Mémoire présenté le :

pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA et l'admission à l'Institut des Actuaires

Par : Jean CAUX

Titre : Modélisation de l'entrée en incapacité de travail en prévoyance collective

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

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

Membre présents du jury de l'Institut des Actuaires

Signature

Entreprise :

Nom : Malakoff-Médéric

Signature :

Membres présents du jury de l'ISFA

Directeur de mémoire en entreprise :

Nom : M. Loc PHAN

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

Signature du candidat

Secrétariat :

Mme Christine DRIGUZZI

Bibliothèque :

Mme Patricia BARTOLO

Malakoff-Médéric

Modélisation de l'entrée en incapacité de travail en prévoyance collective

Jean CAUX

Mémoire ISFA 2017

Remerciements

Je tiens tout d'abord à remercier infiniment l'ensemble de l'équipe de surveillance du portefeuille de Malakoff-Médéric pour cette excellente année d'alternance passée en sa compagnie et pour l'aide apportée dans l'élaboration de ce mémoire. Un grand merci à Philippine pour ses relectures attentives, à Marie-Hélène pour son écoute et ses conseils mais aussi et surtout à Loc Phan pour tout son investissement dans cette étude, pour m'avoir accueilli et fait partager ses connaissances du métier.

Je tiens aussi à remercier Denis Clot, mon tuteur pédagogique, pour ses relectures et Frédéric Planchet pour ses réponses rapides et précises à diverses questions techniques.

Je tiens enfin à remercier Manue pour ses nombreuses relectures qui m'ont permises d'avancer sereinement, mes colocataires Luc et Ronan mais aussi Nadine, Yves, Lère et tous les autres pour leur soutien tout au long de ce mémoire.

Résumé

Mots clés : *Incapacité de travail, fréquence, Poisson, Whittaker-Henderson, variables explicatives, franchise, segmentation, Cox, CART, arbres de décision, forêt aléatoire de régression, approche hybride, outil décisionnel, prédiction*

Ce mémoire répond à des problématiques opérationnelles qui entourent la fréquence du risque d'incapacité de travail en assurance collective. Ces problématiques tournent autour de la compréhension, de l'explication et de la prédiction du risque. Pour y répondre, il est nécessaire d'identifier les variables explicatives de l'incapacité de travail et leur importance, de déterminer des groupes homogènes face à la sinistralité et de construire des lois d'expérience segmentées de l'entrée en incapacité.

Le mémoire se divise en deux grandes parties et neuf chapitres, qui suivent un cheminement logique et participent chacun au développement des réponses à ces problématiques.

Dans un premier temps, une étude et un traitement minutieux des données de sinistres et d'effectifs sont effectués. Malakoff-Médéric réalise une affiliation de chaque assuré à la souscription du contrat, ainsi, des informations telles que le sexe, le collège, la date d'entrée dans le contrat... sont disponibles dans les bases de données. Des études de cohérence des données sont ensuite réalisées sur l'échantillon retenu.

Dans un deuxième temps, une approche basée sur des théories actuarielles classiques est utilisée pour déterminer les variables explicatives de la sinistralité et modéliser des fréquences d'entrée en incapacité. Une comparaison de la sinistralité entre les modalités des variables permet de juger de leur caractère explicatif. Des tables d'entrée en incapacité par âge sont ensuite construites par franchise et sexe, puis par franchise et secteur et enfin par franchise, sexe et catégorie socio-professionnelle. Ces tables donnent la tendance de la sinistralité avec l'âge pour les variables explicatives les plus classiques. Enfin, un modèle mêlant le modèle à risque proportionnel de Cox et des positionnements de populations est paramétré. Il permet d'obtenir des lois d'entrée en incapacité segmentées de façon fine et de prédire avec précision la sinistralité d'un groupe d'individus.
Les résultats obtenus servent de base aux modélisations suivantes, plus "innovantes" et permettent d'en vérifier la cohérence.

Dans un troisième temps, une approche se basant sur les arbres de décision et les algorithmes *CART* et *Random Forest* est développée. Cette méthode, plus souple, est complémentaire à l'approche "classique" dans la recherche des variables explicatives de la sinistralité, elle permet de déterminer des groupes homogènes face au risque et un classement précis des variables par secteur. Elle est aussi utilisée pour créer un modèle prédictif de l'entrée en incapacité. Ce modèle présente des avantages importants face à la modélisation classique présentée précédemment.
Enfin, une approche "hybride" améliorant encore plus la modélisation, est proposée dans une dernière partie. C'est ce modèle qui est finalement retenu pour modéliser les fréquences d'entrée en incapacité.

Ces trois parties fournissent des réponses aux problématiques opérationnelles étudiées. Elles mettent par ailleurs en évidence l'intérêt de l'utilisation des arbres de décision pour modéliser le risque incapacité ainsi que ses avantages et ses inconvénients face à une approche plus classique, utilisant le très connu modèle de Cox.

Abstract

Key words : *sick-leave, frequency, Poisson, Whittaker-Henderson, explanatory variables, deductible, heterogeneity, Cox, CART, decision trees, random forest, hybrid approach, decision-making tool, prediction*

This report answers operational issues about the frequency of sick-leave for collective provident insurance. These issues are about the comprehension, the explanation and the prediction of the risk. To answer these points, it is necessary to identify the explanatory variables of sick-leave and their importance, to determine some homogeneous groups of individuals facing the sinistrality and to build segmented experience tables of the entry in sick-leave.

This report is divided in two main parts and nine chapters. They follow a consistent progression and answer the different issues.

Firstly, a study and a meticulous treatment of the available data are done. Malakoff-Médéric carries out an affiliation of its insured when contracts are subscribed, consequently, informations such as gender, socio-professional category, date of subscription... are available in the database. The consistency of these informations is checked on the chosen sample.

Secondly, an approach built on classical actuarial theories is used to identify the explanatory variables of sick-leave and to model frequencies. A comparison of sinistrality between the modalities of variables enables to determine if a variable explains the risk or not. Then, some frequency tables segmented per age are built per deductible and gender, then per deductible and business sector and finally per deductible, gender and socio-professional category. These tables give a tendency with age of the sinistrality for the most classical explanatory variables. Finally, a model using Cox proportionnal risk model and positionning methods is set. It enables the obtention of accurate sick-leave frequency laws.

Thirdly, an approach using CART and random forest algorithm is used. This approach, more flexible, complements the classical approach in the research of explanatory variables, it enables to identify some homogeneous groups of people facing the risk and to order the explanatory variables. This approach is also used to create a prediction model of sick-leave occurence. This model possesses important advantages.

These three parts give answers to the operationnal issues studied. They also highlight the interest of using decision trees to model sick-leave and the advantages of this approach facing classical models.

Synthèse

Dans un contexte de taux bas et de tension du marché, l'assureur doit estimer au mieux les risques qu'il garantit et se distinguer de ses concurrents. Cette étude porte sur le risque "incapacité de travail" d'ordre privé. L'assureur intervient pour maintenir la rémunération du salarié à un certain taux lorsqu'il ne peut (provisoirement) pas poursuivre son activité professionnelle. La modélisation de ce risque se fait selon deux axes, le coût (la durée de présence en incapacité et l'indemnité journalière versée) et la fréquence d'entrée dans l'état. C'est cette dernière qui est étudiée dans ce mémoire.

Pour comprendre, expliquer et modéliser de façon précise la fréquence, cette étude remplit plusieurs objectifs :

1. Déterminer les variables explicatives de la sinistralité pour différents secteurs d'activité
2. Réaliser un classement par ordre d'influence de ces variables
3. Mettre en évidence des groupes homogènes d'individus (des combinaisons de variables explicatives) ayant une sinistralité supérieure à la moyenne
4. Déterminer des lois d'entrée en incapacité
5. Comparer différentes approches de la modélisation de l'entrée en incapacité

Pour cela, le mémoire se base sur des théories actuarielles classiques et sur des arbres de décision.

L'étude est réalisée sur les données du portefeuille de Malakoff-Médéric. Pour étudier les fréquences, des tables d'effectifs et de sinistres sont utilisées.
Les individus sont affiliés dans les infocentres de Malakoff-Médéric lors de la souscription. Des tables présentant les caractéristiques de chacun des assurés sont donc disponibles. La première étape consiste à éliminer les valeurs aberrantes ou manquantes de ces données, à les mettre en forme et à créer de nouvelles variables utiles à l'étude. Des retraitements similaires sont effectués sur les données des sinistres.
Un échantillon de données composé de quatre franchises avec des effectifs conséquents est retenu. Ces franchises sont les franchises 3 jours continus, 30 jours continus, 90 jours continus et "en relai des obligations conventionnelles de l'employeur". Quatre secteurs d'activité sont présents, le secteur de la santé, le secteur de l'industrie (sera noté IM), le secteur des activités de services et scientifiques (sera noté ASES) qui correspond à une activité tertiaire et un secteur "Autres" regroupant des secteurs moins volumineux du portefeuille.
Pour permettre une étude fiable des variables explicatives de la sinistralité, des couples Secteur/-Franchise sont créés. Notre souhait était de déterminer les variables explicatives par secteur et de comparer l'influence de ces variables entre les secteurs. Une étude des fréquences par secteur, toute franchise confondue aurait été biaisée, en effet certaines modalités des variables explicatives ont souvent des garanties plus avantageuses que d'autres (et donc par exemple des franchises plus courtes). Des tests de cohérence sont ensuite appliqués aux effectifs obtenus. Ces tests se basent sur des comparaisons entre les caractéristiques des effectifs de l'échantillon retenu et les études de l'INSEE sur la population active Française.

Une fois les données retraitées, sélectionnées et validées, deux grandes approches sont utilisées pour modéliser des fréquences et déterminer les variables explicatives de la sinistralité : une approche dite "classique" et une approche par arbres de décision.

Pour déterminer les variables explicatives de la sinistralité par secteur d'activité, l'étude se base tout d'abord sur des comparaisons de fréquences selon les modalités des variables. Cette approche très simple permet, en particulier, de justifier la légitimité d'une recherche des variables explicatives et de leur influence, par secteur d'activité. Les résultats obtenus montrent que, la plupart du temps, les variables sexe, âge, collège et taille de l'entreprise ont une influence sur la valeur de la fréquence. L'influence de la région de domicile du salarié et de sa situation familiale est quant à elle plus difficile à cerner. Par ailleurs, l'influence de ces variables semble varier avec le secteur d'activité étudié. L'exemple le plus évident concerne la variable collège. Pour cette dernière, la différence entre cadres et non cadres est minime pour le secteur ASES, elle est par contre très importante pour le secteur IM.
La principale faiblesse de cette approche est que les variables sont étudiées indépendamment et sans prendre en compte les combinaisons possibles. Ces résultats serviront cependant de base à la seconde approche, plus innovante et permettront de la valider.

Des lois d'entrée en incapacité par âge segmentées selon la franchise et le secteur d'activité, puis la franchise et le sexe et enfin la franchise, le sexe et le collège sont ensuite modélisées. Pour cela, un estimateur poissonien est retenu ainsi que des lissages de Whittaker Henderson. Ces lois d'entrée en incapacité ne se basent sur aucun modèle d'hétérogénéité. Le portefeuille est segmenté en amont selon les modalités des variables explicatives. Les courbes obtenues permettent de donner la tendance de la sinistralité avec l'âge pour des variables très classiques. Ces constatations sont prises en compte lors des autres modélisations et viennent nous rassurer quant à la qualité des autres modèles. Les fréquences obtenues pour les femmes et la franchise 90 jours sont par exemple les suivantes :

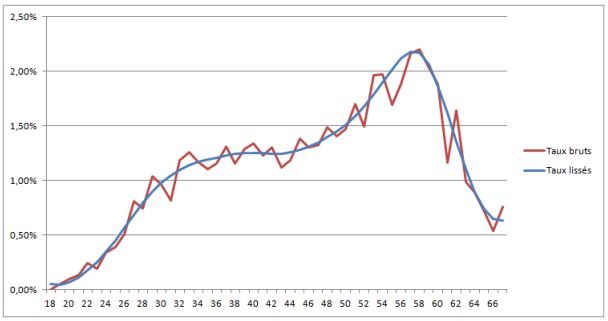


FIGURE 1: Taux bruts et lissés femme, franchise 90 jours

Cette modélisation présente cependant une faiblesse importante, elle ne tient pas compte des combinaisons de variables explicatives. Pour remédier à cela, un modèle utilisant le modèle de Cox et des positionnements de populations est paramétré.

La démarche entreprise permet d'obtenir, pour chaque franchise, des coefficients correcteurs propres à chaque secteur d'activité et des lois par âge segmentées selon le secteur d'activité, le collège, le sexe et la taille de l'entreprise. Comme précédemment, ce modèle est réalisé pour les franchises 3, 30 et 90 jours continues ainsi que pour la franchise en relai. La table de maintien en incapacité du BCAC 2014 permet ensuite de généraliser ces modèles à des franchises plus longues à

partir des résultats des franchises 30 et 90 jours. Enfin une table de correspondance entre franchises continues et discontinues permet de se ramener à ce second type de franchise. Cette généralisation est utilisée pour tous les modèles de l'étude.
L'utilisation du modèle de Cox et de positionnements est cependant critiquable. De plus, les *backtesting* ne sont pas toujours satisfaisants. La principale critique est la non prise en compte d'effets "complexes" lors de la modélisation. A l'aide du modèle de Cox, un coefficient correcteur pour les cadres est par exemple obtenu. Ce coefficient sera appliqué à l'ensemble des tailles d'entreprise, l'ensemble des âges etc... ce que nous ne souhaitons pas. Pour remédier à cela, une nouvelle approche est mise en place, l'approche par "arbres de décision". Cette approche est plus récente, pour la valider, les résultats mis en évidence par les modèles "classiques" seront utilisés.

La modélisation par arbres de décision est tout d'abord utilisée pour compléter les résultats de l'approche classique concernant les variables explicatives de la sinistralité. Des arbres sont construits dans un premier temps à l'aide de l'algorithme CART. Cet algorithme a pour but de séparer un échantillon en plusieurs échantillons constitués d'individus qui se comportent de façon similaire face à une variable réponse. Les individus sont ici caractérisés à l'aide des variables explicitées précédemment et la variable réponse correspond au nombre d'entrées en incapacité au cours d'une année d'âge.
Cet algorithme permet de déterminer des groupes homogènes d'individus face à la sinistralité. Cependant, la robustesse des arbres obtenus est souvent critiquée. En effet, si l'on modifie quelque peu l'échantillon qui permet de construire l'arbre, un arbre différent (en particulier dans les noeuds bas) peut être obtenu. Face à ces constatations, l'algorithme *Random Forest* qui consiste à moyenner les prédictions obtenues par une forêt d'arbres CART est utilisé. A l'aide de calculs de gain d'homogénéité dû aux différentes variables explicatives, un classement robuste, par ordre d'influence, des variables est mis en évidence pour chaque secteur. Ces classements sont différents, cela confirme la nécessité d'isoler les différents secteurs d'activité lors de l'étude de la fréquence. En outre, deux variables se mettent en évidence : l'âge et la taille de l'entreprise cliente.

Ces algorithmes permettent aussi de modéliser des fréquences d'entrée en incapacité par franchise. Cette modélisation est très souple, elle ne nécessite aucune hypothèse (pas d'hypothèse de distribution de loi, de segmentation...) et permet d'obtenir des fréquences segmentées de façon fine et complexe. Cette segmentation est réalisée avec les variables "âge", "sexe", "collège", "taille d'entreprise" et "secteur". Les variables "grande région" et "situation familiale" n'améliorent pas la modélisation. Les différents *backtesting* effectués montrent que le modèle obtenu à l'aide des forêts aléatoires est meilleur que le modèle implémenté précédemment (Cox et positionnements). Cependant, les fréquences obtenues sont constantes sur certains segments d'âge (l'algorithme regroupe plusieurs âges). Cette constatation et des contraintes pratiques nous orientent vers une dernière approche qui améliore encore la modélisation : l'approche "hybride".

L'approche hybride consiste à utiliser un arbre de décision pour segmenter le portefeuille puis à modéliser des lois d'entrée par âge dans les feuilles obtenues à l'aide de l'estimateur poissonien et d'un lissage de Whittaker-Henderson. Ainsi, les variables "sexe", "collège", "taille d'entreprise" et "secteur" sont rentrées dans l'algorithme. L'âge est conservé à part. Une contrainte de volume des feuilles est imposée pour permettre la modélisation d'une loi d'entrée par âge.
Le portefeuille est donc segmenté sans contrainte, à base d'un critère simple et des effets complexes sont mis en évidence à l'aide de l'algorithme CART. Des lois sont ensuite modélisées pour des populations homogènes. Ce modèle est paramétré pour les franchises 30 jours continus, 90 jours continus et "en relai", puis il est adapté aux autres franchises continues et aux franchises discontinues de la

même façon que précédemment. Voici par exemple les lois obtenues pour l'arbre de segmentation de la franchise 90 jours et les feuilles contenant une certaine population féminine :

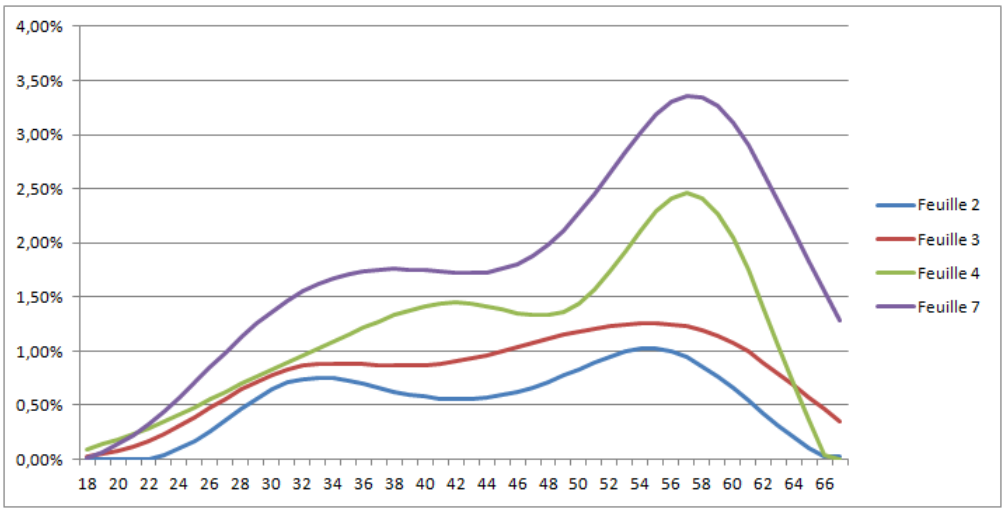


FIGURE 2: Lois d'entrée en incapacité femme, franchise 90 jours

C'est ici qu'il est intéressant de disposer des résultats des modélisations précédentes : la ressemblance des courbes obtenues par cette méthodes avec celles des modélisations précédentes est rassurante et valide cette nouvelle méthode.

Les *backtestings* montrent que cette approche hybride est la meilleure, c'est donc celle-ci qui est utilisée pour prédire les arrêts d'un segment de population et alimente l'outil créé parallèlement à ce mémoire.

L'ensemble des étapes mises en évidence précédemment forme une approche peu commune. Les arbres de décision sont utilisés pour mieux comprendre le risque et isoler des populations homogènes face à la sinistralité. Les résultats obtenus fournissent des outils d'aide à la décision.
Des fréquences sont ensuite modélisées par franchise et pour des groupes homogènes. Ce modèle est inclus dans un outil qui permet de placer la fréquence d'une entité (une CCN, un groupe, un SIREN) sur un benchmark en tenant compte de sa population (ce que ne ferait pas une simple comparaison de fréquence à une fréquence moyenne) et de prédire sa sinistralité future.

Table des matières

Introduction

Le marché français de la prévoyance collective connait depuis plusieurs années de grands bouleversements. Depuis 2010, différentes réformes législatives et réglementaires ont participé à la tension de ce secteur déjà très concurrentiel. On peut citer notamment, le report de l'âge minimum de liquidation de la retraite à 62 ans et l'ANI de 2013 qui ont augmenté la durée de couverture du salarié par l'assureur, mais aussi la décision du conseil constitutionnel du 13 juin 2013, la mise en place des contrats responsables et la généralisation de la couverture santé obligatoire qui ont accentué la concurrence sur le marché de la prévoyance.

Dans ce contexte, le groupe Malakoff Médéric développe depuis quelques années une politique ambitieuse de gestion du risque par la mise en place d'actions d'information et de prévention du risque. Ces actions font parties de la stratégie de différentiation du groupe et permettent de fidéliser les clients, de maitriser les coûts pour les entreprises mais aussi de maitriser le résultat technique pour l'assureur.
Au quotidien ces actions se traduisent par des mesures d'accompagnement et de prévention des salariés, on peut citer :

- La prévention des troubles musculo-squelettiques
- La prévention face au tabagisme
- L'accompagnement à la reprise du travail
- Le coaching sur la nutrition, le stress, l'activité physique...

Ces actions sont davantage efficaces si elles sont ciblées. La présence d'une fréquence d'entrée en incapacité élevée est un premier indicateur pour comprendre et mettre en place des actions adéquates en concertation avec l'entreprise. Comment distinguer des fréquences anormales ? Un résultat technique bon ou mauvais peut-il nous permettre de les identifier ? La réponse à cette question est non. D'une part, parce que le résultat technique est calculé trop tard, d'autre part, car il dépend de nombreux facteurs techniques et commerciaux. Il est donc nécessaire de modéliser directement la fréquence.

Le but de ce mémoire est de trouver une méthode scientifique pour identifier et prédire les fréquences anormales d'entrée en incapacité de travail.

Le modèle retenu sera utilisé pour prédire les arrêts de travail d'un segment de portefeuille (une industrie, un compte...) et anticiper une dérive de la sinistralité, mais aussi pour situer une entreprise cliente sur un benchmark, pour anticiper les actions d'accompagnement du portefeuille et pour vérifier que les engagements pris ou à prendre sont bien réalistes face à un facteur de risque. Il permettra de comprendre, d'évaluer, d'expliquer et de corriger le risque.

Le modèle s'appuiera sur des techniques actuarielles classiques et sur une approche plus souple et récente : l'utilisation d'arbres de décision.

I - Enjeux et contexte du sujet

I.1 - Aspects réglementaires et législatifs

Depuis 2010, le gouvernement a mis en place une succession de réformes liées à la maitrise des dépenses publiques, à l'amélioration de la compétitivité des entreprises, à la sécurisation de l'emploi et à la lutte contre le renoncement aux soins. Certaines de ces réformes impactent de façon importante et durable la prévoyance collective en France.

I.1.a) La réforme des retraites de 2010

Promulguée au journal officiel le 10 novembre 2010, cette réforme est mise en place dans le but de pérenniser les régimes de retraite obligatoires. En effet, depuis maintenant de longues années, le nombre d'actifs de ses régimes devient de plus en plus faible face au nombre de retraités. Si l'évolution de la démographie se poursuit, il y aura en France 1.5 actifs pour 1 retraité en 2030 contre 2.5 actifs pour 1 retraité en 1990. Le point majeur de cette réforme est le report de l'âge de la retraite de 60 à 62 ans. Le relèvement se fait progressivement sur 6 ans à hauteur de 4 mois par an et varie selon l'année de naissance. Il s'accompagne d'une augmentation de la valeur d'achat des points, de dérogations sur les carrières longues ou encore d'un bonus, si le salarié poursuit son activité professionnelle après 64 ans.

Cette réforme a un impact très important sur la prévoyance collective en France. En effet, le report de l'âge de la retraite allonge la durée de couverture des salariés pour l'assureur. Cet allongement impacte les provisions mathématiques de l'organisme d'assurance et la tarification des contrats. Il pose aussi de nombreuses questions sur le niveau de sinistralité des salariés au-delà de 60 ans.

Cette réforme a été complétée voire modifiée depuis, que ce soit par les lois de financement de la sécurité sociale ou par la loi « Hollande » du 20 janvier 2015, mais l'idée originelle reste la même : allonger la durée de travail des salariés pour pérenniser les régimes de retraite obligatoires.

I.1.b) La décision du conseil constitutionnel du 13 juin 2013

Le 13 juin 2013, une décision du conseil constitutionnel vient interdire les désignations des Conventions Collectives Nationales. Les institutions de prévoyance (comme Malakoff-Médéric) ne sont ainsi plus désignées lors des accords collectifs, mais simplement recommandées. Les compagnies d'assurance auparavant écartées de ce marché peuvent maintenant l'attaquer avec des moyens importants. Ces désignations étant à l'origine faites pour 5 ans, la majorité des contrats de prévoyance des CCN sont ou vont être renégociés dans les prochaines années. Pour rappel, plus de 255 CCN existent en France pour la prévoyance.

Cette décision rend le marché de la prévoyance encore plus tendu qu'il ne l'était auparavant. L'arrivée de nouveaux acteurs oblige en particulier les Institutions de prévoyance à être encore plus compétitives sur leurs tarifs et leurs garanties.

I.1.c) La loi de sécurisation de l'emploi du 14 juin 2013

En 2008, les syndicats du patronat et des salariés signent un accord interprofessionnel qui rend obligatoire la portabilité des droits. Ainsi, un salarié licencié pouvait continuer à bénéficier de ses

garanties collectives de santé et de prévoyance durant 9 mois, à condition qu'il fasse partie d'une entreprise adhérente à un des syndicats signataires. Cette portabilité était financée conjointement par le salarié et l'employeur.

La loi de sécurisation de l'emploi du 14 juin 2013 vient renforcer les mesures de l'ANI de 2008 et transpose les mesures de l'ANI du 13 janvier 2013. Ainsi, un salarié licencié (sauf en cas de faute lourde) a maintenant droit à un maintien de ses garanties de santé depuis le 1er juin 2014 et prévoyance depuis le 1er juin 2015, dans une limite de 12 mois. Cette portabilité est valable pour tous les salariés du privé et est mutualisée. C'est une nouvelle mesure de protection des employés.

De la même façon que le report de l'âge de la retraite, cet accord augmente la durée de couverture de l'assureur et par conséquent, les provisions mathématiques et les tarifs en vigueur.

I.1.d) La mise en place des contrats responsables au 1er janvier 2015

Le plan de financement de la sécurité sociale de 2014 refonde profondément la notion de « contrat responsable » et en livre un cahier des charges. Ainsi, depuis le 1er janvier 2015, l'ensemble des contrats d'assurance complémentaire « frais médicaux », individuels ou collectifs, obligatoires ou facultatifs, sont soumis à de nouvelles règles.

Cette nouvelle réglementation impose un panier minimum de soins et des limites de garanties strictes aux contrats santé. La commercialisation et l'achat de contrats non responsables feraient perdre des avantages fiscaux aux assureurs et aux salariés.

Tout ceci n'impacte pas directement le risque « prévoyance », mais il influe sur la stratégie de groupes présents sur les deux segments : la santé et la prévoyance, comme Malakoff-Médéric.

I.1.e) Les accords AGIRC-ARRCO du 30 Octobre 2015

Au moment de l'écriture de ce mémoire, les partenaires sociaux ont conclu un accord très important pour permettre de maintenir l'équilibre des systèmes de retraite complémentaire ARRCO et AGIRC. Cet accord prévoit une fusion des deux régimes en 2019 ainsi que des modifications sur les différents leviers disponibles : le niveau des pensions, les cotisations et l'âge de départ à la retraite. Cet accord encourage en particulier les salariés à travailler plus longtemps.

Comme pour les mesures de l'ANI de 2013, si les salariés décident de travailler plus longtemps, leur durée de couverture est allongée pour l'assureur qui fera face à un risque inconnu : le comportement des seniors face aux arrêts de travail.

I.1.f) La généralisation de la couverture santé obligatoire

Depuis le 1er janvier 2016, les entreprises privées doivent proposer une complémentaire santé à leurs employés. Cette mesure, vouée à offrir une meilleure protection aux salariés, est une conséquence de l'ANI de 2013. Elle s'applique à toutes les entreprises, y compris les TPE/PME, jusqu'à présent peu couvertes collectivement. De nouveaux types de contrats ont ainsi vu le jour.

Cette mesure laisse maintenant place aux négociations pour une généralisation des couvertures

de prévoyance. Cette couverture collective obligatoire viendrait révolutionner le marché de la protection sociale. En effet, actuellement seuls 12 millions de Français sont couverts en prévoyance complémentaire.
Une telle généralisation amènerait les assureurs à retravailler leur vision du risque prévoyance et à se poser de nombreuses questions sur la sinistralité des TPE/PME, en particulier par rapport au risque arrêt de travail.

Toutes ces réformes ont provoqué une tension importante du marché de la prévoyance collective en France. Dans un contexte financier difficile (faibles taux d'intérêts), il est devenu de plus en plus crucial pour l'assureur de mesurer au mieux ses engagements. Il est en effet de plus en plus difficile de rattraper un déficit technique par des produits financiers. En outre, la nouvelle réglementation européenne tant attendue Solvabilité II est entrée en application le 1er janvier 2016. Elle apporte de nombreuses obligations à l'assureur quant à la connaissance et à la gestion de ses risques.

I.2 - Le risque incapacité de travail

I.2.a) Obligations réglementaires du risque incapacité de travail

Plusieurs textes régissent l'indemnisation du risque d'incapacité de travail par les assureurs.

La CCN des cadres de 1947 :

Par ce texte, tout employeur est contraint de cotiser à un régime de prévoyance pour ses salariés cadres. Cette participation patronale s'élève au minimum à 1,5% de la tranche A du salaire dont 0,75% est réservé à la couverture décès. Elle est réservée aux salariés cadres visés par les articles 4 et 4bis de la CCN et facultativement aux salariés visés par l'article 36.

La loi Evin du 31 décembre 1989 :

Cette loi permet une harmonisation des régimes de prévoyance et renforce la protection des salariés. Les mesures phares de cette loi sont :

- L'instauration du caractère collectif de la souscription du contrat de prévoyance collective (Art.2)
- La prévision des sorties de groupe (passage en invalidité, retraite, décès...)
- Le maintien des prestations dues en cas de résiliation du contrat (Art.7)
- Le maintien des garanties décès au niveau atteint pour les assurés en arrêt de travail lors de la résiliation du contrat (Art.7 bis)
- L'obligation pour l'organisme assureur de fournir un compte de résultat avant le 31 août de chaque année (Art.15)

La loi du 8 août 1994 :

Cette loi renforce les mesures de la loi EVIN et contraint les employeurs à organiser la revalorisation des prestations en cas de résiliation du contrat. En général, c'est le nouvel assureur qui prend à sa charge ces revalorisations contre une surprime.

L'accord national interprofessionnel (ANI) :

Il impose à l'employeur d'organiser le maintien des garanties prévoyance et santé aux salariés licenciés (sauf en cas de faute lourde).

I.2.b) Les prestations de la sécurité sociale en cas d'incapacité

En cas d'incapacité de travail du salarié, la sécurité sociale prend à sa charge une partie de son salaire sous forme d'indemnités journalières.

Si l'absence du salarié est due à des raisons personnelles, les indemnités sont limitées à 50% de la tranche A. Elles débutent 4 jours après l'entrée en incapacité (délai de carence de 3 jours) et sont payées tout au long de l'incapacité (par conséquent, dans une limite de 3 ans).

Si l'absence fait suite à un accident de travail ou à une maladie professionnelle, la franchise est nulle, les indemnités sont de 60% du salaire durant les 28 premiers jours puis de 80%.

I.2.c) Les prestations de l'employeur en cas d'incapacité

De son côté, depuis l'accord de mensualisation de 1978, l'employeur est lui aussi obligé de maintenir une partie du revenu du salarié en arrêt. Ce pourcentage du salaire vient en complément des indemnités de la sécurité sociale et est versé suite à un délai de carence de 7 jours si l'ancienneté du salarié est supérieure à 1 an. Il vise à maintenir 90% du salaire les 30 premiers jours de l'arrêt puis à 66.66% les 30 jours suivants. Plus le salarié est ancien dans l'entreprise, plus la durée de maintien est grande. On peut retrouver les différents niveaux de maintien dans le tableau suivant :

Ancienneté	Maintien à 90%	Maintien à 66.66%
$<$1 an	-	-
1-5 ans	30 jours	30 jours
6-10 ans	40 jours	40 jours
...	...	...
$>$31 ans	90 jours	90 jours

Enfin, le salarié est couvert par des prestations complémentaires issues de l'organisme assureur de son entreprise.

I.2.d) Les prestations complémentaires

L'assureur de l'entreprise va verser au salarié en arrêt des indemnités journalières. Il se fie au statut (incapable, invalide) donné par la sécurité sociale à l'individu. Dans les contrats collectifs, les

indemnités versées sont en pourcentage du salaire et sont déduites des prestations de sécurité sociale. La mensualisation peut, quant à elle, être gérée par l'assureur en cas d'accord avec l'entreprise.

Les contrats collectifs comportent des franchises qui peuvent être continues, discontinues ou en relais des obligations conventionnelles et légales de l'employeur.

Dans le cas des franchises continues, l'assureur verse des indemnités après une durée fixe d'arrêt et cela à chaque arrêt de l'assuré. Il est possible qu'une clause de "rechute" existe au contrat. Dans ce cas, si l'assuré retombe en incapacité avant un certain délai et pour les mêmes causes, aucune franchise ne lui sera appliquée.

Dans le cas des franchises discontinues, l'assureur verse des indemnités après une durée d'arrêt cumulée sur une année civile ou une année glissante. Tous les arrêts de l'assuré sont comptabilisés et leurs durées additionnées.

Dans le cas des franchises "en relais", l'assureur verse des indemnités lorsque le niveau de maintien de salaire du salarié issu de la sécurité sociale ou des obligations légales ou conventionnelles de l'employeur descend en dessous d'un certain niveau.

I.2.e) La tarification des contrats d'incapacité

Dans le cadre d'une approche fréquence-coût, la prime pure ligne à ligne de l'incapacité se calcule comme suit :

$$P = q_x(f) \times IJ \times \sum_{i=f}^{36} \mu^{\frac{i-f}{12}} p_{if}$$

f représente la franchise du contrat.
μ représente le facteur d'actualisation.
$q_x(f)$ est la probabilité que l'assuré ait un arrêt de durée supérieure à la franchise.
p_{if} est la probabilité que la durée de l'arrêt dépasse i sachant qu'elle dépasse f.

Pour réaliser un tarif concurrentiel et simulant bien la sinistralité réelle, il faut donc que la table d'entrée en incapacité donnant $q_x(f)$ et la table de maintien donnant p_{if} soient établies avec précisions.

Exemple illustratif : Une entreprise du secteur de l'industrie lourde n'aura pas la même fréquence de sinistre ni la même durée de maintien dans l'état qu'une entreprise du secteur tertiaire. Il en est de même pour une entreprise composée d'une majorité de femme et une autre d'une majorité d'hommes...

Pour déterminer q_x et $p_{i.}$, des tables d'expérience peuvent être utilisées.

I.2.f) Niveau et évolution du risque Arrêt de Travail en France

Au-delà des différents événements législatifs, réglementaires ou financiers présentés précédemment, le risque « incapacité de travail » est lui aussi en évolution.

Les résultats de la sécurité sociale pour l'année 2014 montrent une hausse du nombre d'accidents de travail et de maladies professionnelles en France après plusieurs décennies de baisses. Cette diminution du nombre d'arrêts de travail pour raisons professionnelles était principalement due à la désindustrialisation du territoire Français, mais aussi aux campagnes de sensibilisation sur les produits dangereux.

En parallèle, les dépenses de la sécurité sociale pour les arrêts maladie ne cessent d'augmenter. Le rapport parlementaire du 24 avril 2013 de la député Bérengère Poletti fait état d'une évolution « préoccupante » des dépenses de la sécurité sociale pour ces arrêts depuis les années 2000.
De son côté, la société Réhalto en collaboration avec l'institut Opinion way dévoile, dans une étude de juin 2015, qu'environ 1 salarié sur 3 a connu un arrêt de travail en 2014. Cependant, la majorité de ces arrêts sont des arrêts courts. Ils sont donc souvent absents des bases de données des assureurs en raison des franchises existantes. Seuls 3% de ces arrêts dépasseraient les 3 mois selon la filiale du groupe SCOR.

Une étude de Février 2013 du DARES (la Direction de l'Animation de la Recherche, des Etudes et des Statistiques), organe du ministère du travail, de l'emploi, de la formation professionnelle et du dialogue social, éclaire de son côté sur les caractéristiques et les origines des arrêts de travail pour raison de santé en France. L'étude se déroule sur la période 2003-2011.
Selon cette étude, le taux d'incidence de l'arrêt maladie dépend de variables « socio-démographiques ».

Tout d'abord l'âge : pour les salariés de moins de 25 ans, 2,9% des salariés sont absents au moins un jour durant une semaine de référence, pour raisons médicales. Ce taux atteint les 5.4% pour les 55-59 ans. On remarque cependant une diminution de ce taux pour les 60-64 ans (4,7%). Ce résultat bien que peu intuitif, pourrait s'expliquer par le fait que les personnes prolongeant leur activité professionnelle au-delà des 60 ans seraient en meilleure santé que ceux s'arrêtant.

Ensuite le sexe : l'absentéisme pour raison de santé personnelle ou de celle de ses enfants est à tous les âges plus important chez la femme. On atteint même un pic significatif pour la tranche d'âge 25-34 ans avec une différence de 1.5 points selon le sexe.

Puis la catégorie socioprofessionnelle : les cadres sont en effet moins absents que les non cadres. Ce résultat provient probablement des différences de conditions de travail entre les deux catégories. En lien avec le point précédent, l'étude fait aussi un rapprochement entre le secteur d'activité et la fréquence des arrêts. On observe, par exemple, une incidence de 2,7% dans la finance et l'assurance contre respectivement 4% et 4,6% pour les secteurs de la construction et de la santé. D'autres facteurs explicatifs moins intuitifs sont aussi mis en évidence.

C'est le cas de la composition du foyer, les mères célibataires seraient beaucoup plus absentes que les femmes célibataires sans enfant ou mariées.

C'est aussi le cas du statut et de la sécurité de l'emploi. Un employé en CDI aurait une fréquence d'arrêt maladie supérieure à celle d'un employé en CDD. On observerait les mêmes disparités entre un titulaire de la fonction publique et un employé du privé.

Enfin, le département de travail peut aussi s'avérer être un facteur explicatif. On observe en effet dans plusieurs départements une sinistralité nettement supérieure à la moyenne. C'est le cas

des Hautes-Alpes avec une fréquence de 5,31%. Les caractéristiques de la main d'oeuvre des différents départements pourraient être un facteur explicatif de ces disparités. Certains départements hébergent en effet beaucoup plus d'activités industrielles que d'autres. Le DARES a donc réalisé une nouvelle étude en annulant cet effet. Certaines disparités persistent, comme entre le Pas de Calais et l'Oise par exemple.

I.3 - Objectifs et démarche du mémoire

I.3.a) Objectifs du mémoire

Le contexte actuel de la prévoyance collective française oblige les assureurs à estimer au mieux leurs risques, à connaitre la structure de leur portefeuille et à anticiper ses dérives. Dans cette optique, cette étude va chercher à remplir les objectifs suivants :

1. Déterminer les variables explicatives de la sinistralité pour différents secteurs d'activité
2. Réaliser un classement par ordre d'influence de ces variables
3. Mettre en évidence des groupes homogènes d'individus (des combinaisons de variables explicatives) ayant une sinistralité supérieure à la moyenne
4. Déterminer des lois d'entrée en incapacité
5. Comparer différentes approches de la modélisation de l'entrée en incapacité

Pour se faire, deux approches seront utilisées puis comparées : une approche classique utilisant les outils classiques de la modélisation de lois d'expériences et une approche plus récente et plus souple basée sur l'utilisation d'arbres de décision.

I.3.b) Démarche du mémoire

Pour répondre aux différentes problématiques évoquées, l'analyse va se décomposer en quatre grandes parties.

Dans un premier temps, les données, leurs traitements et leurs validations seront étudiés. Les secteurs d'activités et les franchises retenus seront décrits.
Dans un deuxième temps, des tables d'entrée en incapacité par secteur puis par collège et sexe seront modélisées.
Dans un troisième temps, des théories actuarielles classiques permettront de déterminer des variables explicatives, des coefficients correcteurs et une modélisation de l'entrée en incapacité tenant compte de l'hétérogénéité du portefeuille. Cette partie se décomposera de la façon suivante :

1. Recherche des variables explicatives de la sinistralité pour chaque secteur d'activité
2. Détermination de coefficients correcteurs à l'intérieur des différents secteurs (modèle de Cox)
3. Modélisation d'une loi d'entrée en incapacité segmentée par sexe sur une population de référence (un secteur, une taille d'entreprise et une catégorie socio-professionnelle) pour chaque franchise retenue

4. Positionnement des populations de référence des autres secteurs par rapport à la loi du point 3).

La démarche se décompose de cette façon pour analyser au mieux la sinistralité au sein des secteurs d'activité. Il aurait été possible d'utiliser d'autres approches, par exemple l'utilisation directe d'un modèle de Cox pour chaque franchise, tout secteur confondu (détermination de la loi d'entrée sur la population de référence du modèle puis application des coefficients correcteurs déterminés). Cependant en raison de la faible quantité de données sur certains segments et de la volonté de réaliser une analyse précise de chaque secteur, cette idée a été écartée.
Dans une quatrième temps, le problème sera analysé à l'aide des arbres de décision. Les algorithmes CART et random forest seront utilisés. Cette approche permettra de déterminer les variables explicatives de la sinistralité (de façon plus précise que l'approche classique), de mettre en évidence des groupes d'individus se comportant de la même façon face au risque et de réaliser un modèle prédictif d'entrée en incapacité. Cette partie se décomposera de la façon suivante :

1. Recherche des variables explicatives de la sinistralité et de leurs importances pour chaque secteur d'activité

2. Mise en évidence de groupes homogènes face à la sinistralité pour les différents secteurs d'activité

3. Modélisation de fréquences par franchise et mesure du risque d'estimation

Enfin, dans un dernier temps, les deux approches précédentes seront comparées et une approche "hybride" contournant les faiblesses de ces dernières sera mise en évidence.

II - Les données

Cette étude porte sur les fréquences d'entrée en incapacité. Pour calculer ces taux, il est nécessaire de disposer de tables d'effectifs et de sinistres fiables. Le choix et la validation des échantillons étudiés sont présentés dans cette partie.

II.1 - Choix et traitement des variables

II.1.a) Le périmètre de l'étude

Les données utilisées proviennent des infocentres de Malakoff-Médéric. Dans un souci de fiabilité, l'étude proposée dans ce mémoire se base sur quatre années d'historique, du 01/01/2011 au 31/12/2014. Les effectifs conservés sont les individus affiliés à une garantie "incapacité de travail". Par ailleurs et pour cette même problématique, l'ensemble du portefeuille n'est pas étudié. Seule la partie du portefeuille pour laquelle la franchise est fiable est conservée.

II.1.b) Les variables retenues relatives aux effectifs

L'un des objectifs de ce mémoire est de déterminer les facteurs explicatifs (et leurs combinaisons) qui influent sur le risque incapacité. Par conséquent, les tables de données étudiées doivent disposer de ces champs. Les variables retenues sont présentées ici.

La franchise

Intuitivement, il est impossible de comparer les fréquences d'entrée en incapacité vues par l'assureur si la franchise varie selon les assurés. Par conséquent, il est nécessaire de ramener cette information dans la table des effectifs.

