

Mémoire présenté le :
**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires**

Par : Stella Lugiery

Titre : Hétérogénéité du maintien en incapacité et du passage en invalidité

Confidentialité : NON (Durée : 1 an 2 ans)

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

*Membres présents du jury de Signature
l'Institut des Actuaires*

Célya Meunier
Jean Modry

*Membres présents du jury de
l'ISFA*

Esterina Masiello

Entreprise :

Nom : Sham

Signature :

*Directeur de mémoire en entre-
prise :*

Nom : Sylvain Caraco

Signature :

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

Signature du candidat

Mémoire d'actuaire

Hétérogénéité du maintien en incapacité et du passage en invalidité

Stella Lugiery

Année 2020 - 2021



Je tiens à témoigner ma reconnaissance envers mon tuteur d'entreprise, M. Sylvain Caraco, pour son encadrement, sa bienveillance et ses conseils avisés.

J'adresse ma gratitude à mon tuteur pédagogique, M. Pierre-Emmanuel Thérond pour sa réactivité et son accompagnement.

Je remercie les collaborateurs de Relyens, avec lesquels j'ai pu échanger et évoluer, pour leur accueil et leur participation au bon déroulement de mon alternance.

J'adresse mes sincères remerciements à l'équipe pédagogique de l'ISFA pour la qualité de leurs enseignements.

Enfin, un grand merci à mes professeurs de mathématiques de lycée et de classes préparatoires qui m'ont tour à tour poussée à donner le meilleur de moi-même et portée jusqu'à l'actuariat :

M. Robert Adolle, M. Steeve Bourgade, M. Jérôme Demange et M. Walter Appel.

Résumé

Mots-clés : incapacité, invalidité, Kaplan-Meier, Cox, arbre de décision, survie, hétérogénéité, log-rank.

Ce mémoire s'inscrit dans le besoin de Relyens d'améliorer sa gestion des risques d'incapacité temporaire et permanente de travail auxquels son portefeuille est exposé, en enrichissant sa connaissance de ses clients, afin d'être en mesure de proposer par la suite des modulations tarifaires. Il répond à la problématique suivante : Comment mettre en exergue l'hétérogénéité du portefeuille de Relyens vis-à-vis des risques de maintien en incapacité et d'invalidité en attente afin d'affiner sa tarification ?

Après la description des risques sous-jacents à l'arrêt de travail, des obligations patronales induites et du périmètre d'intervention des assureurs, une première vision des deux risques est donnée par la présentation des variables étudiées à travers le taux de passage en invalidité et le taux d'absentéisme.

L'hétérogénéité de l'évolution des risques de maintien en incapacité et d'invalidité en attente au cours du temps est évaluée par l'estimation des risques instantanés de maintien en incapacité et d'invalidité en attente, par les méthodes de Kaplan-Meier et de Kalbfleisch-Prentice, segmentés selon les variables explicatives étudiées. Dès lors, à partir du test du log-rank, les variables permettant de définir une segmentation significative des courbes de survie en incapacité sont identifiées puis implémentées dans le modèle de Cox mettant en évidence l'effet de leur interférence.

Enfin, le risque de passage en invalidité est résumé à travers un arbre de décision établi par sous-échantillonnage de sa non-survenance puis par modification de la probabilité a priori de sa survenance. Ces deux méthodes seront ensuite comparées et validées en étudiant divers indicateurs de pertinence.

Abstract

Keywords : incapacity, disability, Kaplan-Meier, Cox, decision tree, survival analysis, heterogeneity, log-rank.

This thesis is part of Relyens' need to improve the management of its portfolio of temporary and permanent disability risks, by enriching the client's knowledge, thus being able to offer rate modulations later on. It responds to the following issue : How to highlight the heterogeneity of Relyens' portfolio with regard to the risks of continued incapacity and pending disability to refine the pricing?

After describing the risks underlying work stoppages, the employer's obligations and the scope of intervention of insurers, an initial view of the two risks is given by presenting the variables studied through the rate of passage into disability and the rate of absenteeism.

The heterogeneity in the evolution of the risks of remaining disabled and pending disability over time is evaluated by estimating the instantaneous risks of remaining disabled and pending disability, using the Kaplan-Meier and Kalbfleisch-Prentice methods, segmented according to the explanatory variables studied. From then on, using the log-rank test, the variables allowing to define a significant segmentation of the survival curves in disability are identified and then implemented in the Cox model highlighting the effect of their interference.

