estimated by maximum likelihood.

| Statistique LR pour Analyse de Type 1 | | | | |
|---------------------------------------|---------------------|-----|-------|------------|
| Source | 2*Log-vraisemblance | DDL | Khi-2 | Pr > Khi-2 |
| Intercept | -23033.366 | | | |
| temoin_securicompte_ | -23028.921 | 1 | 4.45 | 0.0350 |
| T_veh_Familiale | -23026.574 | 1 | 2.35 | 0.1255 |
| T_veh_Routiere | -23023.951 | 1 | 2.62 | 0.1053 |
| T_Classe_rep_SRA_G_a | -23023.949 | 1 | 0.00 | 0.9654 |
| T_Classe_rep_SRA_L_a | -23008.685 | 1 | 15.26 | <.0001 |
| T_CRM_51a90 | -23004.107 | 1 | 4.58 | 0.0324 |
| T_CRM_plusde90 | -22993.500 | 1 | 10.61 | 0.0011 |

| Statistique LR pour Analyse de Type 3 | | | |
|---------------------------------------|-----|-------|------------|
| Source | DDL | Khi-2 | Pr > Khi-2 |
| temoin_securicompte_ | 1 | 4.44 | 0.0350 |
| T_veh_Familiale | 1 | 8.05 | 0.0046 |
| T_veh_Routiere | 1 | 6.57 | 0.0104 |
| T_Classe_rep_SRA_G_a | 1 | 6.02 | 0.0142 |
| T_Classe_rep_SRA_L_a | 1 | 14.06 | 0.0002 |
| T_CRM_51a90 | 1 | 8.71 | 0.0032 |
| T_CRM_plusde90 | 1 | 10.61 | 0.0011 |

6. Programme sas permettant le calcul de la prime pure par garantie :

```
data &tablef (compress=yes);
set &tablef;
purenn=exp(logfnn)*exp(logcnn)*&secretenn; /*rcmnida*/
pureii=exp(logfii)*1128; /*rcmida*/
pureav=exp(logfav)*&moycharav; /*avac*/
purerc=exp(logfrc)*&moycharrc; /*rcc*/
if formule ne '1' then do; pureiv=exp(logfiv)*exp(logciv)*&creteiv; /*IV*/
                        pureva=exp(logfva)*exp(logcva); /*VAR*/
                        purebg=exp(logfbg)*exp(logcbg); /*BDG*/
                        pureop=exp(logfop)*exp(logcop); /*OPT*/
                        end;
                        else do; pureiv=0;
                        pureva=0;
                        purebg=0;
                        pureop=0;
                        end;
if formule='3' then do; puredo=exp(logfdo)*exp(logcdo)*&cretedo; /*DOMNIDA*/
                        puredi=exp(logfdi)*(exp(logcdi)-
1128); /*DOMIDA*/
                        puredv=exp(logfdv)*&moychardv; /*VAN*/
                        pureni=0;
                        pureai=0;
                        end;
                        else do; puredo=0;
                        puredi=0;
                        puredv=0;
                        pureni=exp(logfni)*exp(logcni); /*AVANIDA*/
                        pureai=exp(logfai)*(exp(logcai)-
1128); /*AVAIDA*/
                        end;
ppdo=sum(puredo,puredi,puredv);
ppiv=sum(pureva,pureiv);
ppbg=sum(purebg,pureop);
pprc=sum(purenn,pureii,pureav,purerc,pureni,pureai);
run;
```

Cette partie du programme donne la méthode de calcul de l'écrêtement :

```
/******
*****/
/*
MACRO POUR ETUDIER LE SEUIL D'ECRETEMENT
*/
/******
*****/

%macro etude_seuil;

/* Etudier la distribution de la charge */

data toto.char&gar(keep=gtie&gar char&gar som_&gar);
set cout;
if gtie&gar=1;
if char&gar>0;
run;

data temp;
set toto.char&gar;
retain cum 0;
cum = char&gar+cum; output;
```

```

call symput("tot&gar",compress(cum));
run;
%put &&tot&gar;

data temp;
set toto.char&gar(where=(char&gar<=&seuil)) end=eof;
retain cum 0;
cum = char&gar+cum; output;
/* charge sous crête : &&sous&gar */
if eof then call symput("sous&gar",compress(cum));
run;
%put &&sous&gar;

data ratio;
ecrete=&&tot&gar/&&sous&gar;
call symput("ecrete&gar",compress(ecrete));
run;

/* Pourcentage de la charge non modélisée, à ajouter à la prime pure : */
%put &&ecrete&gar;
%mend;