Plusieurs types de franchise existent : les franchises continues, discontinues et "en relais" des obligations légales ou conventionnelles d'indemnisation de l'employé par l'employeur.

Malakoff Médéric intervient ainsi parfois, suite à un nombre de jours d'arrêt fixe suivant la date de survenance, suite à un nombre de jours cumulés sur année civile ou année mobile, ou suite à la période d'indemnisation de l'employeur.

Exemple : Contrat de la CCN du Commerce de gros

Les garanties conventionnelles non-cadres sont les suivantes. Elles sont disponibles sur le site internet de Malakoff-Médéric.

	Franchise
Salarié ayant une ancienneté <1 an	60 jours continus
Salarié ayant une ancienneté >1 an	En relais des obligations conventionnelles de l'employeur

L'information sur la franchise est présente dans la "Garantie tarifée" du contrat, un contrat peut contenir plusieurs garanties tarifées (typiquement, pour certains SIREN, il est possible d'avoir un contrat et deux "garanties tarifées" sur ce contrat, une pour les cadres, l'autre pour les non-cadres).

Les contrats disposant de franchises "en relais des obligations conventionnelles" (sera noté RC) seront traités à part dans cette étude. En effet, les assurés de ces contrats vont avoir une durée de franchise propre qui dépendra de leur ancienneté dans l'entreprise. Si l'assuré n'a pas l'ancienneté suffisante pour disposer du maintien de salaire de l'employeur, une période de franchise continue ou discontinue lui est appliquée. Par ailleurs, cette franchise va dépendre des conventions collectives en vigueur, une CCN peut obliger les employeurs a versé un maintien de salaire pendant des durées supérieures aux durées légales. Des fréquences seront calculées pour cette franchise, mais elles seront à prendre avec beaucoup de précautions.

Seules les lignes pour lesquelles la franchise est fiable sont conservées. Pour vérifier que les franchises présentes ne sont pas erronées, des vérifications aléatoires à partir des contrats papiers scannés sont effectuées pour certains grands groupes.

Les franchises "en relai" seront observées à part. Les franchises continues se répartissent de la façon suivante :

Franchise continue	**Proportion**
3	8,2%
10	1,03%
30	14,7%
45	3,2%
60	2,22%
90	63,9%
120	1,09%
150	2,52%
180	1,56%

Le sexe

Les modalités de cette variable sont les suivantes dans la table initiale :

Sexe	Proportion
Homme	53,16%
Femme	45,19%
Inconnu	1,65%

Si cette information est absente, la ligne de l'assuré est supprimée.

Le collège de l'assuré

Cette variable prend les modalités suivantes :

Collège	Proportion
Cadres	28,31%
Non cadres	70,03%
Inconnu	1,66%

Deux modalités sont retenues : "Cadres" et "Non cadres". Les individus au collège inconnu sont supprimés.

La date de naissance de l'assuré

La date de naissance est globalement bien renseignée.

Date de naissance	Proportion
Présente	97,9%
Absente	2,1%

L'âge de l'assuré est calculé ensuite à partir de cette variable. Cet âge est absolument nécessaire pour l'étude des fréquences d'entrée en incapacité. Les individus dont la date de naissance est absente sont supprimés. Les individus dont cette date est "anormale" seront exclus de l'étude lors du calcul de l'âge de l'assuré.

La situation familiale de l'assuré

Les modalités de cette variable sont renseignées dans 45 % des cas :

Situation familiale	Proportion
Célibataire	12,3%
Concubin	1,4%
Divorcé	0,9%
Marié	28,6%
Pacsé	1,1%
Séparé	0,2%
Veuf	0,1%
Inconnue	55,5%

Les situations de famille "Mariés, concubins et pacsés" sont regroupées ainsi que les situations "Veufs, séparés, divorcés". Cette variable est conservée et sera étudiée. Cependant, en raison de l'absence de cette information pour une grande partie de l'échantillon, les résultats obtenus seront à prendre avec beaucoup de précaution.

Les dates d'affiliation

Les dates de début et de fin d'affiliation sont retenues. Elles correspondent soit à la date de souscription du contrat par l'entreprise, soit à celle du salarié (nouvel arrivant par exemple). Il en est de même pour la date de fin d'affiliation.

Les individus dont la date d'entrée est postérieure au 31/12/2014 sont supprimés. Les individus dont la date de sortie est antérieure au 01/01/2011 sont supprimés. La période d'étude s'étend en effet du 01/01/2011 au 31/12/2014.

Le secteur d'activité

La variable "Secteur d'activité" est ramenée au niveau du Siren à l'aide d'une autre table permettant un croisement avec le code NAF de l'entreprise. Les secteurs présents sont les suivants :

Secteur	Proportion	**Secteur**	Proportion
Activités de services administratifs	2,7%	Enseignement	1,2%
Activités financières et assurance	2%	Hébergement et restauration	22,9%
Activités immobilières	1,3%	Industrie manufacturière	14,9%
Activités spécialisées, scientifique	19,3%	Industries extractives	0,1%
Administration publique	0,6%	Information et communication	9,5%
Agriculture, sylviculture et pêche	0,1%	Production et distribution eau	0,5%
Arts, spectacles	1,1%	Production et distribution	0,1%
Autres activités de services	2,6%	Santé humaine et action sociale	4,8%
Commerce ; réparation automobiles	8,1%	Transports et entreposage	0,6%
Construction	1,1%	Inconnu	6,8%

Seuls 4 secteurs sont conservés, il est en effet souhaitable de disposer d'une quantité de données conséquente pour chaque modalité d'une variable explicative.
Les secteurs "Activités de services administratifs", "Activités spécialisées, scientifique", "Autres activités de services", "Information et communication" et "Activités financières et assurance" sont regroupés pour former le secteur "Activités de service et scientifique" (ASES).
Les secteurs retenus sans regroupement sont les secteurs "Industrie manufacturière" (IM) et "Santé humaine et action sociale". Les autres secteurs sont regroupés au sein du secteur "Autres".

Remarque : Le secteur "Hébergement et restauration" n'est pas retenu, les prestations de la quasi totalité des effectifs du secteur, sont en gestion déléguée.

Le département

Dans nos tables deux informations sont présentes, le département de l'assuré et celui de son entreprise (le département du Siren). Généralement ces deux information coïncident. Par défaut, c'est le département de vie de l'assuré qui est conservé. Si cette information est absente, le département du Siren est ramené. Cette information est alors toujours renseignée.

II.1.c) Les autres variables

Les autres variables présentes sont :

- Le nom du groupe de l'assuré (certains Siren appartiennent à un même groupe)
- Le nom de la CCN pour les contrats relevant de la CCN et non de l'entreprise
- Le numéro du contrat
- Le numéro de la garantie tarifée
- Un numéro d'identification propre à chaque assuré
- Un indicateur de gestion déléguée des prestations
- Le type de garantie (Mensualisation ou incapacité)

Ces champs seront particulièrement utiles lors du croisement de la table des sinistres et de la table des effectifs. L'indicateur de gestion déléguée permet ensuite de ne garder que les effectifs pour lesquels la gestion des prestations n'est pas déléguée, sinon les prestations n'apparaissent pas dans nos infocentres.

II.1.d) Les variables calculées

Certaines variables sont calculées à partir des variables présentées précédemment.

La grande région de l'assuré

Il est nécessaire pour cette étude de disposer d'effectifs conséquents pour chaque modalité d'une variable explicative. En conséquence, la variable "Grande région" est créée à partir du département de la table d'origine (les effectifs par département ou région n'étaient pas toujours suffisants).
Cinq grandes régions sont créées : Nord, Est, Ouest, Sud et Ile de France.

La taille de l'entreprise

La taille de l'entreprise est une moyenne par Siren du nombre d'affiliés différents sur les 4 années d'observation.

Les tailles sont ensuite regroupées en 4 "classes", ces classes sont propres aux secteurs d'activité. Ainsi, une même classe pour les secteurs ASES et Santé ne regroupera pas les mêmes tailles d'entreprises. Cela est dû à la spécificité des secteurs. Le secteur ASES est constitué de beaucoup d'entreprises de taille faible, tandis que le secteur de la santé est composé de cliniques ou hôpitaux, ainsi le nombre d'entreprises de taille inférieure à 20 salariés est très faible.

Date et âge de début d'observation

Les individus ne sont observés qu'à partir du 01/01/2011, par conséquent, la date de début d'observation est égale à :

$$\max(01/01/2011 \text{ ;Date début d'affiliation})$$

L'âge de début d'observation est ensuite calculé par différence entre la date d'anniversaire et la date de début d'observation.

Date et âge de fin d'observation

Les individus ne sont observés que jusqu'au 31/12/2014, par conséquent, la date de fin d'observation est égale à :

$$\min(31/12/2014 ; \text{Date de fin d'affiliation})$$

L'âge de fin d'observation est ensuite calculé par différence entre la date d'anniversaire et la date de fin d'observation.

L'indicateur de censure/troncature

Lors de l'étude des comportements humains, l'âge de l'assuré est une variable incontournable. L'étude des fréquences d'entrée en incapacité par âge nécessite de connaitre la durée d'étude de l'assuré à chaque âge entier.

Rappel des notions de censure et troncature :

Les données de durée utilisées pour la création d'une table d'expérience sont souvent incomplètes. Au début de l'étude, un échantillon de personnes et une période d'étude sont choisis. La durée d'étude est limitée, le résultat peut n'être que partiellement connu. C'est un problème de censure. Pour créer une table d'expérience par âge, il est nécessaire d'observer les individus entre des âges entiers. Des censures "à droite" sont présentes dans les cas suivants :

- La période d'étude se termine entre deux âges entiers pour certains assurés.
- Le contrat d'une entité est résilié. Certains assurés sont entre deux âges entiers à la date de résiliation.
- L'assuré décède entre deux âges entiers.

Par ailleurs, les informations sont récoltées sur une sous partie de $[0; +\infty[$, il n'est pas possible de savoir si une information existe avant la période d'étude. C'est un problème de Troncature. L'indicateur prend la valeur :

- 0 si ni censure ni troncature
- 1 si censure ou troncature
- 2 si censure et troncature

L'âge et la durée de présence par âge

A chaque âge, la durée d'observation de l'assuré est calculée pour chacun de ses contrats. Si l'individu n'est ni censuré, ni tronqué, sa durée de présence par âge sera égale à 1, l'assuré est observable à tout moment entre les âges x et x+1. Sinon, sa durée de présence par âge à l'âge x est égale à la différence entre la date d'anniversaire x+1 et la date de début d'observation.

Exemple illustratif :
L'assuré est né le 18/11/1966, il est affilié le 01/01/2010 et son contrat n'est à ce jour pas résilié. Sa date de début d'observation est donc le 01/01/2011. Sa date de fin d'observation est le 31/12/2014. Au 01/01/2011 l'assuré a 44 ans, il a 45 ans le 18/11/2011, il est donc observable à 44 ans sur une fraction d'année égale à :

(18/11/2011 - 01/01/2011 + 1)/365

Le même raisonnement s'applique à l'âge de sortie d'observation. Dans la suite de l'étude, l'expression "Exposition" désignera la somme des durées d'âge, elle s'apparente à un nombre d'années d'étude ou à un nombre d'individus proratisés selon leur durée de présence à chaque âge.

Dans la table finale des effectifs, chaque ligne donne les caractéristiques d'un assuré à un âge auquel il est observable. Ainsi, si un assuré est observable entre 20 et 22 ans, 3 lignes seront présentes dans la table (20 ans, 21 ans et 22 ans) pour chacun de ses contrats.

Si l'âge présent dans une ligne est inférieur à 16 ans, la ligne est supprimée, il en est de même si l'âge est supérieur à 75 ans. Ces âges sont des âges limites, il est fait l'hypothèse que les âges en dehors du segment [16 ;75] sont issus d'erreurs dans nos bases de données. Ce filtre entraine la suppression de 48 lignes seulement.

Remarque : Pour simplifier la gestion des données, les individus en arrêt sont toujours exposés au risque. Cette simplification est étudiée. Elle entraine une sous-estimation moyenne de la fréquence d'entrée en incapacité de l'ordre de 0,01%. Elle est considérée comme acceptable.

II.2 - Choix de l'échantillon étudié

Avant tout traitement et sélection des données, la table des affiliations tout risque confondu contenait plus de 17 millions de lignes. Après retraitements et sélections des données(sur l'âge, l'indicateur de gestion déléguée...), la table étudiée contient 5 869 613 lignes. 1 387 787 assurés différents sont observés pour la période d'étude. La durée de présence totale des assurés pendant la période d'étude est de 2 983 633 années.

Suite à ces retraitements, il est maintenant nécessaire de choisir l'échantillon étudié. **L'ensemble de l'étude reposera sur quatre secteurs d'activité et 4 franchises. Une généralisation des résultats à tous les types et toutes les durées de franchise sera possible par la suite à l'aide d'une loi de maintien et de coefficients correcteurs.**

II.2.a) Le secteur d'activité

Quatre secteurs d'activité sont conservés, les effectifs par secteur se décomposent en terme d'exposition totale de la façon suivante :

Secteur	Exposition	Proportion
Activités de services et scientifiques	1 250 886	41,9%
Industrie manufacturière	920 382	30,8%
Santé humaine et action sociale	161 436	5,4%
Autres secteurs	650 929	21,8%

Les secteurs d'activité les plus représentatifs de notre étude sont : les activités de services spécialisés scientifiques et l'industrie manufacturière. Ils représentent à eux deux 73% de l'exposition totale. Cela est dû en particulier à la présence dans le périmètre de l'étude des CCN des bureaux d'études techniques, de l'habillement et de l'industrie textile.

II.2.b) La franchise

Comme il a été évoqué précédemment, la franchise est une variable très importante de cette étude. Les franchises continues et "en relai" sont conservées pour l'étude. Les effectifs par type de franchise se répartissent comme suit :

Franchise	**Exposition**
Continue	1 993 729
"En relai"	989 898

Trois franchises continues ont des expositions conséquentes et une sinistralité généralisable :

Franchise continue (en jours)	**Exposition**
3	120 431
30	475 941
90	1 179 946

La franchise continue 75 jours possède aussi une exposition importante, cependant, la quasi-totalité de cette exposition repose sur une même entreprise. Un modèle propre à cette franchise ne serait donc pas généralisable.

Les effectifs décrits à partir de maintenant sont ceux ayant une franchise continue de 3, 30 ou 90 jours ou une franchise "En relai".

Pour le secteur de la santé, les franchises se répartissent comme suit :

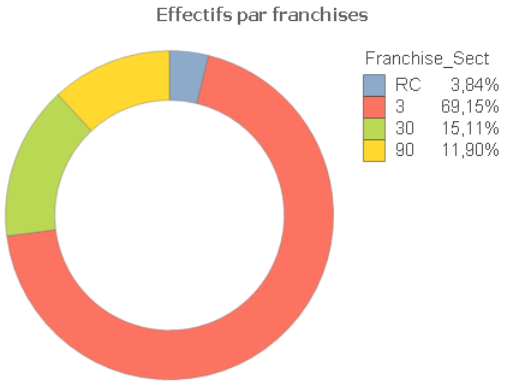


FIGURE II.3: Répartition des effectifs par franchise, secteur de la santé

Pour le secteur des "activités de services et scientifiques", les franchises se répartissent comme suit :

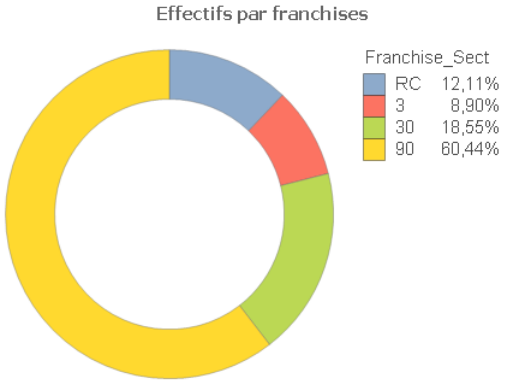


FIGURE II.4: Répartition des effectifs par franchise, secteur ASES

Pour le secteur de l'industrie, les franchises se répartissent comme suit :

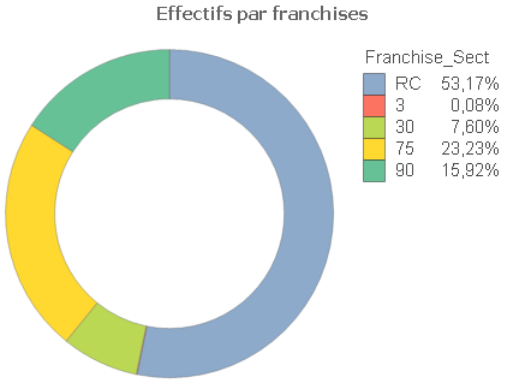


FIGURE II.5: Répartition des effectifs par franchise, secteur IM

Pour les autres secteurs, les franchises se répartissent comme suit :

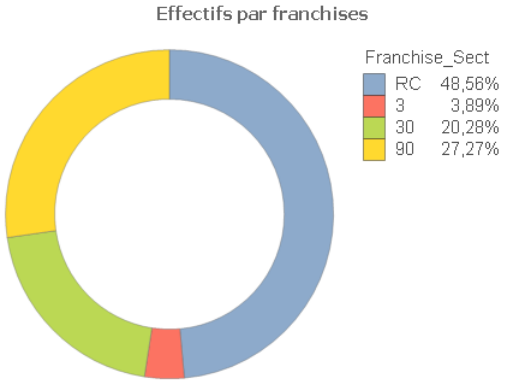


FIGURE II.6: Répartition des effectifs par franchise, secteur AUTRES

La franchise et le secteur d'activité sont deux variables très dépendantes. Cela se voit particulièrement sur le secteur de la santé (où 70% de l'effectif étudié a une franchise continue 3 jours) et du secteur de l'industrie manufacturière (où 83% de l'effectif a une franchise "en relai").

Au vu de ces différents résultats, l'étude proposée va s'appuyer sur des couples "Secteur-Franchise" pour approcher au mieux les caractéristiques de chacun des secteurs :

- Le secteur de la santé sera étudié avec la franchise 3 jours
- Le secteur "activités de services et scientifiques", qui regroupe des activités tertiaires, sera étudié avec la franchise 90 jours
- Le secteur de l'industrie sera étudié avec la franchise "En relai des obligations légales ou conventionnelles de l'employeur"
- Les autres secteurs seront étudiés avec la franchise "En relai des obligations légales ou conventionnelles de l'employeur"

Ces couples permettront de déterminer des coefficients correcteurs propres à chaque secteur pour les variables explicatives. Il n'est pas possible de considérer uniquement le secteur d'activité, toute franchise confondue. En effet, des variables explicatives comme le collège ou la taille de l'entreprise sont corrélées à ces franchises, ce qui entrainerait un biais important lors du calcul de coefficients correcteurs (les cadres auront par exemple de meilleurs contrats avec des franchises plus courtes, de même pour les grandes entreprises).

Ensuite, des lois d'entrée en incapacité seront modélisées pour chaque franchise.

Pour la franchise 3 jours, les secteurs d'activité se répartissent comme suit :

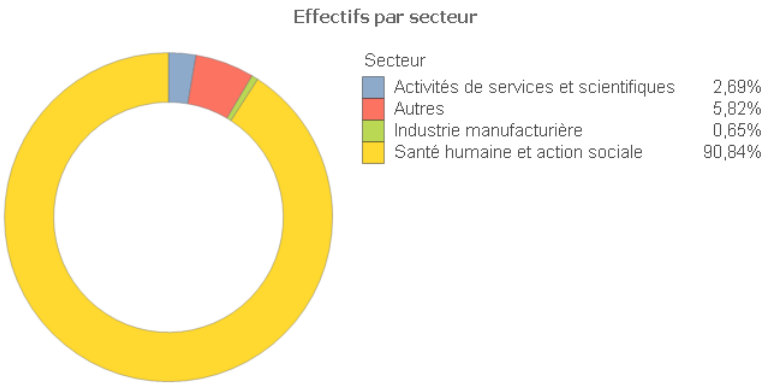


FIGURE II.7: Répartition des effectifs par secteur, franchise 3 jours

Pour la franchise 30 jours, les secteurs d'activité se répartissent comme suit :

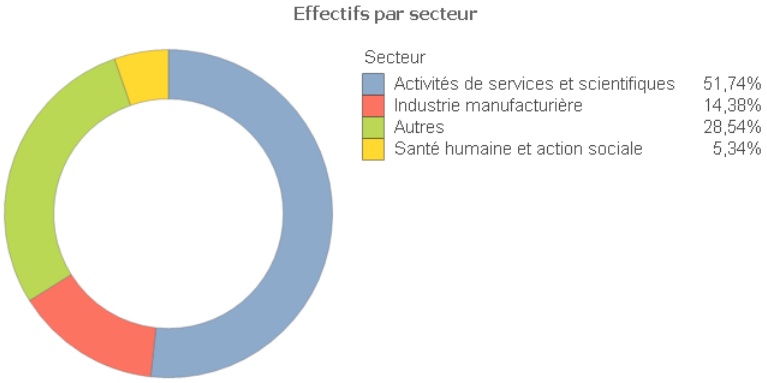


FIGURE II.8: Répartition des effectifs par secteur, franchise 30 jours

Pour la franchise 90 jours, les secteurs d'activité se répartissent comme suit :

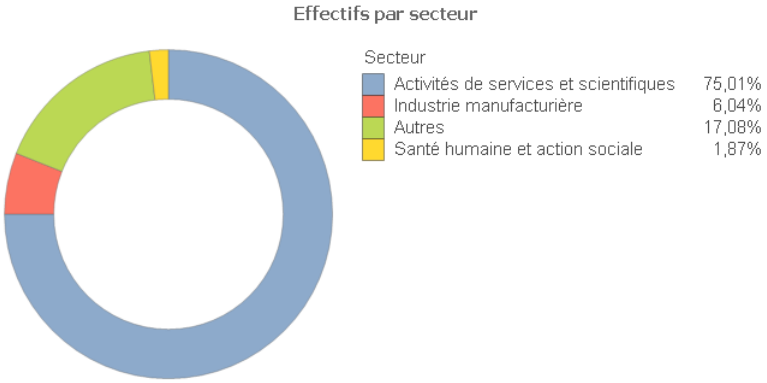


FIGURE II.9: Répartition des effectifs par secteur, franchise 90 jours

Pour la franchise "En relai", les secteurs d'activité se répartissent comme suit :

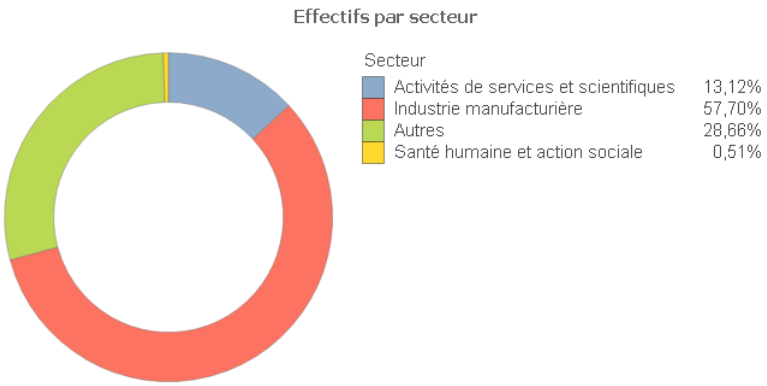


FIGURE II.10: Répartition des effectifs par secteur, franchise "en relai"

II.3 - Détails des effectifs et validation

Cette partie contient une description des effectifs des couples cités précédemment. Des comparaisons entre ces effectifs et les moyennes françaises par secteurs d'activité sont effectuées pour permettre la validation de la cohérence des données. Ces statistiques nationales sont issues d'études de l'INSEE.

II.3.a) Couple Santé/3jours

Le secteur de la **"Santé et de l'action sociale"** regroupe dans notre portefeuille un certain nombre de cliniques, maisons de retraites et associations. Les caractéristiques de ce secteur sont les suivantes :

Age et sexe

L'âge moyen de ce segment est de 39,29 ans. Cela est dans la moyenne des âges des salariés français (selon une étude de pôle emploi de 2012, elle était de 39,4 ans).

Ce secteur est composé majoritairement de femmes :

Secteur	Femme	Homme
Santé humaine et action sociale	84%	16%

L'échelle des âges par sexe est la suivante :

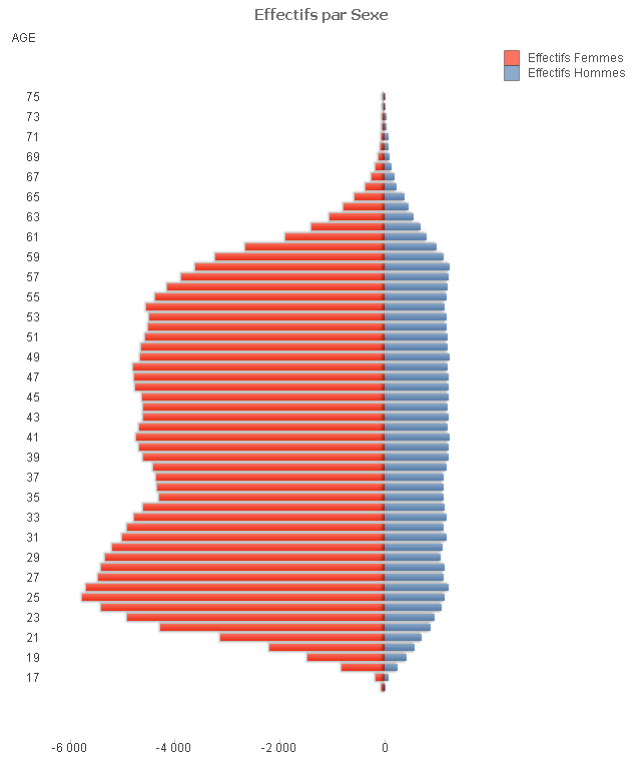


FIGURE II.11: Échelle des âges par sexe, secteur de la santé

Les jeunes femmes (aux alentours de 25 ans) sont majoritaires. La proportion homme/femme est en accord avec les statistiques nationales (les femmes représentés plus de 85% du personnel infirmier ou aide soignant en 2011).

Catégorie socio-professionnelle

Les individus de ce segment sont en très grande majorité des non cadres :

Secteur	Cadres	Non cadres
Santé humaine et action sociale	5%	95%

L'échelle des âges est la suivante :

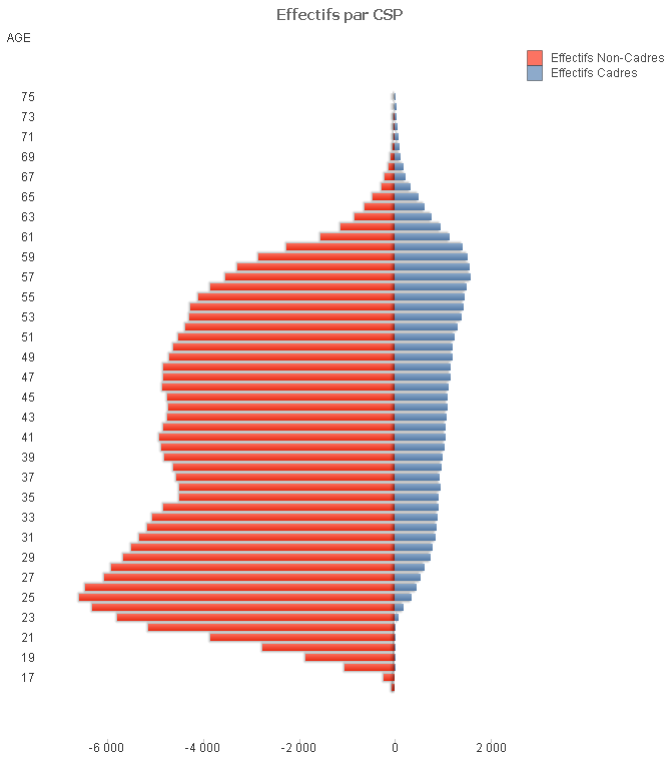


FIGURE II.12: Échelle des âges par CSP, secteur de la santé

Ce graphique est cohérent avec l'intuition, peu de jeunes salariés ont un statut de cadre.

Grande région

La majorité des assurés de ce segment se situent en île de France. Cependant les autres grandes régions ont une partie de l'effectif non négligeable.

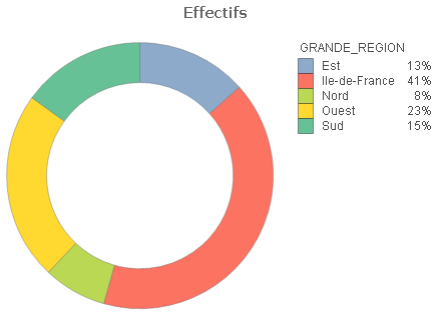


FIGURE II.13: Répartition des effectifs par région, secteur de la santé

Situation familiale

La situation familiale des individus est la suivante :

Situation Familiale	Effectif
Inconnue	39,51%
Célibataires	3,14%
Mariés, concubins ou pacsés	56,66%
Veufs, séparés, divorcés	0,69%

Pour 40% des effectifs, cette variable est malheureusement non renseignée.

La taille de l'entreprise

Quatre classes d'entreprises sont créées, de sorte que les expositions de chaque classes soient proches et assez importantes pour être exploitables.

Les classes suivantes sont obtenues :

Classe	Taille	Effectif	Nombre d'entreprises
A	< 167	26%	165
B	[167 ;438[	24%	61
C	[438 ;724[	25%	26
D	>724	26%	14

La majorité des entreprises sont de taille assez importante.

Détails des sinistres

La franchise de ce couple est de durée très faible, en conséquence, de nombreux sinistres sont visibles.

25 679 sinistres sont répertoriés. Ces sinistres se divisent selon la cause de la façon suivante :

Cause événement	Exposition
Vie privée	23 791
Vie professionnelle	1 888

Ce qu'il faut retenir du couple Santé/3jours :

- Beaucoup d'individus femmes et jeunes
- Majorité de non cadres
- Entreprises de taille assez importante

II.3.b) Couple ASES/90jours

Le secteur ASES : **"Activités de services et scientifiques"** est composé majoritairement de la CCN des bureaux d'études techniques. Les caractéristiques de ce secteur sont les suivantes :

Age et sexe

L'âge moyen de ce segment est de 36,42 ans. C'est un secteur très jeune, bien en dessous de la moyenne française. Ce résultat n'est pas incohérent avec le type d'activité. Ce dernier fait en effet appel à des techniques scientifiques modernes.

Ce secteur est composé majoritairement d'hommes :

Secteur	Femme	Homme
ASES	41,35%	58,65%

L'échelle des âges par sexe est la suivante :

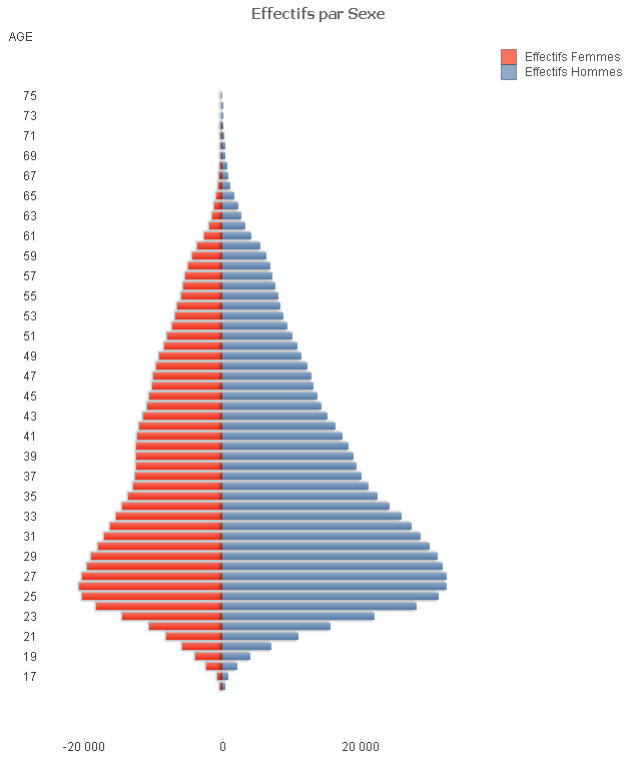


FIGURE II.14: Échelle des âges par sexe, secteur ASES

Catégorie socio-professionnelle

Les individus de ce segment sont en majorité des non cadres :

Secteur	Cadres	Non cadres
ASES	43%	57%

L'échelle des âges est la suivante :

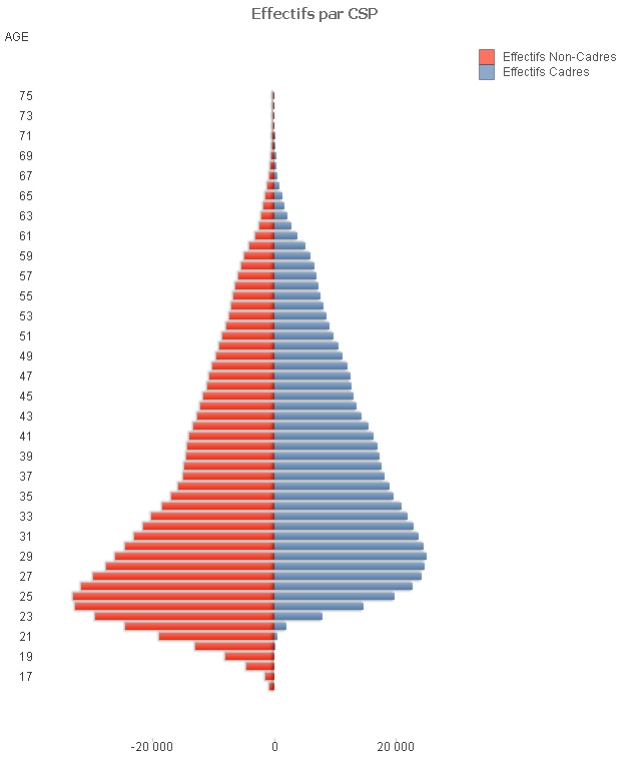


FIGURE II.15: Échelle des âges par CSP, secteur ASES

Ce graphique est à nouveau cohérent avec l'intuition, peu de jeunes salariés ont un statut de cadres. Par ailleurs, une étude plus poussée montre que 33% des femmes sont cadres. Un pourcentage identique (34,7%) était mis en évidence dans une étude sur les inégalités hommes/femmes de l'INSEE en 2013.

Grande région

La majorité des assurés de ce segment se situent en île de France. Les grandes régions Est et Ouest sont aussi fortement représentées.

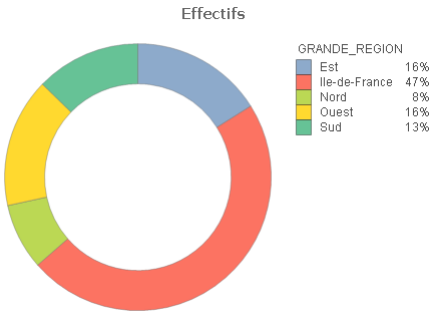


FIGURE II.16: Répartition des effectifs par région, secteur ASES

Situation familiale

La situation familiale des individus est la suivante :

Situation Familiale	Pourcentage
Inconnue	48,98%
Célibataires	18,53%
Mariés, concubins ou pacsés	31,26%
Veufs, séparés, divorcés	1,23%

La taille de l'entreprise

Les classes suivantes sont obtenues :

Code	Classe	Pourcentage	Nombre d'entreprises
A	< 8	23%	30 744
B	[8 ;27[	27%	5 908
C	[27 ;110[	26%	1 556
D	>110	24%	212

La majorité des entreprises sont des TPE/PME.

Détails des sinistres

6 594 sinistres d'incapacité pure sont répertoriés. Ces sinistres se divisent selon la cause de la façon suivante :

Cause événement	Exposition
Vie privée	6 201
Vie professionnelle	393

Ce qu'il faut retenir du couple ASES/90jours :

- Individus très jeunes
- Équilibre cadres/non cadres
- Majoritairement en Ile de France
- Entreprises de tailles faibles voir très faibles majoritaires

II.3.c) Couple IM/RC

Le secteur de l'industrie manufacturière (IM) est composé en particulier de la CCN de la métallurgie et de plusieurs grands groupes connus. Les caractéristiques de ce secteur sont les suivantes :

Age et sexe

L'âge moyen de ce segment est de 42,35 ans. Il est bien au dessus de la moyenne française. Ce résultat était attendu en raison de l'activité du secteur.

Ce secteur est composé très majoritairement d'hommes :

Secteur	Femme	Homme
Industrie manufacturière	33%	67%

L'échelle des âges par sexe est la suivante :

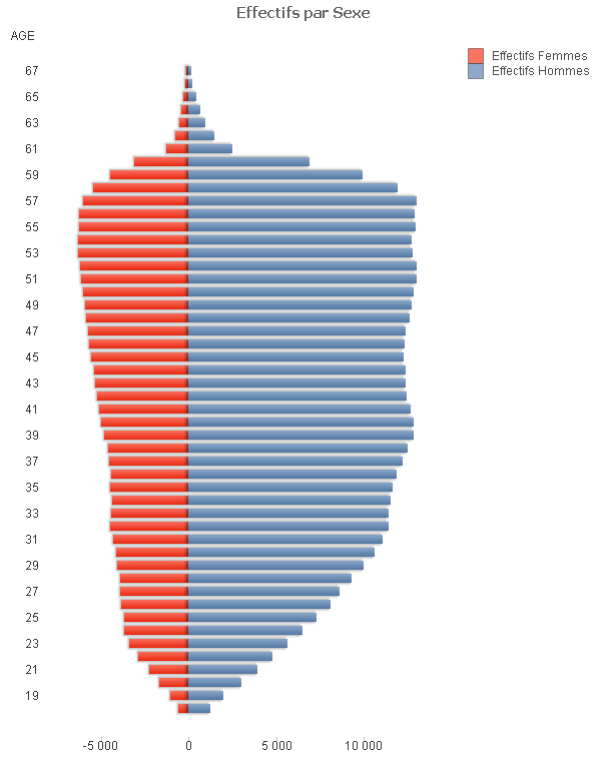


FIGURE II.17: Échelle des âges par sexe, secteur IM

Ces statistiques sont à nouveau cohérentes avec les études au niveau nationale, selon le rapport du ministère des droits des femmes de 2014, 30% des salariés de l'industrie étaient des femmes.

Catégorie socio-professionnelle

Les individus de ce segment sont en très grande majorité des non-cadres :

Secteur	Cadres	Non cadres
ASES	19%	81%

L'échelle des âges est la suivante :

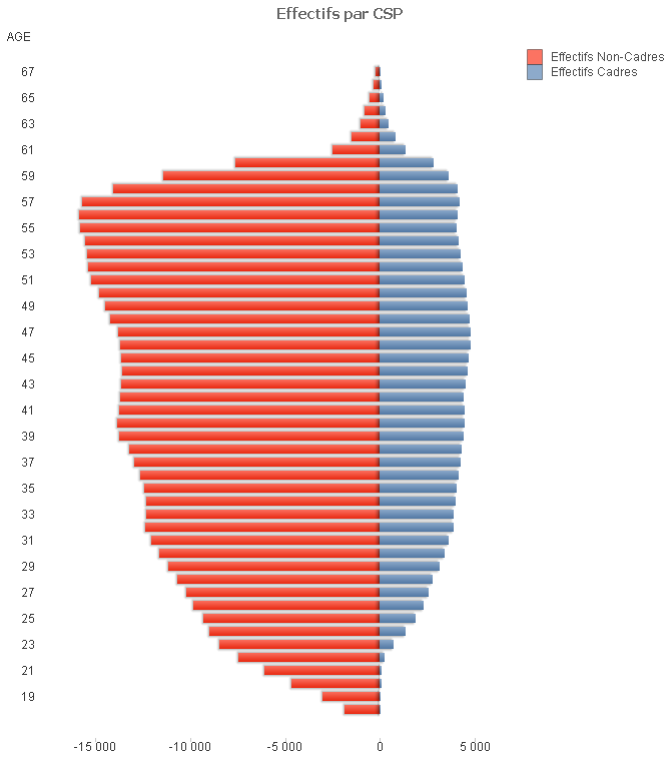


FIGURE II.18: Échelle des âges par CSP, secteur IM

Grande région

Cette variable ne peut être exploitée pour ce couple, en effet, la grande région dans la base de donnée est incorrecte pour deux grands groupes en particulier. La région renseignée est celle du siège du groupe et non les lieux des usines.

Situation familiale

La situation familiale des individus est la suivante :

Situation Familiale	Pourcentage
Inconnue	40,90%
Célibataires	13,73%
Mariés, concubins ou pacsés	49,47%
Veufs, séparés, divorcés	1,90%

La taille de l'entreprise

Les entreprises sont majoritairement de tailles importantes voire très importantes.

Code	**Classe**	**Pourcentage**	**Nombre d'entreprises**
A	< 369	25%	2 600
B	[369 ;2988[	26%	85
C	[2988 ;8516[	22%	14
D	>8516	27%	5

Les entreprises de ce secteur sont de grandes entreprises.

Détails des sinistres

8 727 sinistres d'incapacité pure sont répertoriés. Ces sinistres se divisent selon la cause de la façon suivante :

Cause événement	**Exposition**
Vie privée	7 719
Vie professionnelle	1 008

Ce qu'il faut retenir du couple IM/RC :

- Moyenne d'âge élevée
- Majorité de non cadres hommes
- Entreprises de tailles importantes

II.3.d) Couple Autres/RC

Les autres secteurs couplés à la franchise "En relai" ont les caractéristiques suivantes :

Age et sexe

L'âge moyen de ce segment est de 39,95 ans.

Ce secteur est équilibré entre hommes et femmes :

Secteur	**Femme**	**Homme**
Autres	48,15%	51,85%

L'échelle des âges par sexe est la suivante :

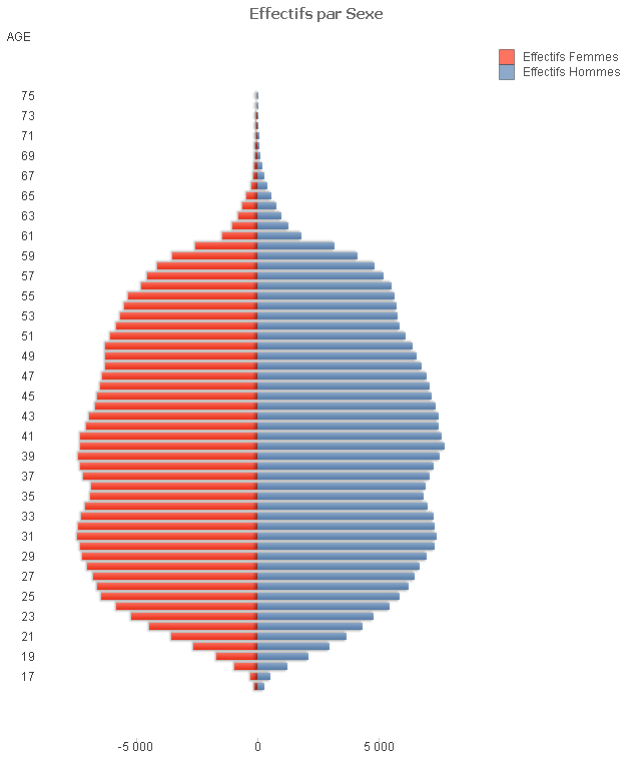


FIGURE II.19: Échelle des âges par sexe, secteur AUTRES

La proportion homme/femme est dans la moyenne nationale tout secteur confondus.