Finally, the risk of becoming disabled is summarized through a decision tree established by undersampling its non-survival and then modifying the prior probability of occurrence. These two methods will then be compared and validated by studying various relevance indicators.

Table des matières

Résumé	v
Abstract	vi
Table des matières	vii
Introduction	1
CONTEXTE DE L'ÉTUDE, DONNÉES ET ANALYSE DESCRIPTIVE	3
1 La protection sociale en France	4
1.1 Le risque d'arrêt de travail	4
1.1.1 L'incapacité	4
1.1.2 L'invalidité	5
1.2 Obligation de maintien de salaire par l'employeur	6
1.3 La loi Évin	7
1.3.1 La loi Évin du 31 décembre 1989	7
1.3.2 L'arrêté du 28 mars 1996	7
1.4 Salaire de remplacement en incapacité	8
1.4.1 En cas d'arrêt non lié à l'activité professionnelle, le salaire de remplacement peut fortement diminuer au cours du temps	8
1.4.2 En cas d'arrêt lié à l'activité professionnelle, la sécurité sociale indemnise l'assuré de façon plus conséquente	9
1.5 Les tables réglementaires de provisionnement	10
2 Présentation des données	11
2.1 Table des franchises	11
2.2 Table des effectifs	12
2.3 Table des masses salariales	13
2.4 Table des cotisations appelées	13
2.5 Table des clients	13
2.6 Tables des arrêts en incapacité et en invalidité	14
3 Hétérogénéité du taux de passage en invalidité et du taux d'absentéisme	16
3.1 Le taux de passage en invalidité	16
3.1.1 Périmètre d'observation	16
3.1.2 Critère de significativité	16
3.2 Le taux d'absentéisme	17
3.2.1 Périmètre d'observation	18
3.2.2 Redressement des données	18
3.2.3 Critère de significativité	19

3.3	Corrélation entre le taux de passage en invalidité et le taux d'absentéisme	20
3.3.1	L'influence de la taille des entreprises (en termes d'effectif)	23
3.3.2	L'influence de l'âge des assurés	24
3.3.3	L'influence du sexe des assurés	25
3.3.4	L'influence de la catégorie socioprofessionnelle des assurés	26
3.3.5	L'influence de la zone géographique	26
3.3.6	L'influence de la catégorie de l'activité des entreprises	28
3.3.7	L'influence de la direction client des entreprises	28
3.3.8	L'influence du statut juridique des entreprises	29
3.3.9	L'influence de la convention	29
ÉTUDE DES LOIS DE MAINTIEN EN INCAPACITÉ ET D'INVALIDITÉ EN ATTENTE		30
4	Traitement des données	31
4.1	Méthodologie appliquée	31
4.2	Variables étudiées	34
4.3	Notations	34
4.3.1	Application : Fonction de survie en incapacité	36
4.4	Fonction de hasard	37
4.5	Censures et troncatures de l'information	38
4.5.1	Censure à droite de type I	38
4.5.2	Censure à droite de type III	38
4.5.3	Troncature à gauche	38
5	Estimation des fonctions de hasard	40
5.1	L'estimateur de Kaplan-Meier	40
5.1.1	Construction de l'estimateur de Kaplan-Meier	40
5.1.2	Application : Estimation de la fonction de hasard du passage en invalidité par Kaplan-Meier	41
5.2	L'estimateur de Kalbfleisch et Prentice	43
5.2.1	Les risques compétitifs	43
5.2.2	Construction de l'estimateur de Kalbfleisch et Prentice	43
5.2.3	Application : Estimation de la fonction de hasard du passage en invalidité par Kalbfleisch et Prentice	44
5.3	Hétérogénéité de la fonction de hasard du risque d'invalidité en attente	45
5.3.1	Statut juridique	45
5.3.2	Convention	46
5.3.3	Catégorie socioprofessionnelle	46
5.3.4	Âge	47
5.4	Comparaison des courbes de survie	49
5.4.1	Le test du log-rank	49
5.4.2	Application du test du log-rank	49
6	Modèle de Cox	50
6.1	Présentation du modèle de Cox	50