```

Cette partie du programme indique comment nous pouvons calculer les coûts moyens écrétés

```

/* Création du fichier des coûts
*/

%macro creation_fichier_cout;

%let et=;
%let ecret=;
%let siida=;
%if &seuil ne %then %let ecret=if char&gar>&seuil then char&gar=&seuil;
%if &cout_machine ne %then %let et=and char&gar ne %eval(&cout_machine);

data cout (compress=yes rename=(cdoida=chardi gdoida=gtiedi gavaida=gtieai
cavaida=charai gavanida=gtieni
cavanida=charni grcmnida=gtienn crcmida=charnn
grcmida=gtieii
crcmida=charii grcc=gtierc crcc=charrc gvac=gtieav
cavac=charav giv=gtieiv civ=chariv gvar=gtieva cvar=charva
gdonida=gtiedo cdonida=chardo cvan=chardv gvan=gtiedv
cbdg=charbg gbdg=gtiebg copt=charop gopt=gtieop)
keep=formule cfract mtcothta annees &unique cdoida cavaida cdonida crcc
cavac civ cvar cvan copt crcmida grcmnida cbdg cavanida gdoida gavaida
gdonida grcc gvac giv grcmida grcmnida gavanida gbdg gvar gvan gopt);
set &tablef;
*if agecond=. then delete;
*if agecond<0 or agecond>2000 then delete;
if agevehic=. then delete;
if annees<0 then delete;
format &for;
run;

%if &seuil ne %then %etude_seuil;

data cout;
set cout;

```

```
if gtie&gar ne 0;  
char&gar=char&gar/gtie&gar;  
%if &gar=ai or &gar=di %then char&gar=char&gar+%eval(&forfait);  
&secret;  
run;  
  
data toto.cm_&gar;  
set cout;  
if char&gar>0 &et;  
wghta=1;  
logexp=log(annees);  
run;  
%mend;
```

BIBLIOGRAPHIE

Optimisez l'évaluation des risques pour personnaliser votre tarification AUTO.
Formation IIR Form'action assurance dispensée par Pascal GONZALVEZ et Sophie O'KELLY

Bulletin Français d'actuariat, Vol 4, N°7, 2000, pp 61-82:
Assurance Automobile Etudes actuarielles et politique tarifaire en 2001.
Gilbert THIRY Membre d'honneur de l'IAF

Assurance automobile toute la vérité sur le bonus-malus français.
Gilbert THIRY Membre d'honneur de l'IAF
E.S.R.A - Paris

Mémoire d'actuariat CEA février 1994 :
Tarif, concurrence & rentabilité : le cas de l'assurance automobile
Détermination d'un tarif commercial qui maximise la valeur de l'entreprise dans un environnement concurrentiel
Frédéric BOULANGER et Laurent TRAN VAN LIEU

Systèmes Bonus-Malus
Philippe BIENAIME Actuaire I.S.F.A., GPA
Laboratoire de Sciences Actuarielle et Financière, I.S.F.A.,
Université Claude Bernard Lyon 1
Nathalie RICHARD GPA
Laboratoire de Sciences Actuarielle et Financière, I.S.F.A.,
Université Claude Bernard Lyon 1

BESSON J.L. et PARTRAT C., Trend et système bonus-Malus, 1992,
ASTIN Bulletin, VOL. 22, N 1, 1992

The mixed bivariate Hofmann distribution by J.F. Walhin and J. Paris Université Catholique
de Louvain Secura Belgian Re
ASTIN BULLETIN, Vol. 31, N°. 1, 2001, pp. 123-138

BIENAIME P. et RICHARD N., Fréquence de sinistres et Bonus -Malus, 1998,
Mémoire d'actuariat, Institut des Sciences Financière et d'Assurances

KOCHERLAKOTA S. et KOCHERLAKOTA N. Bivariate discrete distributions, 1992,
Marcel Dekker

Séminaire TARIFICATION AUTOMOBILE ET SYSTEMES BONUS-MALUS du 7 et 8
novembre 2002 PARIS ISFA UCL

Produits multigaranties : Techniques de réévaluation des risques en cours.
Etude réalisée pour Winterthur Belgium en mai 2001.
UCL Institut de statistique Consulting report N°01-03.
Michel Denuit et Jean Pinquet

Institut de statistique Université Catholique de Louvain

Discussion paper 0143 October 19, 2001

Personnalisation des primes-fréquence en assurance automobile par régression poissonienne en présence de données longitudinales.

Michel Denuit, Sandra Pitrebois & Jean-François WALHIN

Site <http://www.crest.fr/pageperso/lfa/charpent.htm>

Actuariat non-vie Mathématiques de l'assurance Non-Vie.

Arthur Charpentier ENSAE - CREST

Site <http://www.crest.fr/pageperso/lfa/charpent.htm>

Actuariat non-vie Provisionnement et tarification. Approche technique

Arthur Charpentier ENSAE - CREST

Denuit, M. & Charpentier, A. (2004). Mathématiques de l'Assurance Non-Vie, Principes fondamentaux de théorie du risque (Tome 1).

Charpentier, A. & Denuit, M. (2004). Mathématiques de l'Assurance Non-Vie, Théorie du risque, tarification, provisionnement (Tome 2).

Gouriéroux, C.& Monfort, A. (1996). Statistique et modèles Econométriques. Vol 1.

Dobson, A.J. (2002). An introduction to Generalized Linear Models. Chapman & Hall.

La Réassurance Approche technique Jacques Blondeau et Christian Partrat.

Pierre PETAUTON, 2000, Théorie de l'assurance dommages, Dunod, Paris.

Quelques relations entre les principaux moments représentatifs du risque en assurance automobile, sur la base d'un échantillon segmenté. Charles HELBRONNER Institut des actuaires Bulletin Français d'actuariat. Volume 4 - N°8 - 2000