Catégorie socio-professionnelle

Les individus de ce segment sont en très grande majorité des non cadres :

Secteur	Cadres	Non cadres
AUTRES	15%	85%

Cette proportion est très déséquilibrée et peu générale. L'échelle des âges est la suivante :

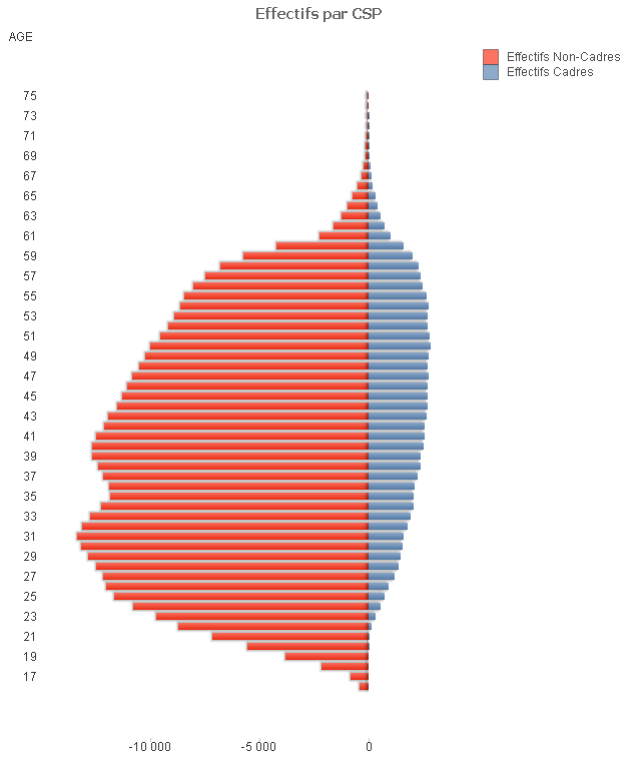


FIGURE II.20: Échelle des âges par CSP, secteur AUTRES

Grande région

La majorité des assurés de ce segment se situent en île de France. Cependant les autres grandes régions ont une partie de l'effectif non négligeable.

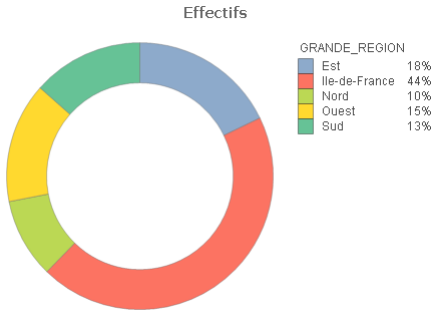


FIGURE II.21: Répartition des effectifs par région, secteur AUTRES

Situation familiale

La situation familiale des individus est la suivante :

Situation Familiale	Pourcentage
Inconnue	47,35%
Célibataires	13,01%
Mariés, concubins ou pacsés	37,36%
Veufs, séparés, divorcés	2,28%

La taille de l'entreprise

Quatre classes d'entreprises sont créées, de sorte que les expositions de chaque classes soient proches et assez importantes pour être exploitables.

Les classes suivantes sont obtenues :

Code	Classe	Pourcentage	Nombre d'entreprises
A	< 16	25%	10 300
B	[16 ;120[	25%	982
C	[120 ;2737[	25%	87
D	>2737	25%	2

De très petites entreprises représentent 1/4 de l'effectif mais de très grandes représentent la même proportion.

Détails des sinistres

5 752 sinistres d'incapacité pure sont répertoriés. Ces sinistres se divisent selon la cause de la façon suivante :

Cause événement	Exposition
Vie privée	5 513
Vie professionnelle	239

Ce qu'il faut retenir du couple AUTRES/RC :

- Age moyen de 40 ans, dans la moyenne française
- Équilibre hommes/femmes
- Grande majorité de non cadres
- Majoritairement en IDF
- La classe A est composée de très petites entreprises tandis que la classe D est composée d'entreprises de tailles très importantes

III - Approche classique, influence des variables explicatives

Dans cette partie, les couples précédemment décrits sont étudiés un à un. Leur structure est maintenant connue et leur cohérence a été validée par comparaison avec des moyennes nationales. Pour chacun d'eux, les différences de sinistralité entre les modalités des hypothétiques variables explicatives sont maintenant mises en évidence.

Rappel : La sinistralité étudiée est celle des arrêts de travail "incapacité" de vie privée.

III.1 - Le taux d'entrée en incapacité

Dans un premier temps il faut définir l'expression "taux d'entrée en incapacité".
Deux approches seront retenues dans cette étude :

- L'entrée en incapacité à un âge x est considérée comme un évènement unique. Dans ce cas, il est nécessaire de calculer les probabilités qu'un individu ait **au moins un arrêt à un âge x**, puis la probabilité qu'il ait **au moins deux arrêts** etc...
- Pour un individu d'âge x, tous ses arrêts de travail sont retenus et considérés à part entière. Le taux d'entrée en incapacité correspondra au nombre total d'arrêt de travail divisé par la somme des durées d'exposition des individus.
  C'est cette seconde approche qui est retenue dans cette partie.

III.2 - Recherche des variables explicatives de la sinistralité

La variation du taux d'entrée en incapacité en fonction des modalités des variables explicatives est observée ici. Si pour des modalités différentes, les taux sont différents, la variable explicative sera considérée comme ayant de l'influence sur l'entrée en incapacité.

En considérant le nombre d'entrée en incapacité à un âge x comme une variable aléatoire de loi de Poisson, la méthode exposée revient à comparer les moyennes de deux lois de Poisson.

III.2.a) Couple Santé/3jours

Age

Il est très intuitif de supposer que l'âge de l'assuré joue un rôle prépondérant sur la fréquence d'entrée en incapacité. Le calcul du taux d'entrée en incapacité à des âges différents semble révéler que les individus ont un taux d'entrée en incapacité croissant avec l'âge.

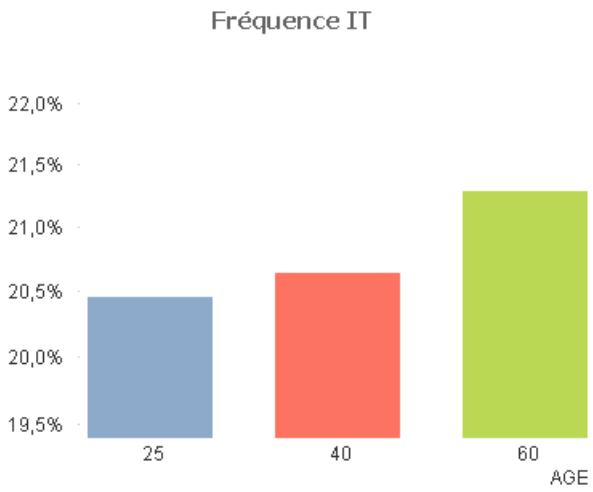


FIGURE III.22: Fréquence par âge, secteur de la santé

Sexe de l'assuré

Il existe une différence forte entre la fréquence d'entrée en incapacité des hommes et des femmes :

Fréquence Homme	Fréquence Femme
15,22%	21,79%

La catégorie socio-professionnelle

De la même façon que pour le sexe de l'assuré, la catégorie socio-professionnelle impacte la fréquence d'entrée en incapacité.

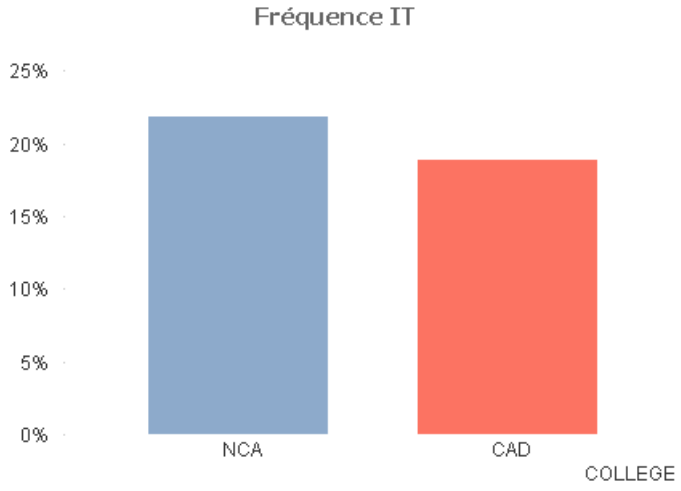


FIGURE III.23: Fréquence par CSP, secteur de la santé

Les non-cadres ont une fréquence d'entrée en incapacité supérieure à celle des cadres, mais cette différence n'est ici que de 5 points.

La situation familiale de l'assuré

Cela a été évoqué précédemment, pour la moitié de notre effectif, cette situation familiale est inconnue. Par ailleurs, elle a pu changer depuis son écriture dans les bases de données de Malakoff-Médéric. Il faut par conséquent être très prudent quant à l'étude de cette variable.
Cette fréquence se décompose comme suit :

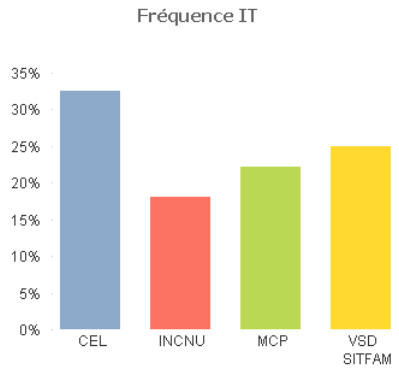


FIGURE III.24: Fréquence par situation familiale, secteur de la santé

Pour les individus dont la situation est renseignée, les mariés, concubins ou pacsés ont la fréquence la plus faible. Les célibataires ont quant à eux la fréquence la plus élevée. Ce qui n'est pas intuitif, en effet, l'intuition laisserait plutôt penser que ce sont les veufs, séparés ou divorcés qui ont la sinistralité la plus élevée.

La "Grande région" de domicile du salarié :

Dans l'étude du DARES présentée précédemment, la zone géographique est un facteur explicatif de la sinistralité. Certaines régions sont plus sinistrées que d'autres. Pour vérifier ce phénomène, la variable "Grande Région" est créée dans le but de réduire le nombre de modalité pour cette variable géographique. Les résultats sont les suivants :

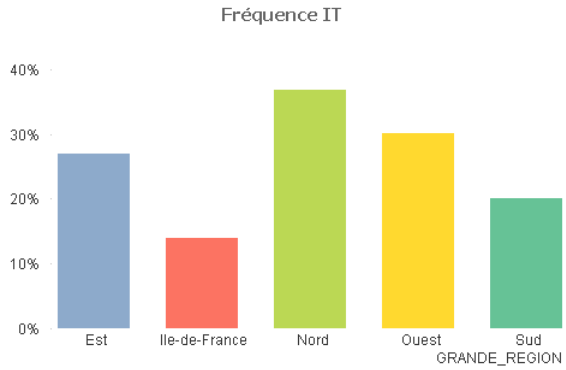


FIGURE III.25: Fréquence par grande région, secteur de la santé

La région Ile-de-France a une sinistralité bien plus faible que celle des autres "grandes régions". Pour ce couple, c'est la grande région Nord qui a la sinistralité la plus importante.

La taille de l'entreprise cliente

Cette variable est pertinente dans cette étude. En assurance collective, il est important de déterminer quel type d'entreprise présente le meilleur profil de risque, il est alors légitime de se demander si la taille de l'entreprise cliente joue un rôle sur la fréquence d'entrée en incapacité d'un salarié.
Pour rappel, la codification de la taille de l'entreprise pour le secteur de la santé est la suivante :

Codification	**Classe**
A	< 167
B	[167 ;438[
C	[438 ;724[
D	>724

Les fréquences obtenues sont les suivantes :

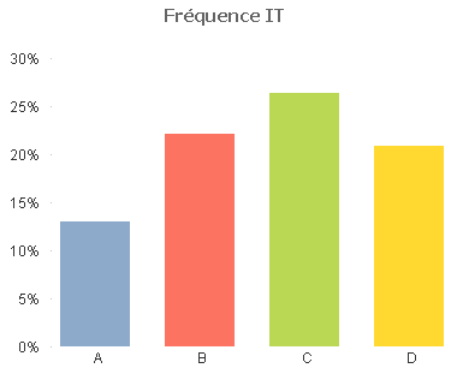


FIGURE III.26: Fréquence par taille, secteur de la santé

Les plus petites entreprises ont une fréquence d'entrée en incapacité plus faible que les autres. Par ailleurs, le pic de sinistralité n'est pas atteint pour les plus grandes entreprises mais pour celles de classe C.

III.2.b) Couple ASES/90jours

Le secteur d'activité des 'Activités de Services et Scientifiques' est maintenant étudié. C'est une activité tertiaire, cette information est à prendre en compte dans la recherche des variables explicatives de la sinistralité.

Les fréquences obtenues pour le sexe sont les suivantes :

Fréquence Homme	**Fréquence Femme**
0,56%	1,01%

Le sexe joue un rôle important pour ce secteur d'activité. Les femmes ont en moyenne deux fois plus de sinistres que les hommes.

Les résultats sont bien moins différents pour la catégorie socio-professionnelle :

Fréquence cadres	Fréquence non cadres
0,69%	0,77%

Ici la différence Cadres/Non cadres est d'environ 10%, c'est à nouveau très faible. Cela peut s'expliquer par la tache effectuée. Dans ce secteur la majorité des salariés ont un emploi de bureau, qu'ils soient cadres ou non cadres.

La grande région Ile-de-France montre à nouveau des fréquences plus faibles :

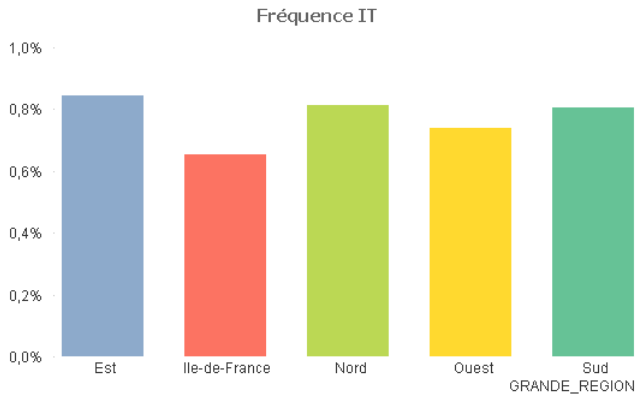


FIGURE III.27: Fréquence par grande région, secteur ASES

Pour la taille de l'entreprise, les entreprises de taille faible sont à nouveau moins sinistrées :

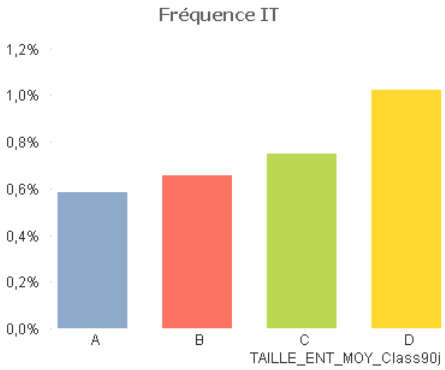


FIGURE III.28: Fréquence par taille d'entreprise, secteur ASES

Pour rappel, la codification de la taille de l'entreprise est la suivante :

Codification	**Classe**
A	< 8
B	[8 ;27[
C	[27 ;110[
D	$>$110

La sinistralité est croissante avec la taille de l'entreprise. On peut s'imaginer que les salariés d'entreprises de taille faible se sentent plus concernés par leur travail et moins en sécurité que dans les grandes structures. Cette tendance sera étudiée avec attention pour les autres secteurs d'activité.

Enfin pour la situation familiale, les résultats suivants sont visibles :

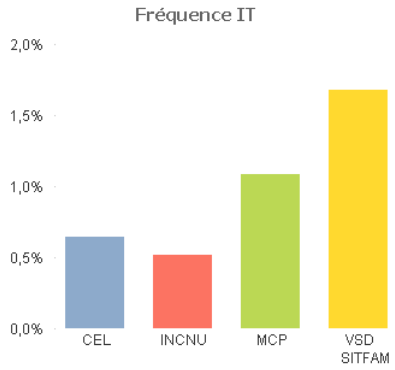


FIGURE III.29: Fréquence par situation familiale, secteur ASES

III.2.c) Couple IM/RC

Le secteur de l'industrie manufacturière est maintenant étudié au sein de la franchise "en relai". Les employés de ce secteur sont pour certains soumis à des travaux physiques importants. Cela sera à prendre en compte.

Les fréquences obtenues pour le sexe sont les suivantes :

Fréquence Homme	**Fréquence Femme**
1,66%	2,04%

Pour la CSP :

Fréquence cadres	**Fréquence non cadres**
0,67%	2,09%

A l'inverse des deux secteurs étudiés précédemment, la différence de sinistralité est très importante ici (environ trois fois plus élevée pour les non cadres). Cela peut être du à la différence de taches

entre les salariés cadres et non cadres, ces derniers effectuant des travaux plus physiques.

Pour la taille de l'entreprise :

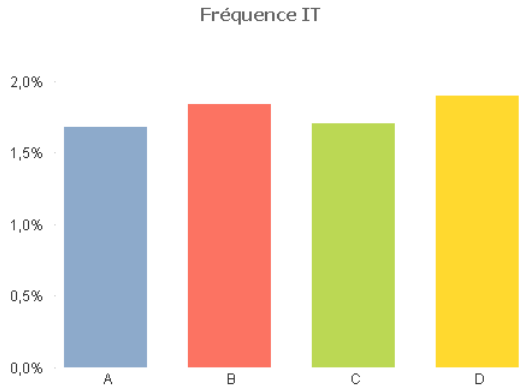


FIGURE III.30: Fréquence par taille d'entreprise, secteur IM

Pour rappel, la codification de la taille de l'entreprise est la suivante :

Codification	**Classe**
A	< 153
B	[153 ;990[
C	[990 ;3250[
D	>3250

Contrairement aux autres secteurs, la taille de l'entreprise semble jouer un rôle moins important. Cela peut être du au fait que les classes constituées regroupent toutes globalement des entreprises de tailles assez importantes.

Enfin, pour la situation familiale, des résultats similaires à ceux du secteur précédent sont visibles :

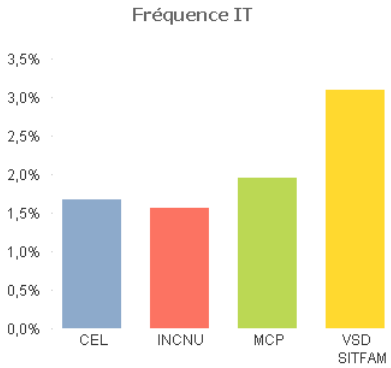


FIGURE III.31: Fréquence par situation familiale, secteur IM

Ce sont bien les individus "Veufs, séparés ou divorcés" qui ont la sinistralité la plus forte.

III.2.d) Couple Autres/RC

Pour finir cette analyse rapide de la sinistralité, les secteurs restants sont étudiés avec la franchise "En relai".

Les fréquences obtenues pour le sexe sont les suivantes :

Fréquence Homme	**Fréquence Femme**
1,29%	2,00%

Les hommes sont toujours moins sinistrés que les femmes, cependant l'écart est bien moins important que pour le secteur ASES. Pour la catégorie socio-professionnelle, les fréquences suivantes sont obtenues :

Fréquence cadres	**Fréquence non cadres**
1,10%	1,73%

La catégorie socio-professionnelle joue un rôle plus important que pour le secteur ASES mais bien moins important que pour le secteur de l'industrie.

Pour la taille de l'entreprise, une tendance croissante est encore visible :

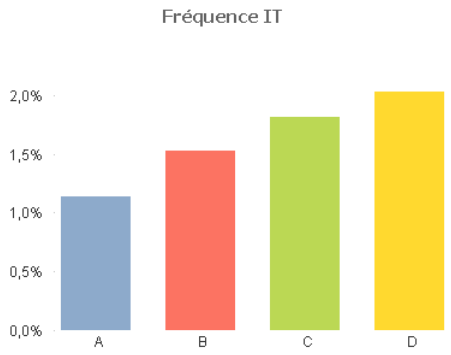


FIGURE III.32: Fréquence par taille d'entreprise, secteur AUTRES

Pour rappel, la codification de la taille de l'entreprise est la suivante :

Codification	**Classe**
A	< 16
B	[16 ;120[
C	[120 ;2737[
D	$>$2737

Cette tendance est d'autant plus intéressante que les quatre classes regroupent des entreprises de tailles très différentes, de la TPE à la grande entreprise.

Enfin pour la situation familiale :

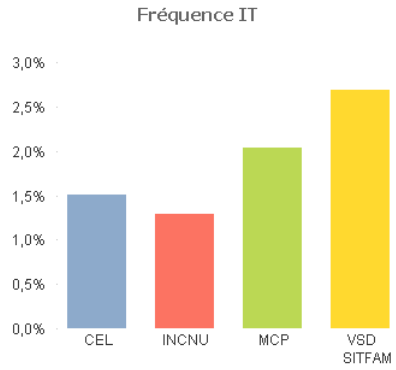


FIGURE III.33: Fréquence par situation familiale, secteur AUTRES

III.2.e) Conclusions

Ces calculs sont nécessaires à l'assureur. Ils permettent de prouver la significativité des variables et justifient le fait que l'assureur devra segmenter sa table d'entrée en incapacité selon tel ou tel critère. Il est par ailleurs intéressant de signaler que certains effets mis en évidences dans cette partie étaient visibles dans l'étude du DARES évoquée dans le chapitre I. Cela vient confirmer la qualité de nos données et le caractère généralisable de la sinistralité de notre portefeuille.
Plusieurs choses sont à retenir pour les quatre secteurs :

- Les variables aléatoires étudiées semblent avoir une influence sur la sinistralité
- Les femmes non cadres, les salariés des grandes entreprises et les individus domiciliés en dehors de la région IDF semblent plus exposées que les autres
- Les célibataires sont moins exposés que les veufs, divorcés ou séparés.

Par ailleurs, cette partie et la fin du chapitre précédent mettent en évidence des différences inter-secteurs significatives (on peut citer par exemple l'influence différente du collège de l'assuré pour les secteurs ASES et IM...). Cela justifie la nécessité d'étudier les secteurs un à un.

Cependant, ces calculs simples de fréquences donnent une information limitée. En effet, les interactions entre les variables explicatives ne sont pas étudiées. Par exemple, il n'est pas possible de constater ici si la variable "Sexe" joue un rôle plus ou moins important pour les entreprises de telle ou telle taille etc... Cela sera possible par la suite lors de l'approche par arbre de décision.

IV - Approche classique, théorie des tables d'entrée en incapacité

Une table d'entrée en incapacité est une table de fréquences qui se construit de la même façon qu'une table de mortalité du moment. Une table du moment est une table de mortalité non prospective. Elle rend compte de la sinistralité actuelle sans prendre en compte la génération des assurés. Pour construire une table d'entrée en incapacité, on peut avoir recours à des estimateurs paramétriques (on fait une hypothèse sur la loi du nombre d'entrées) ou non-paramétriques (on ne fait pas d'hypothèse sur les lois).

IV.1 - Les estimateurs

IV.1.a) L'estimateur des moments de Hoem

Idée et hypothèses :

Ce modèle est intuitif. Il s'agit de considérer que l'assuré i n'est observable dans une classe d'âge $[x, x+1[$ que sur une période $[a_i, b_i]$.
ai correspond au début de période d'observation de l'assuré d'âge x :

- Si la date de début d'observation (date de début d'affiliation ou 01/01/2011 pour cette étude) est antérieure à la date d'anniversaire de l'âge x de l'assuré, alors ai = x
- Si la date de début d'observation est postérieure à la date d'anniversaire de l'âge x de l'assuré, ai correspond à la date de début d'observation

bi correspond à la fin de période d'observation de l'assuré d'âge x :

- Si la date de fin d'observation (date de fin d'affiliation ou 31/12/2014 pour cette étude) est postérieure à la date d'anniversaire de l'âge x+1 de l'assuré, alors bi = x+1
- Si la date de début d'observation est antérieure à la date d'anniversaire de l'âge x+1 de l'assuré, bi correspond à la date de fin d'observation

La durée $(b_i - a_i)$ correspond à la durée de présence effective d'un individu.

Construction

Cet estimateur est paramétrique, par hypothèse, l'entrée en incapacité X d'un assuré d'âge x pendant une année suit une loi de Bernoulli de paramètre :

$$ {}_{b_i-a_i}q_{x+a_i} = P(T_i < x + b_i | T_i > x + a_i). $$

Avec T_i la variable aléatoire d'âge d'entrée en incapacité de l'individu i. Les T_i sont supposées indépendantes.
Ainsi le nombre total d'entrée en incapacité D_x "au moins une fois" à l'âge x suit une loi binomiale : $\mathcal{B}(n_x, {}_{b_i-a_i}q_{x+a_i})$ en temps que somme de Bernoulli indépendantes avec n_x le nombre d'individus d'âge x observés.

Pour simplifier l'expression de ${}_{b_i-a_i}q_{x+a_i}$, il est utile de faire une hypothèse de linéarité des taux d'entrée en incapacité à un âge x, ie :

$$ {}_tq_x = t \times q_x $$

On fait l'approximation (vérifiée par aucune loi) :

$$_{b_i-a_i}q_{x+a_i} \approx_{b_i} P_x -_{a_i} P_x$$

On en déduit :

$$_{b_i-a_i}q_{x+ai} \approx (b_i - a_i).q_x$$

On pose ensuite :

$$Z_i = \frac{X_i}{b_i - a_i}$$

Avec la loi des grands nombres sur Z_i on obtient :

$$\hat{q_x} = \frac{D_x^{obs}}{\sum_{i=1}^{n_x}(b_i - a_i)}$$

Avec D_x^{obs} une réalisation de D_x.

Propriétés de l'estimateur :

On montre que l'estimateur est sans biais.

$$E(\hat{q_x}) = \frac{\sum_{i=1}^{n_x} E(X_i)}{\sum_{i=1}^{n_x}(b_i - a_i)}$$

D'où d'après ce qui précède :

$$E(\hat{q_x}) = \frac{q_x.\sum_{i=1}^{n_x}(b_i - a_i)}{\sum_{i=1}^{n_x}(b_i - a_i)}$$

Donc :

$$E(\hat{q_x}) = q_x$$

La variance de l'estimateur est donnée par :

$$Var(\hat{q_x}) = \frac{\sum_{i=1}^{n_x} Var(X_i)}{(\sum_{i=1}^{n_x}(b_i - a_i))^2}$$

$$Var(\hat{q_x}) = \frac{\sum_{i=1}^{n_x}(b_i - a_i).q_x \times (1 - (b_i - a_i).q_x)}{(\sum_{i=1}^{n_x}(b_i - a_i))^2}$$

Détermination de l'intervalle de confiance de q_x :

Il peut être intéressant de disposer d'un intervalle de confiance pour nos taux d'entrée en incapacité. L'estimateur q_x est sans biais et convergent. Par ailleurs le critère de Cochran : $n_x \times \hat{q}_x > 5$ et $n_x \times (1 - \hat{q}_x) > 5$ justifie l'approximation normale de la loi de D_x.

L'intervalle de confiance est le suivant :

$$[\hat{q}_x - p_{97.5\%} \cdot \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{n_x}}; \hat{q}_x + p_{97.5\%} \cdot \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{n_x}}]$$

Avec $p_{97.5\%}$ le quantile à 97,5% de la loi normale centrée réduite.

IV.1.b) L'estimateur Poissonien

Idée et hypothèses :

Pour cette estimateur, on suppose que le nombre d'arrêt de travail pour un individu i suit une loi de poisson de paramètre : λ_i.

On fait aussi une hypothèse forte de linéarité des sinistres dans l'année. On pose ainsi $\lambda_i = \lambda_x * (b_i - a_i)$

Construction

On prend N_i le nombre d'arrêt de travail à un âge x d'un individu i et un échantillon de n individus pour l'âge x. Les variables aléatoires $(N_i)_i$ sont supposées indépendantes et identiquement distribuées. On a N_i -> $\mathcal{P}(\lambda_i)$. L est la fonction de vraisemblance.

$$L(\lambda_i) = \prod_{i=1}^{n_x} P(N_i = x_i)$$

$$L(\lambda_i) = \prod_{i=1}^{n_x} \frac{{\lambda_i}^{x_i}}{x_i!} \exp(-\lambda_i)$$

Grâce à l'hypothèse de linéarité des sinistres et en remplaçant λ_i par λ_x , il vient :

$$L(\lambda_x) = \prod_{i=1}^{n_x} \frac{(\lambda_x.(b_i - a_i))^{x_i}}{x_i!} \exp(-\lambda_x.(b_i - a_i))$$

En passant au logarithme puis en dérivant par rapport à λ_x, il vient alors :

$$\sum_{i=1}^{n_x} \frac{x_i}{\hat{\lambda x}} - \sum_{i=1}^{n_x} (b_i - a_i) = 0$$

D'où :

$$\hat{\lambda_x} = \frac{\sum_{i=1}^{n_x} x_i}{\sum_{i=1}^{n_x} (b_i - a_i)}$$

$$\hat{\lambda_x} = \frac{D_x^{obs}}{\sum_{i=1}^{n_x} (b_i - a_i)}$$

Cet estimateur autorise plusieurs arrêts de travail pendant un âge d'étude pour un individu.

Propriétés de l'estimateur :

Cet estimateur est sans biais. En effet :

$$E(\hat{\lambda_x}) = \frac{E(\sum_{i=1}^{n_x} N_i)}{\sum_{i=1}^{n_x}(b_i - a_i)}$$

$$E(\hat{\lambda_x}) = \frac{\sum_{i=1}^{n_x} E(N_i)}{\sum_{i=1}^{n_x}(b_i - a_i)}$$

$$E(\hat{\lambda_x}) = \frac{\sum_{i=1}^{n_x} \lambda_x \times (b_i - a_i)}{\sum_{i=1}^{n_x}(b_i - a_i)}$$

$$E(\hat{\lambda_x}) = \lambda_x$$

La variance de cet estimateur est :

$$Var(\hat{\lambda_x}) = Var(\frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n_x}(b_i - a_i)})$$

Or les $(N_i)_i$ sont iid.

$$Var(\hat{\lambda_x}) = \frac{1}{(\sum_{i=1}^{n}(b_i - a_i))^2} \times \sum_{i=1}^{n_x} Var(N_i)$$

$$Var(\hat{\lambda_x}) = \frac{1}{\sum_{i=1}^{n_x}(b_i - a_i)} \times \lambda x$$

Détermination de l'intervalle de confiance de λ_x :

L'estimateur de maximum de vraisemblance est asymptotiquement gaussien et efficace, par ailleurs, il est ici sans biais. On a donc pour n_x ->$+\infty$:

$$\sqrt{n_x}.\frac{|\hat{\lambda}_x - \lambda_x|}{\sqrt{Var(\hat{\lambda}_x)}} -> \mathcal{N}(0;1)$$

On obtient donc l'intervalle de confiance à 95% suivant :

$$[\hat{\lambda}_x - q_{97.5\%}.\sqrt{\frac{Var(\hat{\lambda x})}{n_x}}; \hat{\lambda}_x + q_{97.5\%}.\sqrt{\frac{Var(\hat{\lambda x})}{n_x}}]$$

Avec $q_{97.5\%}$ le quantile à 97.5% de la loi normale centrée réduite.

IV.2 - Test d'adéquation

Pour bien choisir l'estimateur, il faut dans un premier temps vérifier les hypothèses qui lui sont propres. Pour l'estimateur de Hoem, l'hypothèse X suit une loi de Bernoulli est très intuitive, à partir du moment où l'on prend la définition du taux d'entrée en incapacité appropriée.

L'hypothèse de l'estimateur Poissonien : le nombre d'arrêt à un âge x suit une loi de poisson, doit quant à elle, être vérifiée rigoureusement.

IV.2.a) Graphiquement

Les âges 30, 40, 50 et 60 ans sont retenus. On illustrera ici notre démarche avec les âges 40 et 50 pour lesquels notre effectif est conséquent.

Pour les franchises 3 jours :

Les effectifs par nombre d'arrêts se décomposent comme suit :

Nombre d'arrêts	Effectif
0	1 447
1	180
2	49
3 et +	4

Age x = 40 ans

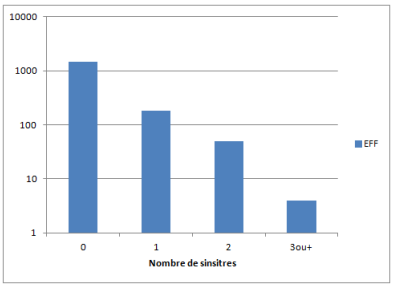


FIGURE IV.34: Adéquation Poisson, Age x = 40 ans

Nombre d'arrêts	Effectif
0	1 341
1	196
2	47
3 et +	3

Age x = 50 ans

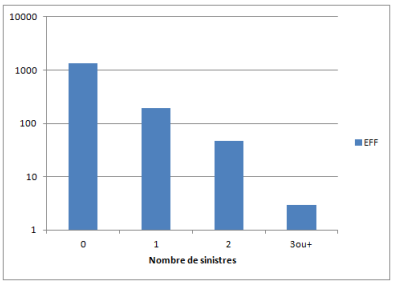


FIGURE IV.35: Adéquation Poisson, Age x = 50 ans

Les deux graphiques ont une échelle logarithmique. En observant la décroissance des histogrammes, il est légitime de penser que le nombre de sinistres par individus suit une loi de Poisson.

Pour les franchises 30 jours :

Cette hypothèse est aussi légitime pour les franchises 30 jours :

Nombre d'arrêts	Effectif
0	8 631
1	164
2 et +	7

Age x = 40 ans

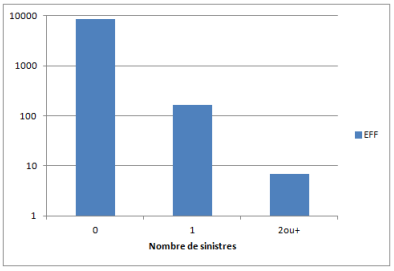


FIGURE IV.36: Adéquation Poisson, Age x = 40 ans

Nombre d'arrêts	Effectif
0	6 859
1	178
2 et +	13

Age x = 50 ans

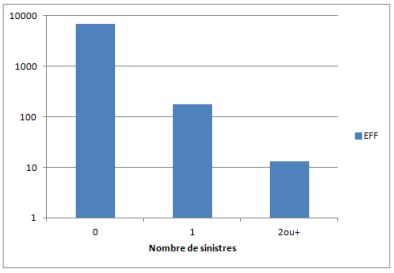


FIGURE IV.37: Adéquation Poisson, Age x = 50 ans

Il en est de même pour les franchises 90 jours et "en relai". Graphiquement, l'hypothèse semble légitime, un test d'adéquation permet de la valider ou de l'infirmer rigoureusement.

IV.2.b) Le test d'adéquation du χ^2

Le test du χ^2 est un test d'adéquation de données à une loi de probabilité. Ce test se base sur la loi du χ^2 et sur la distance entre les valeurs observées d'une suite de données et les valeurs théoriques que l'échantillon devrait prendre s'il suivait la loi de probabilité testée.

Notations et définitions :

- L'hypothèse H_0 du test est : N_x (le nombre d'arrêt de travail pour un individu d'âge x) suit une loi de Poisson. On souhaite tester H_0 avec un risque de 5%
- n le nombre de valeurs prises par N_x
- O_i le nombre d'incapacité observé
- E_i le nombre d'incapacité théorique
- La statistique du test est notée T

$$T = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Sous l'hypothèse H_0, T suit une loi du χ^2 à (n-1) degrés de liberté.

Méthodologie :

1. Dans un premier temps, la valeur empirique de la moyenne de la loi de Poisson est calculée sur l'échantillon observé
2. Les probabilités d'avoir un nombre k de sinistres par individu sont calculées à partir de la loi de Poisson de moyenne empirique
3. Les effectifs théoriques sont reconstitués pour chaque k avec la formule :

$$\text{Effectif total} \times P[N_x = k]$$

4. La distance (du χ^2) est calculée par la formule :

$$\frac{(O_i - E_i)^2}{E_i}$$

5. La statistique T est calculée puis comparée au quantile à 95% de la loi du χ^2

Rappel :
Le quantile à 95% d'un χ^2 à 3 degrés de liberté est 7,81.
Le quantile à 95% d'un χ^2 à 2 degrés de liberté est 5,99.

Les résultats obtenus sont les suivants pour les franchises 3 jours :

Pour l'âge x = 40 ans :

Nombre d'arrêts	0	1	2	3 et +
Effectif observé	1 447	180	49	4
Effectif théorique	1410	247	22	3
Distance	1	18	35	1

La valeur de T est 55. Elle est largement supérieure au quantile de la loi du χ^2. L'hypothèse nulle est rejetée.

Pour l'âge x = 50 ans :

Nombre d'arrêts	0	1	2	3 et +
Effectif observé	1 341	196	47	3
Effectif théorique	1 315	248	23	3
Distance	1	11	24	0

La valeur de T est 35. Elle est largement supérieure au quantile de la loi du χ^2. L'hypothèse nulle est rejetée.

Il n'est pas possible de valider l'hypothèse de distribution du nombre de sinistres par âge pour ces deux âges et pour la franchise 3 jours.
On observe d'ailleurs une sur-dispersion de ces échantillons. La variance empirique est supérieure à l'espérance empirique (respectivement : 0,219 contre 0,1573 et 0,2106 contre 0,1749).

Les résultats obtenus sont les suivants pour les franchises 90 jours :

Pour l'âge x = 40 ans :

Nombre d'arrêts	0	1	2 et +
Effectif observé	16 890	148	2
Effectif théorique	16 741	150	1
Distance	1	0	3

La valeur de T est 4. Elle est inférieure au quantile de la loi du χ^2. L'hypothèse nulle n'est pas rejetée.

Pour l'âge x = 50 ans :

Nombre d'arrêts	0	1	2 et +
Effectif observé	11 432	141	4
Effectif théorique	11 288	145	1
Distance	2	0	10

La valeur de T est 12. Elle est supérieure au quantile de la loi du χ^2. L'hypothèse nulle est rejetée.

L'hypothèse de distribution est validée pour l'âge 40 ans et rejetée pour l'âge 50.
A 40 ans, il est possible d'observer la principale caractéristique de la loi de Poisson : Espérance = Variance. A 50 ans, on observe par contre une sur-dispersion.

Globalement, pour les franchises "en relai", 3 jours, 30 jours et 90 jours, il n'est pas possible de valider l'adéquation à une loi de Poisson pour tous les âges.
Attention, dans cette partie, seules les individus visibles pendant toute une année d'âge ont été étudiés. Pour les autres, ce sera la linéarité de la moyenne de la loi de Poisson qui sera testée.

IV.2.c) La linéarité de la moyenne

Pour les individus tronqués et/ou censurés, on cherche à vérifier l'hypothèse de linéarité de la moyenne de la loi de poisson :

$$\lambda_i = \lambda_x \times (b_i - a_i)$$

Méthodologie :

Pour vérifier cette hypothèse, on crée une variable de présence par trimestre et on trace le rapport entre le nombre de sinistres et le nombre de personnes distinctes présentes par trimestre. Cela revient à vérifier qu'à chaque âge, le taux d'entrée en incapacité augmente avec le temps de présence dans l'étude.

Résultats :

Rigoureusement, il faudrait tracer la fréquence d'entrée en incapacité à chaque âge en fonction du nombre de trimestre de présence. Cependant des effectifs insuffisants créent parfois des problèmes d'échantillonnage. Il a donc été choisi de tracer la fréquence d'entrée tout âge confondu pour les franchises "en relai", 3 jours, 30 jours et 90 jours.

Pour les franchises "en relai" :

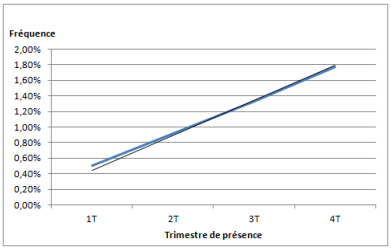


FIGURE IV.38: Linéarité de la moyenne, franchise "en relai"

Le coefficient de corrélation linéaire est satisfaisant ($R^2 = 0,987$)

Pour les franchises 3 jours :

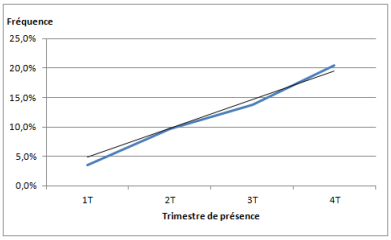


FIGURE IV.39: Linéarité de la moyenne, franchise 3 jours

Le coefficient de corrélation linéaire est satisfaisant ($R^2 = 0,9776$).

Pour les franchises 30 jours :

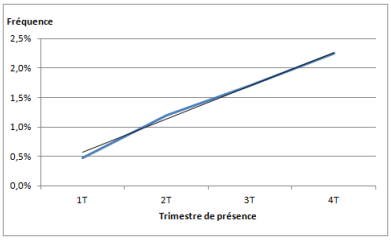


FIGURE IV.40: Linéarité de la moyenne, franchise 30 jours

Le coefficient de corrélation linéaire est satisfaisant ($R^2 = 0,9928$).

Pour les franchises 90 jours :

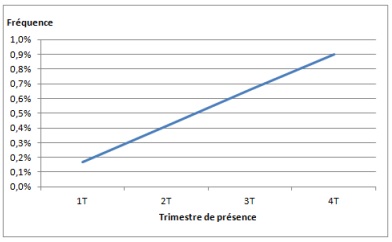


FIGURE IV.41: Linéarité de la moyenne, franchise 90 jours

Le coefficient de corrélation linéaire est satisfaisant ($R^2 = 0,9885$). L'hypothèse de linéarité des taux d'entrée en incapacité avec la durée de présence est vérifiée.

IV.3 - Choix de l'estimateur

L'hypothèse d'adéquation à une loi de Poisson n'est pas toujours vérifiée. Cependant c'est cet estimateur qui est conservé. En effet, le recours à l'estimateur de Hoem impliquerait la modélisation d'une loi d'entrée en incapacité "au moins une fois" puis "au moins deux fois" etc... Or pour les franchises supérieures ou égale à 30 jours, très peu d'individus ont plus d'un arrêt au delà de la franchise à un âge donné.

D'autres estimateurs auraient pu être utilisés, c'est le cas du très classique estimateur de "Kaplan-Meier", cependant sa mise en oeuvre était plus complexe et moins intuitive que celle des estimateurs proposés. Par ailleurs, comme pour l'estimateur des moments de Hoem, il serait nécessaire de calculer une loi d'entrée en incapacité "au moins une fois" puis "au moins deux fois" etc...

Remarque : Par la suite, les fréquences d'entrée en incapacité à un âge x seront notées q_x

IV.4 - Positionnement d'une population spécifique

Une fois les taux bruts par âge et par sexe déterminés pour une population de référence, il peut être nécessaire de positionner des populations spécifiques par rapport à ces tables.

Si l'on sélectionne les assurés ayant des modalités de plusieurs variables explicatives communes, les échantillons obtenus sont souvent de taille faible. Cela rend impossible la création d'une table d'entrée en incapacité pour ces échantillons avec les estimateurs exposés précédemment. Pour se ramener à une table propre à ces populations, deux méthodes sont employées : l'utilisation d'un modèle de Brass et le calcul d'un coefficient de réduction/majoration.