6.2	Tests de validation du modèle	51
6.2.1	Tests de significativité globale des coefficients de régression	51
6.2.2	Test de l'hypothèse de hasard proportionnel	51
6.3	Application du modèle de Cox	52
6.3.1	Coefficients de régression	52
6.3.2	Validation du modèle	53
6.4	Modèle de Cox stratifié	54
6.5	Présentation du modèle	54
6.5.1	Application du modèle	54
6.5.2	Validation du modèle	54
CLASSIFICATION DU RISQUE D'ENTRÉE EN INVALIDITÉ		56
7	Enjeux de la classification	57
7.1	Contexte	57
7.2	Variables étudiées	57
7.3	Structure d'un arbre de décision	58
8	Arbre de décision du risque de passage en invalidité	59
8.1	Méthode 1 : Sous-échantillonnage	59
8.2	Méthode 2 : Modification de la probabilité a priori	61
8.3	Autre méthode : Spécification d'une matrice de perte	61
9	Validation des modèles	63
9.1	Matrice de confusion	63
9.2	Précision du modèle	64
9.2.1	Expression	64
9.2.2	Limites	64
9.3	Sensibilité et spécificité du modèle	65
9.3.1	Expression	65
9.3.2	Taux de faux négatifs en fonction du seuil d'affectation	65
9.3.3	Courbe ROC et AUC	67
Synthèse		69
Conclusion		72
Bibliographie		75
Annexes		78
1	Création des tranches de taille d'entreprise par Maan-Whhitney	78
2	Création des tranches d'âge par Maan-Whhitney	80
Lexique		82

Table des figures

1.1	Le champ d'intervention de l'assureur est important pour un arrêt non lié à l'activité professionnelle	8
1.2	Pour un arrêt lié à l'activité professionnelle, le champ d'intervention de l'assureur est inférieur à 20% du salaire	9
3.1	Le taux de passage en invalidité et le taux d'absentéisme sont corrélés positivement	20
3.2	La régression du taux de passage en invalidité en fonction du taux d'absentéisme montre une corrélation linéaire légèrement positive	21
3.3	Le taux de passage en invalidité et le taux d'absentéisme évoluent de façon analogue en fonction de la taille des entreprises (en termes d'effectif)	24
3.4	Le taux de passage en invalidité et le taux d'absentéisme augmentent de façon significative avec l'âge des assurés	25
3.5	Le taux de passage en invalidité et le taux d'absentéisme doublent pour les femmes	25
3.6	Le taux de passage en invalidité et le taux d'absentéisme sont plus élevés pour les non-cadres	26
3.7	Le taux de passage en invalidité et le taux d'absentéisme par rapport à la zone géographique ne sont pas corrélés	26
3.8	Les proportions des modalités de chaque variable sont similaires d'une région à une autre	27
3.9	Le taux de passage en invalidité et le taux d'absentéisme des deux directions client sont similaires	28
3.10	Le taux de passage en invalidité et le taux d'absentéisme par rapport au statut juridique sont antagoniques	29
3.11	Le taux de passage en invalidité et le taux d'absentéisme par rapport à la convention sont analogues	29
4.1	Proportion de passage de l'incapacité à l'invalidité par âge	32
4.2	Schéma explicatif du calcul de l'exposition	33
4.3	Comparaison de la fonction de survie en incapacité du portefeuille de l'entreprise et du BCAC	36
4.4	Ratio entre la fonction de survie en incapacité du portefeuille de l'entreprise et celle du BCAC	36
5.1	Les fonctions de hasard de passage en invalidité du portefeuille de l'entreprise et du BCAC estimées par Kaplan-Meier passent de moins de 10% à environ 60% au cours du dernier mois	41
5.2	Les taux de censure des portefeuilles de l'entreprise et du BCAC sont très élevés dès le 8 ^e mois	42
5.3	La fonction de hasard du passage de l'incapacité à l'invalidité du BCAC estimées par Kalbfleisch et Prentice s'éloigne de celle du portefeuille de l'entreprise dès le 8 ^e mois	44