Communément, ces méthodes sont utilisées pour positionner la mortalité d'un portefeuille par rapport à la mortalité des tables INSEE. Ici, c'est l'entrée en incapacité d'une population spécifique qui va être positionnée par rapport à une population de référence de taille plus conséquente.

IV.4.a) Coefficient de réduction/majoration

Notations

Dans cette partie, les notations suivantes sont employées :

- $S(x)$ la fonction de survie (ici survie signifie ne pas entrer en incapacité) d'un individu à l'âge x
- $\mu(x)$ le taux instantané d'entrée en incapacité à l'âge x
- $\mu(x)^{spe}$ le taux instantané d'entrée en incapacité pour la population étudiée
- $\mu(x)^{ref}$ le taux instantané d'entrée en incapacité pour la population de référence
- q_x la probabilité d'entrée en incapacité entre x et x+1
- D_x le nombre d'individu ayant eu au moins un arrêt à un âge x
- L_x l'exposition totale à l'âge x, correspond à la somme des durées de présence effective des individus

Idée

L'hypothèse suivante est réalisée :

$$\mu_x^{spe} = \alpha.\mu_x^{ref}$$

Or on a :

$$\mu_x = \frac{S^{'}(x)}{S(x)}$$

D'où :

$$S(x) = e^{-\int_0^x \mu(t)dt}$$

or :

$$q_x = 1 - \frac{S(x+1)}{S(x)}$$

$$q_x = 1 - e^{-\int_0^x \mu(t)dt + \int_0^x \mu(t)dt + \int_x^{x+1} \mu(t)dt}$$

Ainsi, avec l'hypothèse de constance du taux de hasard entre deux âges entiers, il vient :

$$\mu_x = -ln(1 - q_x)$$

En supposant que les q_x sont petits et en réalisant un développement limité au premier ordre, il vient alors :

$$q_x^{spe} \approx \alpha.q_x^{ref}$$

Estimation de α

Intuitivement, une estimation de α peut être obtenue à partir de l'écart entre le nombre observé d'individus ayant eu au moins un arrêt et leur nombre théorique. Par égalité de ces deux nombres il vient :

$$\sum_x D_x^{obs} = \sum_x \alpha.q_x^{ref}.L_x^{obs}$$

$$\hat{\alpha} = \frac{\sum_x D_x^{obs}}{\sum_x q_x^{ref}.L_x^{obs}}$$

Une autre approche consiste à estimer le α qui minimise la statistique de type χ^2 suivante :

$$\sum_x L_x^{obs}.\frac{(q_x^{obs} - \alpha.q_x^{ref})^2}{\alpha.q_x^{ref}}$$

D'un point de vue pratique, les deux approches sont testées et la valeur de α conservée est celle qui minimise l'erreur quadratique entre le q_x^{spe} observé et le q_x^{spe} estimé.

IV.4.b) Modèle de Brass

Ce modèle permet de lier deux populations. Il se base sur les logit des taux d'entrée en incapacité. Pour rappel :

$$logit(q_x) = ln(\frac{q_x}{1-q_x})$$

Rappel sur les régressions linéaires simples

L'objectif est de déterminer $\hat{\beta}_0$ et $\hat{\beta}_1$ tel que l'on ait :

$$Y = \beta.X + \epsilon$$

Avec :

- Y le vecteur aléatoire de taille n de la variable à expliquer $Y = (y_i, ..., y_n)$
- X la matrice des variables explicatives, ici : $X = \begin{pmatrix} 1 & ... & 1 \\ x_1 & ... & x_n \end{pmatrix}$
- $\beta = (\beta_0, \beta_1)$ les paramètres du modèle à estimer

L'hypothèse fondamentale réalisée est que $\epsilon -> \mathbb{N}(0, \sigma^2)$, par conséquent $Y -> N(\beta.X, \sigma^2.I_n)$.

β est ensuite estimé par la méthode des moindres carrés ordinaires. Après calcul il vient :

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{\sum_1^n (y_i - \overline{y}).x_i}{\sum_1^n (x_i - \overline{x}).x_i}$$

Le modèle de Brass

Brass propose le modèle suivant :

$$logit(q_x^{spe}) = \beta_1.logit(q_x^{ref}) + \beta_0 + \epsilon_x$$

Le logit des taux d'entrée en incapacité d'une population spécifique est ainsi lié linéairement au logit des taux d'entrée en incapacité d'une population de référence.

Une fois $\hat{\beta}_0$ et $\hat{\beta}_1$ calculés, on se ramène aux taux spécifiques à l'aide de la formule :

$$q_x^{spe} = (\frac{q_x^{ref}}{1-q_x^{ref}})^{\hat{\beta}_1}.\frac{e^{\hat{\beta}_0}}{(1+(\frac{q_x^{ref}}{1-q_x^{ref}})^{\hat{\beta}_1}.e^{\hat{\beta}_0})}$$

IV.5 - Lissage des tables d'entrée en incapacité par la méthode de Whittaker Henderson

Les tables des taux bruts obtenus présentent des irrégularités. Il est légitime de penser que ces irrégularités ne sont pas dues au phénomène étudié mais à des fluctuations d'échantillonnages. Il est alors nécessaire de lisser les tables des taux bruts.
La méthode utilisée ici pour le lissage est la méthode de Whittaker Henderson. Cette méthode est non paramétrique et est utilisée par le BCAC pour le lissage de ces barèmes de provisions mathématiques incapacité/invalidité.

IV.5.a) Idée de la méthode et définitions en dimension un

L'idée de cette méthode est de trouver un compromis entre la fidélité aux taux brut et la régularité (le lissage) voulue. Les notations suivantes sont utilisées :

- L'opérateur de différence avant Δ
- q_i les taux d'entrée en incapacité lissés
- $\hat{q}_i$ les taux bruts d'entrée en incapacité estimés
- F le critère de fidélité
- S le critère de régularité
- z un paramètre du modèle
- h le paramètre d'équilibre entre régularité et fidélité
- w_i des poids
- p le nombre de taux à lisser

Opérateur de différence :
L'opérateur de différence avant correspond à la différentiation :

$$\Delta f(x) = f(x+1) - f(x)$$

De manière récursive, la différence avant d'ordre n est :

$$\Delta^n f(x) = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} f(x+j)$$

Critère de fidélité :

$$F = \sum_{i=1}^{p} w_i (q_i - \hat{q}_i)^2$$

Critère de régularité :

$$S = \sum_{i=1}^{p-z} (\Delta^z q_i)^2$$

IV.5.b) Démarche en dimension un

L'objectif est de trouver un compromis entre fidélité aux données et régularité du lissage. Cela revient à minimiser la fonction suivante :

$$M = F + h \times S.$$

Pour déterminer la solution de ce problème, on résout le système d'équation :

$$\frac{\delta M}{\delta q_i} \text{ avec } 1 \leq i \leq n$$

La résolution de ce système d'équation nécessite un certain nombre de manipulations matricielles. On a :

$$F = (q - \hat{q})' w (q - \hat{q}) \text{ et } S = (\Delta^z q_i)' (\Delta^z q_i)$$

Avec $q = (q_i)_{1 \leq i \leq n}$ et $\hat{q} = (\hat{q}_i)_{1 \leq i \leq n}$ et $w = diag(w_i)_{1 \leq i \leq n}$ Par ailleurs, on peut vérifier que :

$$(\Delta^z q_i) = K_z q$$

avec K_z, une matrice de taille (p-z,p) avec pour diagonale, sur-diagonale et sous-diagonale les coefficients binomiaux d'ordre z. Les signes s'alternent et commencent positivement. Ainsi avec l'expression de M donnée précédemment, on obtient :

$$M = (q - \hat{q})' w (q - \hat{q}) + h q' K_z' K_z q$$

D'où :

$$\frac{\delta M}{\delta q} = 2wq - 2w\hat{q} + 2hK_z' K_z q$$

La résolution du système d'équation $\frac{\delta M}{\delta q} = 0$ fournit le résultat suivant :

$$q^* = (w + hK_z' K_z)^{-1} w\hat{q}$$

IV.5.c) Validation du lissage

Pour valider les différents lissages, un test du changement de signe sera réalisé. Si l'on suppose que la différence entre les taux lissés et les taux bruts suit une loi normale, alors cette différence à la même probabilité d'être positive ou négative (1/2). Par conséquent, en considérant n âges le nombre de changement de signe devrait suivre une loi binomiale de paramètre $(p-1{,}1/2)$ en tant que somme de Bernoulli indépendantes.

Le test consistera à vérifier que le nombre de changement de signe ne s'éloigne pas trop de la moyenne de cette loi, ie $\frac{p-1}{2}$. A noter que dans certains cas, le lissage peut volontairement surestimer certains taux pour épouser au mieux certains pics. Dans ce cas, il se peut que ce test soit inutile pour juger de la qualité du lissage.

IV.6 - Mesure de l'hétérogénéité

Dans le chapitre précédent, une certaine hétérogénéité a été mise en évidence au sein des secteurs d'activité. Il est par exemple possible de constater que les cadres ont une sinistralité différente voire très différente des non cadres, il en est de même pour les petites et les grandes entreprises etc...

Pour modéliser cette hétérogénéité, un modèle de Cox est utilisé. La spécification de la fonction de hasard du modèle ne nous intéresse pas ici. On cherche uniquement à mesurer l'effet des covariables et à positionner des populations les unes par rapport aux autres.

IV.6.a) Le modèle

Les notations suivantes sont utilisées :

- Z le vecteur des variables explicatives
- $\lambda(t|Z=z)$ la fonction de hasard du modèle pour une population de caractéristiques $Z=z$
- $\lambda_0(t)$, la fonction de hasard de base du modèle (elle ne sera pas spécifiée)
- δ un vecteur de paramètres de même taille que Z

Le modèle de Cox peut s'écrire :

$$\lambda(t|Z=z) = \lambda_0(t).e^{{}^t\delta.z}$$

Par conséquent, si l'on prend deux populations de caractéristiques z_1 et z_2, le rapport des fonctions de hasard ne dépend que de z_1 et z_2 :

$$\frac{\lambda(t|Z=z_1)}{\lambda(t|Z=z_2)} = e^{{}^t\delta.(z_1-z_2)}$$

Pour déterminer les paramètres du modèle, Cox se base sur une vraisemblance partielle. Les détails de ce modèle et les tests nécessaires pour en juger la qualité sont présentés en annexe 5.

IV.6.b) Application dans R

Le package survival est utilisé. Pour chaque assuré, une ligne est présente dans la table en entrée pour chaque année d'âge avec un âge exact de début et un âge exact de fin. Cet âge de fin correspond à :

(Age de début + durée de présence à cet âge)

Si l'assuré n'a pas de sinistre dans l'année.

(Age exact auquel l'assuré a son premier sinistre)

Si l'assuré a "au moins un sinistre" dans l'année.

Un indicateur de sinistre est aussi présent. Ainsi, le modèle de Cox compare la sinistralité de deux populations par rapport au premier sinistre. On supposera que les coefficients correcteurs obtenus sont valables pour les sinistres suivants. Cette structure des données permet d'utiliser le package survival.

IV.7 - Problèmes de volumes

Dans la partie suivante, des lois d'entrée en incapacité vont être modélisées. Ces lois doivent être segmentées selon différentes variables explicatives. Cependant, dans certains cas, on se heurte très rapidement à des problèmes de volumes de données.

Les tables d'entrée en incapacité sont modélisées par franchise. Certaines disposent d'un échantillon assez conséquent pour modéliser des lois d'entrée en incapacité par âge pour une population spécifique. Un lissage de Whittaker-Henderson est alors utilisé pour lisser les taux bruts par âge obtenus. Cependant, pour d'autres franchises, des échantillons aussi spécifiques sont de tailles trop faibles, le calcul des taux par âge puis un lissage de Whittaker-Henderson entrainerait alors des erreurs d'estimation trop élevées.

Dans ce contexte, l'approche suivante est retenue : pour palier à l'insuffisance de données, une moyenne mobile de 3 ans autour de chaque âge est calculée. Les taux bruts "moyens" obtenus sont ensuite lissés par une méthode de Whittaker-Henderson (bien que la moyenne mobile puisse faire office de lissage, ce dernier n'est la plupart du temps pas suffisant).

La moyenne mobile

Pour une moyenne mobile symétrique sur 2t+1 années, la formule de base est la suivante :

$$q_x^m = \sum_{i=-t}^{t} \alpha_i q_{x+i}$$

Les poids choisis dans cette étude sont les suivants :

$$\alpha_i = \frac{(bi - ai)}{\sum_{j=-t}^{t}(b_j - a_j)}$$

Remarque :

Pour les âges extrêmes (inférieurs à 21 ans et supérieurs à 63 ans) la quantité de données est parfois extrêmement faible. De part sa structure, l'utilisation de la moyenne mobile n'est pas possible pour ces valeurs extrêmes. Certaines lois seront donc "tirées" pour ces valeurs extrêmes à partir (respectivement) du premier et du dernier point où la quantité de données est suffisante.

IV.8 - Généralisation du modèle

Des fréquences d'entrée en incapacité vont être calculées pour les franchises "en relai", 3, 30 et 90 jours continus. Il est cependant souhaitable de disposer de fréquences d'entrée en incapacité pour d'autres franchises continues (par exemple 45, 60, 75, 120 et 150 jours) et discontinues. Pour se ramener à de nouvelles franchises continues, une loi de maintien sera utilisée. Ensuite, une table de correspondance entre franchises continues et discontinues sera utilisée pour généraliser les résultats aux franchises discontinues.

IV.8.a) Utilisation d'une loi de maintien en incapacité

Une fois la loi d'entrée en incapacité modélisée, il est possible d'extrapoler une loi d'entrée pour une franchise plus longue à l'aide d'une table de maintien en incapacité. L'idéal serait pour cela de modéliser des tables d'expérience de maintien en incapacité sur le même portefeuille que celui étudié pour les fréquences d'entrée. Cependant, par manque de temps, il a été décidé de ne pas modéliser ces lois de maintien. C'est donc les tables de maintien de 1996 et 2014 du BCAC qui sont utilisée (ces tables sont disponibles en annexe).

Théorie

Les tables de maintien en incapacité sont des tables à deux dimensions avec l'âge de l'assuré en lignes et la durée dans l'état en colonnes. Cette durée est en mois. 10 000 assurés sont présents pour chaque âge à la durée 0, cette quantité diminue ensuite avec le temps et la sortie de l'état des individus.

On note T_x, la durée passée en incapacité par un assuré d'âge x. La probabilité que cet assuré soit présent dans l'état après une durée y+t sachant qu'il l'est après une durée y est :

$$P(T_x > y + t | T_x > y) = \frac{P(T_x > y + t)}{P(T_x > y)} = \frac{l^x_{y+t}}{l^x_y}$$

Avec l^x_y le nombre d'assuré toujours en incapacité après une durée y.

Ainsi, dans notre cas, pour passer de la franchise 30 jours à la franchise 60 jours, il suffit de multiplier les fréquences par le coefficient correcteur $\frac{l^x_2}{l^x_1}$.

Pour obtenir une franchise "demi-mensuelle" comme la 45 jours, une hypothèse de linéarité (et donc une interpolation linéaire) des sorties dans le mois est réalisée.

Validation de l'approche

Pour que cette approche soit valable, il est nécessaire que la population étudiée se comporte globalement de la même façon que la population utilisée pour élaborer les tables du BCAC. Une validation simple de cette hypothèse est de vérifier que, partant de la fréquence d'entrée en incapacité 30 jours calculée sur le portefeuille, on retombe bien à l'aide de la loi de maintien du BCAC sur les fréquences d'entrée de la franchise 90 jours calculées sur le portefeuille.

Si cette approche est validée, il sera possible quelque soit le modèle de passer de la franchise continue 30 jours aux franchises continues 45, 60 et 75 jours et de la franchise 90 jours aux franchises 120 et 150 jours.

IV.8.b) Utilisation d'une table de correspondance pour les franchises discontinues

Un certain nombre de contrats ont des franchises discontinues. Cet échantillon n'est cependant pas assez conséquent pour envisager la construction de lois d'entrée en incapacité d'expérience sur cette partie du portefeuille. Pour se ramener à des lois propres à ces franchises, une table des correspondances entre franchises continues et discontinues est utilisée. La relation entre les deux types de franchise est la suivante :

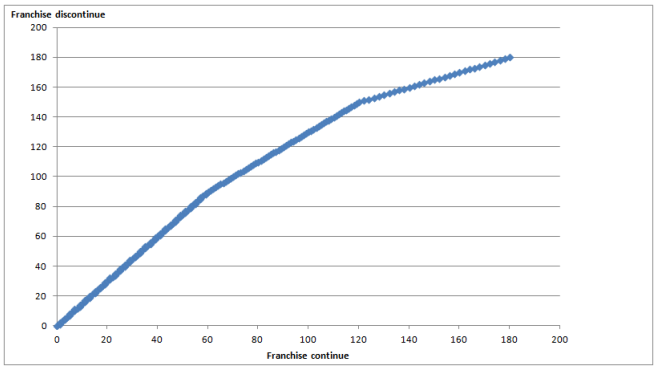


FIGURE IV.42: Évolution de la différence des taux H/F, franchise 90 jours

Une franchise continue équivaut à 2/3 d'une franchise discontinue jusqu'à la franchise continue 60 jours. La pente de la courbe diminue ensuite, ce qui est compréhensible, les franchises très longues (180 jours par exemple) sont très peu souvent atteinte que ce soit pour une franchise continue ou discontinue. De plus, il est difficilement imaginable que la somme des durées de petits arrêts atteignent 180 jours au cours d'une année.

V - Approche classique, tables d'entrée en incapacité

Dans un premier temps, des lois d'entrée en incapacité segmentées par sexe puis par secteur et enfin par sexe et collège sont modélisées pour les fréquences "en relai", 3, 30 et 90 jours. Ces lois sont ensuite adaptées aux franchises 45, 60, 75, 120 et 150 jours continues ou discontinues à l'aide de la table de maintien du BCAC 2014 et d'équivalences entre franchises continues et discontinues. Les lois obtenues par sexe pour les franchises 30 et 90 jours seront présentées et commentées. Les autres lois sont disponibles en annexe 1, 2 et 3.

Dans un second temps, un modèle mêlant hétérogénéité et positionnement est calculé. Ce modèle, qui permettra d'obtenir des lois segmentées d'entrée en incapacité selon les variables présentées dans le chapitre précédent, sera critiqué et comparé à d'autres approches par la suite.

Remarque : Dans la suite de ce mémoire, le secteur de la santé ne sera étudié que par la franchise 3 jours. Les effectifs de ce secteur pour les autres franchises sont trop faibles.

V.1 - Lois d'entrée en incapacité

Dans un premier temps, des lois d'entrée en incapacité sont modélisées par franchise et sexe, puis par franchise et secteurs, enfin par franchise, sexe et collège. La segmentation du portefeuille selon ces variables se fait en amont. La quantité de données disponibles est assez conséquente pour ne pas avoir à positionner des populations les unes par rapport aux autres.

Seules les lois pour les franchises 30 jours et 90 jours segmentées par sexe sont présentées ici, les autres lois sont disponibles en annexe 1, 2 et 3.

V.1.a) Table d'entrée pour la franchise 90 jours

Taux bruts et lissés par sexe

Deux tables de taux bruts d'entrée en incapacité sont calculées. Les effectifs aux âges inférieurs à 18 et supérieurs à 67 sont très faibles. Les tables sont donc calculées sur des âges variants de 18 à 67 ans. Par ailleurs, 67 ans est l'âge de liquidation de la retraite à taux plein sans condition.
Les taux bruts hommes et femmes sont les suivants :

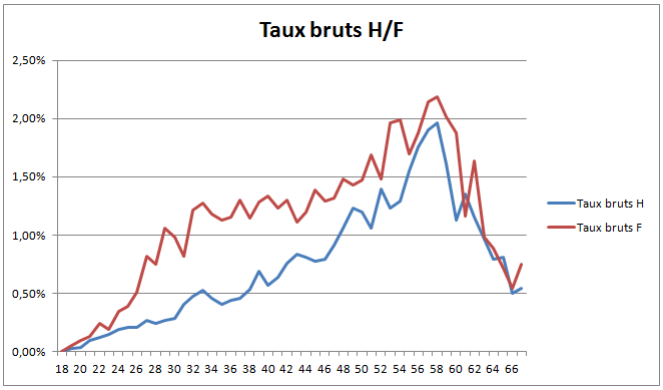


FIGURE V.43: Taux bruts par sexe, franchise 90 jours

Pour chacune des tables, l'intervalle de confiance à 95% est tracé (son expression est disponible

dans la section précédente), cet intervalle de confiance est très étroit. La largeur maximale est de 0.05% et est obtenu à l'âge 67 ans (âge pour lequel on dispose du moins de données).
Les intervalles de confiance à 95% sont représentés en 3D, ils sont invisibles à l'oeil nu en 2D sur de si petits graphiques :

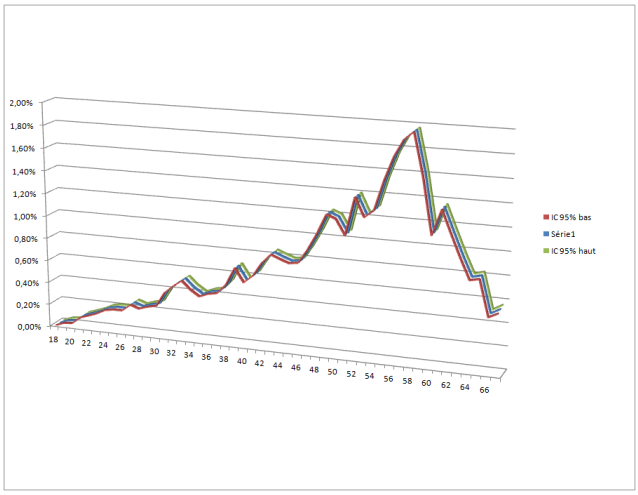


FIGURE V.44: Taux bruts et IC homme

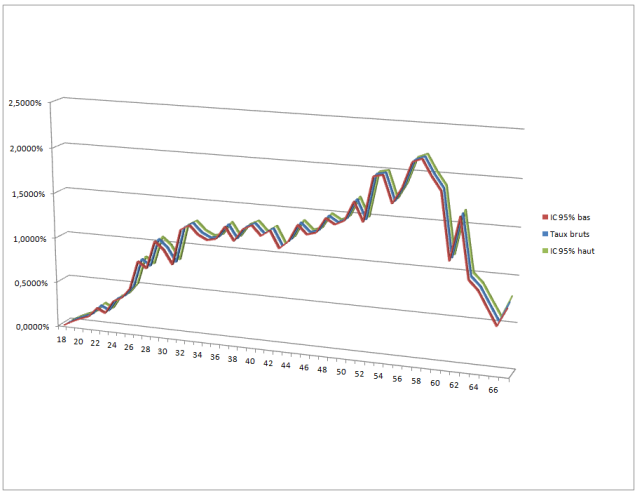


FIGURE V.45: Taux bruts et IC femme

Ces intervalles de confiance sont très étroits. On pouvait s'y attendre en regardant l'expression obtenue en partie IV. Bien que cela reflète la qualité du modèle, il est tout de même nécessaire de lisser cette courbe, en effet, il est légitime de penser que (au moins pour les hommes), les taux d'entrée en incapacité sont croissants avec l'âge. Cela n'est pas vrai ici, on voit un certain nombre de pics.

Un lissage de Wittaker-Henderson fournit les courbes suivantes :

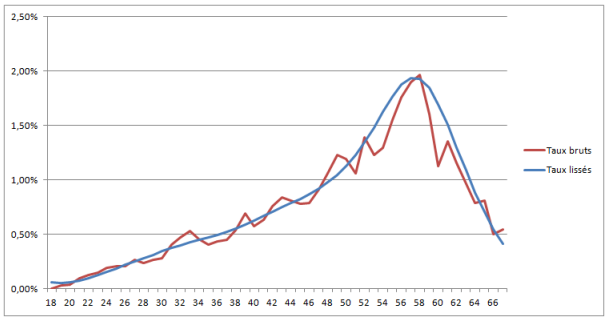


FIGURE V.46: Taux bruts et lissés homme

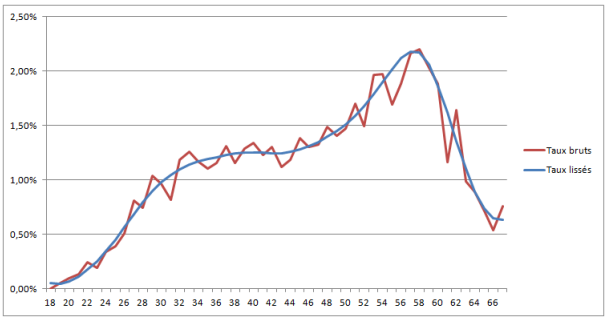


FIGURE V.47: Taux bruts et lissés femme

Cela avait été mis en évidence dans la partie précédente : il existe de fortes disparités entre les fréquences d'entrée en incapacité des hommes et des femmes. Ces disparités sont plus ou moins importantes avec l'âge :

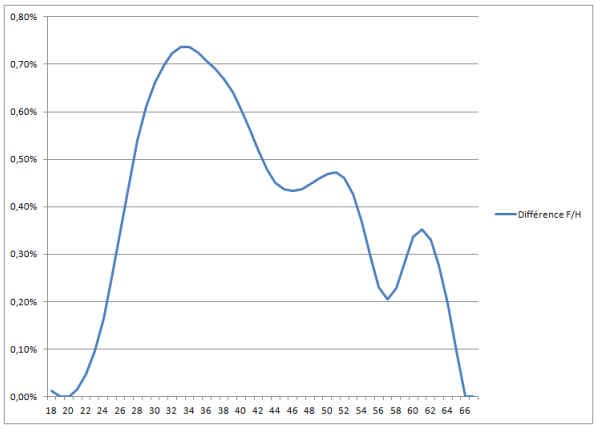


FIGURE V.48: Évolution de la différence des taux H/F, franchise 90 jours

La différence est maximale à 33 ans (0.7369% de différence). Elle est croissante jusqu'à cet âge puis décroit progressivement jusque 67 ans (avec cependant deux pics aux alentours de 52 et 64 ans). Les différences importantes se situent aux âges où le taux de maternité est élevé.

Des lois d'entrée en incapacité sont aussi modélisées par secteurs et par sexe et collège. Ces modélisations se font sur les données réelles et sans modèle d'hétérogénéité. Ces lois sont disponibles en annexe 2 et 3 ainsi que les lois pour les franchises continues ou discontinues 120 et 150 jours qui en découlent.

V.1.b) Table d'entrée pour la franchise continue 30 jours

Taux bruts et lissés par sexe

Les lois obtenues par sexe sont les suivantes :

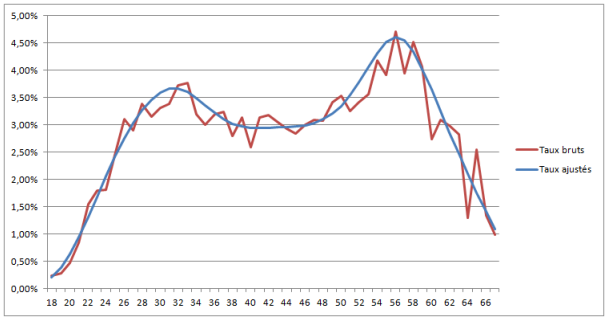


FIGURE V.49: Taux lissés femmes, franchise 30 jours

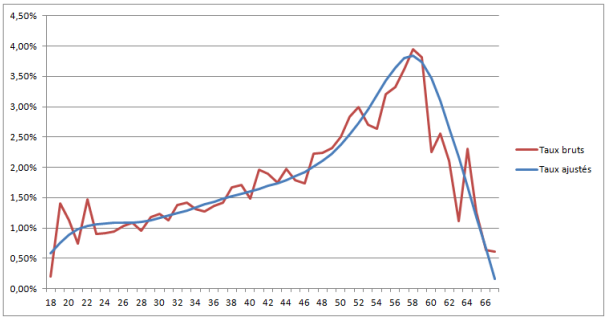


FIGURE V.50: Taux lissés hommes, franchise 30 jours

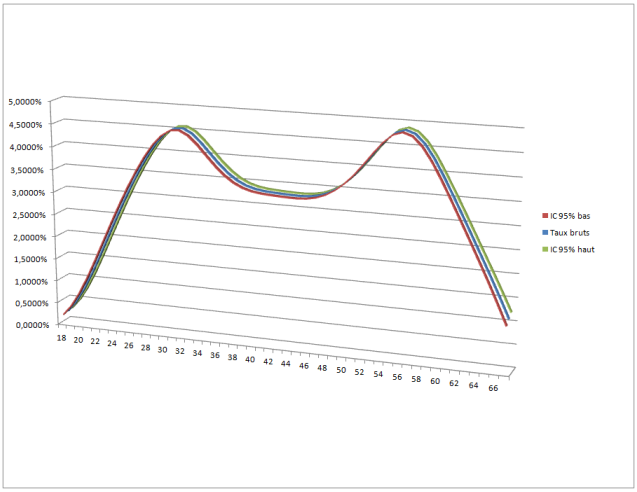


FIGURE V.51: Taux lissés et IC femmes, franchise 30 jours

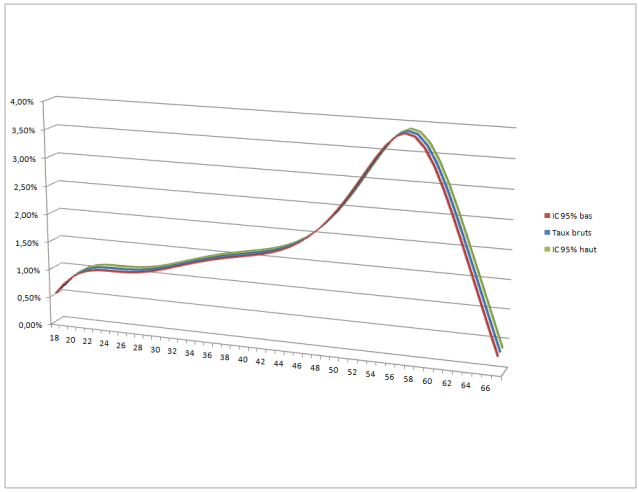


FIGURE V.52: Taux lissés et IC hommes, franchise 30 jours

Les lois par secteurs d'activité puis par sexe et collège sont à nouveau disponibles en annexe 2 et 3 ainsi que les lois pour les franchises continues ou discontinues 45, 60 et 75 jours qui en découlent.

V.1.c) Forme des lois obtenues

Plusieurs lois d'entrée en incapacité ont été calculées. Les lois obtenues ont des formes similaires quelle que soit la franchise.

Pour les hommes, quelle que soit la franchise, une croissance des taux d'entrée en incapacité en fonction de l'âge est constatée jusque 58 ou 60 ans. Les taux diminuent ensuite fortement.

Pour les femmes, deux pics de sinistralité sont observés. Un aux alentours de 30 ans et l'autre aux alentours de 60 ans. Le second, comme pour les hommes, est très surement dû à la dégradation de la santé du salarié avec l'âge. Le premier intervient aux âges où la natalité est la plus forte. D'après une étude de l'INSEE, en 2015, le taux de fécondité des femmes était maximum à 30 ans avec un taux de 144,3 enfants pour 1000 femmes, le nombre de naissance dépassant 100 enfants pour 1000 femmes entre 26 et 34 ans. Le pic de sinistralité mis en évidence correspond à ce segment d'âge. Il est possible d'imaginer que cela est du à des problèmes prénataux, ou à des absences pour enfant malade (pour rappel, l'absence pour congé de maternité/paternité n'est pas étudiée dans ce mémoire). Il est important de signaler que ce premier pic est absent pour la franchise 90 jours, ce qui corrobore l'hypothèse évoquée (il est plus rare que des absences pour enfant malade ou problèmes prénataux durent plus de 3 mois).

La décroissance observée aux alentours de 60 ans peut s'expliquer par plusieurs facteurs. La première hypothèse est que les individus qui continuent leur activité professionnelle après 60 ans sont globalement en meilleure santé que ceux qui s'arrêtent, seuls des "bons risques" seraient donc conservés en portefeuille à ces âges. Une seconde hypothèse beaucoup plus pratique serait le manque de données à ces âges élevés en raison du report récent de l'âge de la retraite à 62 ans. Dans ce cas, cette décroissance devrait se décaler progressivement autour de 62 ans dans les prochaines années.

Les informations fournies par ces lois sont intéressantes, elles permettent de voir comment évolue la sinistralité selon les modalités des variables secteurs, sexe et collège. Cependant, elles ne permettent pas de segmenter parfaitement le portefeuille et de prendre en compte des combinaisons de variables. Pour cela, un modèle d'hétérogénéité est utilisé.

V.2 - Modélisation de l'hétérogénéité du portefeuille

Dans la partie précédente, des lois d'entrée en incapacité sont calculées. Elles s'appuient uniquement sur l'estimateur Poissonien des taux bruts et un lissage. Cette approche a cependant la faiblesse de ne pas permettre une segmentation très fine, le volume de données serait insuffisant. Pour cela, l'approche suivante est employée.

Dans un premier temps, l'hétérogénéité au sein de chaque couple secteur d'activité/franchise majoritaire est étudié. L'hypothèse faite par la suite est que les populations au sein d'un même secteur d'activité ont un comportement similaire quelle que soit la franchise. Cela signifie par exemple que le coefficient correcteur appliqué aux cadres du secteur ASES est le même si la franchise étudiée et de 30, 90 jours ou "en relai".

Pour la franchise 90 jours, la démarche est ensuite la suivante :

1. Modélisation de deux lois d'entrée en incapacité (une homme et une femme) pour la population de référence du modèle de Cox du secteur ASES (le secteur pris pour référence)
2. Positionnement des populations des modèles de Cox des autres secteurs par rapport à la population de référence du secteur ASES
3. Utilisation des coefficients correcteurs intra-secteurs

La même démarche est ensuite entreprise pour les autres franchises, la seule modification portera sur la référence sur laquelle est calculée les lois d'entrée. Les tables et la segmentation des franchises 30 et 90 jours sont présentées et comparées ici, celles pour la franchise "en relai" et 3 jours sont présentées en annexe.

V.2.a) Modélisation de Cox

Les couples mis en évidence dans les chapitres précédents sont étudiés un à un. Pour le secteur ASES, la population de référence est constituée des individus d'une entreprise de taille D et non cadres de la région Ile-de-France, pour les autres segments la population de référence est constituée des mêmes individus mais d'une entreprise de taille A. Le modèle à hasard proportionnel de Cox est utilisé ici.

Lors de l'étude, il a été montré que le modèle sans la variable "Grande région" était plus précis que le modèle avec cette variable. Les résultats présentés n'incluent donc pas cette variable.

Secteur ASES

Le modèle est calculé à l'aide du logiciel R et du package survival. Le test de validation de l'hypothèse de risques proportionnels (avec l'hypothèse $H0$ en tant qu'hypothèse de proportionnalité) donne les p value suivantes :

Variable explicative et modalité	p value
Collège Cadre	0.002
Taille A	0.7
Taille B	0.16
Taille C	0.27

De fortes présomptions contre cette hypothèse existent pour le collège cadre. Cela est bien visible si l'on trace les variations du rapport $\frac{q_x^{cad}}{q_x^{nca}}$ (Ici les âges sont regroupés par classes de 5 ans pour éviter des variations d'échantillonnages importantes) :

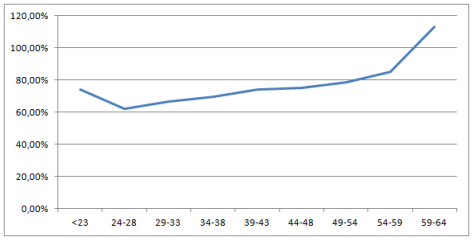


FIGURE V.53: Évolution du rapport $\frac{q_x^{cad}}{q_x^{nca}}$

Une tendance croissante avec l'âge est très visible. Le test précédent était effectué sur l'ensemble des données disponibles. Si l'on restreint les données pour ne garder que les âges où la population est conséquente (entre 30 et 55 ans), le test fournit les résultats suivants :

Variable explicative et modalité	p value
Collège Cadre	0.182
Taille A	0.408
Taille B	0.147
Taille C	0.693

L'hypothèse nulle n'est alors plus rejetée directement. On trace les fonction $ln(q_x^{cad})$ et $ln(q_x^{nca})$ en fonction de l'âge de l'assuré :

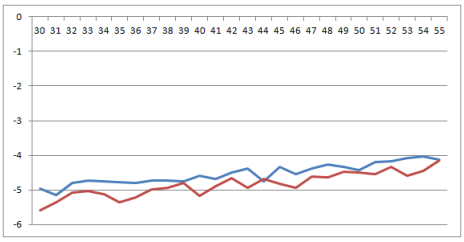


FIGURE V.54: $log(q_x^{cad})$ et $log(q_x^{nca})$

Les deux courbes sont plutôt parallèles, cela confirme la qualité du modèle. Il est aussi possible de vérifier cela pour les différentes tailles d'entreprises. C'est ce modèle qui est donc conservé. Les coefficients obtenus seront utilisés pour tous les âges (pas uniquement pour la tranche d'âge 30-55 ans).
Pour ce secteur, le modèle de Cox fournit donc les coefficients correcteurs suivants :

Variable explicative et modalité	Coefficient correcteur
Collège Cadre	0.73
Taille A	0.565
Taille B	0.72
Taille C	0.83

Secteur de l'industrie manufacturière

L'hypothèse de risques proportionnels est ici rejetée pour toutes les modalités.
Pour la covariable "Collège", une tendance croissante est présente.

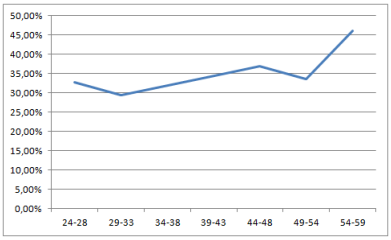


FIGURE V.55: Évolution du rapport $\frac{q_x^{cad}}{q_x^{nca}}$

Tandis que pour la covariable "Taille de l'entreprise", des oscillations du rapport $\frac{q_x^{spe}}{q_x^{ref}}$ sont visibles, par exemple :

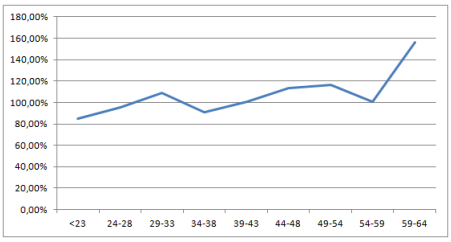


FIGURE V.56: Évolution du rapport $\frac{q_x^{spe}}{q_x^{ref}}$ pour la taille d'entreprise

La non validation de cette hypothèse nous encourage à tester cette hypothèse sur un autre segment d'âge. Plusieurs segments sont testés, l'hypothèse est acceptable pour le segment 25-50 ans. Cela se vérifie aussi graphiquement. Les coefficients correcteurs obtenus sont alors :

Variable explicative et modalité	**Coefficient correcteur**
Collège Cadre	0.28
Taille B	1.28
Taille C	1.50
Taille D	1

La modalité "Taille D" n'est pas significative, le coefficient 1 lui est donc attribué.

Secteur de la santé

L'hypothèse de risques proportionnels est vérifiée sur le segment d'âge [27-47]. Les coefficients correcteurs suivants sont obtenus :

Variable explicative et modalité	**Coefficient correcteur**
Collège Cadre	0.8784
Taille B	1.50
Taille C	1.64
Taille D	1.68

Les covariables sont toutes significatives.

Autres secteurs

L'hypothèse de hasard proportionnel est ici vérifiée sur le segment d'âge [30-60]. Pour ces autres secteurs, le modèle de Cox fournit les coefficients correcteurs suivants :

Variable explicative et modalité	Coefficient correcteur
Collège Cadre	0.529
Taille B	1.462
Taille C	1.665
Taille D	2.225

Conclusion

Des différences importantes entre les valeurs des coefficients correcteurs des différentes modalités des variables explicatives sont visibles. **Cette partie confirme bien la nécessité d'étudier les secteurs un à un et isole de façon rigoureuse certaines caractéristiques des secteurs.**

V.2.b) Modélisation de l'hétérogénéité, franchise 90 jours

Taux lissés de la population de référence

Pour la franchise 90 jours, la population de référence est la population des hommes ou femmes, non cadres, d'une entreprise de taille D du secteur ASES. Les lois d'entrée en incapacité pour cette population sont les suivantes :

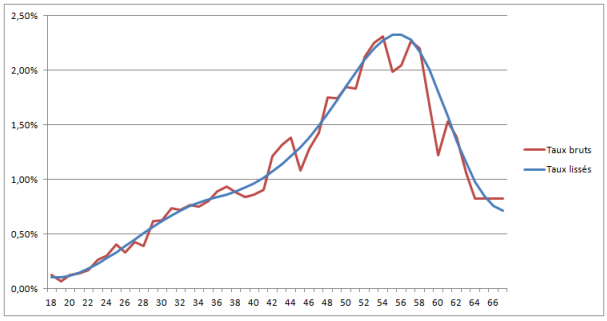


FIGURE V.57: Taux moyennés et lissés, population de référence hommes, franchise 90 jours

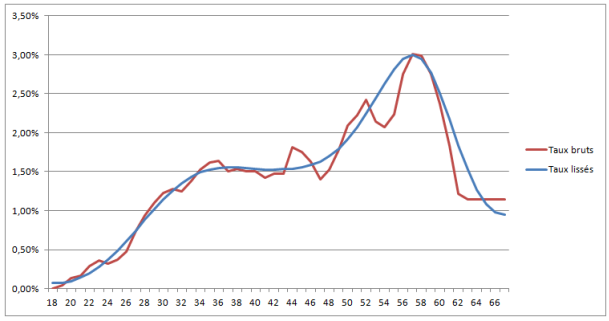


FIGURE V.58: Taux moyennés et lissés, population de référence femmes, franchise 90 jours

Les population suivantes sont ensuite positionnées par rapport à ces courbes (le positionnement se fait indépendamment du sexe pour garder des échantillons conséquents) :

- Les individus non cadres du secteur "Autres" d'une entreprise de taille A de ce secteur (Attention, les classes sont différentes entre chaque secteur)
- Les individus non cadres du secteur de l'industrie d'une entreprise de taille A de ce secteur
- Les individus non cadres du secteur de la santé d'une entreprise de taille A de ce secteur

Positionnement avec coefficient de réduction/majoration

Pour justifier ce positionnement, il est nécessaire de vérifier que la relation entre les deux populations ne varie pas trop avec l'âge. Pour vérifier cette hypothèse, des classes d'âges de 5 ans sont créées. Les graphiques suivants sont tracés. Les taux bruts de référence sont placés en abscisse, les taux spécifiques sont en ordonnée :

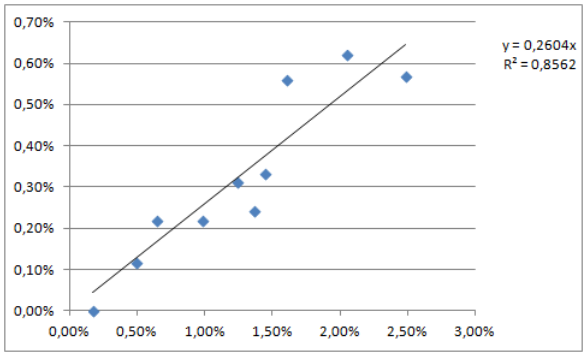


FIGURE V.59: Vérification hypothèses, population spécifique AUTRES

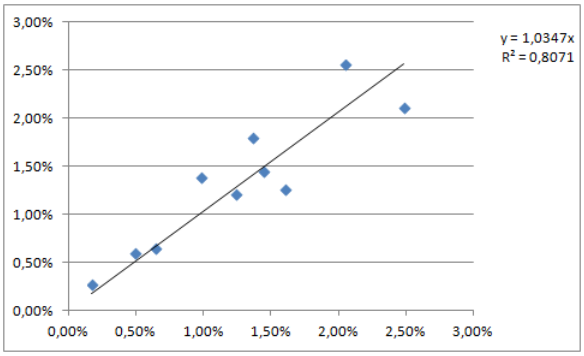


FIGURE V.60: Vérification hypothèses, population spécifique IM

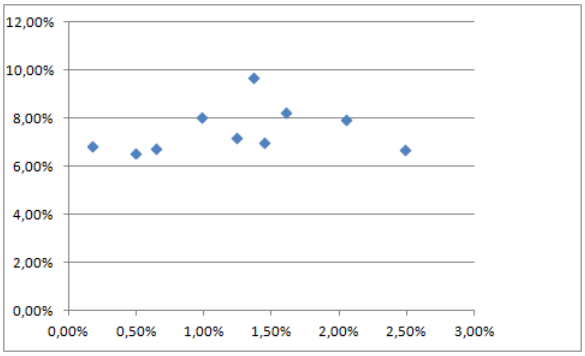


FIGURE V.61: Vérification hypothèses, population spécifique santé

L'hypothèse est admissible pour les deux premiers secteurs. Elle ne l'est pas pour le secteur de la santé. Trop peu de données sont présentes pour modéliser des lois fiables pour ce secteur et cette franchise, il est sorti de l'étude.
Les valeurs des coefficients correcteurs obtenus sont regroupées dans le tableau suivant, ils sont calculés par les deux méthodes présentées précédemment :

Secteur	Méthode 1	Méthode 2
AUTRES	0.2028	0.2359
IM	0.9652	0.9801

Les sommes des erreurs quadratiques d'estimation des sinistres obtenues sont les suivantes :

Secteur	Méthode 1	Méthode 2
AUTRES	297,4	322,3
IM	1 140,1	1 171,5

Les coefficients obtenus par la méthode 1 sont retenus.