5.4	Le rapport entre la fonction de hasard par Kalbfleisch et Prentice du passage en invalidité du portefeuille de l'entreprise et celle du BCAC converge vers 0 à partir du 8 ^e mois	44
5.5	Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction du statut juridique	45
5.6	Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction de la convention	46
5.7	Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction de la catégorie socioprofessionnelle	46
5.8	Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction de l'âge	47
5.9	Rapport entre la fonction de hasard par Kalbfleisch et Prentice du passage en invalidité du portefeuille de l'entreprise et celle du BCAC par âge	48
6.1	Test de l'hypothèse de hasard proportionnel du modèle de Cox	53
6.2	Test de l'hypothèse de hasard proportionnel du modèle de Cox stratifié	55
8.1	Exemple : Détermination de ξ pour $i = 150$	59
8.2	Méthode 1 : Arbre de décision par sous-échantillonnage	60
8.3	Méthode 2 : Arbre de décision par modification de la probabilité a priori	61
9.1	Taux de faux négatifs en fonction du taux d'affectation négative	66
9.2	Taux de faux négatifs en fonction du seuil d'affectation	66
9.3	Courbes ROC	67

Introduction

Fondée en 1927, la société hospitalière d'assurances mutuelles (Sham) accompagne les acteurs de la santé (médecins, centres de soins, établissements de santé, hôpitaux ...), de l'action sociale et de l'action territoriale dans la gestion de leurs risques d'activité et le management de leur performance.

Elle a connu un développement soutenu depuis 2013 avec l'acquisition de Sofaxis - courtier du groupe - suivie de son implantation en Espagne en 2014 puis en Italie en 2015. Après avoir créé Neeria - prestataire de services en France, elle s'est également implantée en Allemagne. Enfin, en 2019, le Groupe Sham devient Relyens.

Avec la création de son identité de groupe, Relyens a pour ambition d'être reconnu par les acteurs de la santé, de l'action sociale et de l'action territoriale, comme le groupe mutualiste européen de référence en management et couverture des risques.

Avec le report du début de la vie active et de l'âge de départ à la retraite, le risque de perdre sa capacité de travail, et donc de ne plus être à même de subvenir à ses besoins au cours de sa période d'activité professionnelle, s'avère grandissant. Les prestations de la sécurité sociale étant insuffisantes pour maintenir son niveau de vie en cas d'incapacité temporaire ou permanente de travailler, les assureurs ont un grand rôle à jouer afin de garantir un niveau de revenu minimal à leurs clients tout au long de leur carrière.

La loi encadre le provisionnement de ce risque par les assureurs en imposant notamment qu'il soit basé sur des tables réglementaires construites à partir des données de plusieurs organismes.

Toutefois, dans un cadre concurrentiel, les assureurs se doivent de proposer la tarification la plus adaptée possible à leur portefeuille, dont le comportement peut différer de celui des tables réglementaires. Ainsi, une fine connaissance du comportement des assurés s'avère nécessaire afin d'obtenir le tarif le plus adapté que possible au profil des clients.

Ce mémoire s'inscrit dans le besoin de Relyens d'améliorer sa gestion des risques d'incapacité temporaire et permanente de travail du portefeuille de Relyens, en enrichissant sa connaissance de ses clients, afin d'être en mesure de proposer par la suite des modulations tarifaires.

Objectifs de Sham :

A partir de ce mémoire, l'objectif de Sham est d'affiner la tarification, d'une part, de la garantie incapacité temporaire, et d'autre part, de la garantie invalidité, au travers de la mesure du risque de passage en invalidité.

Les principaux facteurs à étudier sont :

- les franchises souscrites en incapacité;
- les conventions collectives;
- la démographie assurée (collège, âge et sexe);
- et, plus globalement, les caractéristiques de l'entreprise et sa situation géographique.

D'autre part, le maintien en incapacité d'expérience de Sham et le risque de passage en invalidité propres aux clients de Sham seront comparés aux tables du BCAC, à titre exploratoire, mais sans visée à constituer des tables d'expériences. Sham considère en effet que le volume de données est trop faible, en particulier sur le risque invalidité.

Déroulé du mémoire :

Nous commencerons par expliciter le contexte et l'encadrement législatif dans lequel s'instaure ce mémoire en définissant les risques sous-jacents à l'arrêt de travail, les obligations patronales induites et le périmètre d'intervention des assureurs. Ensuite, nous baliserons les possibilités et les limites d'analyse par la présentation des données utilisées pour mener à bien l'étude ainsi que des traitements préalables effectués afin de fiabiliser les données. Il s'en suivra une présentation des variables a priori explicatives à travers deux indicateurs des risques étudiés - à savoir le taux de passage en invalidité et le taux d'absentéisme - afin d'avoir une première vision des deux risques et de leur dépendance mutuelle.