Positionnement avec modèle de Brass

Un modèle de Brass de positionnement est maintenant paramétré, sa précision sera comparée avec celle de la méthode ci-dessus.
Dans un premier temps, il faut vérifier l'hypothèse de linéarité des logit des taux de la population spécifique en fonction des logits des taux de la population de référence.
Des classes d'âges de 5 ans sont utilisées pour vérifier cette hypothèse. Les résultats obtenus sont les suivants :

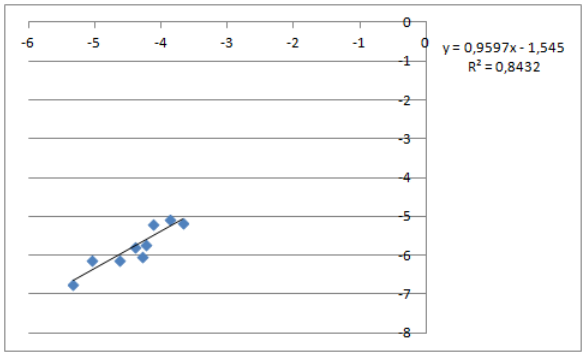


FIGURE V.62: Vérification hypothèses Brass, population spécifique AUTRES

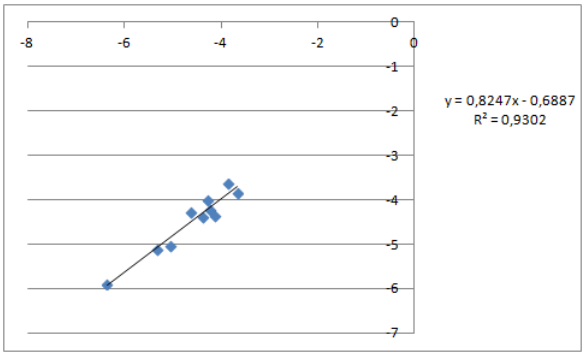


FIGURE V.63: Vérification hypothèses Brass, population spécifique IM

Les fréquence d'entrée en incapacité sont calculées et les erreurs quadratiques d'estimation des sinistres sont regroupées dans le tableau suivant :

Secteur	Erreur quadratique
AUTRES	323
IM	952

L'estimation du modèle de Brass est moins bonne qu'un simple coefficient de réduction/majoration pour le secteur "AUTRES", mais meilleure pour le secteur IM. Cependant la somme des erreurs quadratiques des deux secteurs est inférieure avec le modèle de Brass. C'est cette approche qui est conservée.

V.2.c) Validation du modèle

Résumé de la démarche :

1. Mesure de l'hétérogénéité interne à chaque secteur

2. Calcul d'une loi d'entrée par sexe pour la population de référence : les individus non-cadres d'une entreprise de taille D du secteur ASES

3. Positionnement des populations de référence des secteurs "AUTRES" et "IM" (les non cadres d'une entreprise de taille A) par rapport à la référence décrite au point précédent

4. Application aux tables positionnées des coefficients correcteurs mis en évidence dans le 1

Le modèle réalisé cherche à approcher au mieux la sinistralité de chaque secteur d'activité. C'est pour cette raison que les modèles de Cox sont paramétrés sur les données propres à chaque secteur. Les tables d'entrée en incapacité homme et femme calculées sur la population de référence choisie pour cette franchise (les non-cadres d'une entreprise de taille A du secteur ASES) permettent de donner "la tendance" des fréquences. Ces tendances sont ensuite adaptées à l'aide de coefficients correcteurs.

Une vérification "Backtesting" est utilisée sur ce modèle : le ratio "nombre de sinistres observés" sur "nombre de sinistres prédits" (ratio O/A). Pour que le modèle soit validé, il faut que ce ratio se rapproche de 100%.

Ratio Observé/Prédit

Le nombre d'incapacités prédit est calculé sur un échantillon de 600 000 lignes ayant une franchise continue de 90 jours. L'observé est de 3 995 sinistres.

Les résultats obtenus sont les suivants :

Modèle	Prédiction	Ratio O/A
Modèle 1	3 594	111,2%

Les sinistres sont sous estimés de 11% par rapport à l'observé. Secteur par secteur les ratios sont les suivants :

Secteur	**Observé**	**Prédiction**	**Ratio O/A**
ASES	2847	2844	100%
AUTRES	602	310	194,2%
IM	547	438	125%

Le modèle est très bon pour le secteur de référence. Il l'est beaucoup moins pour les deux autres. Cela laisse penser que, pour la franchise 90 jours, les populations des secteurs "AUTRES" et "IM" ne se comportent pas de la même façon que ce qui avait été déterminé lors de la modélisation de l'hétérogénéité. C'est une faiblesse de ce modèle.

V.2.d) Modélisation de l'hétérogénéité, franchise 30 jours

La population de référence est ici constituée des individus d'une entreprise de taille D cadres du secteur ASES. L'exposition de ce type d'individu est très faible (10 000 années d'étude pour les femmes et 20 000 pour les hommes), il est donc nécessaire de recalculer les taux bruts d'entrée en incapacité par âge à l'aide d'une moyenne mobile symétrique. Cette moyenne est réalisée sur 3 années d'âges (une avant l'âge étudié et une après). Les lois d'entrée en incapacité obtenues sont les suivantes :

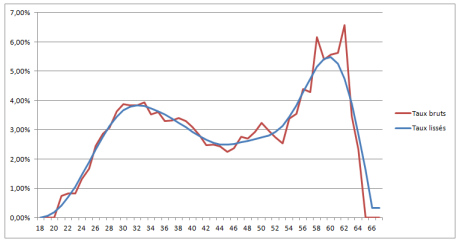


FIGURE V.64: Taux moyennés et lissage femmes

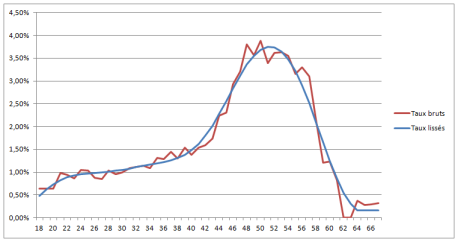


FIGURE V.65: Taux moyennés et lissage hommes

Un lissage est effectué pour éliminer les pics restants.

Un positionnement des populations de référence des autres secteurs par rapport à cette population est réalisé. Le modèle de Brass est à nouveau légèrement meilleur. Par ailleurs, les hypothèses ne sont à nouveau pas vérifiées pour le secteur de la santé. Il est sorti de l'étude.

A titre d'indication, les coefficients de majoration obtenus sont :

Secteur	Coefficient
AUTRES	0,9822
IM	1,6189

Validation du modèle

Le nombre d'incapacités prédit est calculé sur un échantillon de 600 000 années d'âges ayant une franchise continue 30 jours. L'observé est de 9 288 sinistres.

Les résultats obtenus sont les suivants :

Observé	Prédiction	Ratio O/A
9 288	9 222	100,7%

Par secteur, les ratios O/A sont les suivants :

Secteur	Observé	Prédit	Ratio O/A
ASES	4 295	4 669	92%
Autres	2 842	2 719	104,5%
IM	2 150	1 833	117,3%

Remarque : Le secteur de la santé ne sera étudié que par la franchise 3 jours continus.

V.2.e) Critique du modèle

La démarche entreprise présente certaines faiblesses, cependant, c'est cette approche qui nous a paru la plus appropriée dans notre analyse.

En effet, il aurait aussi été possible d'utiliser des modèles comme par exemple le modèle de Cox, en y incluant le secteur d'activité et de calculer des taux bruts sur la population de référence de ce modèle. Cependant, cette approche ne nous permet pas d'analyser en profondeur les secteurs d'activités et n'est pas satisfaisante en raison de la structure des données. Il est aussi possible d'envisager une approche sans modèle d'hétérogénéité, en modélisant des courbes d'entrée en incapacité pour toutes les combinaisons des variables explicatives. Face à l'insuffisance des données de certaines combinaisons, des taux moyennés seraient calculés. Cependant cette approche serait trop propre au portefeuille étudié et peu généralisable.

Pour apporter une meilleur estimation, une approche par arbres de décision de la modélisation de fréquences sera employée par la suite et comparée avec cette approche "classique".

V.3 - Généralisation de la modélisation

On dispose de lois d'entrée en incapacité et d'un modèle d'hétérogénéité pour les franchises 30 et 90 jours continues. Il est maintenant nécessaire de généraliser ces résultats aux autres franchises existantes. La table du BCAC 2014 est utilisée ici pour cela. Les taux de passages d'une franchise 30 jours continue aux franchises 45, 60, 75 et d'une franchise 90 jours continue aux franchises 120 et 150 jours sont données en annexe 4. Le test suivant est ensuite réalisé :

On considère les lois d'entrée en incapacité toute population confondue pour les franchise 30 et 90 jours continus. On suppose que l'on dispose de 10 000 assurés à chaque âge. En utilisant la loi de maintien du BCAC 2014, les résultats suivants sont obtenus :

Nombre d'entrée en incapacité à 30 jours	**Nombre d'entrée en incapacité à 90 jours**
11 019	6 026

En utilisant les fréquences calculées sur le portefeuille, on obtient :

Nombre d'entrée en incapacité à 30 jours	**Nombre d'entrée en incapacité à 90 jours**
11 019	4 890

Environ 80% des sinistres sont retrouvés. Ce qui est peu satisfaisant. De fortes disparités existent entre les secteurs d'activités, si on réalise ce test sur le secteur ASES, on retrouve plus de 90% des sinistres, par contre, sur les secteurs de l'industrie et "AUTRES", ce pourcentage descend à 80% et 70%.

La loi de maintien du BCAC sera tout de même utilisée, mais il sera gardé en mémoire que les fréquences obtenues pour les franchises 45, 60, 75, 120 et 150 devraient être légèrement surestimées par rapport à l'expérience du portefeuille.

V.4 - Conclusion

Dans cette partie, des lois d'entrée en incapacité ont tout d'abord été modélisées sans utilisation de modèle d'hétérogénéité. Ces lois permettent de connaitre et d'analyser la tendance de la sinistralité avec l'âge pour différentes populations.

Plusieurs choses sont à retenir pour les différentes franchises étudiées :

- Les populations homme et femme ont une évolution très différente de la sinistralité avec l'âge, un pic aux âges de maternité élevés est présent pour les femmes
- Les hommes non cadres voit leur pic de sinistralité arriver plus tôt que les cadres (56 ans contre 58 ans), cet effet est moins évident pour les femmes
- Le secteur de l'industrie a une sinistralité supérieure à celle des autres secteurs à tout âge. La sinistralité du secteur ASES est la plus faible à tout âge.

Un modèle mêlant hétérogénéité et positionnement est ensuite construit pour approcher au mieux la sinistralité du portefeuille. Nous disposons ainsi de lois d'entrée en incapacité segmentées selon les combinaisons des variables secteur, collège, sexe et taille d'entreprise. En outre, les coefficients correcteurs obtenus par le modèle de Cox viennent confirmer certains effets vus dans le chapitre III.

Ces deux approches sont classiques, bien que leurs limites soient connues, elles sont un outil de validation très solide des approches suivantes. En effet, l'approche du risque incapacité de travail par les arbres de décision est récente, il est rassurant pour l'actuaire de pouvoir comparer les résultats obtenus à ceux de techniques qu'il connait bien.

VI - Approche par arbres, influence des variables explicatives

L'approche classique permet de déterminer les variables explicatives qui ont une influence sur la fréquence d'entrée en incapacité. Cependant, cette approche ne permet pas d'effectuer un "classement" par ordre d'importance entre nos variables explicatives, ile ne permet pas non plus de déterminer des groupes homogènes (des combinaisons de variables explicatives) face à la sinistralité sans réaliser d'hypothèses fortes.

Pour répondre à ces besoins, les arbres de décision (au travers des algorithmes CART et des forêts aléatoires) vont être utilisés dans cette partie. L'algorithme CART va en particulier permettre de mettre en évidence des groupes homogènes face à la sinistralité, puis les forêts aléatoires fourniront un classement rigoureux des variables explicatives.

A noter que cette partie n'a pas pour objectif de démontrer les résultats des algorithmes d'apprentissage statistique (le lecteur pourra se reporter aux documents présents dans la bibliographie pour une présentation plus mathématique), son objectif est de donner des informations qui permettront au lecteur de comprendre le fonctionnement des algorithmes et leur utilisation.

VI.1 - Théorie de l'apprentissage

VI.1.a) Principe d'estimation

L'objectif est d'estimer une variable aléatoire Y (ici, le nombre de sinistres), à l'aide de variables explicatives X. Pour cela, on dispose de réalisations des variables. Cela revient à chercher une fonction $\phi : \mathbb{R}^p \to \mathbb{R}$ telle que :

$$Y = \phi(X)$$

Pour choisir la meilleure fonction ϕ, un critère d'erreur d'estimation (une mesure de la distance entre l'estimation et la réalité) est nécessaire. Dans cette étude, c'est la minimisation de la somme des erreurs quadratiques qui sera utilisée. Ainsi, on cherche à résoudre :

$$\hat{\phi} = argmin_{\phi} E((Y - \phi(X))^2)$$

Si on dispose d'un échantillon de taille n de couples (x_i, y_i), réalisations de X et Y, alors :

$$\hat{\phi} = argmin_{\phi} \frac{1}{n} \sum_{i=1}^{n} (y_i - \phi(x_i)^2)$$

Ce critère sera utilisé pour comparer les modèles. Par ailleurs, un autre critère de mesure de la qualité de l'estimation sera utilisé : le ratio "Nombre de sinistres observés" sur "Nombre de sinistres prédits par le modèle".

Remarque : Il est possible de montrer que dans un espace hilbertien :

$$\hat{\phi} = E[Y|X]$$

On rejoint alors la théorie de l'espérance conditionnelle.

VI.1.b) Notion de sur-apprentissage

L'objectif est de déterminer la meilleure fonction ϕ, intuitivement, la meilleure fonction serait celle qui reproduit exactement les données. On aurait tendance à proposer la fonction :

$$x \to \sum_{i=1}^{n} y_i 1(x = x_i)$$

Le risque est alors de connaitre un effet de "sur-apprentissage". La prédiction de nouvelles données serait alors mauvaise dans la majorité des cas.

La fonction $\hat{\phi}$ doit par conséquent apprendre de l'échantillon initial tout en gardant un pouvoir prédictif sur d'autres échantillons. Pour déterminer la fonction $\hat{\phi}$ il sera donc nécessaire de tester les estimations sur un échantillon indépendant de l'échantillon d'apprentissage. Pour ce faire, l'échantillon est généralement découpé en trois parties distinctes :

- Un échantillon d'apprentissage sur lequel la modélisation est réalisée
- Un échantillon de validation indépendant pour ajuster les paramètres de la modélisation
- Un échantillon test indépendant pour vérifier le pouvoir prédictif de la modélisation

VI.2 - Algorithme CART

CART signifie "Classification And Regression Trees", c'est un algorithme par arbre binaire de décision crée en 1984 par Leo Breiman. L'objectif de ces algorithmes est de séparer un échantillon en plusieurs échantillons homogènes pour une variable réponse. On distingue deux type d'arbre, les arbres de classification qui servent à prédire l'appartenance d'un "objet" à une classe, et les arbres de régression qui eux vont chercher à expliquer et prédire les valeurs d'une variable quantitative en fonction de variables explicatives.

C'est l'algorithme CART pour les arbres de régression qui est utilisé ici.

Un arbre binaire de décision est constitué d'une racine, de branches, de noeuds et de feuilles. Chaque noeud correspond à une question sur une variable explicative. La réponse à cette question (oui ou non) entraine une division de l'échantillon en deux échantillons (deux nouveaux noeuds) selon un critère de division, la procédure est ensuite itérée. Une fois l'arbre maximal obtenu (tous les noeuds sont terminaux selon une règle d'arrêt), il est ensuite élagué pour obtenir un arbre "optimal", l'arbre maximal subit en effet des problèmes de "sur-apprentissage".

Pour définir l'algorithme, il faut :

- Déterminer le critère de division à chaque noeud
- Déterminer un critère d'arrêt ie une règle pour dire si le noeud est terminal ou non

Attention, chaque individu présent dans la table initiale doit appartenir à une unique feuille de l'arbre.

La réalisation de l'arbre binaire de décision suit ensuite le processus suivant :

- Construction de l'arbre maximal
- Élagage de l'arbre pour déterminer l'arbre optimal

VI.2.a) Le critère de division

Pour diviser l'espace (l'échantillon), une covariable est sélectionnée. Si elle est numérique, la partition de l'espace se fera entre deux de ses valeurs successives (exemple : $Age \leq 40$ ans et $Age > 40$ ans). Si elle est catégorielle, on partitionne selon une modalité de la variable (exemple : Catégorie Socio-Professionnelle = Cadre).

La variable réponse est notée Y. L'échantillon initial à chaque noeud est noté J, les échantillons obtenus par division sont notés I et K.

La valeur moyenne de la variable réponse sur les noeuds J, I et K sont notées respectivement μ_J, μ_I et μ_K. Pour les arbres de régression, l'échantillon initial est divisé en deux tel que la réduction d'hétérogénéité :

$$\sum_{j\in J}(Y_j - \mu_J)^2 - (\sum_{i\in I}(Y_i - \mu_I)^2 + \sum_{k\in K}(Y_k - \mu_K)^2)$$

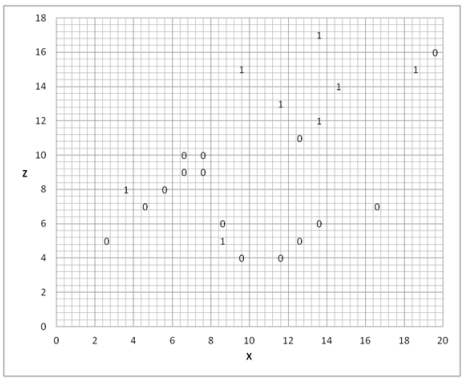
soit maximale. Cela revient à chercher la division qui rend minimale le désordre des noeuds "fils" I et K.

Pour chacune des divisions possibles (une variable et une modalité), l'homogénéité globale est calculée. Une liste d'homogénéité est obtenue, la division selon la covariable et sa modalité qui maximise l'homogénéité globale est retenue et l'algorithme est itéré sur les noeuds "fils".

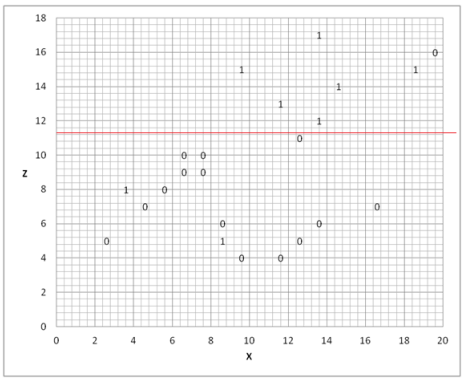
Exemple :

Pour illustrer ce critère graphiquement, on suppose que Y est une variable binaire pouvant prendre les valeurs 0 ou 1. On suppose par ailleurs que l'on ne dispose que de deux variables explicatives X et Z.

On a alors le schéma suivant :



L'algorithme va alors séparer l'espace tel que l'on ait la meilleur séparation entre les 0 et les 1. Tous les partitionnements possibles de la variable X et de la variable Z sont testés. Ici la variable explicative qui sépare le mieux les 0 et les 1 est la variable Z pour la modalité 11.5.

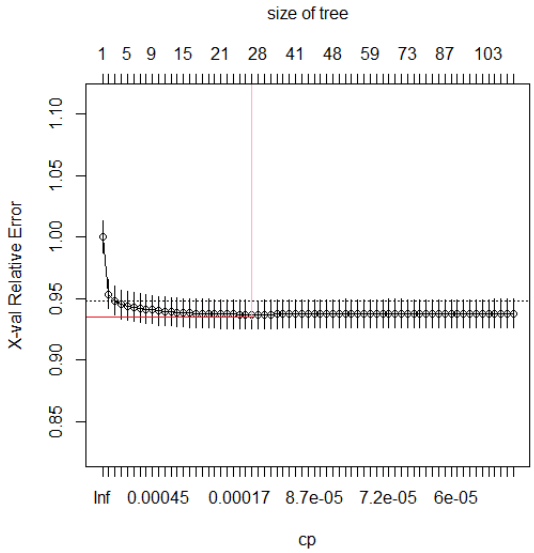


On obtient donc deux échantillons $Z > 11.5$ et $Z < 11.5$ sur lesquels on itère l'algorithme.

VI.2.b) Critère d'arrêt et élagage

La division de l'échantillon initial en sous échantillons s'arrête en fonction d'une règle fixée à l'avance. L'arbre "maximal" est alors obtenu. Un noeud est généralement considéré comme "terminal" lorsque l'échantillon obtenu est de taille inférieure à une certaine valeur ou lorsque l'amélioration relative de l'homogénéité résultant d'une nouvelle division devient inférieure à un certain seuil.

Une fois l'arbre "maximal" obtenu il faut déterminer l'arbre "optimal" par élagage de l'arbre "maximal". Plusieurs méthodes sont possibles pour déterminer l'optimalité. Le processus qui sera utilisé par la suite (présent dans le package rpart de R) est une validation croisée à k-blocs sur tous les sous-arbres possibles.
Pour chaque sous-arbre possible, l'échantillon est divisé en k blocs, (k-1) blocs servent d'échantillon d'apprentissage. L'erreur d'estimation est calculée sur le dernier bloc. Ce processus est ensuite itéré sur les autres blocs puis les erreurs d'estimations sont moyennées. Dans le package rpart, c'est l'erreur relative d'estimation par rapport à l'erreur d'origine (au sein du noeud initial) qui est calculée.
Il est possible de tracer cette erreur relative d'estimation en fonction du nombre de feuilles de l'arbre :



L'arbre optimal est celui qui minimise l'erreur d'estimation. Ici l'arbre optimal est celui de taille 27.
Remarque : Un paramètre important de l'algorithme est le gain minimal d'homogénéité demandé.

VI.2.c) Application à l'arrêt de travail

De la même façon que pour l'approche classique de recherche des variables explicatives, les deux définitions de la fréquence d'entrée en incapacité sont envisageables. Si la première définition de la page 42 est retenue, l'étude s'inspire de la méthode employée par Walter Olbricht dans son article *Tree-based methods : a useful tool for life insurance*. La variable réponse est alors une variable binaire : l'individu aura au moins un arrêt à un certain âge ou zéro.

Cependant, dans un souci de cohérence avec la définition utilisée dans la partie III.2, c'est la seconde définition de la page 43 qui est retenue : le taux d'entrée en incapacité est égale au nombre total de sinistres divisé par l'exposition totale.

Une table composée des caractéristiques des assurés et du nombre de sinistres est mise en entrée. Le logiciel R est utilisé ainsi que le package *rpart*. Il est important de signaler que ce package ne prend pas en compte l'exposition (les troncatures et censures) des individus. **Seules les données non censurées seront donc étudiées dans ce chapitre. Cela ne pose pas de problème, il est en effet légitime de supposer que l'influence des modalités des variables explicatives ne varie pas si l'assuré est présent pendant une année d'âge entière ou seulement une fraction de cette année.** Les résultats de ce chapitre pourront donc être appliqués à l'ensemble des données.

L'algorithme CART va séparer le groupe étudié en sous groupes homogènes par rapport à la variable réponse. L'exemple suivant illustre le fonctionnement de l'algorithme :

Exemple : Une table composée de 4 colonnes (l'âge de l'assuré, son sexe, sa catégorie socio-professionnelle et l'indicateur de sinistralité) est mise en entrée. L'arbre obtenu après élagage est le suivant :

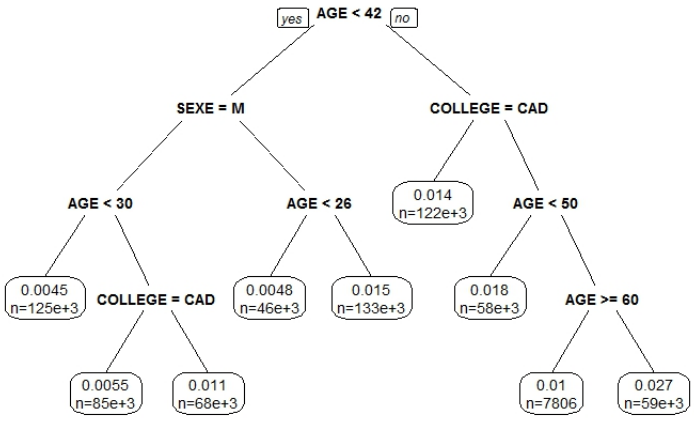


FIGURE VI.66: Exemple arbre CART

Sur chaque noeud apparait une question, par exemple, sur le noeud initial la question est "l'âge de l'individu est-il inférieur à 42 ans ?". Si oui, la question suivante est "l'individu est-il de sexe masculin ?" et ainsi de suite.
Cela signifie donc que la variable qui sépare le mieux l'espace (selon le critère de division) est la variable âge avec la modalité : 42 ans. Ensuite, parmi les individus de moins de 42 ans, la variable qui sépare le mieux l'espace est le sexe etc...
Dans chaque feuille de l'arbre, il est possible de distinguer la taille de l'échantillon notée n et la fréquence d'entrée en incapacité d'un individu. L'algorithme CART donne ainsi un classement entre les différentes variables explicatives de la sinistralité selon leurs degrés d'importance (qui correspond à leur emplacement sur l'arbre). Il permet aussi de déterminer des groupes homogènes de personnes face à l'arrêt de travail, par exemple ici, le premier groupe homogène est constitué des hommes de moins de 30 ans :

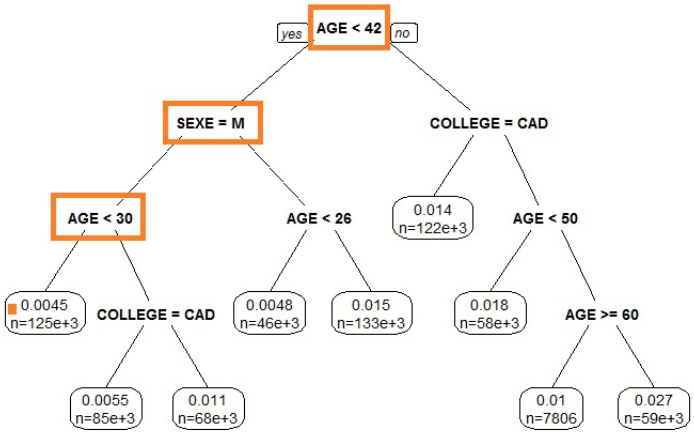


FIGURE VI.67: Arbre CART, groupe homogène 1

Le second groupe homogène est constitué des hommes de plus de 30 ans cadres etc...

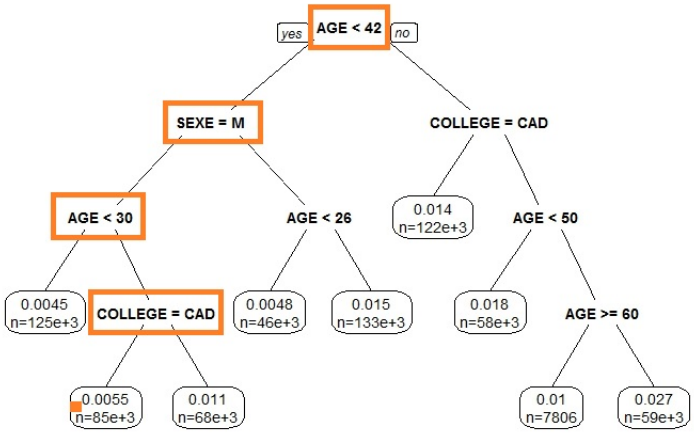


FIGURE VI.68: Arbre CART, groupe homogène 2

VI.2.d) Résultats

Les résultats obtenus à partir des arbres CART sont présentés dans cette partie, ces résultats mettent en évidence un classement entre les différentes variables explicatives et permettent de déterminer des populations homogènes face à la sinistralité.

Les couples du chapitre III vont être étudiés à tour de rôle.

Pour chaque couple, une table composée des caractéristiques des assurés non tronqués ou censurés et du nombre de sinistres par assuré est mise en entrée de l'algorithme. La situation familiale ne sera pas étudié ici en raison de son manque de fiabilité.

Rappel : La taille de l'arbre dépend du paramètre de gain de complexité demandé.

Couple Santé/3jours

L'arbre qui fournit la meilleure estimation (au sens de l'erreur quadratique) est composé de 4 variables, l'âge, le sexe, le collège et la taille de l'entreprise. La grande région n'apporte pas une estimation plus précise.

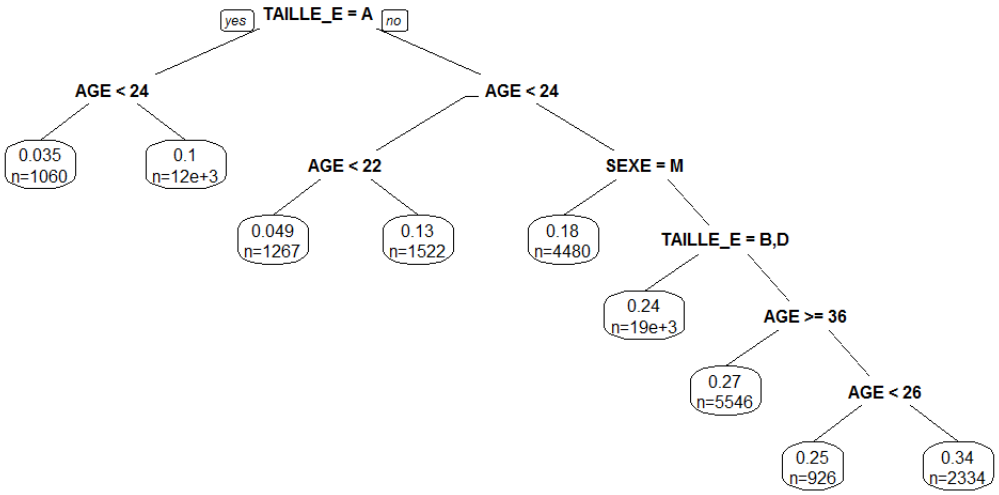


FIGURE VI.69: Arbre CART couple Santé/3 jours

La variable "Taille de l'entreprise" se place en première position avec la modalité A, il est donc nécessaire de différencier les petites entités des grandes ou moyennes.
Dans ces deux arbres , des effets invisibles lors de l'approche classique sont mis en évidence. Par exemple, il apparait qu'il convient de différencier les entreprises de taille B et D, des entreprises de taille C, uniquement pour les femmes de plus de 24 ans. L'approche classique inciterait de son côté à différencier les tailles des entreprises quelles que soient les modalités des autres variables aléatoires.

Les indicateurs de qualité du modèle sont testés sur l'échantillon test :

Erreur Quadratique	Ratio O/A
2738.678	95.24%

Couple ASES/90jours

Le meilleur arbre obtenu pour ce couple est le suivant :

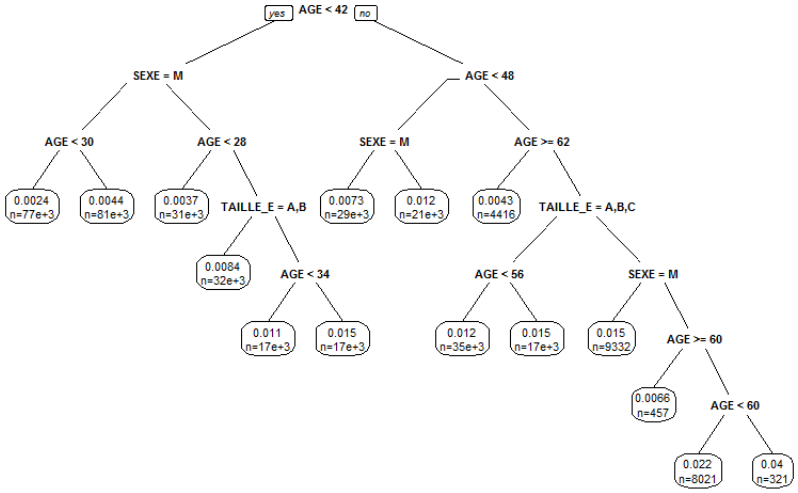


FIGURE VI.70: Arbre CART couple ASES/90 jours

Les résultats des indicateurs de qualité sont les suivants :

Erreur Quadratique	Ratio O/A
769,69	106,27%

Des groupes homogènes face à la sinistralité sont mis en évidence, c'est le cas par exemple des femmes d'entreprises C ou D et âgées de 35 à 41 ans. Des effets peu intuitifs peuvent être observés par exemple le fait que ce type d'individu a une sinistralité équivalente à celle des individus d'une entreprise A, B ou C et âgés de plus de 56 ans quel que soit le sexe.

Couple IM/RC

Le meilleur arbre est le suivant :

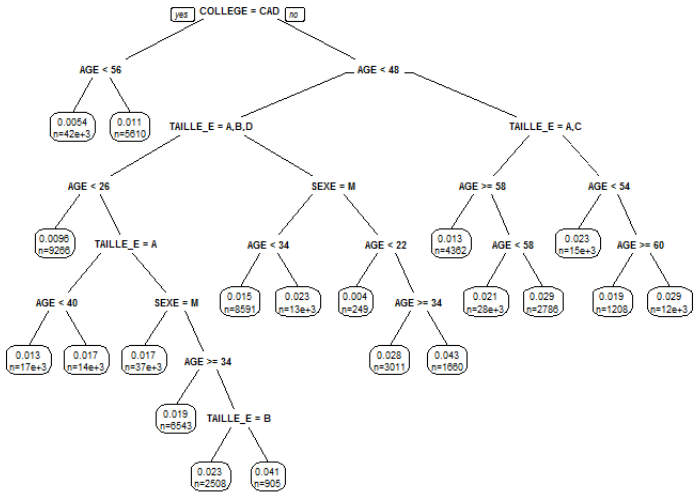


FIGURE VI.71: Arbre CART couple IM/RC

Les résultats des indicateurs de qualité du modèle sont les suivants :

Erreur Quadratique	Ratio O/A
923,24	98,07%

Le groupe d'individus les plus sinistrés est composé des femmes non cadres, d'une entreprise de taille C et d'âge compris entre 24 et 38 ans. Par ailleurs, des groupes ont des sinistralités similaires en moyenne, on peut citer par exemple : les hommes non cadres d'une entreprise de taille C et âgés de moins de 52 ans et les individus hommes ou femmes, non cadres, d'une entreprise de taille A, B ou C et âgés de plus de 58 ans.

Couple AUTRES/RC

Le meilleur arbre est le suivant :

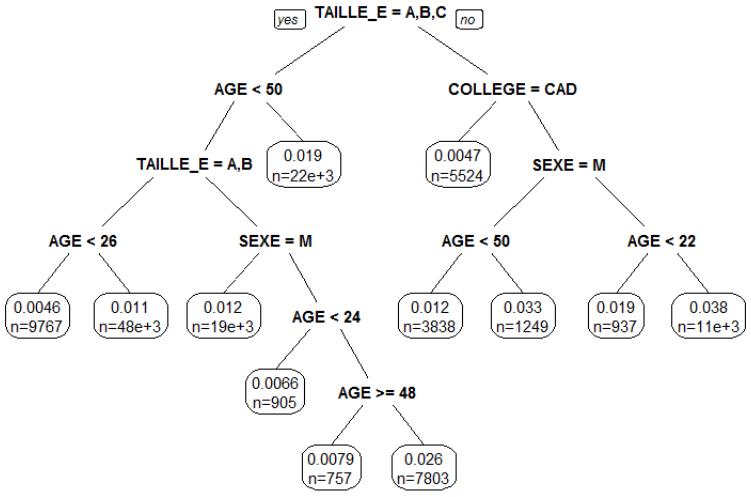


FIGURE VI.72: Arbre CART couple AUTRES/RC

Les résultats des indicateurs de qualité du modèle sont les suivants :

Erreur Quadratique	Ratio O/A
483.69	98,16%

Les individus les plus sinistrés sont les femmes non cadres d'une entreprise de taille D et d'âge supérieur à 22 ans. Par ailleurs, des effets non intuitifs sont mis en évidence comme la proximité de la sinistralité des individus hommes ou femmes d'âge compris entre 26 et 49 ans d'une entreprise de taille A ou B et les hommes d'une entreprise de taille C ou D d'âge inférieur à 50 ans.

VI.2.e) Avantages et inconvénients des arbres CART

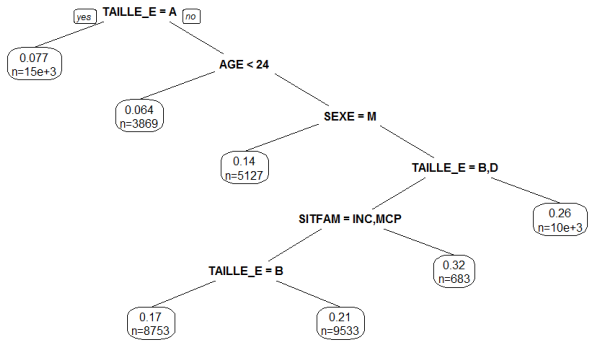
Pour l'étude des variables explicatives de la sinistralité arrêt de travail et la mise en évidence de groupes homogènes, l'algorithme CART présente de nombreux avantages :

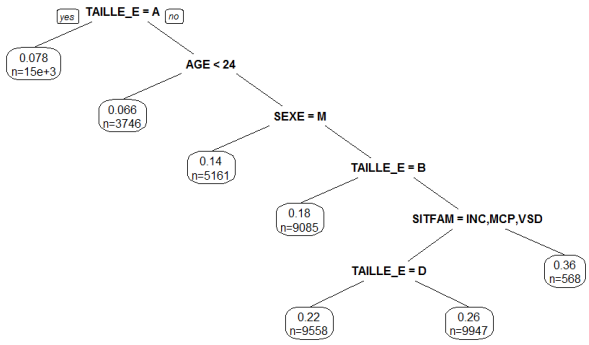
- Aucune hypothèse de distribution n'est faite sur les variables
- L'algorithme supporte des variables quantitatives ou qualitatives
- L'algorithme procure une vision intuitive de la sinistralité de combinaisons de variables aléatoires.

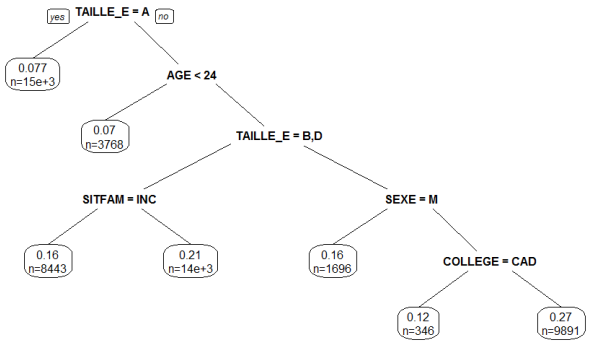
Cependant, l'algorithme CART possède plusieurs points faibles. Le plus important est la robustesse limitée de la prédiction de la variable réponse. Pour illustrer ce point, trois tirages aléatoires avec

remises sont réalisés dans les effectifs de la franchise 30 jours pour créer trois échantillons d'apprentissage et de validation.

Les arbres obtenus sont les suivants :







La structure des arbres varie à partir d'un certain niveau.

Bien qu'ils procurent une vision intuitive des variables explicatives de la sinistralité et que cet algorithme soit très intuitif dans son déroulement, les arbres obtenus sont instables. Par ailleurs, le classement par ordre d'importance des variables explicative est difficilement décelable, chaque variable est séparée selon ces modalités, parfois à des niveaux différents des arbres.

Pour atténuer ces effets et renforcer la robustesse des arbres, il est possible de recourir à des forêts.

VI.3 - Les forêts aléatoires

L'algorithme est présenté ici dans ses grandes lignes, plus de détails et les démonstrations techniques le lecteur pourra se reporter à différentes références bibliographiques présentes en annexe.

VI.3.a) L'algorithme

Une *random forest* ou forêt aléatoire est un ensemble d'arbres CART. L'objectif de ces forêts est de fournir un estimateur se basant sur une moyenne des estimations des différents arbres CART qui la composent.

Si une forêt est composée de N arbres CART, la prédiction de l'observation Y_i sera :

$$Y_i^{pred} = \frac{1}{N} . \sum_{k=1}^{N} Y_i^{CART_k}$$

Dans une forêt aléatoire de régression, chaque arbre est construit de la manière suivante :

1. Création d'un échantillon bootstrap (tirage avec remise) dans l'échantillon d'apprentissage
2. Construction de l'arbre de régression sur cet échantillon bootstrap
3. A chaque noeud, m variables parmi les variables présentes sont tirées aléatoirement. La division optimale est cherchée sur ces m variables.
4. Un arbre est obtenu, il n'est pas élagué.

Les forêts de régression permettent de donner une prédiction plus robuste de la variable réponse. Ils permettent par ailleurs de classer les variables aléatoires par ordre d'importance. Pour ce faire le gain d'homogénéité résultant d'une division selon chaque variable est mesuré sur chaque arbre de la forêt. Ces montants sont ensuite sommés sur tous les arbres de la forêt.