Dans un second temps, les lois de maintien en incapacité et d'invalidité en attente seront étudiées sur l'ensemble des arrêts de travail en incapacité indemnisés par la mutuelle afin d'évaluer l'hétérogénéité de l'évolution des risques de maintien en incapacité et d'invalidité en attente au cours du temps. Nous poserons tout d'abord le cadre de cette étude en détaillant les traitements effectués afin d'élaborer la base permettant de créer les modèles de durée. Puis, nous estimerons les risques instantanés de maintien en incapacité et d'invalidité en attente, par les méthodes de Kaplan-Meier et de Kalbfleisch-Prentice, segmenté selon les variables explicatives étudiées. Dès lors, les variables permettant de définir une segmentation significative des courbes de survie en incapacité seront identifiées à partir du test du log-rank. Un modèle de durée du risque d'invalidité en attente fonction de l'ensemble de ces dernières sera explicité en utilisant le modèle de Cox et permettra de déterminer leur effet croisé.

Enfin, nous effectuerons la classification du risque de passage en invalidité par la construction d'un arbre de décision en utilisant des méthodes adaptées à la classification du risque de sa survenance d'un évènement rare, à savoir le sous-échantillonnage de sa non-survenance, ou encore la modification de la probabilité a priori de survenance. Ces deux méthodes seront ensuite comparées et validées en étudiant divers indicateurs de pertinence.

**CONTEXTE DE L'ÉTUDE, DONNÉES ET
ANALYSE DESCRIPTIVE**

La protection sociale fait référence à l'ensemble des mécanismes permettant de faire face financièrement aux situations pouvant entraîner une diminution des ressources ou une augmentation des dépenses. On distingue six risques sociaux principaux : les risques liés à la santé, le risque d'arrêt de travail, le risque de décès, la retraite, le risque de dépendance et le chômage.

En 2018, les prestations versées dans le cadre de la protection sociale en France s'élèvent à 742 milliards d'euros, soit 31% du produit intérieur brut (PIB) dont :

- 72% par la sécurité sociale;
- 19% par les autres administrations publiques;
- 9% par le secteur privé (dont les mutuelles et les institutions de prévoyance). [9]

1.1 Le risque d'arrêt de travail

Le risque d'arrêt de travail correspond à la baisse de revenu qu'un individu peut subir du fait de son incapacité à poursuivre tout ou partie de son activité professionnelle. La reconnaissance d'un arrêt de travail pouvant être sujette à interprétation, les assureurs s'alignent sur les critères de la sécurité sociale. Notamment, selon l'article L321-1 du code de la sécurité sociale [5], l'incapacité physique de l'assuré de poursuivre son activité professionnelle doit être constatée par son médecin traitant.

Certains facteurs influent particulièrement sur ce risque :

- Le risque a tendance à augmenter avec l'âge, c'est l'un des facteurs les plus discriminants;
- Les femmes ont tendance à avoir plus d'arrêts de travail tandis que les arrêts des hommes ont tendance à durer plus longtemps;
- Le risque est en général plus élevé pour les non-cadres;
- Pour les non-cadres, le milieu industriel est plus à risque que le milieu tertiaire;

Le code de la sécurité sociale distingue deux types d'incapacité de travail :

- l'incapacité temporaire de travail (ITT), communément appelée incapacité;
- l'incapacité permanente de travail (IPT), communément appelée invalidité.

1.1.1 L'incapacité

L'incapacité temporaire de travail - ou incapacité - ne peut dépasser trois ans. En effet, à l'issue de trois ans en incapacité, l'impossibilité de reprendre le travail induit systématiquement au passage en incapacité permanente de travail - ou invalidité.

Au cours de ces trois ans, trois causes de sortie de l'incapacité sont possibles :

- le rétablissement (et donc la reprise du travail);

- le passage en invalidité ;
- le décès.