VI.3.b) Les résultats

De la même façon que pour l'algorithme CART, les différents couples sont étudiés séparément. L'algorithme est lancé sans la variable "situation familiale".

Les résultats présentés ici sont les gains d'homogénéité dus à chaque variable. Ainsi, un classement entre les variables aléatoires est effectué. La connaissance de ce classement est un outil d'aide à la décision et un moyen de mieux connaitre le portefeuille. Par exemple, si l'algorithme montre que la taille de l'entreprise est la variable explicative la plus importante, l'assureur a tout intérêt à sélectionner ses futurs contrats ou à proposer des contrats différents selon la taille de l'entreprise cliente.

Les forêts sont toutes composées de 200 arbres et la taille minimale des noeuds de chaque arbre est fixé arbitrairement à 2000 (pour conserver des effectifs conséquents et diminuer le risque d'erreur du aux spécificités des effectifs étudiés). Plusieurs nombres m (évoqué précédemment) sont testés. La valeur de m telle que l'erreur quadratique de la prédiction sur l'échantillon de validation soit minimale est conservée.

Par ailleurs, pour vérifier l'utilité du recours à ces forêts, une comparaison des erreurs de prédictions entre l'arbre optimal CART et la forêt aléatoire est réalisée. Ces comparaisons sont réalisées sur la prédiction de l'échantillon test.

Couple Santé/3jours

Les valeurs prises par les indicateurs de qualité du modèle sont les suivantes :

Algorithme	**Erreur quadratique totale**	**Ratio O/A**
CART	2738,678	95,24%
Random Forest	2735,631	95,4%

La modélisation par forêt aléatoire est légèrement meilleure qu'un unique arbre CART. Les gains d'homogénéité sont les suivants :

Variable	Sexe	Age	Collège	Taille entreprise
Gain	20,7	64,05	3,57	190,1

En conclusion, le classement des variables explicatives est le suivant :

1. Taille entreprise
2. Age
3. Sexe
4. Collège

Il est possible de conclure que pour le secteur de la santé, la taille de l'entité joue un rôle important. Le classement entre les variables restantes est quant à lui intuitif.

Couple ASES/90jours

Les indicateurs de qualité du modèle prennent les valeurs suivantes :

Algorithme	**Erreur quadratique totale**	**Ratio O/A**
CART	769,95	106,27%
Random Forest	768,51	105,1%

Les gains d'homogénéité obtenus sont les suivants :

Variable	Sexe	Age	Collège	Taille entreprise
Gain	2,185	6,87	0,32	1,13

En conclusion, le classement est le suivant :

1. Age
2. Sexe
3. Taille entreprise
4. Collège

Pour le secteur ASES, on en déduit donc que la variable la plus importante est l'âge de l'assuré.

Couple IM/RC

Les indicateurs de qualité du modèle prennent les valeurs suivantes :

Algorithme	**Erreur quadratique totale**	**Ratio O/A**
CART	923,24	98,07%
Random Forest	922,69	98,3%

Les gains d'homogénéité sont les suivants :

Variable	Sexe	Age	Collège	Taille entreprise
Gain	2,67	4,48	4,64	4,21

En conclusion, le classement est le suivant :

1. Collège
2. Age
3. Taille entreprise
4. Sexe

Couple AUTRES/RC

Les indicateurs de qualité du modèle prennent les valeurs suivantes :

Algorithme	**Erreur quadratique totale**	**Ratio O/A**
CART	483,69	98,16%
Random Forest	483,28	98,217%

Les gains d'homogénéité sont les suivants :

Variable	Sexe	Age	Collège	Taille entreprise
Gain	2,37	4,40	2,88	5,32

En conclusion, le classement est le suivant :

1. Taille entreprise
2. Age
3. Collège
4. Sexe

VI.4 - Conclusion

Dans cette partie, la structure du portefeuille et un classement entre les variables aléatoires ont été mis en évidence. L'algorithme CART a permis de donner une vision intuitive des causes (et du lien entre les variables explicatives) de l'entrée en incapacité. Les forêts aléatoires sont quant à elles venues confirmer certains effets mis en avant par les arbres CART et ont renforcé la robustesse de l'étude des variables explicatives.

De cette partie, il faut retenir les points suivants :

- L'âge de l'assuré et la taille de l'entreprise sont deux variables très importantes
- Chaque secteur possède ses particularités (un classement des facteurs explicatifs différent)
- Chaque secteur possède des populations à la sinistralité très élevée et d'autres à sinistralité très faible (visibles dans les arbres CART)

VII - Approche par arbres, calcul de fréquences

La partie précédente a permis de déterminer une hiérarchie entre les variables explicatives au sein des secteurs et de mettre en évidence des groupes homogènes face à la sinistralité. Cette approche est un complément de l'approche classique. Elle apporte plus d'information.

A partir des arbres de décision, il est aussi possible de déterminer des fréquences d'entrée en incapacité. Le modèle approchant au mieux la sinistralité réelle sera retenu.

Comme précédemment, les franchises continues 90, 30 et 3 jours et "en relai" seront analysées l'une après l'autre. L'algorithme *Random Forest* est utilisé en raison de sa robustesse face à l'algorithme CART.

VII.1 - Les spécificités du modèle

VII.1.a) Les données censurées

A nouveau, le logiciel R et le package *Random Forest* sont utilisés. Cela a été évoqué précédemment, les censures et troncatures ne sont pas gérées avec ces outils. Cela pose t-il problème dans l'estimation des fréquences d'entrée en incapacité ?
Pour répondre à cette question, les fréquences moyennes pour les différentes franchises sont calculées sur l'ensemble des individus puis sur les individus non censurés :

Franchise	**Ensemble des individus**	**Individus non censurés/tronqués**
RC	1,705%	1,607%
3 jours	20,936%	18,974%
30 jours	2,561%	2,494%
90 jours	0,815%	0,799%

Rappel : un individu est censuré/tronqué lorsque son exposition à un âge donné est inférieure à 1.

Les résultats de ce test montrent bien que la sinistralité est sous estimée si l'on ne considère que les individus non censurés et non tronqués. Cela est presque négligeable pour les franchises longues, ça ne l'est pas pour les franchises plus courtes.

Par conséquent, il est possible que les fréquences modélisées dans cette partie sous estiment légèrement le risque réel.

VII.1.b) Les variables

Les variables explicatives âge, collège, sexe, taille de l'entreprise et secteur d'activité sont ici intégrées dans l'algorithme.

La taille de l'entreprise cliente n'est pas regroupée en classes. Ces classes sont en effet spécifiques à chaque secteur, cela poserait des problèmes de logique au sein du modèle. Cette variable est donc traitée comme une variable continue. L'algorithme mettra alors en évidence des modalités pivots importantes (par exemple : entreprise de taille $<$ 130 salariés etc...).

VII.2 - Les résultats

Les résultats pour les franchises 30 et 90 jours sont présentés dans cette partie. Les résultats pour les franchises 3 jours continues et "en relai" sont présents en annexe. Pour chaque franchise, la somme des erreurs quadratiques et le ratio observé/attendu sont calculés sur l'échantillon test. Ce ratio est ensuite calculé pour les différents secteurs d'activité.

VII.2.a) Franchise 90 jours

Au global, les résultats obtenus sont les suivants :

Observés	**Prédit**	**Ratio O/A**
1243	1231	100,97%

L'estimation est sous estimée de 1%.

Par secteur, on obtient :

Secteur	**Ratio O/A**
ASES	104,02%
AUTRES	85,57%
IM	98,75%
Santé	104,7%

L'estimation est bonne pour les secteurs ASES, IM et santé. Le modèle sur-estime de 15% le nombre de sinistres pour les autres secteurs.

VII.2.b) Franchise 30 jours

Au global, les résultats obtenus sont les suivants :

Observés	Prédit	Ratio O/A
1 809	1 853,75	97,6%

Par secteur, on obtient :

Secteur	**Ratio O/A**
ASES	94,3%
AUTRES	87,7%
IM	111,7%
Santé	110,6%

VII.3 - Comparaison des approches classiques et par arbres

Il a été mis en évidence précédemment que les algorithmes par arbres (CART ou forêts aléatoires) présentaient de nombreux avantages face à une approche classique par simple comparaison de fréquence lors de la recherche des variables explicatives de la sinistralité. Il en était de même lors de la recherche de groupes homogènes d'individus face à la sinistralité.

Cette partie va maintenant comparer les deux approches du point de vue de la modélisation de fréquences d'entrée en incapacité. Cette comparaison sera dans un premier temps théorique puis pratique. Enfin, les inconvénients de l'approche par arbres pour la modélisation de fréquences d'entrée en incapacité seront mis en évidence.

VII.3.a) Comparaison théorique

D'un point de vue théorique, les algorithmes par arbres sont à préférer à l'approche classique utilisée lors du chapitre 5. En effet, contrairement à cette dernière :

- Les algorithmes par arbre se basent sur un critère de division simples, l'erreur quadratique
- Ils ne font aucune hypothèse sur la distribution du nombre d'arrêt de travail par individus à un âge x
- Ils n'obligent pas la réalisation de segmentations selon toutes les modalités des variables explicatives, ils ne gardent que les segmentations "nécessaires"

Cette approche est donc beaucoup plus souple, elle n'a pas recours à des hypothèses complexes (parfois non vérifiées) et permet une segmentation fine et justifiée.

A titre d'illustration, utilisons l'exemple suivant :
Que ce soit le calcul de coefficient de réduction/majoration, le modèle de Brass ou le modèle de Cox, les non-cadres verront toujours leur taux d'entrée en incapacité (et leur tarif pour aller plus loin) multiplié par un nombre supérieur à 1. Cependant, on peut observer dans notre étude que les jeunes non-cadres (25 ans) se comportent comme les cadres de 35 ans ! Par conséquent il est injustifié d'appliquer à tous les non cadres ce coefficient correcteur. Les algorithmes par arbres peuvent isoler des effets comme cela.

VII.3.b) Comparaison pratique

Théoriquement, l'approche par arbre est préférée à l'approche classique. L'échantillon test de la modélisation de la forêt aléatoire pour la franchise 90 jours est intégré dans le modèle classique. Les projections des deux modèles sont comparées à l'aide du ratio "Observé/Attendu". Ce premier test s'effectue sur des données non censurées/tronquées :

Test	O/A classique	O/A par arbres
Données complètes	104,4%	99,05%

La prédiction des forêts aléatoires est toujours meilleure que celle de l'approche classique mêlant positionnement et modèle de Cox. Cela est aussi vérifié si l'on compare les prédictions des secteurs

d'activité une à une.

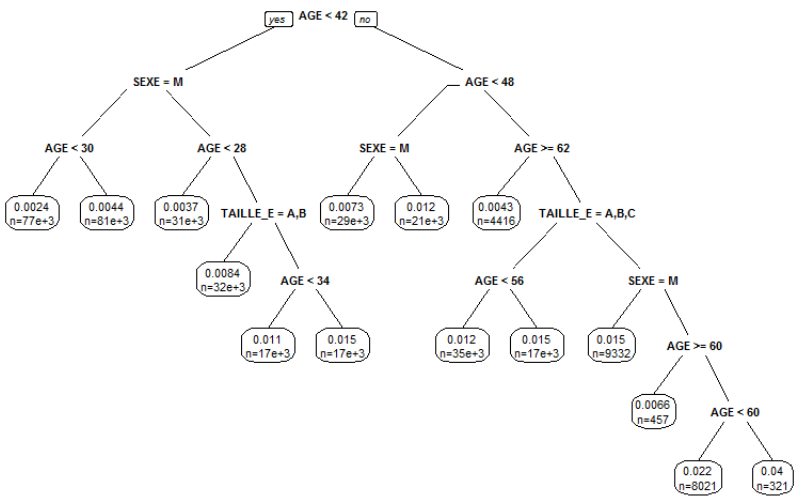
Un échantillon est ensuite créé à partir de l'ensemble des données, il est donc composé de données censurées/tronquées. Les résultats sont les suivants :

Test	O/A classique	O/A par arbres
Ensemble des données	110,3%	103,9%

Le modèle par arbre sous-estime la sinistralité. Ce résultat était attendu (cela est explicité précédemment). Cependant il reste le meilleur modèle.

VII.3.c) Faiblesses de l'approche par arbres

Pour le calcul de fréquence et la projection de la sinistralité, l'approche par arbres est donc meilleure que l'approche classique (tout du moins pour la franchise 90 jours). Elle présente cependant des inconvénients. L'inconvénient le plus marquant lors de la recherche d'une loi d'entrée en incapacité est la modélisation de cette loi "en escalier". Pour éviter les problèmes de sur-apprentissage, l'algorithme a en effet tendance à regrouper un certain nombre d'âge entre eux (de la même façon, pour les forêts aléatoires, l'utilisateur aura tendance à limiter le nombre de feuilles après différentes mesures de l'erreur de prédiction de l'algorithme). Pour illustrer ce point, prenons l'exemple de l'arbre CART suivant (à noter que l'inconvénient persiste pour les forêts aléatoires qui sont composées de plusieurs arbres CART) :



A partir de cet arbre, il est possible de déterminer une loi d'entrée en incapacité pour les tous les individus. Prenons l'exemple des hommes d'une entreprise de taille A, B ou C. La loi obtenue est la suivante :

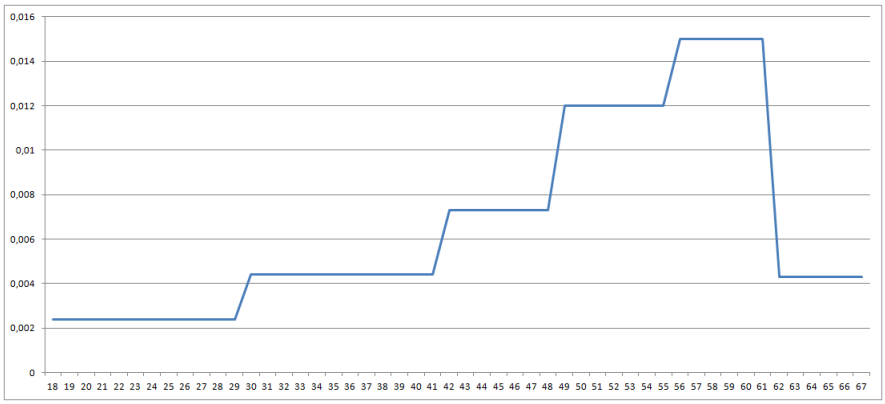


FIGURE VII.73: Loi d'entrée en incapacité CART

Les taux d'entrée en incapacité sont constants sur différentes tranches d'âges, la prédiction sera biaisée, en effet, pour les hommes, on a vu précédemment que les taux d'entrée en incapacité étaient croissants avec l'âge.

Par ailleurs, cela a été montré précédemment, cette modélisation devrait sous-estimer légèrement l'entrée en incapacité (les données non censurées ne sont pas utilisées). Une meilleure estimation serait obtenue à l'aide d'une modification des algorithmes dans R, cependant plutôt que d'entreprendre cette démarche coûteuse en temps et pour remédier à la faiblesse présentée précédemment, une nouvelle approche est privilégiée.

VIII - Approche hybride

Face aux inconvénients des modèles précédents, une troisième approche est étudiée ici. Elle se base sur l'approche par arbres pour la segmentation du portefeuille et sur les tables d'entrée en incapacité de l'approche classique.

Les modèles et résultats obtenus pour les franchises 30 jours et 90 jours sont présentés ici. Le modèle obtenu pour la franchise "en relai" est disponible en annexe. La franchise 3 jours n'est pas étudiée ici, les effectifs ne sont pas assez conséquents pour appliquer cette approche.

VIII.1 - Démarche et avantages

VIII.1.a) Démarche

Dans ce modèle, un arbre CART est calculé sur l'ensemble des données non censurées ou tronquées d'une franchise. Les variables en entrée de l'arbre sont : le sexe, le collège, la taille de l'entreprise (qui est une variable continue, aucune classe n'est calculée) et le secteur d'activité. L'âge n'est pas inclus. Ainsi, une segmentation est effectuée sans hypothèses contraignantes. L'algorithme est paramétré de telle sorte que les feuilles disposent d'une quantité de données conséquente.

Une table des taux bruts est ensuite calculée sur l'ensemble des données puis lissée pour chaque feuille avec les estimateurs et outils du chapitre IV.

VIII.1.b) Avantages et critique de cette approche

Cette approche permet de segmenter le portefeuille de manière simple et sans faire d'hypothèse importante, à l'aide de la théorie des arbres de décision. Toute segmentation "superflue" est ainsi évitée. De plus, elle permet de calculer des lois continues d'entrée en incapacité et d'apprécier au mieux la sinistralité à chaque âge. Enfin, cette approche permet d'intégrer l'ensemble des données dans la modélisation (l'arbre est réalisé sur les données non censurées, les tables sont réalisées sur l'ensemble des données).

La principale critique concernant cette approche provient de la stabilité limitée des arbres CART, cela a déjà été évoqué précédemment et posait problème lorsque la prédiction ne se basait que sur l'algorithme. Dans cette approche hybride, l'algorithme est utilisé pour segmenter le portefeuille, il faut donc vérifier sur différents échantillons que cette segmentation ne varie pas trop. Pour se faire, la démarche est la suivante :

- Calcul d'un arbre sur l'ensemble des données non censurées par franchise
- Calcul d'une dizaine d'arbres CART sur des échantillons bootstrap différents
- Vérification que la structure de la segmentation reste stable
- Modélisation d'une forêt aléatoire de 400 arbres, vérification de la stabilité de l'ordre des variables (et donc de la segmentation)

VIII.2 - Lois obtenues

Pour chaque franchise, un arbre et les lois d'entrée en incapacité propres à chaque feuille sont présentés ici.

VIII.2.a) Franchise 90 jours

L'arbre de segmentation obtenu est le suivant, le secteur de la santé n'est à nouveau pas étudié pour cette franchise :

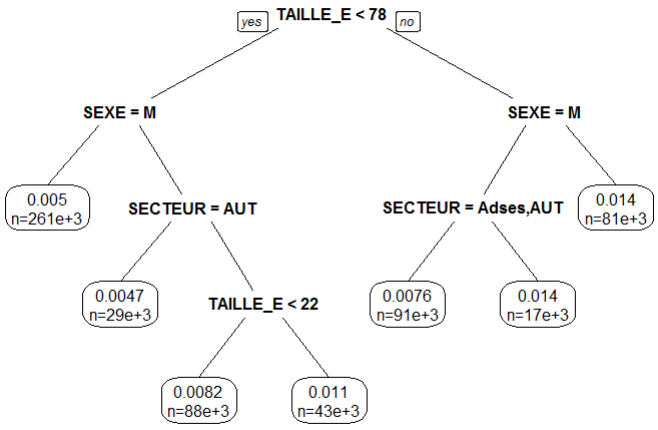


FIGURE VIII.74: Approche hybride, arbre CART franchise 90 jours

Cet arbre dispose de 8 feuilles. 8 lois sont donc obtenues. Pour les hommes, trois lois sont calculées et lissées :

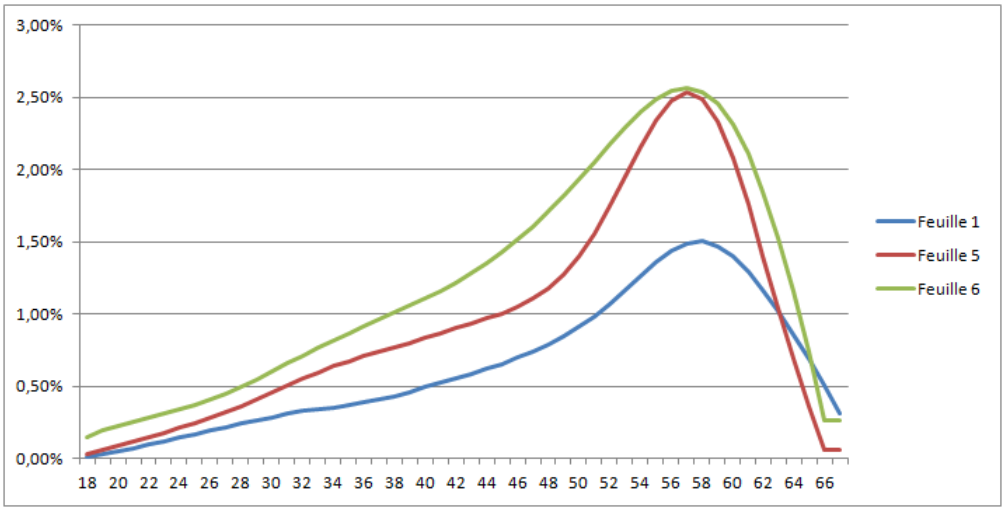


FIGURE VIII.75: Lois d'entrée en incapacité hommes

Les feuilles regroupent les individus suivants :

- Feuille 1 : Les hommes d'une entreprise de taille inférieure à 78 salariés
- Feuille 5 : Les hommes des secteurs "Autres" et "ASES" d'une entreprise de taille supérieure ou égale à 78 salariés
- Feuille 6 : Les hommes du secteur de l'industrie manufacturière d'une entreprise de taille supérieure ou égale à 78 salariés

Pour les femmes, cinq lois sont calculées :

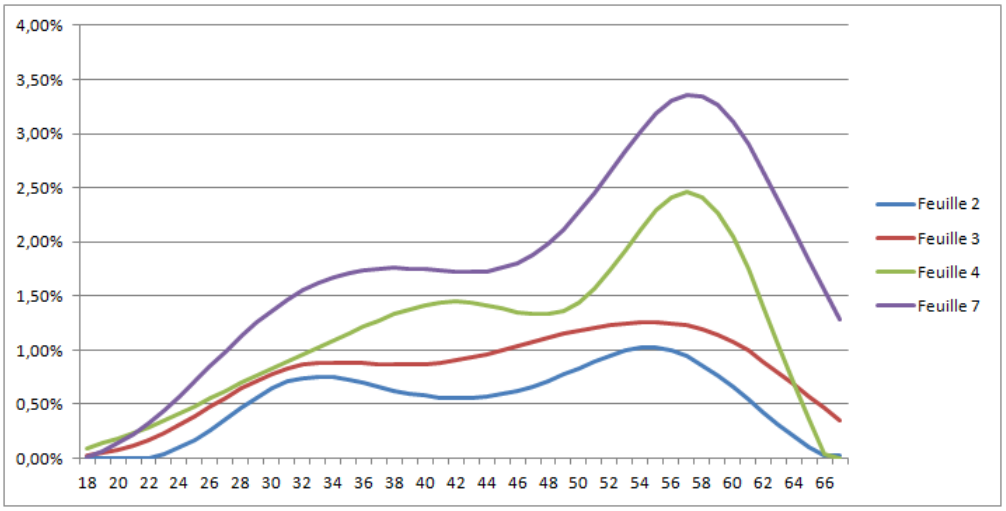


FIGURE VIII.76: Lois d'entrée en incapacité femmes

Les feuilles regroupent les individus suivants :

- Feuille 2 : Les femmes du secteur "Autres" d'une entreprise de taille inférieure à 78 salariés
- Feuille 3 : Les femmes des secteurs "ASES" et "Industrie manufacturière" d'une entreprise de taille inférieure à 22 salariés
- Feuille 4 : Les femmes des secteurs "ASES" et "Industrie manufacturière" d'une entreprise de taille comprise entre 22 et 77 salariés
- Feuille 7 : Les femmes d'une entreprise de taille supérieure ou égale à 78 salariés

Que ce soit pour les hommes ou les femmes, la forme des courbes est proche de celles modélisées sous l'ensemble de l'échantillon de la franchise 90 jours au chapitre V). Cela est rassurant quant à la légitimité de cette approche.

VIII.2.b) Franchise 30 jours

L'arbre de segmentation obtenu est le suivant, à nouveau, le secteur de la santé n'est pas inclus dans l'étude :

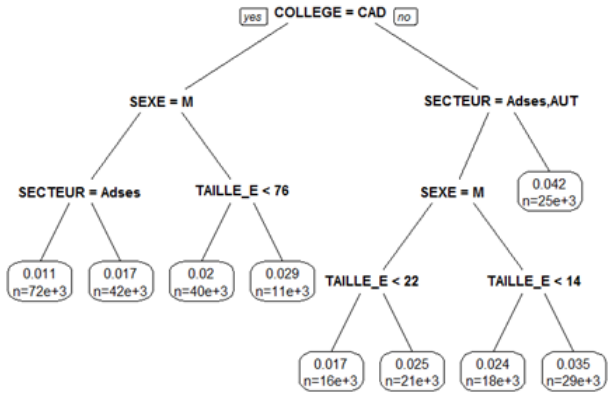


FIGURE VIII.77: Approche hybride, arbre CART franchise 30 jours

Pour les hommes, quatre lois sont obtenues :

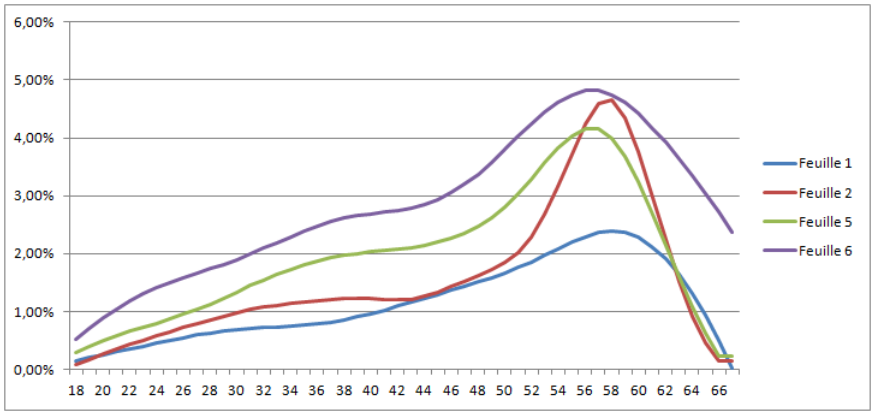


FIGURE VIII.78: Lois d'entrée en incapacité hommes 30 jours

Les feuilles regroupent les individus suivants :

- Feuille 1 : Les hommes cadres du secteur "ASES"
- Feuille 2 : Les hommes cadres des secteurs "IM" et "AUTRES"
- Feuille 5 : Les hommes non cadres des secteurs "ASES" et "AUTRES" d'une entreprise de taille inférieure à 22 salariés
- Feuille 6 : Les hommes non cadres des secteurs "ASES" et "AUTRES" d'une entreprise de taille supérieure ou égale à 22 salariés

Pour les femmes cinq lois sont obtenues :

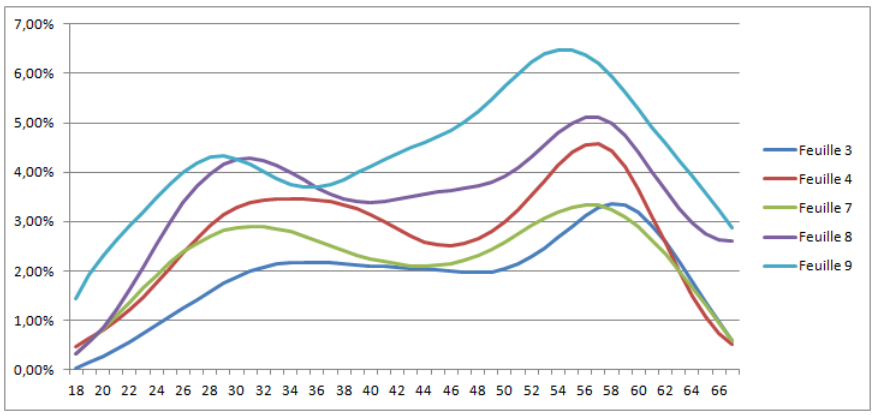


FIGURE VIII.79: Lois d'entrée en incapacité femmes et non cadres du secteur IM, 30 jours

- Feuille 3 : Les femmes cadres d'une entreprise de taille inférieure à 76 salariés
- Feuille 4 : Les femmes cadres d'une entreprise de taille supérieure ou égale à 76 salariés
- Feuille 7 : Les femmes non cadres d'une entreprise des secteurs ASES et AUTRES de taille inférieure à 14 salariés
- Feuille 8 : Les femmes non cadres d'une entreprise des secteurs ASES et AUTRES de taille supérieure ou égale à 14 salariés
- Feuille 9 : Les individus non cadres du secteur IM

A nouveau, la forme des courbes est proche de celles obtenues au chapitre V. En effet, la croissance des fréquences avec l'âge est visible pour les hommes et le pic de sinistralité aux alentours de 30 ans est visible pour les femmes.

VIII.3 - Test du modèle

Le modèle hybride est testé et comparé aux deux autres modèles (Approche classique et approche par arbre). Ce test s'effectue sur un échantillon de lignes (censurées/tronquées ou non) tirées aléatoirement dans les effectifs ayant une franchise 90 jours.

VIII.3.a) Franchise 90 jours

L'observé est de 1 960 sinistres. L'échantillon testé contient 300 000 lignes.

Approche	Prédit	Ratio O/A
Classique	1 785	109,8%
Forêts aléatoires	1 836	106,2%
Hybride	1 998	98,1%

Le modèle hybride fournit la meilleure estimation de la sinistralité en terme de ratio "O/A". Ce test est réitéré plusieurs fois sur des données différentes et le classement entre les modèles reste identique.

VIII.3.b) Franchise 30 jours

L'observé est de 3 530 sinistres. L'échantillon testé contient 200 000 lignes

Approche	**Prédit**	**Ratio O/A**
Classique	3 421	103,2%
Forêts aléatoires	3 098	114%
Hybride	3 458	102%

Il est bien visible ici que la réalisation d'un modèle par arbre sans données censurées sous estime fortement la sinistralité réelle lorsqu'il est testé sur des données censurées pour une franchise assez courte. Le modèle hybride est à nouveau le meilleur modèle.

VIII.4 - Conclusion

La modélisation "hybride" contourne les inconvénients des deux approches précédentes, elle sera retenue pour prédire les fréquences d'entrée en incapacité d'un groupe et comparer la sinistralité d'une entité à celle du portefeuille étudié dans son ensemble.

Les modèles hybrides obtenus pour les franchises 30 et 90 jours peuvent être adaptés aux autres franchises continues et discontinues avec les tables de correspondance et la loi de maintien du BCAC. On suppose alors que la segmentation réalisée par l'arbre à 30 jours est identique à 45, 60 et 75 jours et que celle obtenue à 90 jours est identique à 120 et 150 jours. A noter que pour être totalement généralisable, il serait nécessaire de trouver des solutions à apporter pour que la segmentation pour la franchise 30 jours et celle pour la franchise 90 jours soient plus proches.

IX- Utilisation pratique de l'étude

IX.1 - Connaitre les causes de l'entrée en incapacité

Les différentes approches proposées apportent chacune des informations complémentaires, qui permettent de mieux connaître la structure du portefeuille et l'influence de variables explicatives sur l'entrée en incapacité. Ces informations fournissent des outils décisionnels intéressants.

Exemple :

Le secteur AUTRES est étudié. Le classement des variables explicatives mis en évidence dans le chapitre VI est le suivant :

Classement	Variable
1	Taille entreprise
2	Age
3	Collège
4	Sexe

La taille de l'entreprise cliente est donc la variable la plus importante pour ce secteur. A partir de ce constat, il est intéressant de revenir aux résultats du chapitre III.

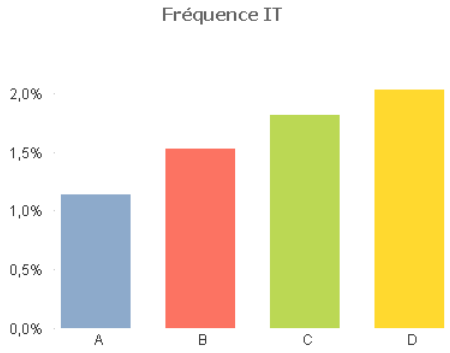


FIGURE IX.80: Fréquence par taille d'entreprise, secteur AUTRES

Ce graphique montre que la fréquence d'entrée en incapacité est croissante avec la taille de l'entreprise. Il serait donc intéressant de souscrire des entreprises de tailles plus faibles ou tout du moins, de tenir compte de cette information dans la tarification.

Par ailleurs, l'arbre CART pour le couple Autres/RC permet d'isoler des populations ayant une sinistralité particulièrement élevée :

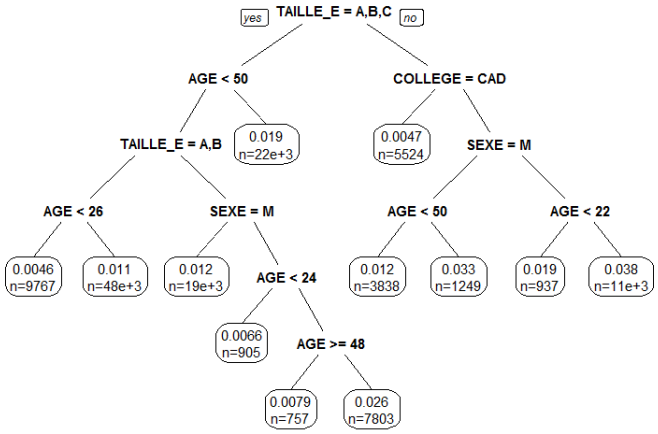


FIGURE IX.81: Arbre CART couple AUTRES/RC

On peut distinguer par exemple les femmes d'âge compris entre 24 et 48 ans qui ont une sinistralité plus élevée que les individus de plus de 50 ans d'une entreprise de taille A,B ou C. Cette observation non intuitive peut inciter Malakoff-Médéric à lancer des campagnes de sensibilisation ou d'aide à ces individus.

IX.2 - Tarification

La tarification d'un contrat arrêt de travail se base sur un calcul ligne à ligne "fréquence-coût". Les fréquences calculées dans ce mémoire peuvent être utilisées pour cette tarification. Il serait cependant nécessaire de mener une étude complémentaire (segmentée selon les mêmes variables explicatives que dans ce mémoire) sur le coût de l'incapacité.

IX.3 - Déterminer un benchmark de l'incapacité

La modélisation de ce mémoire est utilisée dans un outil. Ce dernier permet de déterminer (indépendamment d'autres indicateurs comme le coût moyen ou le ratio sinistres sur primes) des entreprises pour lesquelles la sinistralité est anormalement élevée par rapport au reste du portefeuille. L'outil permet, à partir de la modélisation de l'approche hybride et d'une table des effectifs, de déterminer la fréquence d'entrée en incapacité que devrait avoir une entreprise en fonction de sa population. Ce taux est ensuite comparé au taux réel observé sur une période.

IX.3.a) Exemple 1, étude d'une convention collective

Une convention collective du secteur ASES est sélectionnée. On souhaite déterminer quels SIREN parmi les 250 plus importants de cette convention, ont des taux d'entrée en incapacité supérieurs au reste du portefeuille. Utiliser un simple taux d'entrée sans tenir compte de la population assurée donnerait une information inexacte. Il est donc nécessaire d'utiliser la modélisation réalisée.

La sinistralité réelle est comparée à la sinistralité prédite à l'aide du ratio O/A. Le graphique

suivant présente le ratio en fonction de la taille des SIREN.

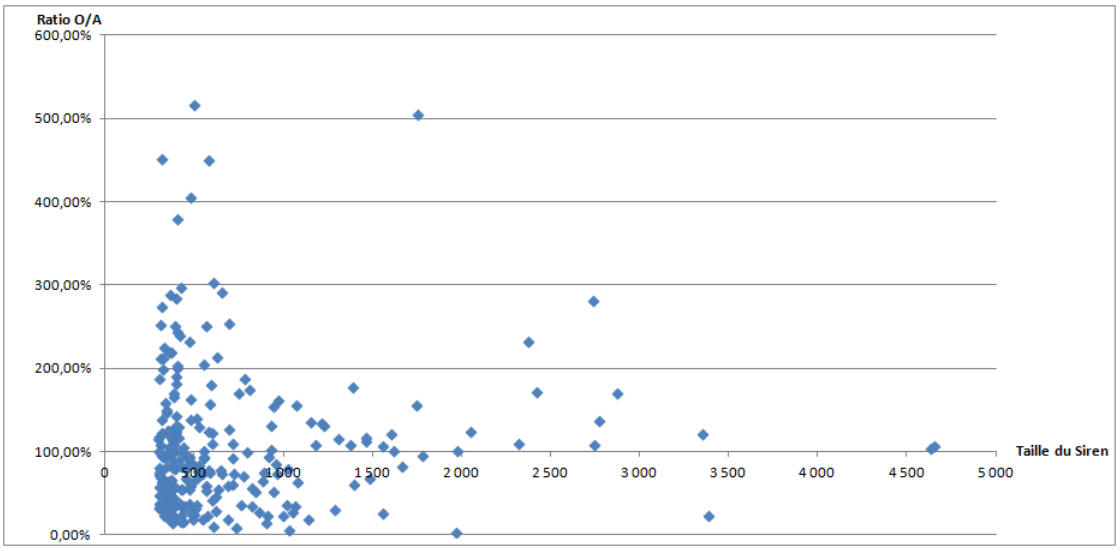


FIGURE IX.82: Positionnement des différents SIREN par rapport à la sinistralité prédite

La majorité des SIREN ont un ratio O/A compris entre 0 et 200%. Une quantité non négligeable a cependant une sinistralité plus de deux fois supérieure à la moyenne.

A partir de cette information et en s'appuyant sur d'autres indicateurs comme le coût moyen et le ratio S/P de ces SIREN, il sera possible de déterminer des actions à entreprendre les concernant.

IX.3.b) Exemple 2, étude d'un grand compte

Un grand compte du secteur ASES est sélectionné. Son ratio O/A est de 268%. Ce compte a donc une sinistralité 2,5 fois supérieure à la sinistralité qu'il aurait dû avoir si sa population se comportait comme le reste du portefeuille. Face à ce constat, il est nécessaire de déterminer plus en détail quel type d'individu entraine cette sinistralité importante.

Les deux tableaux suivants montrent le ratio O/A de l'entreprise par collège et sexe :

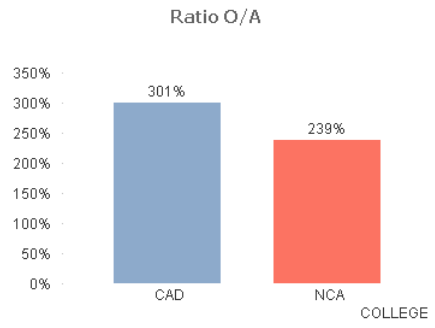


FIGURE IX.83: Ratio O/A en fonction du collège de l'assuré

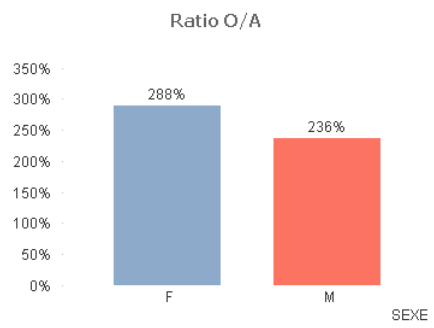


FIGURE IX.84: Ratio O/A en fonction du sexe de l'assuré

On déduit de ces graphiques que ce sont les cadres et les femmes qui ont la sinistralité la plus anormale. Il est aussi possible de s'intéresser à l'âge :

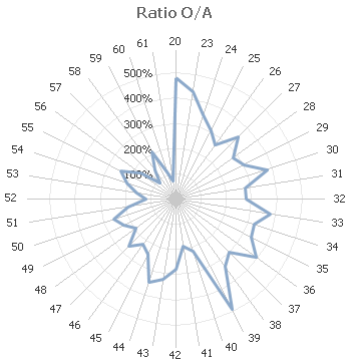


FIGURE IX.85: Ratio O/A en fonction de l'âge de l'assuré

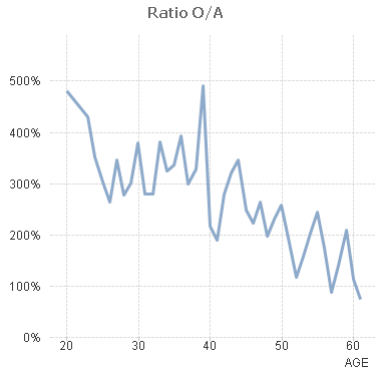


FIGURE IX.86: Ratio O/A en fonction de l'âge de l'assuré

On constate ici que le ratio O/A a tendance à diminuer avec l'âge. Les individus dont la sinistralité est la plus anormale sont les individus âgés de 20 à 40 ans. Cela est aussi visible sur ce graphique :

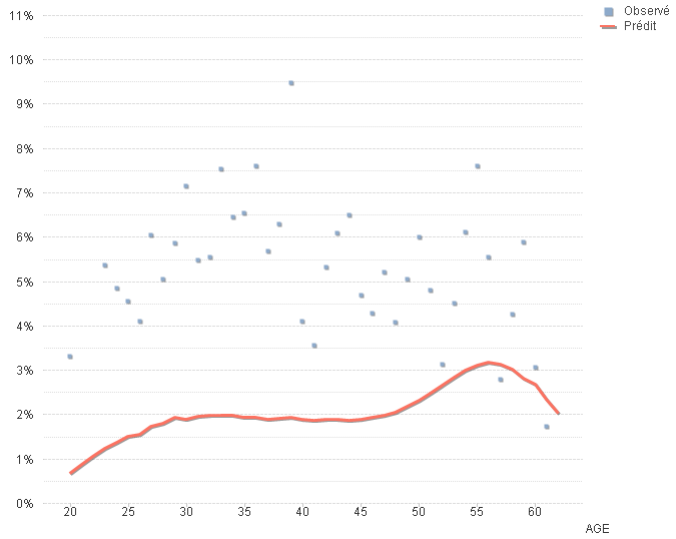


FIGURE IX.87: Fréquence d'entrée en incapacité en fonction de l'âge

La courbe correspond à la prédiction du modèle tandis que les points sont les fréquences d'entrée en incapacité observées à chaque âge pour le groupe étudié. On constate bien que ces fréquences observées sont globalement très supérieures à la prédiction.

Conclusion

Dans le contexte très tendu du marché de la prévoyance collective en France, ce mémoire apporte des réponses à des problématiques opérationnelles et permet à Malakoff-Médéric de mieux connaitre son risque incapacité de travail.

Cette étude a permis de mettre en évidence des variables ayant une influence sur la sinistralité. Parmi ces variables, on peut citer les très classiques sexe, collège et âge, mais aussi la taille de l'entreprise du salarié, son secteur d'activité et même sa situation familiale et sa grande région de domicile (bien que ces deux dernières soient à considérer avec beaucoup de précautions). Ces variables, à l'aide des arbres de décision, ont pu être classées par ordre d'influence au sein des différents secteurs et des groupes homogènes face à la sinistralité ont été déterminés. Ces informations sont des outils décisionnels importants tant lors de la souscription d'une entreprise cliente qu'au cours de la vie du contrat.

Des tables d'entrée en incapacité ont ensuite montré que la sinistralité était bien croissante avec l'âge du salarié pour les hommes mais que cela n'était pas forcément vrai pour les femmes. De la même façon des disparités de forme entre cadres et non cadres (décalage du pic de sinistralité) et entre les secteurs d'activité ont été mis en évidence.

Enfin, l'hétérogénéité du portefeuille a été modélisée pour permettre la construction de lois d'entrée en incapacité segmentées à une maille fine et reflétant au mieux la sinistralité réelle. Plusieurs approches ont été proposées et comparées. La meilleure estimation vient de l'approche hybride. Cette dernière consiste à segmenter le portefeuille à l'aide d'un arbre de décision puis à modéliser des lois d'entrée en incapacité (à l'aide de l'estimateur Poissonien et d'un lissage de Wittaker-Henderson) pour chaque feuille de l'arbre. Cette approche contourne les faiblesses du modèle de Cox et de l'utilisation exhaustive des arbres de décision. C'est ce modèle qui est utilisé pour déterminer un *benchmark* du risque arrêt de travail du portefeuille et pour prédire la sinistralité future.

Ces différentes étapes ont ainsi démontré :

- Que les variables explicatives n'ont pas la même importance pour tous les secteurs.
- Que la sinistralité peut varier de façon importante en intensité et en tendance (avec l'âge) en fonction des caractéristiques d'une population.
- Que les approches algorithmiques fournissent des outils intéressants et précis pour la modélisation de ce risque.

Elles ont permis le développement d'un outil d'identification et de prédiction des fréquences anormales d'entrée en incapacité. Les modélisations retenues sont maintenant utilisées pour vérifier si les lois, barèmes et hypothèses retenues lors de la tarification sont adaptés au risque et orienter des plans d'actions ciblés (redressements, campagnes de sensibilisation...) si besoin.

Suite à cette étude, quatre axes de prolongement et d'amélioration se distinguent :

- La réalisation d'une étude comparable sur la durée de l'arrêt de travail (et donc son coût), elle serait nécessaire pour déterminer des tarifs uniquement basés sur l'expérience du portefeuille de Malakoff-Médéric.

- L'amélioration de la connaissance de la franchise "relai aux conventions collectives", elle permettrait une modélisation plus précise des fréquences sur ce secteur.
- L'automatisation de l'ensemble des étapes de calcul et de reporting pour permettre une modélisation et une mise à jour de l'étude rapide.
- La comparaison des modèles issus des algorithmes par arbre avec des améliorations du modèle de Cox (Aalen, Lin-Ying...).

Lexique

ASES = Secteur des activités de services et scientifiques

IM = Secteur de l'industrie manufacturière

Backtesting = Test de validité d'une modélisation sur une période antérieure

Ratio O/A = Ratio du nombre de sinistres observés sur le nombre de sinistres prédits (outil de *backtesting*)

CART = *Classification and regression trees*

Annexe 1

Lois d'entrée en incapacité, franchises "en relai"

	Franchise "en relai"		
Age	ASES	Industrie	Autres
18	0,77%	0,84%	0,33%
19	0,77%	0,84%	0,33%
20	0,77%	0,84%	0,33%
21	0,77%	1,00%	0,38%
22	0,79%	1,15%	0,46%
23	0,82%	1,26%	0,57%
24	0,86%	1,39%	0,70%
25	0,91%	1,14%	0,83%
26	0,97%	1,35%	0,95%
27	1,04%	1,66%	1,06%
28	1,13%	1,64%	1,16%
29	1,23%	1,43%	1,24%
30	1,33%	1,63%	1,30%
31	1,43%	1,62%	1,34%
32	1,52%	1,65%	1,36%
33	1,58%	1,44%	1,37%
34	1,62%	1,76%	1,38%
35	1,62%	1,85%	1,39%
36	1,60%	1,51%	1,39%
37	1,57%	1,66%	1,40%
38	1,54%	1,65%	1,40%
39	1,52%	1,67%	1,40%
40	1,51%	1,58%	1,39%
41	1,50%	1,68%	1,39%
42	1,51%	1,61%	1,38%
43	1,51%	1,70%	1,38%
44	1,53%	1,68%	1,39%
45	1,56%	1,68%	1,41%
46	1,62%	1,81%	1,45%
47	1,72%	1,77%	1,51%
48	1,87%	1,89%	1,59%
49	2,04%	2,07%	1,68%
50	2,24%	1,79%	1,79%
51	2,43%	1,86%	1,91%
52	2,59%	1,86%	2,04%
53	2,67%	2,10%	2,15%
54	2,66%	2,44%	2,25%
55	2,55%	2,31%	2,31%
56	2,38%	2,40%	2,33%
57	2,17%	2,40%	2,29%
58	1,95%	2,63%	2,19%
59	1,74%	1,82%	2,05%
60	1,54%	1,30%	1,87%
61	1,36%	0,96%	1,67%
62	1,20%	1,46%	1,46%
63	1,08%	1,68%	1,24%
64	0,99%	1,40%	1,01%
65	0,99%	0,86%	0,77%
66	0,99%	0,86%	0,77%
67	0,99%	0,86%	0,77%

Figure IX.88: Lois par secteur, franchise "en relai"

	Franchise "en relai"			
	H		F	
Age	CAD	NCA	CAD	NCA
18	0,25%	0,66%	0,43%	0,70%
19	0,25%	0,66%	0,43%	0,70%
20	0,25%	0,66%	0,43%	0,70%
21	0,25%	0,66%	0,43%	0,70%
22	0,25%	0,77%	0,49%	0,88%
23	0,27%	0,89%	0,57%	1,10%
24	0,29%	1,01%	0,66%	1,34%
25	0,32%	1,12%	0,75%	1,60%
26	0,34%	1,21%	0,85%	1,86%
27	0,35%	1,29%	0,95%	2,10%
28	0,36%	1,37%	1,05%	2,31%
29	0,37%	1,43%	1,14%	2,46%
30	0,38%	1,49%	1,24%	2,56%
31	0,39%	1,54%	1,33%	2,60%
32	0,40%	1,58%	1,43%	2,58%
33	0,42%	1,62%	1,51%	2,52%
34	0,44%	1,66%	1,57%	2,43%
35	0,46%	1,70%	1,63%	2,32%
36	0,48%	1,73%	1,65%	2,22%
37	0,49%	1,76%	1,66%	2,13%
38	0,50%	1,79%	1,64%	2,08%
39	0,51%	1,81%	1,61%	2,07%
40	0,52%	1,82%	1,57%	2,08%
41	0,52%	1,84%	1,52%	2,11%
42	0,53%	1,85%	1,48%	2,15%
43	0,54%	1,86%	1,44%	2,20%
44	0,56%	1,87%	1,42%	2,25%
45	0,60%	1,90%	1,41%	2,29%
46	0,64%	1,94%	1,41%	2,33%
47	0,70%	1,98%	1,44%	2,38%
48	0,78%	2,04%	1,49%	2,45%
49	0,87%	2,11%	1,55%	2,53%
50	0,97%	2,20%	1,62%	2,62%
51	1,08%	2,33%	1,69%	2,72%
52	1,19%	2,48%	1,76%	2,81%
53	1,28%	2,67%	1,84%	2,88%
54	1,35%	2,86%	1,91%	2,91%
55	1,39%	3,02%	1,99%	2,90%
56	1,40%	3,12%	2,05%	2,82%
57	1,37%	3,13%	2,09%	2,69%
58	1,32%	3,00%	2,11%	2,50%
59	1,24%	2,76%	2,08%	2,27%
60	1,15%	2,42%	2,01%	2,02%
61	1,04%	2,06%	1,90%	1,77%
62	0,91%	1,69%	1,76%	1,55%
63	0,78%	1,35%	1,62%	1,39%
64	0,63%	1,02%	1,50%	1,28%
65	0,63%	1,02%	1,50%	1,24%
66	0,63%	1,02%	1,50%	1,24%
67	0,63%	1,02%	1,50%	1,24%

Figure IX.89: Lois par sexe et collège, franchise "en relai"

Modèle avec prise en compte de l'hétérogénéité, franchise "en relai"

Modèle de Cox

La population de référence du modèle est la population non cadres d'une entreprise de taille D du secteur de l'industrie manufacturière.

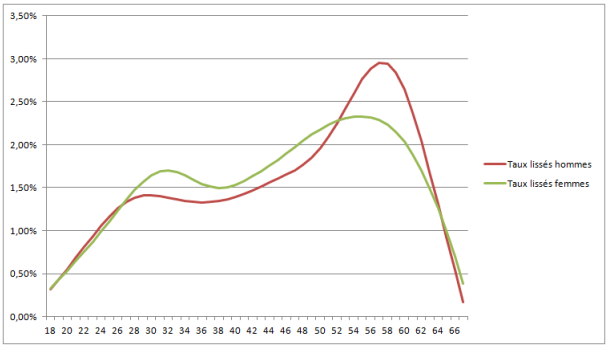


FIGURE IX.90: Taux lissés, franchise "en relai"

Résultats

Le nombre d'incapacités prédit est calculé pour chaque individu non censuré ayant une franchise "En relai". L'observé est de 6772 sinistres.

Les résultats obtenus sont les suivants :

Prédiction	Ratio O/A
7084	95,6%

Au global, la prédiction est proche de l'observé. Par secteur et pour ce modèle, les ratios O/A sont les suivants :

Secteur	Observé	Prédit	Ratio O/A
ASES	908	1 268	71,6%
Autres	2 103	2 021	104,1%
IM	3 761	3 794	99,1%

Ce modèle est de qualité pour les secteurs "Autres" et "IM". Il surestime cependant de façon importante la sinistralité du secteur "ASES".

Approche par arbres de décision

Le modèle est calculé sur les données non censurées. :

Observés	Prédit	Ratio O/A
2 422	2 348	103,14%

Par secteur les résultats suivants sont obtenus :

Secteur	**Ratio O/A**
ASES	123,3%
AUTRES	106,7%
IM	98,8%

Approche hybride

L'arbre de segmentation obtenu est le suivant :

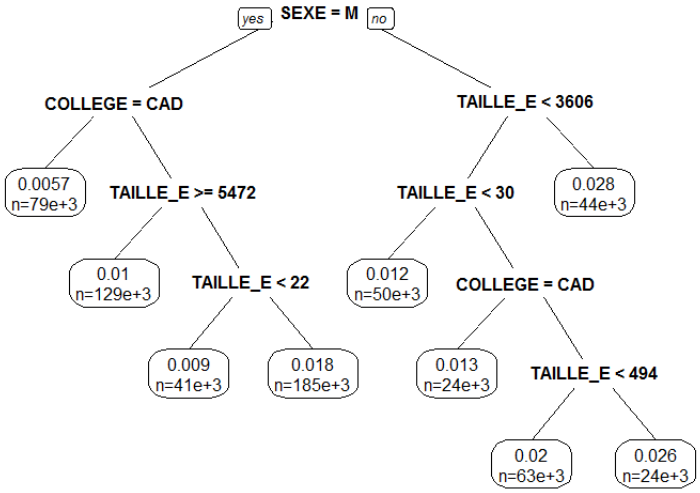


FIGURE IX.91: Approche hybride, arbre CART franchise "En relai"

Les lois obtenues pour chaque feuille sont les suivantes :

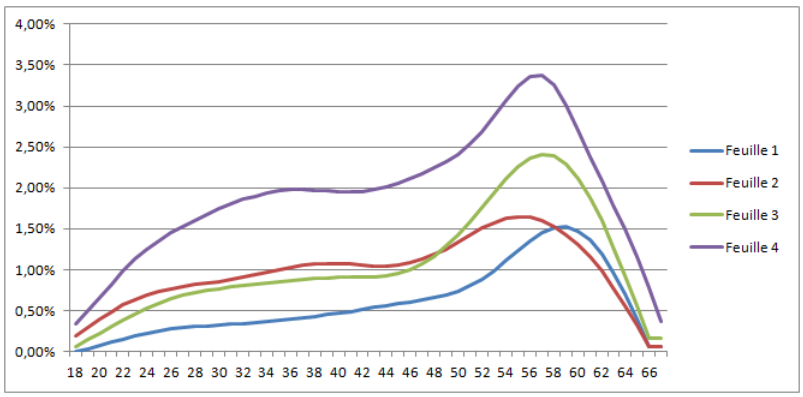


FIGURE IX.92: Lois d'entrée en incapacité hommes, franchise "en relai"

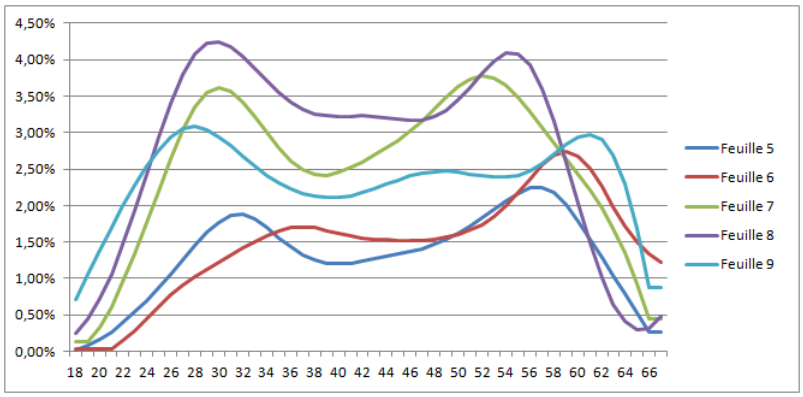


FIGURE IX.93: Lois d'entrée en incapacité femmes, franchise "en relai"

Lois d'entrée en incapacité, franchises 3 jours

	Franchise "3 jours"	
	H	F
Age	Toute population	Toute population
18	9,14%	9,11%
19	9,14%	9,11%
20	9,14%	9,11%
21	9,14%	9,11%
22	9,14%	12,35%
23	10,38%	15,82%
24	11,46%	19,28%
25	12,36%	22,48%
26	13,09%	25,22%
27	13,68%	27,33%
28	14,15%	28,71%
29	14,51%	29,37%
30	14,80%	29,40%
31	15,03%	28,94%
32	15,22%	28,15%
33	15,39%	27,17%
34	15,54%	26,12%
35	15,68%	25,07%
36	15,82%	24,08%
37	15,96%	23,20%
38	16,10%	22,48%
39	16,23%	21,92%
40	16,36%	21,52%
41	16,48%	21,23%
42	16,58%	21,02%
43	16,68%	20,89%
44	16,77%	20,87%
45	16,86%	20,99%
46	16,98%	21,26%
47	17,16%	21,65%
48	17,40%	22,13%
49	17,72%	22,61%
50	18,12%	23,06%
51	18,58%	23,52%
52	19,05%	24,02%
53	19,49%	24,62%
54	19,82%	25,28%
55	19,99%	25,93%
56	19,92%	26,45%
57	19,58%	26,63%
58	18,95%	26,32%
59	18,06%	25,49%
60	16,96%	24,18%
61	15,70%	22,49%
62	14,33%	20,53%
63	12,91%	18,41%
64	11,45%	16,20%
65	11,45%	13,94%
66	11,45%	13,94%
67	11,45%	13,94%

FIGURE IX.94: Lois par sexe, franchise 3 jours

Modèle avec prise en compte de l'hétérogénéité, franchise 3 jours

Modèle de Cox

La franchise 3 jours est une franchise très spécifique. L'étude ne portera ici que sur le secteur de la santé.
Les populations de références sont composées des individus non cadres d'une entreprise de taille A. Les effectifs de sexe masculin sont très faibles pour cette population de référence. En conséquence, la loi n'est pas segmentée par sexe.

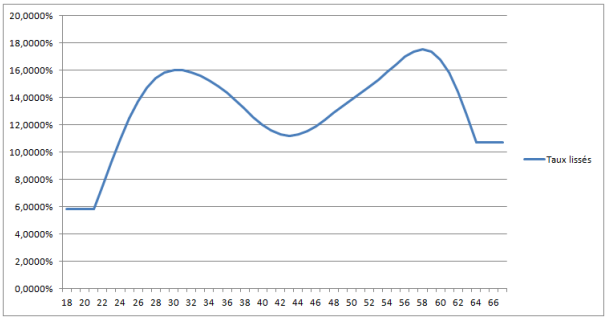


FIGURE IX.95: Taux lissés, franchise 3 jours

Résultats

Le nombre d'incapacités prédit est calculé pour chaque individu non censuré ayant une franchise continue 3 jours. L'observé est de 23 742 sinistres.

Les résultats obtenus sont les suivants :

Prédiction	Ratio O/A
21 835	108,7%

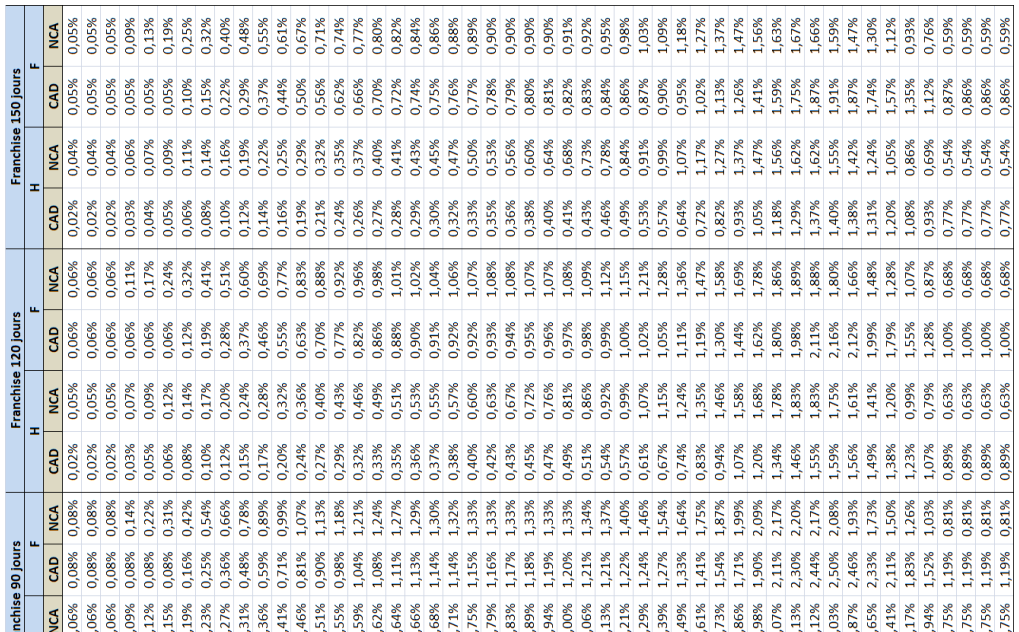
Au global, la prédiction est 8,7% en dessous de l'observé.

Approche par arbres de décision

Le modèle est calculé sans les variables "Grande région" et "Situation familiale". Les résultats obtenus sont les suivants pour des données non censurées/tronquées :

Observés	**Prédit**	**Ratio O/A**
2 833	2 932,66	96,6%

Annexe 2



Age	Franchise 30 jours						Franchise 45 jours						Franchise 60 jours						Franchise 75 jours						Franchise 90 jours						Franchise 120 jours						Franchise 150 jours					
	H			F			H			F			H			F			H			F			H			F			H			F			H		
	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA	CAD	NCA	NCA
18	0,08%	0,95%	0,19%	0,19%	0,64%	0,64%	0,07%	0,76%	0,15%	0,15%	0,52%	0,52%	0,05%	0,57%	0,12%	0,12%	0,39%	0,39%	0,04%	0,49%	0,10%	0,10%	0,34%	0,34%	0,03%	0,06%	0,08%	0,08%	0,08%	0,08%	0,02%	0,05%	0,06%	0,06%	0,06%	0,06%	0,02%	0,04%	0,05%	0,05%	0,05%	0,05%
19	0,08%	0,95%	0,19%	0,19%	0,64%	0,64%	0,07%	0,76%	0,15%	0,15%	0,52%	0,52%	0,05%	0,57%	0,12%	0,12%	0,39%	0,39%	0,04%	0,49%	0,10%	0,10%	0,34%	0,34%	0,03%	0,06%	0,08%	0,08%	0,08%	0,08%	0,02%	0,05%	0,06%	0,06%	0,06%	0,06%	0,02%	0,04%	0,05%	0,05%	0,05%	0,05%
20	0,08%	0,95%	0,19%	0,19%	0,64%	0,64%	0,07%	0,76%	0,15%	0,15%	0,52%	0,52%	0,05%	0,57%	0,12%	0,12%	0,39%	0,39%	0,04%	0,49%	0,10%	0,10%	0,34%	0,34%	0,03%	0,06%	0,08%	0,08%	0,08%	0,08%	0,02%	0,05%	0,06%	0,06%	0,06%	0,06%	0,02%	0,04%	0,04%	0,05%	0,05%	0,05%
21	0,15%	1,12%	0,35%	0,35%	1,08%	0,90%	0,12%	0,90%	0,28%	0,28%	0,87%	0,87%	0,09%	0,69%	0,21%	0,21%	0,66%	0,66%	0,08%	0,59%	0,18%	0,18%	0,57%	0,57%	0,04%	0,09%	0,08%	0,08%	0,14%	0,14%	0,03%	0,07%	0,06%	0,06%	0,11%	0,11%	0,03%	0,06%	0,05%	0,05%	0,09%	0,09%
22	0,23%	1,26%	0,53%	0,53%	1,59%	1,59%	0,18%	1,02%	0,43%	0,43%	1,29%	1,29%	0,14%	0,79%	0,33%	0,33%	0,99%	0,99%	0,12%	0,68%	0,29%	0,29%	0,85%	0,85%	0,06%	0,12%	0,08%	0,08%	0,22%	0,22%	0,05%	0,09%	0,06%	0,06%	0,17%	0,17%	0,04%	0,07%	0,05%	0,05%	0,13%	0,13%
23	0,31%	1,39%	0,74%	0,74%	2,15%	2,15%	0,25%	1,13%	0,60%	0,60%	1,75%	1,75%	0,19%	0,87%	0,47%	0,47%	1,35%	1,35%	0,17%	0,75%	0,40%	0,40%	1,16%	1,16%	0,08%	0,15%	0,08%	0,08%	0,31%	0,31%	0,06%	0,12%	0,06%	0,06%	0,24%	0,24%	0,05%	0,09%	0,05%	0,05%	0,19%	0,19%
24	0,39%	1,49%	0,97%	0,97%	2,72%	2,72%	0,32%	1,22%	0,79%	0,79%	2,22%	2,22%	0,25%	0,95%	0,62%	0,62%	1,73%	1,73%	0,21%	0,82%	0,53%	0,53%	1,49%	1,49%	0,10%	0,19%	0,16%	0,16%	0,42%	0,42%	0,08%	0,14%	0,12%	0,12%	0,32%	0,32%	0,06%	0,11%	0,10%	0,10%	0,25%	0,25%
25	0,47%	1,59%	1,21%	1,21%	3,26%	3,26%	0,39%	1,31%	0,99%	0,99%	2,68%	2,68%	0,30%	1,02%	0,78%	0,78%	2,10%	2,10%	0,26%	0,88%	0,67%	0,67%	1,81%	1,81%	0,13%	0,23%	0,25%	0,25%	0,54%	0,54%	0,10%	0,17%	0,19%	0,19%	0,41%	0,41%	0,08%	0,14%	0,15%	0,15%	0,32%	0,32%
26	0,55%	1,68%	1,46%	1,46%	3,73%	3,73%	0,45%	1,39%	1,20%	1,20%	3,08%	3,08%	0,36%	1,09%	0,95%	0,95%	2,42%	2,42%	0,31%	0,94%	0,82%	0,82%	2,09%	2,09%	0,16%	0,27%	0,36%	0,36%	0,66%	0,66%	0,12%	0,20%	0,28%	0,28%	0,51%	0,51%	0,10%	0,16%	0,22%	0,22%	0,40%	0,40%
27	0,62%	1,77%	1,71%	1,71%	4,11%	4,11%	0,51%	1,46%	1,42%	1,42%	3,40%	3,40%	0,40%	1,16%	1,12%	1,12%	2,69%	2,69%	0,35%	1,00%	0,97%	0,97%	2,33%	2,33%	0,19%	0,31%	0,48%	0,48%	0,78%	0,78%	0,15%	0,24%	0,37%	0,37%	0,60%	0,60%	0,12%	0,19%	0,29%	0,29%	0,48%	0,48%
28	0,67%	1,86%	1,96%	1,96%	4,40%	4,40%	0,56%	1,54%	1,62%	1,62%	3,65%	3,65%	0,44%	1,22%	1,29%	1,29%	2,89%	2,89%	0,38%	1,06%	1,11%	1,11%	2,51%	2,51%	0,23%	0,36%	0,59%	0,59%	0,89%	0,89%	0,17%	0,28%	0,46%	0,46%	0,69%	0,69%	0,14%	0,22%	0,37%	0,37%	0,55%	0,55%
29	0,72%	1,95%	2,17%	2,17%	4,60%	4,60%	0,60%	1,62%	1,81%	1,81%	3,82%	3,82%	0,48%	1,29%	1,44%	1,44%	3,04%	3,04%	0,42%	1,12%	1,25%	1,25%	2,64%	2,64%	0,26%	0,41%	0,71%	0,71%	0,99%	0,99%	0,20%	0,32%	0,46%	0,46%	0,77%	0,77%	0,16%	0,25%	0,44%	0,44%	0,61%	0,61%
30	0,76%	2,04%	2,36%	2,36%	4,73%	4,73%	0,64%	1,70%	1,96%	1,96%	3,93%	3,93%	0,51%	1,36%	1,57%	1,57%	3,14%	3,14%	0,44%	1,18%	1,36%	1,36%	2,73%	2,73%	0,30%	0,46%	0,81%	0,81%	1,07%	1,07%	0,24%	0,36%	0,53%	0,53%	0,83%	0,83%	0,19%	0,29%	0,50%	0,50%	0,67%	0,67%
31	0,80%	2,14%	2,50%	2,50%	4,79%	4,79%	0,66%	1,79%	2,08%	2,08%	4,00%	4,00%	0,53%	1,43%	1,67%	1,67%	3,20%	3,20%	0,46%	1,25%	1,45%	1,45%	2,79%	2,79%	0,34%	0,51%	0,90%	0,90%	1,13%	1,13%	0,27%	0,40%	0,70%	0,70%	0,88%	0,88%	0,21%	0,32%	0,56%	0,56%	0,71%	0,71%
32	0,82%	2,24%	2,59%	2,59%	4,78%	4,78%	0,69%	1,87%	2,17%	2,17%	4,00%	4,00%	0,55%	1,51%	1,74%	1,74%	3,21%	3,21%	0,48%	1,31%	1,52%	1,52%	2,80%	2,80%	0,38%	0,55%	0,98%	0,98%	1,18%	1,18%	0,29%	0,43%	0,77%	0,77%	0,92%	0,92%	0,24%	0,35%	0,62%	0,62%	0,74%	0,74%
33	0,85%	2,35%	2,63%	2,63%	4,69%	4,69%	0,71%	1,96%	2,21%	2,21%	3,93%	3,93%	0,57%	1,58%	1,78%	1,78%	3,16%	3,16%	0,50%	1,38%	1,55%	1,55%	2,76%	2,76%	0,40%	0,59%	1,04%	1,04%	1,21%	1,21%	0,32%	0,46%	0,82%	0,82%	0,96%	0,96%	0,26%	0,37%	0,66%	0,66%	0,77%	0,77%
34	0,87%	2,45%	2,64%	2,64%	4,52%	4,52%	0,73%	2,05%	2,21%	2,21%	3,79%	3,79%	0,59%	1,65%	1,78%	1,78%	3,05%	3,05%	0,51%	1,45%	1,56%	1,56%	2,67%	2,67%	0,42%	0,62%	1,08%	1,08%	1,24%	1,24%	0,33%	0,49%	0,86%	0,86%	0,98%	0,98%	0,27%	0,40%	0,70%	0,70%	0,80%	0,80%
35	0,89%	2,55%	2,61%	2,61%	4,30%	4,30%	0,75%	2,14%	2,18%	2,18%	3,60%	3,60%	0,60%	1,72%	1,76%	1,76%	2,90%	2,90%	0,53%	1,50%	1,54%	1,54%	2,53%	2,53%	0,44%	0,64%	1,11%	1,27%	1,27%	1,27%	0,35%	0,51%	0,88%	0,88%	1,01%	1,01%	0,28%	0,41%	0,72%	0,72%	0,82%	0,82%
36	0,92%	2,65%	2,57%	2,57%	4,07%	4,07%	0,77%	2,22%	2,14%	2,14%	3,40%	3,40%	0,62%	1,78%	1,72%	1,72%	2,73%	2,73%	0,54%	1,51%	1,51%	1,51%	2,39%	2,39%	0,45%	0,66%	1,13%	1,29%	1,29%	1,29%	0,36%	0,53%	0,90%	0,90%	1,02%	1,02%	0,29%	0,43%	0,74%	0,74%	0,84%	0,84%
37	0,95%	2,74%	2,51%	2,51%	3,86%	3,86%	0,80%	2,29%	2,10%	2,10%	3,22%	3,22%	0,64%	1,84%	1,68%	1,68%	2,58%	2,58%	0,56%	1,61%	1,47%	1,47%	2,26%	2,26%	0,46%	0,68%	1,14%	1,30%	1,30%	1,30%	0,37%	0,55%	0,91%	0,91%	1,04%	1,04%	0,30%	0,45%	0,75%	0,75%	0,86%	0,86%
38	0,99%	2,82%	2,46%	2,46%	3,69%	3,69%	0,83%	2,36%	2,05%	2,05%	3,08%	3,08%	0,66%	1,89%	1,65%	1,65%	2,47%	2,47%	0,58%	1,65%	1,44%	1,44%	2,16%	2,16%	0,48%	0,71%	1,14%	1,32%	1,32%	1,32%	0,38%	0,57%	0,92%	0,92%	1,06%	1,06%	0,32%	0,47%	0,76%	0,76%	0,88%	0,88%
39	1,03%	2,89%	2,41%	2,41%	3,58%	3,58%	0,86%	2,41%	2,01%	2,01%	2,99%	2,99%	0,69%	1,94%	1,62%	1,62%	2,40%	2,40%	0,61%	1,70%	1,42%	1,42%	2,10%	2,10%	0,50%	0,75%	1,15%	1,33%	1,33%	1,33%	0,40%	0,60%	0,92%	0,92%	1,07%	1,07%	0,33%	0,50%	0,77%	0,77%	0,89%	0,89%
40	1,07%	2,95%	2,37%	2,37%	3,53%	3,53%	0,90%	2,47%	1,98%	1,98%	2,94%	2,94%	0,72%	1,98%	1,59%	1,59%	2,36%	2,36%	0,63%	1,74%	1,39%	1,39%	2,07%	2,07%	0,52%	0,79%	1,16%	1,33%	1,33%	1,33%	0,42%	0,63%	0,93%	0,93%	1,08%	1,08%	0,35%	0,53%	0,78%	0,78%	0,90%	0,90%
41	1,11%	3,00%	2,33%	2,33%	3,51%	3,51%	0,93%	2,51%	1,95%	1,95%	2,94%	2,94%	0,75%	2,02%	1,56%	1,56%	2,36%	2,36%	0,65%	1,77%	1,37%	1,37%	2,07%	2,07%	0,54%	0,83%	1,17%	1,33%	1,33%	1,33%	0,43%	0,67%	0,94%	0,94%	1,08%	1,08%	0,36%	0,56%	0,79%	0,79%	0,90%	0,90%
42	1,15%	3,05%	2,29%	2,29%	3,53%	3,53%	0,96%	2,55%	1,92%	1,92%	2,95%	2,95%	0,77%	2,06%	1,54%	1,54%	2,38%	2,38%	0,68%	1,80%	1,36%	1,36%	2,09%	2,09%	0,56%	0,89%	1,18%	1,33%	1,33%	1,33%	0,45%	0,72%	0,95%	0,95%	1,07%	1,07%	0,38%	0,60%	0,80%	0,80%	0,90%	0,90%
43	1,19%	3,10%	2,26%	2,26%	3,56%	3,56%	0,99%	2,60%	1,89%	1,89%	2,98%	2,98%	0,80%	2,10%	1,52%	1,52%	2,41%	2,41%	0,70%	1,84%	1,34%	1,34%	2,11%	2,11%	0,58%	0,94%	1,19%	1,33%	1,33%	1,33%	0,47%	0,76%	0,96%	0,96%	1,07%	1,07%	0,40%	0,64%	0,81%	0,81%	0,90%	0,90%
44	1,23%	3,16%	2,22%	2,22%	3,60%	3,60%	1,03%	2,65%	1,86%	1,86%	3,02%	3,02%	0,83%	2,14%	1,50%	1,50%	2,44%	2,44%	0,73%	1,88%	1,32%	1,32%	2,15%	2,15%	0,61%	1,00%	1,20%	1,33%	1,33%	1,33%	0,49%	0,81%	0,97%	0,97%	1,08%	1,08%	0,41%	0,68%	0,82%	0,82%	0,91%	0,91%
45	1,27%	3,25%	2,18%	2,18%	3,65%	3,65%	1,07%	2,72%	1,83%	1,83%	3,07%	3,07%	0,87%	2,20%	1,48%	1,48%	2,48%	2,48%	0,76%	1,94%	1,30%	1,30%	2,19%	2,19%	0,63%	1,06%	1,21%	1,34%	1,34%	1,34%	0,51%	0,86%	0,98%	0,98%	1,09%	1,09%	0,43%	0,73%	0,83%	0,83%	0,92%	0,92%
46	1,33%	3,36%	2,15%	2,15%	3,72%	3,72%	1,12%	2,82%	1,81%	1,81%	3,13%	3,13%	0,91%	2,29%	1,47%	1,47%	2,54%	2,54%	0,80%	2,02%	1,30%	1,30%	2,25%	2,25%	0,66%	1,13%	1,21%	1,37%	1,37%	1,37%	0,54%	0,92%	0,99%	0,99%	1,12%	1,12%	0,46%	0,78%	0,84%	0,84%	0,95%	0,95%
47	1,39%	3,51%	2,13%	2,13%	3,81%	3,81%	1,18%	2,96%	1,80%	1,80%	3,22%	3,22%	0,96%	2,41%	1,47%	1,47%	2,62%	2,62%	0,85%	2,13%	1,30%	1,30%	2,32%	2,32%	0,70%	1,21%	1,22%	1,40%	1,40%	1,40%	0,57%	0,99%	1,00%	1,00%	1,15%	1,15%	0,49%	0,84%	0,86%	0,86%	0,98%	0,98%
48	1,47%	3,69%	2,13%	2,13%	3,91%	3,91%	1,25%	3,13%	1,81%	1,81%	3,32%	3,32%	1,02%	2,57%	1,48%	1,48%	2,72%	2,72%	0,91%	2,28%	1,31%	1,31%	2,41%	2,41%	0,75%	1,29%	1,24%	1,46%	1,46%	1,46%	0,61%	1,07%	1,02%	1,02%	1,21%	1,21%	0,53%	0,91%	0,87%	0,87%	1,03%	1,03%
49	1,57%	3,92%	2,16%	2,16%	4,04%	4,04%	1,34%	3,34%	1,84%	1,84%	3,44%	3,44%	1,10%	2,76%	1,52%	1,52%	2,84%	2,84%	0,98%	2,45%	1,35%	1,35%	2,53%	2,53%	0,81%	1,39%	1,27%	1,54%	1,54%	1,54%	0,67%	1,15%	1,05%	1,05%	1,28%	1,28%	0,57%	0,99%	0,90%	0,90%	1,09%	1,09%
50	1,70%	4,19%	2,22%	2,22%	4,20%	4,20%	1,45%	3,58%	1,90%	1,90%	3,60%	3,60%	1,21%	2,98%	1,58%	1,58%	2,99%	2,99%	1,08%	2,66%	1,41%	1,41%	2,67%	2,67%	0,89%	1,49%	1,33%	1,64%	1,64%	1,64%	0,74%	1,24%	1,11%	1,11%	1,36%	1,36%	0,64%	1,07%	0,95%	0,95%	1,18%	1,18%
51	1,85%	4,48%	2,32%	2,32%	4,40%	4,40%	1,59%	3,85%	2,00%	2,00%	3,78%	3,78%	1,33%	3,22%	1,67%	1,67%	3,17%	3,17%	1,20%	2,89%	1,50%	1,50%	2,84%	2,84%	0,99%	1,61%	1,41%	1,75%	1,75%	1,75%	0,83%	1,35%	1,19%	1,19%	1,47%	1,47%	0,72%	1,17%	1,02%	1,02%	1,27%	1,27%
52	2,04%	4,77%	2,47%	2,47%	4,63%	4,63%	1,76%	4,13%	2,14%	2,14%	4,01%	4,01%	1,49%	3,48%	1,80%	1,80%	3,38%	3,38%	1,34%	3,13%	1,62%	1,62%	3,04%	3,04%	1,12%	1,73%	1,54%	1,87%	1,87%	1,87%	0,94%	1,46%	1,30%	1,30%	1,58%	1,58%	0,82%	1,27%	1,13%	1,13%	1,37%	1,37%
53	2,26%	5,05%	2,66%	2,66%	4,90%	4,90%	1,97%	4,39%	2,31%	2,31%	4,26%	4,26%	1,67%	3,73%	1,96%	1,96%	3,61%	3,61%	1,51%	3,36%	1,77%	1,77%	3,26%	3,26%	1,26%	1,86%	1,71%	1,99%	1,99%	1,99%	1,07%	1,58%	1,44%	1,44%	1,69%	1,69%	0,93%	1,37%	1,26%	1,26%	1,47%	1,47%
54	2,52%	5,29%	2,86%	2,86%	5,17%	5,17%	2,20%	4,62%	2,50%	2,50%	4,51%	4,51%	1,88%	3,95%	2,14%	2,14%	3,86%	3,86%	1,70%	3,57%	1,93%	1,93%	3,49%	3,49%	1,42%	1,98%	1,90%	2,09%	2,09%	2,09%	1,20%	1,68%	1,62%	1,62%	1,78%	1,78%	1,05%	1,47%	1,41%	1,41%	1,56%	1,56%
55	2,78%	5,47%	3,06%	3,06%	5,40%	5,40%	2,44%	4,80%	2,69%	2,69%	4,74%	4,74%	2,10%	4,13%	2,31%	2,31%	4,07%	4,07%	1,90%	3,75%	2,10%	2,10%	3,70%	3,70%	1,57%	2,07%	2,11%	2,17%	2,17%	2,17%	1,34%	1,78%	1,80%	1,80%	1,86%	1,86%	1,18%	1,56%	1,59%	1,59%	1,63%	1,63%
56	3,03%	5,56%	3,23%	3,23%	5,56%	5,56%	2,67%	4,89%	2,84%	2,84%	4,89%	4,89%	2,31%	4,23%	2,45%	2,45%	4,23%	4,23%	2,10%	3,85%	2,24%	2,24%	3,85%	3,85%	1,70%	2,13%	2,30%	2,20%	2,20%	2,20%	1,46%	1,83%	1,98%	1,98%	1,89%	1,89%	1,29%	1,62%	1,75%	1,75%	1,67%	1,67%
57	3,24%	5,53%	3,32%	3,32%	5,58%	5,58%	2,86%	4,88%	2,93%	2,93%	4,93%	4,93%	2,48%	4,24%	2,54%	2,54%	4,27%	4,27%	2,27%	3,87%	2,32%	2,32%	3,90%	3,90%	1,80%	2,12%	2,44%	2,17%	2,17%	2,17%	1,55%	1,83%	2,11%	2,11%	1,88%	1,88%	1,37%	1,62%	1,87%	1,87%	1,66%	1,66%
58	3,36%	5,37%	3,30%	3,30%	5,44%	5,44%	2,97%	4,75%	2,92%	2,92%	4,81%	4,81%	2,58%	4,12%	2,53%	2,53%	4,18%	4,18%	2,36%	3,76%	2,31%	2,31%	3,81%	3,81%	1,84%	2,03%	2,50%	2,08%	2,08%	2,08%	1,59%	1,75%	2,16%	2,16%	1,80%	1,80%	1,40%	1,55%	1,91%	1,91%	1,59%	1,59%
59	3,37%	5,08%	3,17%	3,17%	5,14%	5,14%	2,97%	4,48%	2,80%	2,80%	4,53%	4,53%	2,57%	3,88%	2,42%	2,42%	3,92%	3,92%	2,35%	3,53%	2,21%	2,21%	3,58%	3,58%	1,82%	1,87%	2,46%	1,93%	1,93%	1,93%	1,56%	1,61%	2,12%	2,12%	1,66%	1,66%	1,38%	1,42%	1,87%	1,87%	1,47%	1,47%
60	3,24%	4,67%	2,94%	2,94%	4,72%	4,72%	2,85%	4,10%	2,59%	2,59%	4,15%	4,15%	2,45%	3,53%	2,23%	2,23%	3,57%	3,57%	2,23%	3,21%	2,02%	2,02%	3,24%	3,24%	1,74%	1,65%	2,33%	1,73%	1,73%	1,73%	1,49%	1,41%	1,99%	1,99%	1,48%	1,48%	1,31%	1,24%	1,74%	1,74%	1,30%	1,30%
61	2,98%	4,18%	2,64%	2,64%	4,24%	4,24%	2,61%	3,65%	2,31%	2,31%	3,71%	3,71%	2,24%	3,13%	1,98%	1,98%	3,18%	3,18%	2,02%	2,84%	1,80%	1,80%	2,88%	2,88%	1,62%	1,41%	2,11%	1,50%	1,50%	1,50%	1,38%	1,20%	1,79%	1,79%	1,28%	1,28%	1,20%	1,05%	1,57%	1,57%	1,12%	1,12%
62	2,61%	3,62%	2,31%	2,31%	3,74%	3,74%	2,28%	3,16%	2,02%	2,02%	3,27%	3,27%	1,95%	2,70%	1,72%	1,72%	2,79%	2,79%	1,76%	2,43%	1,55%	1,55%	2,51%	2,51%	1,46%	1,17%	1,83%	1,26%	1,26%	1,26%	1,23%	0,99%	1,55%	1,55%	1,07%	1,07%	1,08%	0,86%	1,35%	1,35%	0,93%	0,93%
63	2,17%	3,03%	1,97%	1,97%	3,24%	3,24%	1,89%	2,64%	1,72%	1,72%	2,82%	2,82%	1,61%	2,25%	1,46%	1,46%	2,41%	2,41%	1,45%	2,03%	1,32%	1,32%	2,17%	2,17%	1,26%	0,94%	1,52%	1,03%	1,03%	1,03%	1,07%	0,79%	1,28%	1,28%	0,87%	0,87%	0,93%	0,69%	1,12%	1,12%	0,76%	0,76%
64	1,67%	2,41%	1,66%	1,66%	2,73%	2,73%	1,45%	2,10%	1,44%	1,44%	2,37%	2,37%	1,24%	1,79%	1,23%	1,23%	2,02%	2,02%	1,12%	1,61%	1,11%	1,11%	1,82%	1,82%	1,05%	0,75%	1,19%	0,81%	0,81%	0,81%	0,89%	0,63%	1,00%	1,00%	0,68%	0,68%	0,77%	0,54%	0,87%	0,87%	0,59%	0,59%
65	1,67%	2,41%	1,66%	1,66%	2,73%	2,73%	1,45%	2,10%	1,44%	1,44%	2,37%	2,37%	1,24%	1,79%	1,23%	1,23%	2,02%	2,02%	1,12%	1,61%	1,11%	1,11%	1,82%	1,82%	1,05%	0,75%	1,19%	0,81%	0,81%	0,81%	0,89%	0,63%	1,00%	1,00%	0,68%	0,68%	0,77%	0,54%	0,86%	0,86%	0,59%	0,59%
66	1,67%	2,41%	1,66%	1,66%	2,73%	2,73%	1,45%	2,10%	1,44%	1,44%	2,37%	2,37%	1,24%	1,79%	1,23%	1,23%	2,02%	2,02%	1,12%	1,61%	1,11%	1,11%	1,82%	1,82%	1,05%	0,75%	1,19%	0,81%	0,81%	0,81%	0,89%	0,63%	1,00%	1,00%	0,68%	0,68%	0,77%	0,54%	0,86%	0,86%	0,59%	0,59%
67	1,67%	2,41%	1,66%	1,66%	2,73%	2,73%	1,45%	2,10%	1,44%	1,44%	2,37%	2,37%	1,24%	1,79%	1,23%	1,23%	2,02%	2,02%	1,12%	1,61%	1,11%	1,11%	1,82%	1,82%	1,05%	0,75%	1,19%	0,81%	0,81%	0,81%	0,89%	0,63%	1,00%	1,00%	0,68%	0,68%	0,77%	0,54%	0,86%	0,86%	0,59%	0,59%

FIGURE IX.96: Lois par franchise, sexe et CSP

Annexe 3

	Franchise 30 jours			Franchise 45 jours			Franchise 60 jours			Franchise 75 jours			Franchise 90 jours			Franchise 120 jours			Franchise 150 jours		
Age	ASES	Industrie	Autres	ASES	Industrie	Autres	ASES	Industrie	Autres	ASES	Industrie	Autres	ASES	Industrie	Autres	ASES	Industrie	Autres	ASES	Industrie	Autres
18	0,37%	2,23%	0,65%	0,30%	1,79%	0,52%	0,23%	1,35%	0,39%	0,20%	1,17%	0,34%	0,07%	0,11%	0,01%	0,05%	0,09%	0,00%	0,04%	0,07%	0,00%
19	0,37%	2,23%	0,65%	0,30%	1,79%	0,52%	0,23%	1,35%	0,39%	0,20%	1,17%	0,34%	0,07%	0,11%	0,01%	0,05%	0,09%	0,00%	0,04%	0,07%	0,00%
20	0,37%	2,23%	0,65%	0,30%	1,79%	0,52%	0,23%	1,35%	0,39%	0,20%	1,17%	0,34%	0,07%	0,11%	0,01%	0,05%	0,09%	0,00%	0,04%	0,07%	0,00%
21	0,53%	2,50%	0,86%	0,43%	2,02%	0,69%	0,33%	1,54%	0,53%	0,28%	1,33%	0,45%	0,10%	0,21%	0,03%	0,08%	0,17%	0,02%	0,06%	0,13%	0,02%
22	0,70%	2,69%	1,07%	0,57%	2,18%	0,87%	0,43%	1,68%	0,67%	0,37%	1,44%	0,57%	0,14%	0,32%	0,06%	0,11%	0,25%	0,04%	0,09%	0,20%	0,04%
23	0,86%	2,82%	1,28%	0,70%	2,30%	1,04%	0,54%	1,78%	0,81%	0,47%	1,53%	0,69%	0,18%	0,44%	0,10%	0,14%	0,34%	0,08%	0,11%	0,27%	0,06%
24	1,03%	2,91%	1,47%	0,84%	2,38%	1,21%	0,65%	1,85%	0,94%	0,56%	1,59%	0,81%	0,23%	0,55%	0,15%	0,17%	0,42%	0,12%	0,14%	0,33%	0,09%
25	1,18%	2,96%	1,65%	0,97%	2,44%	1,36%	0,76%	1,91%	1,06%	0,66%	1,64%	0,92%	0,28%	0,66%	0,22%	0,21%	0,50%	0,17%	0,17%	0,40%	0,13%
26	1,33%	3,00%	1,80%	1,10%	2,48%	1,49%	0,86%	1,95%	1,17%	0,74%	1,68%	1,01%	0,33%	0,75%	0,28%	0,25%	0,58%	0,22%	0,20%	0,45%	0,17%
27	1,45%	3,02%	1,93%	1,20%	2,50%	1,60%	0,95%	1,98%	1,26%	0,82%	1,71%	1,09%	0,39%	0,83%	0,35%	0,30%	0,64%	0,27%	0,24%	0,51%	0,21%
28	1,56%	3,03%	2,03%	1,29%	2,51%	1,69%	1,03%	1,99%	1,34%	0,89%	1,73%	1,16%	0,44%	0,90%	0,41%	0,34%	0,69%	0,32%	0,27%	0,55%	0,25%
29	1,64%	3,01%	2,11%	1,37%	2,50%	1,75%	1,09%	1,99%	1,40%	0,94%	1,73%	1,21%	0,49%	0,95%	0,47%	0,38%	0,74%	0,36%	0,31%	0,59%	0,29%
30	1,71%	2,98%	2,17%	1,42%	2,48%	1,80%	1,13%	1,98%	1,44%	0,99%	1,72%	1,25%	0,54%	0,99%	0,51%	0,42%	0,78%	0,40%	0,34%	0,62%	0,32%
31	1,75%	2,93%	2,20%	1,46%	2,44%	1,84%	1,17%	1,96%	1,47%	1,02%	1,70%	1,28%	0,59%	1,02%	0,55%	0,46%	0,80%	0,43%	0,37%	0,64%	0,35%
32	1,77%	2,87%	2,22%	1,48%	2,40%	1,85%	1,19%	1,93%	1,49%	1,04%	1,68%	1,30%	0,63%	1,02%	0,58%	0,49%	0,80%	0,46%	0,40%	0,65%	0,37%
33	1,77%	2,81%	2,22%	1,48%	2,35%	1,86%	1,19%	1,90%	1,50%	1,04%	1,66%	1,31%	0,66%	1,02%	0,61%	0,52%	0,80%	0,48%	0,42%	0,65%	0,39%
34	1,76%	2,76%	2,20%	1,48%	2,31%	1,85%	1,19%	1,86%	1,49%	1,04%	1,63%	1,30%	0,69%	1,01%	0,62%	0,55%	0,80%	0,49%	0,45%	0,65%	0,40%
35	1,76%	2,71%	2,18%	1,47%	2,27%	1,83%	1,18%	1,82%	1,47%	1,03%	1,60%	1,29%	0,72%	1,02%	0,63%	0,57%	0,81%	0,50%	0,47%	0,66%	0,41%
36	1,75%	2,68%	2,15%	1,47%	2,24%	1,80%	1,18%	1,80%	1,45%	1,03%	1,57%	1,27%	0,74%	1,04%	0,64%	0,59%	0,83%	0,51%	0,49%	0,68%	0,42%
37	1,76%	2,66%	2,12%	1,47%	2,22%	1,77%	1,18%	1,78%	1,42%	1,03%	1,56%	1,25%	0,77%	1,09%	0,65%	0,61%	0,87%	0,52%	0,51%	0,72%	0,43%
38	1,77%	2,67%	2,09%	1,48%	2,23%	1,75%	1,19%	1,79%	1,40%	1,04%	1,57%	1,23%	0,79%	1,17%	0,66%	0,63%	0,94%	0,53%	0,53%	0,77%	0,44%
39	1,79%	2,70%	2,07%	1,50%	2,25%	1,72%	1,20%	1,81%	1,38%	1,05%	1,58%	1,21%	0,82%	1,25%	0,66%	0,66%	1,00%	0,53%	0,55%	0,83%	0,44%
40	1,82%	2,74%	2,04%	1,52%	2,29%	1,70%	1,22%	1,83%	1,37%	1,07%	1,61%	1,20%	0,84%	1,32%	0,67%	0,68%	1,07%	0,54%	0,57%	0,89%	0,45%
41	1,85%	2,78%	2,02%	1,55%	2,33%	1,69%	1,25%	1,87%	1,36%	1,09%	1,64%	1,19%	0,87%	1,39%	0,67%	0,70%	1,12%	0,54%	0,59%	0,94%	0,45%
42	1,89%	2,83%	2,00%	1,58%	2,37%	1,68%	1,27%	1,91%	1,35%	1,12%	1,67%	1,18%	0,89%	1,45%	0,68%	0,72%	1,17%	0,55%	0,61%	0,99%	0,46%
43	1,93%	2,88%	1,99%	1,61%	2,42%	1,67%	1,30%	1,95%	1,34%	1,14%	1,71%	1,18%	0,92%	1,51%	0,70%	0,74%	1,22%	0,57%	0,62%	1,03%	0,48%
44	1,96%	2,94%	1,99%	1,64%	2,47%	1,67%	1,33%	1,99%	1,35%	1,17%	1,75%	1,18%	0,94%	1,56%	0,73%	0,76%	1,27%	0,59%	0,64%	1,07%	0,50%
45	2,00%	3,01%	2,00%	1,68%	2,53%	1,68%	1,36%	2,05%	1,36%	1,19%	1,80%	1,20%	0,96%	1,62%	0,77%	0,78%	1,32%	0,63%	0,66%	1,11%	0,53%
46	2,04%	3,11%	2,04%	1,72%	2,61%	1,72%	1,39%	2,12%	1,40%	1,23%	1,87%	1,23%	0,99%	1,68%	0,82%	0,81%	1,37%	0,67%	0,69%	1,17%	0,57%
47	2,09%	3,23%	2,11%	1,76%	2,73%	1,78%	1,44%	2,22%	1,45%	1,27%	1,97%	1,29%	1,03%	1,76%	0,87%	0,84%	1,44%	0,72%	0,72%	1,23%	0,61%
48	2,14%	3,38%	2,22%	1,82%	2,86%	1,88%	1,49%	2,35%	1,54%	1,32%	2,08%	1,37%	1,07%	1,84%	0,93%	0,88%	1,52%	0,77%	0,76%	1,30%	0,66%
49	2,22%	3,56%	2,36%	1,89%	3,03%	2,01%	1,56%	2,50%	1,66%	1,38%	2,23%	1,48%	1,13%	1,93%	1,00%	0,94%	1,60%	0,83%	0,81%	1,37%	0,71%
50	2,30%	3,77%	2,56%	1,97%	3,23%	2,19%	1,64%	2,68%	1,82%	1,46%	2,39%	1,62%	1,21%	2,01%	1,07%	1,01%	1,68%	0,90%	0,87%	1,45%	0,77%
51	2,41%	3,99%	2,79%	2,08%	3,43%	2,40%	1,74%	2,88%	2,01%	1,56%	2,58%	1,80%	1,31%	2,09%	1,17%	1,10%	1,75%	0,98%	0,95%	1,51%	0,85%
52	2,55%	4,22%	3,08%	2,20%	3,65%	2,66%	1,86%	3,08%	2,24%	1,67%	2,77%	2,02%	1,43%	2,16%	1,29%	1,21%	1,82%	1,09%	1,05%	1,58%	0,94%
53	2,70%	4,43%	3,39%	2,34%	3,85%	2,95%	1,99%	3,27%	2,50%	1,80%	2,95%	2,26%	1,57%	2,24%	1,42%	1,33%	1,90%	1,21%	1,16%	1,65%	1,05%
54	2,86%	4,61%	3,72%	2,50%	4,03%	3,25%	2,14%	3,44%	2,77%	1,93%	3,11%	2,51%	1,71%	2,33%	1,57%	1,45%	1,98%	1,34%	1,27%	1,73%	1,17%
55	3,02%	4,75%	4,02%	2,65%	4,16%	3,53%	2,28%	3,58%	3,03%	2,07%	3,25%	2,75%	1,84%	2,41%	1,72%	1,58%	2,06%	1,47%	1,39%	1,81%	1,30%
56	3,16%	4,81%	4,26%	2,78%	4,23%	3,75%	2,40%	3,65%	3,24%	2,19%	3,33%	2,95%	1,95%	2,47%	1,85%	1,68%	2,12%	1,59%	1,48%	1,88%	1,41%
57	3,25%	4,77%	4,39%	2,87%	4,21%	3,87%	2,49%	3,65%	3,36%	2,27%	3,34%	3,07%	2,01%	2,48%	1,94%	1,73%	2,14%	1,67%	1,54%	1,90%	1,48%
58	3,25%	4,63%	4,37%	2,88%	4,09%	3,86%	2,50%	3,55%	3,35%	2,28%	3,25%	3,06%	2,00%	2,41%	1,95%	1,72%	2,08%	1,69%	1,53%	1,85%	1,49%
59	3,17%	4,38%	4,19%	2,79%	3,86%	3,69%	2,42%	3,34%	3,20%	2,20%	3,05%	2,91%	1,92%	2,26%	1,88%	1,65%	1,94%	1,62%	1,45%	1,71%	1,43%
60	2,97%	4,03%	3,87%	2,61%	3,54%	3,40%	2,25%	3,05%	2,93%	2,04%	2,77%	2,66%	1,77%	2,03%	1,70%	1,51%	1,73%	1,46%	1,32%	1,52%	1,28%
61	2,69%	3,61%	3,45%	2,35%	3,16%	3,02%	2,02%	2,71%	2,59%	1,82%	2,45%	2,34%	1,56%	1,76%	1,46%	1,33%	1,50%	1,24%	1,16%	1,31%	1,09%
62	2,33%	3,15%	2,97%	2,04%	2,75%	2,59%	1,74%	2,35%	2,21%	1,57%	2,11%	2,00%	1,33%	1,48%	1,19%	1,12%	1,26%	1,01%	0,98%	1,10%	0,88%
63	1,94%	2,66%	2,47%	1,69%	2,32%	2,15%	1,44%	1,97%	1,83%	1,30%	1,78%	1,65%	1,08%	1,24%	0,93%	0,91%	1,04%	0,78%	0,79%	0,91%	0,68%
64	1,53%	2,15%	1,96%	1,33%	1,88%	1,71%	1,13%	1,60%	1,46%	1,02%	1,44%	1,31%	0,82%	1,03%	0,69%	0,69%	0,87%	0,58%	0,60%	0,75%	0,50%
65	1,53%	2,15%	1,96%	1,33%	1,88%	1,71%	1,13%	1,60%	1,46%	1,02%	1,44%	1,31%	0,82%	1,03%	0,69%	0,69%	0,87%	0,58%	0,60%	0,75%	0,50%
66	1,53%	2,15%	1,96%	1,33%	1,88%	1,71%	1,13%	1,60%	1,46%	1,02%	1,44%	1,31%	0,82%	1,03%	0,69%	0,69%	0,87%	0,58%	0,60%	0,75%	0,50%
67	1,53%	2,15%	1,96%	1,33%	1,88%	1,71%	1,13%	1,60%	1,46%	1,02%	1,44%	1,31%	0,82%	1,03%	0,69%	0,69%	0,87%	0,58%	0,60%	0,75%	0,50%

Figure IX.97: Lois par secteur

ANNEXE 4

Table de maintien en incapacité BCAC 1996

Age à l'entrée en incapacité	Ancienneté en mois en incapacité																																			
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
20	10000	2842	1743	1144	838	625	455	339	291	253	215	187	173	152	138	129	123	114	102	98	94	91	87	84	80	76	76	74	72	68	68	65	63	62	58	55	15
21	10000	2842	1743	1144	838	625	455	339	291	253	215	187	173	152	138	129	123	114	102	98	94	91	87	84	80	76	76	74	72	68	68	65	63	62	58	55	15
22	10000	2842	1743	1144	838	625	455	339	291	253	215	187	173	152	138	129	123	114	102	98	94	91	87	84	80	76	76	74	72	68	68	65	63	62	58	55	15
23	10000	2842	1743	1144	838	625	455	339	291	253	215	187	173	152	138	129	123	114	102	98	94	91	87	84	80	76	76	74	72	68	68	65	63	62	58	55	15
24	10000	2931	1848	1215	894	657	478	343	291	256	217	183	166	143	130	121	114	105	95	91	88	87	84	82	79	74	72	68	67	62	62	58	57	55	52	46	14
25	10000	3080	2001	1345	997	739	536	382	327	289	251	216	195	172	159	149	140	129	116	113	110	106	102	97	92	87	83	78	76	73	73	70	67	66	63	58	16
26	10000	3177	2112	1461	1087	812	591	431	372	325	285	249	226	201	186	171	161	150	137	129	124	119	114	107	102	95	91	89	87	82	81	78	76	73	69	63	23
27	10000	3251	2180	1540	1156	869	643	476	407	360	320	285	263	237	222	207	192	179	168	159	151	144	140	134	128	118	111	108	104	97	93	90	88	85	81	74	28
28	10000	3298	2243	1600	1209	915	688	524	448	400	359	322	297	270	255	238	222	210	199	189	180	172	167	160	153	143	132	128	120	112	105	103	99	96	90	82	33
29	10000	3348	2273	1640	1246	956	726	559	476	425	384	352	327	298	280	262	247	233	220	208	199	190	184	175	168	159	147	143	133	125	118	113	109	106	98	91	35
30	10000	3386	2275	1659	1264	964	744	583	494	439	396	363	338	308	287	267	252	240	227	214	202	193	185	177	171	161	149	143	134	125	117	111	108	105	97	89	34
31	10000	3388	2228	1618	1249	965	756	595	501	449	406	375	347	318	295	276	261	250	236	223	212	204	194	186	179	172	159	154	141	131	121	114	111	108	101	93	30
32	10000	3433	2238	1617	1254	975	772	612	522	468	421	388	357	325	302	279	264	252	235	222	211	202	192	183	176	170	159	153	137	127	118	110	106	102	96	89	25
33	10000	3466	2235	1627	1260	983	782	628	540	484	431	395	364	332	310	286	270	256	238	223	212	202	191	181	172	162	154	146	134	122	117	105	100	98	94	88	18
34	10000	3567	2298	1684	1321	1033	828	684	597	535	477	436	401	366	344	319	298	282	265	247	233	220	207	197	186	175	167	158	146	134	126	117	110	106	101	96	21
35	10000	3645	2331	1705	1357	1082	876	732	647	586	528	481	443	402	377	351	331	309	294	275	261	246	234	220	207	199	191	179	166	153	146	135	126	121	115	109	24
36	10000	3701	2390	1747	1390	1106	905	771	682	617	560	508	469	428	397	370	347	323	308	287	273	255	246	230	217	208	199	186	174	160	153	142	132	128	120	114	23
37	10000	3822	2458	1804	1430	1148	932	801	704	635	579	526	487	443	406	379	357	335	319	298	279	263	252	235	222	212	204	191	181	167	161	149	135	130	123	114	19
38	10000	3958	2526	1851	1479	1193	980	841	739	671	616	564	521	477	439	411	384	358	340	319	299	282	270	252	242	235	229	217	203	188	180	167	154	148	141	131	21
39	10000	4035	2600	1923	1541	1266	1055	915	807	739	680	623	572	530	486	455	427	400	381	364	343	329	314	294	279	268	260	248	234	215	207	189	177	170	162	148	24
40	10000	4073	2652	1973	1575	1303	1097	965	853	783	719	659	607	565	521	490	458	428	404	384	362	349	332	313	295	281	272	263	246	228	214	195	184	178	171	156	21
41	10000	4214	2776	2096	1680	1408	1193	1054	937	866	798	731	676	626	582	552	519	483	455	433	407	393	372	352	330	314	304	295	276	260	244	224	213	205	194	182	19
42	10000	4364	2930	2237	1814	1540	1314	1162	1039	971	895	825	764	710	666	630	593	553	521	499	476	457	432	411	381	364	353	340	322	300	280	257	247	236	223	213	26
43	10000	4473	3046	2341	1907	1633	1400	1243	1120	1045	965	892	830	774	726	691	654	614	582	558	532	513	488	464	432	409	396	378	362	337	311	290	278	263	244	231	35
44	10000	4621	3155	2417	1974	1676	1441	1282	1158	1077	1000	928	872	809	760	725	682	643	608	581	555	531	503	479	453	431	417	396	379	353	323	302	287	273	254	241	26
45	10000	4780	3318	2557	2097	1776	1529	1361	1240	1148	1069	1001	938	872	825	791	745	705	675	642	614	588	559	536	509	488	466	442	421	393	363	344	328	310	286	264	33
46	10000	4895	3392	2641	2190	1860	1609	1437	1319	1218	1132	1066	997	929	882	843	793	756	728	690	658	632	602	573	547	520	492	463	441	412	380	360	343	320	297	273	45
47	10000	5015	3486	2742	2284	1933	1696	1527	1403	1294	1207	1138	1067	994	947	904	854	818	786	741	705	675	636	601	574	543	509	483	462	435	404	387	369	347	321	292	46
48	10000	5161	3662	2911	2441	2076	1836	1659	1534	1418	1328	1259	1179	1099	1047	991	937	898	864	813	779	744	697	655	623	588	549	520	494	470	438	414	389	359	338	311	45
49	10000	5140	3702	2995	2536	2181	1939	1772	1642	1523	1423	1352	1271	1191	1137	1073	1018	968	929	881	837	798	749	699	667	629	586	557	525	497	470	448	419	384	358	334	51
50	10000	5245	3801	3093	2637	2305	2057	1875	1736	1618	1518	1440	1358	1285	1220	1148	1087	1037	988	945	898	847	794	739	697	649	609	567	538	505	483	455	427	391	367	345	51
51	10000	5310	3904	3198	2746	2414	2175	1984	1838	1715	1614	1527	1447	1374	1302	1226	1158	1096	1040	995	943	883	827	776	732	685	646	607	572	536	514	482	451	414	389	366	42
52	10000	5297	3931	3260	2828	2506	2276	2082	1941	1815	1709	1623	1543	1467	1391	1318	1239	1165	1109	1063	1009	950	895	843	796	741	705	652	615	572	543	512	480	432	404	383	49
53	10000	5336	3992	3361	2939	2618	2384	2198	2055	1920	1813	1724	1643	1568	1491	1407	1324	1241	1176	1121	1058	994	932	879	829	771	735	672	632	582	556	521	490	443	412	387	48
54	10000	5316	3998	3395	2976	2673	2440	2252	2120	1987	1882	1793	1706	1631	1550	1457	1368	1282	1208	1145	1090	1023	956	903	840	779	739	677	638	591	565	532	504	463	433	409	44
55	10000	5336	3875	3271	2878	2582	2367	2202	2075	1947	1842	1758	1671	1592	1514	1426	1332	1246	1175	1111	1062	1001	939	885	830	776	737	685	637	589	564	535	506	470	439	414	59
56	10000	5375	3714	3123	2753	2474	2265	2115	2003	1890	1791	1710	1627	1546	1478	1396	1314	1236	1171	1112	1071	1013	957	902	852	798	762	711	670	622	600	569	540	508	475	450	107
57	10000	5422	3502	2930	2581	2322	2125	1991	1889	1788	1700	1623	1547	1469	1407	1330	1258	1187	1127	1070	1032	979	930	882	839	795	762	722	682	642	624	595	571	547	520	496	174
58	10000	5426	3437	2876	2544	2297	2108	1986	1894	1798	1710	1636	1558	1476	1416	1339	1263	1192	1131	1071	1041	990	942	892	848	804	770	734	691	650	633	606	582	562	533	509	181
59	10000	5449	3311	2762	2450	2217	2039	1931	1849	1762	1679	1608	1530	1448	1392	1317	1245	1177	1117	1058	1034	986	942	892	851	811	778	747	705	666	650	624	601	587	559	535	212
60	10000	5472	3184	2649	2356	2138	1970	1876	1804	1726	1647	1579	1503	1420	1368	1296	1226	1161	1104	1044	1027	982	941	893	854	817	785	760	718	681	667	643	621	613	585	561	244
61	10000	5496	3058	2536	2262	2059	1900	1821	1759	1690	1615	1551	1476	1391	1344	1274	1208	1146	1090	1031	1020	978	941	893	857	824	793	774	731	696	684	661	641	638	611	587	275
62	10000	5519	2931	2422	2168	1980	1831	1766	1714	1654	1584	1522	1449	1363	1320	1253	1189	1131	1077	1017	1013	974	941	894	860	831	801	787	744	711	701	680	661	663	636	612	307
63	10000	5542	2805	2309	2074	1901	1762	1711	1669	1618	1552	1494	1422	1335	1296	1231	1170	1115	1063	1004	1006	970	940	894	864	837	809	801	757	726	719	698	681	689	662	638	338
64	10000	5565	2679	2195	1980	1822	1693	1656	1624	1581	1520	1465	1395	1306	1272	1210	1152	1100	1050	990	999	966	940	895	867	844	816	814	771	741	736	717	700	714	688	664	370
65	10000	5588	2553	2082	1886	1743	1624	1601	1579	1545	1489	1437	1368	1278	1248	1188	1133	1085	1036	977	992	962	940	895	870	851	824	828	784	756	753	735	720	739	714	690	401
66	10000	5611	2427	1968	1792	1664	1555	1546	1534	1509	1457	1408	1341	1250	1224	1167	1114	1070	1023	963	985	958	940	896	873	858	832	841	797	771	770	754	740	764	740	716	433

FIGURE IX.98: BCAC 1996

Table de maintien en incapacité BCAC 2014

Âge à l'entrée en incapacité	ancienneté en mois en incapacité																																					
	0	1	1,5	2	2,5	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
20	10000	4414	3545	2676	2308	1939	1536	1256	1036	856	709	587	489	417	364	320	280	247	223	208	196	184	168	150	135	120	104	89	80	76	74	70	64	57	50	42	33	24	18
21	10000	4638	3746	2855	2460	2064	1615	1305	1068	881	731	613	519	446	389	342	302	270	246	228	213	198	184	168	154	139	124	110	100	93	87	80	72	62	53	43	33	23	17
22	10000	4848	3936	3024	2598	2173	1678	1343	1091	896	745	628	535	461	402	355	317	286	260	238	219	202	187	172	158	145	132	120	110	100	94	85	75	64	54	44	33	22	17
23	10000	5026	4098	3169	2722	2274	1746	1386	1120	917	763	644	550	476	415	367	328	295	266	241	220	202	186	172	159	148	137	127	118	110	103	95	85	75	65	55	43	32	25
24	10000	5159	4223	3286	2827	2367	1815	1434	1155	946	789	667	572	495	433	381	338	301	268	241	219	201	186	173	161	150	141	132	123	115	109	101	92	82	73	63	51	40	33
25	10000	5273	4334	3395	2925	2456	1886	1489	1198	982	819	694	596	517	452	398	352	312	278	249	226	207	192	179	167	156	146	137	128	120	113	104	95	85	75	64	53	41	33
26	10000	5357	4417	3478	3003	2528	1944	1534	1233	1009	843	716	617	536	470	415	369	328	292	262	238	219	202	189	176	164	153	143	134	125	117	108	98	87	76	66	53	41	33
27	10000	5427	4488	3550	3070	2590	1996	1580	1272	1044	875	745	643	561	493	436	389	348	311	281	255	234	217	202	189	176	163	152	141	132	123	113	103	92	81	71	59	46	39
28	10000	5477	4541	3604	3123	2641	2043	1623	1313	1083	910	779	675	591	521	462	413	371	334	302	273	250	231	215	201	187	174	161	150	140	130	120	109	98	86	76	63	50	42
29	10000	5526	4590	3654	3171	2687	2087	1666	1354	1123	948	813	705	618	546	486	434	391	352	319	289	264	242	225	209	195	181	168	156	145	135	124	113	102	90	79	67	54	46
30	10000	5567	4633	3699	3215	2731	2129	1704	1391	1156	977	838	728	639	565	503	450	405	365	329	298	272	250	231	214	198	184	170	157	145	135	124	113	101	90	79	67	54	46
31	10000	5614	4681	3749	3264	2780	2173	1744	1427	1189	1009	868	755	663	587	523	469	421	379	342	309	282	260	241	223	205	189	174	160	148	137	126	114	102	90	79	66	53	45
32	10000	5645	4718	3792	3307	2823	2214	1784	1464	1224	1042	899	783	688	609	543	487	438	395	357	323	296	273	253	234	216	198	182	168	154	142	130	117	104	92	80	67	53	45
33	10000	5670	4746	3823	3340	2857	2250	1821	1501	1260	1077	931	812	713	631	562	505	456	413	376	343	315	291	269	249	230	212	195	179	164	151	138	124	110	97	84	69	55	46
34	10000	5654	4735	3815	3337	2859	2262	1840	1525	1286	1103	956	834	733	648	578	520	472	430	394	361	333	307	283	262	243	224	206	189	174	160	146	132	117	103	90	75	60	51
35	10000	5629	4710	3791	3317	2844	2259	1848	1541	1305	1122	974	853	752	668	598	540	492	450	412	379	349	322	297	275	255	236	217	201	185	171	156	141	126	111	97	81	65	55
36	10000	5581	4664	3747	3280	2813	2242	1844	1545	1317	1136	991	871	771	687	617	559	510	466	427	392	360	332	307	284	263	244	225	208	193	179	164	148	133	117	103	86	69	58
37	10000	5548	4633	3717	3255	2792	2231	1842	1551	1329	1154	1012	892	792	708	637	578	527	483	443	407	374	345	319	295	273	252	232	215	199	186	171	155	139	123	107	89	71	60
38	10000	5531	4617	3703	3244	2784	2231	1849	1563	1345	1174	1034	915	813	728	655	594	543	497	458	422	389	359	333	309	286	263	242	223	207	193	177	161	144	127	110	92	73	60
39	10000	5532	4618	3705	3248	2791	2243	1867	1588	1372	1200	1060	942	841	753	678	613	559	513	471	434	401	371	345	320	296	272	250	230	212	197	180	163	145	127	110	90	71	59
40	10000	5543	4629	3715	3259	2803	2259	1888	1611	1396	1225	1086	969	868	780	702	635	579	530	486	448	414	384	357	331	306	282	260	240	222	205	187	168	149	130	112	92	72	59
41	10000	5572	4657	3741	3283	2824	2279	1909	1635	1420	1250	1112	996	895	805	728	662	604	552	506	467	432	401	372	345	320	296	274	255	236	218	199	179	159	140	122	101	80	67
42	10000	5606	4690	3774	3314	2853	2305	1935	1660	1448	1278	1139	1022	920	831	754	688	630	576	528	487	451	418	388	360	335	311	289	269	249	230	210	189	169	149	129	107	85	71
43	10000	5637	4723	3809	3347	2885	2336	1965	1691	1477	1308	1167	1048	945	855	778	711	652	597	548	505	467	433	402	374	348	324	302	280	259	239	218	196	175	154	133	111	88	73
44	10000	5652	4739	3827	3366	2906	2360	1989	1715	1502	1332	1190	1069	964	874	796	728	667	611	561	517	479	445	414	386	360	335	311	288	266	245	223	201	179	158	138	115	93	78
45	10000	5666	4756	3846	3389	2931	2389	2019	1746	1535	1365	1222	1099	993	901	821	751	688	631	580	536	496	461	430	402	375	348	323	298	275	252	230	207	185	163	142	119	96	81
46	10000	5716	4811	3905	3447	2989	2443	2075	1802	1589	1418	1273	1148	1039	944	861	789	724	665	612	564	522	485	451	421	394	367	339	313	287	264	241	217	194	171	150	125	101	85
47	10000	5786	4885	3984	3524	3065	2517	2147	1872	1657	1483	1337	1210	1099	1000	914	838	770	709	653	603	558	518	481	448	418	390	362	334	307	283	258	234	209	185	162	136	110	93
48	10000	5872	4975	4078	3617	3157	2604	2226	1945	1726	1550	1403	1275	1162	1061	971	892	821	756	699	647	599	555	515	480	448	418	388	358	330	304	278	251	225	199	174	147	118	101
49	10000	5983	5091	4200	3737	3275	2713	2327	2038	1813	1633	1482	1352	1236	1132	1039	955	879	810	749	695	645	598	556	518	483	450	417	385	354	326	297	269	241	214	189	160	131	112
50	10000	6097	5216	4335	3872	3409	2841	2447	2150	1918	1731	1575	1440	1320	1212	1115	1027	947	874	809	752	699	651	607	566	528	491	455	419	385	353	321	290	261	232	205	175	144	124
51	10000	6217	5348	4479	4014	3550	2974	2571	2267	2029	1837	1674	1533	1408	1296	1193	1100	1015	937	868	807	751	701	654	610	568	528	488	449	412	378	343	310	277	246	217	185	152	131
52	10000	6349	5490	4631	4165	3699	3116	2706	2395	2151	1952	1784	1638	1507	1387	1278	1178	1087	1003	929	863	803	750	699	651	606	561	518	476	436	399	363	327	292	260	229	194	159	137
53	10000	6477	5628	4778	4311	3843	3254	2837	2520	2270	2066	1892	1742	1607	1483	1368	1261	1164	1074	994	921	856	796	740	687	637	589	543	499	456	416	376	338	301	267	235	199	163	140
54	10000	6592	5755	4917	4450	3983	3388	2966	2646	2392	2182	2005	1852	1713	1584	1463	1349	1244	1146	1059	980	908	841	779	720	666	615	567	520	474	431	388	348	309	273	240	202	165	142
55	10000	6690	5866	5043	4580	4117	3524	3102	2780	2522	2309	2129	1973	1830	1694	1566	1444	1329	1223	1128	1043	964	892	823	760	702	649	598	548	497	451	406	363	322	284	248	209	171	146
56	10000	6778	5966	5154	4695	4237	3647	3225	2899	2636	2420	2235	2073	1925	1785	1651	1522	1399	1285	1184	1093	1010	932	861	794	734	680	627	574	522	475	429	385	342	302	265	224	182	156
57	10000	6845	6043	5241	4784	4327	3736	3310	2979	2713	2492	2302	2134	1981	1837	1699	1565	1437	1318	1213	1119	1034	956	884	818	759	704	651	598	548	501	456	412	369	328	290	247	204	176
58	10000	6872	6074	5275	4817	4358	3762	3331	2997	2728	2503	2308	2134	1974	1824	1685	1552	1426	1308	1204	1111	1029	954	885	822	765	713	663	614	568	525	483	440	399	359	321	277	232	203
59	10000	6838	6030	5222	4759	4295	3694	3261	2924	2653	2425	2227	2049	1885	1734	1598	1474	1358	1250	1154	1069	993	924	860	800	747	699	655	613	573	535	496	457	419	381	345	303	259	230
60	10000	6777	5954	5130	4659	4187	3576	3138	2799	2526	2296	2096	1921	1763	1618	1490	1377	1275	1182	1099	1025	958	898	839	784	733	689	649	612	576	542	506	470	434	398	362	321	277	247
61	10000	6747	5906	5065	4583	4101	3488	3049	2714	2443	2216	2019	1849	1698	1562	1444	1338	1243	1155	1077	1009	949	892	837	784	735	690	651	614	578	543	507	471	435	400	365	326	283	255
62	10000	6710	5854	4998	4506	4014	3396	2965	2639	2377	2159	1973	1808	1663	1534	1420	1319	1228	1141	1062	993	934	880	828	776	725	676	634	597	563	528	493	457	420	384	349	307	264	235
63	10000	6698	5837	4975	4479	3983	3358	2926	2604	2347	2134	1952	1790	1648	1517	1405	1306	1213	1120	1034	958	895	843	791	736	678	625	580	545	516	487	455	421	385	349	313	271	226	197
64	10000	6668	5809	4950	4457	3963	3335	2895	2569	2313	2099	1915	1754	1616	1487	1373	1270	1166	1064	968	880	807	748	695	641	587	538	500	469	442	422	397	370	340	309	277	239	198	171
65	10000	6647	5789	4931	4444	3957	3325	2880	2537	2271	2046	1856	1693	1560	1443	1335	1227	1116	1004	895	793	702	633	586	548	509	474	440	405	383	371	356	341	323	304	285	260	233	213
66	10000	6647	5789	4931	4444	3957	3325	2880	2537	2271	2046	1856	1693	1560	1443	1335	1227	1116	1004	895	793	702	633	586	548	509	474	440	405	383	371	356	341	323	304	285	260	233	213

FIGURE IX.99: BCAC 2014

Correcteurs BCAC 2014

Age	30-45	30-60	30-75	75-90	90-120	90-150
20	80%	61%	52%	79%	65%	44%
21	81%	62%	53%	78%	63%	45%
22	81%	62%	54%	77%	62%	45%
23	82%	63%	54%	77%	61%	45%
24	82%	64%	55%	77%	61%	46%
25	82%	64%	55%	77%	61%	47%
26	82%	65%	56%	77%	61%	47%
27	83%	65%	57%	77%	61%	48%
28	83%	66%	57%	77%	61%	48%
29	83%	66%	57%	78%	62%	49%
30	83%	66%	58%	78%	62%	49%
31	83%	67%	58%	78%	63%	50%
32	84%	67%	59%	78%	63%	50%
33	84%	67%	59%	79%	64%	50%
34	84%	67%	59%	79%	64%	51%
35	84%	67%	59%	79%	65%	51%
36	84%	67%	59%	80%	66%	50%
37	83%	67%	59%	80%	66%	50%
38	83%	67%	59%	80%	66%	50%
39	83%	67%	59%	80%	67%	50%
40	84%	67%	59%	81%	67%	51%
41	84%	67%	59%	81%	68%	51%
42	84%	67%	59%	81%	68%	51%
43	84%	68%	59%	81%	68%	51%
44	84%	68%	60%	81%	68%	51%
45	84%	68%	60%	81%	69%	52%
46	84%	68%	60%	82%	69%	52%
47	84%	69%	61%	82%	70%	53%
48	85%	69%	62%	82%	71%	54%
49	85%	70%	62%	83%	71%	55%
50	86%	71%	64%	83%	72%	56%
51	86%	72%	65%	84%	72%	57%
52	86%	73%	66%	84%	73%	58%
53	87%	74%	67%	85%	74%	59%
54	87%	75%	68%	85%	74%	60%
55	88%	75%	68%	86%	75%	62%
56	88%	76%	69%	86%	76%	63%
57	88%	77%	70%	86%	76%	63%
58	88%	77%	70%	86%	76%	63%
59	88%	76%	70%	86%	76%	63%
60	88%	76%	69%	85%	75%	62%
61	88%	75%	68%	85%	74%	61%
62	87%	74%	67%	85%	74%	60%
63	87%	74%	67%	84%	73%	59%
64	87%	74%	67%	84%	73%	59%
65	87%	74%	67%	84%	73%	60%
66	87%	74%	67%	84%	73%	60%

FIGURE IX.100: Taux de passage d'une franchise 30 ou 90 jours à une franchise plus longue

Annexe 5

Le modèle à risques proportionnels de Cox

Les coefficients recherchés sont obtenus par maximum de vraisemblance. Ce maximum est recherché sur la "vraisemblance partielle de Cox". Elle correspond au produit des probabilités d'observer un décès (ou le premier arrêt de travail d'un individu) à un instant t_i.

On se place dans le cas simple où l'on considère que sur un groupe de N individus, à un instant t_i, il ne peut y avoir qu'un décès (un premier arrêt de travail). Dans ce cas, la vraisemblance partielle de Cox s'écrit de la façon suivante :

$$L(\delta) = \prod_{i=1}^{N} \frac{e^{\delta^T z_i}}{\sum_{j=1}^{N_i} e^{\delta^T z_j}}$$

Avec z_i les caractéristiques de l'individu sinistré en t_i, z_j les caractéristiques d'un individu j en t_i et N_i le nombre d'individus encore sous risques à l'instant t_i. On a alors pour la log-vraisemblance :

$$ln(L(\delta)) = \sum_{i=1}^{N} \delta^T z_i - \sum_{i=1}^{N} ln(\sum_{j=1}^{N_i} e^{\delta^T z_j})$$

La résolution du maximum de vraisemblance se fait à l'aide d'algorithmes de résolution informatique. Une valeur de $\hat{\delta}_{ML}$ est obtenue. On a donc :

$$-ln(1 - q_x^{spe}(\hat{\delta}_{ML})) = -ln(1 - q_x^{ref}).e^{\hat{\delta}_{ML}}$$

Pour des taux d'entrée en incapacité faibles, on obtient :

$$q_x^{spe}(\hat{\delta}_{ML}) = q_x^{ref}.e^{\hat{\delta}_{ML}}$$

Sinon, le développement nous amène à :

$$q_x^{spe}(\hat{\delta}_{ML}) = 1 - (1 - q_x^{ref})^{e^{\hat{\delta}_{ML}}}$$

Tests du modèle

Tests de significativité des paramètres

Deux tests doivent être réalisés, un test de significativité globale et un test de significativité de chaque paramètre. Pour cela, il est possible d'utiliser le test de Wald, le test des rapports de vraisemblance ou le test du score.

Test de significativité globale

On s'appuie ici sur le test des rapports de vraisemblance. L'hypothèse nulle du test est $H_0 : \delta = 0$ (tous les coefficients sont nulles). Sous cette hypothèse, on obtient :

$$\chi^2(p) = 2[L(\hat{\delta}) - L(0)]$$

Avec p, le nombre de degrés de liberté (la taille du vecteur δ).

Test de significativité de chaque paramètre

L'hypothèse nulle du test est $H_0 : \delta_j = 0$, $1 \leq j \leq p$ (le j^e coefficient est nul). Sous cette hypothèse, on obtient :

$$\chi^2(p) = 2[L(\hat{\delta}) - L(\delta, \delta_j = 0)]$$

Test de l'hypothèse de hasard proportionnel

De la même façon que pour la significativité des coefficients, un test global et un test modalité par modalité doit être effectué.
Les tests se basent à nouveau sur une statistique du χ^2 et les hypothèses nulles : $H_0 : \delta(t) = \delta = cste$ et $H_0 : \delta(t)x^j = \delta^j = cste$.

En toute rigueur, il est aussi nécessaire de vérifier l'hypothèse de log linéarité en vérifiant que la différence des logarithmes des fonctions de hasard de base et de la population spécifique de caractéristiques z est linéaire avec z.

Références

Rémi BELLINA (2014), *Méthodes d'apprentissage appliquées à la tarification non-vie*, Mémoire d'actuariat

Michel LEBLANC, John CROWLEY (1992), *Relative Risk Trees for Censored Survival Data*, International Biometric Society

DARES (2013), *Les absences au travail des salariés pour raison de santé*, Publication de la direction de l'animation de la recherche, des études et des statistiques

DARES (2013), *La répartition des hommes et des femmes par métiers*, Publication de la direction de l'animation de la recherche, des études et des statistiques

Aurélie GAUMET (2001), *Construction de tables d'expérience pour l'entrée et le maintien en incapacité*, Mémoire d'actuariat

Aymric KAMEGA, Frédéric PLANCHET (2011), *Hétérogénéité : mesure du risque d'estimation dans le cas d'une modélisation intégrant des facteurs observables*, Bulletin Français d'actuariat, Vol. 11, n°21, janvier ? juin 2011, pp. 99 - 129

Tom LEURENT (2010), *Construction de tables d'expérience des risques incapacité et invalidité*, Mémoire d'actuariat

Xavier MILHAUD (2016), *Cours d'introduction au big data*, ISFA

Ministère des droits des femmes (2014), *Vers l'égalité réelle entre les femmes et les hommes*

Walter OLBRICHT (2012), *Tree-based methods : a useful tool for life insurance*, Springer

Antoine PAGLIA, Martial V. PHELIPPE-GUINVARCH (2011), *Tarification des risques en assurance non-vie, une approche par modèle d'apprentissage statistique*, Bulletin Français d'actuariat , Vol. 11, n°22, juillet-décembre 2011, pp. 49 - 81.

Frédéric PLANCHET, Pierre THEROND (2006), *Modèles de durée - applications actuarielles*, Economica

Julien TOMAS, Frédéric PLANCHET, *Méthode de positionnement : aspects méthodologiques*, Note de travail II1291-12 v1.5, Institut des actuaires

Table des figures