En cas d'incapacité temporaire de travail liée à une maladie non professionnelle, la Caisse Primaire d'Assurance Maladie (CPAM) verse une indemnité journalière à l'employeur, correspondant à 50% du salaire * journalier de base, i.e. 50% de la somme des salaires bruts des trois mois précédant l'arrêt plafonnée à 1,8 SMIC et divisée par 91,25 (l'indemnité journalière est donc plafonnée à 45,55€ bruts en 2020). Cette indemnisation est effective après un délai de carence de trois jours : les trois premiers jours d'un arrêt de travail maladie ne sont pas indemnisés par la sécurité sociale †. [18]

Si l'arrêt de travail est lié à l'activité professionnelle, l'indemnité journalière est alors de 60% du salaire journalier de référence du 1^{er} au 28^e jour d'arrêt et de 80% du salaire journalier de référence à partir du 29^e jour d'arrêt. Le salaire de référence correspond au salaire du mois précédent le début de l'arrêt divisé par 30,42 dans la limite de 343,07€ et peut être revalorisé pour les arrêts de plus de trois mois. [18]

1.1.2 L'invalidité

Un individu salarié est considéré comme **invalide** au sens de la sécurité sociale [17], si sa capacité de travail est inférieure à 1/3 de celle de la moyenne des salariés de même catégorie.

Un salarié reconnu invalide peut prétendre à une compensation de sa perte de salaire par la sécurité sociale, pouvant être complétée par la garantie invalidité souscrite par son employeur.

L'invalidité se décompose en trois catégories. En effet, selon l'article L341-4 du Code de la sécurité sociale [7],

« En vue de la détermination du montant de la pension, les invalides sont classés comme suit :

1. Invalides capables d'exercer une activité rémunérée ;
2. Invalides absolument incapables d'exercer une profession quelconque ;
3. Invalides qui, étant absolument incapables d'exercer une profession, sont, en outre, dans l'obligation d'avoir recours à l'assistance d'une tierce personne pour effectuer les actes ordinaires de la vie. »

En cas d'invalidité non liée à l'activité professionnelle jusqu'à l'âge légal de la retraite, la Caisse Primaire d'Assurance Maladie (CPAM) indemnise mensuellement le salarié invalide à hauteur de :

- 30% du PMSS ‡ en cas d'invalidité de 1^{re} catégorie ;
- 50% du PMSS en cas d'invalidité de 2^e catégorie ;
- 50% du PMSS, majoré de 40% par la **majoration pour tierce personne (MTP)** permettant de rembourser les frais d'assistance, en cas d'invalidité de 3^e catégorie.

Autrement, la rente d'invalidité est évaluée en fonction du taux d'invalidité de l'individu. [20]

*. si le salarié a au moins trois enfants à charge, cette indemnité passe à 2/3 du salaire journalier de base à partir du 31^e jour d'arrêt [18]

†. sauf en cas de congé parental

‡. Le Plafond Mensuel de la Sécurité Sociale (PMSS) correspondant à 3 428€ en 2020 et en 2021. [3]

1.2 Obligation de maintien de salaire par l'employeur

La loi du 19 janvier 1978 relative à la mensualisation et à la procédure conventionnelle [21], mise à jour par la loi du 25 juin 2008 portant sur la modernisation du marché du travail [24], oblige les employeurs à assurer un certain niveau de salaire à leurs employés en arrêt de travail en incapacité ayant plus d'un an d'ancienneté dans l'entreprise. Cette obligation prend effet à partir du 8^e jour d'arrêt s'il n'est pas lié à l'activité professionnelle et dès le premier jour dans le cas contraire.

Cette obligation consiste à compléter l'indemnisation journalière de la sécurité sociale, afin de maintenir un niveau de salaire de :

- 90% du salaire de référence pendant 30 jours (+10 jours par tranches de 5 années complètes d'ancienneté, dans la limite de 90 jours);
- puis $\frac{2}{3}$ du salaire de référence pendant les 30 jours suivants (+10 jours par tranches de 5 années complètes d'ancienneté, dans la limite de 90 jours). [1, 18, 19]

La convention collective[§], correspondant au secteur d'activité de l'entreprise, peut par ailleurs imposer des conditions de maintien de salaire plus favorable pour le salarié en diminuant l'ancienneté nécessaire, en réduisant le délai de carence, en augmentant le niveau de maintien de salaire ou en augmentant sa durée.

Il est à noter que l'obligation de maintien de salaire par l'employeur peut être prise en charge totalement ou partiellement par un organisme assureur mais cette assurance ne rentre pas dans le cadre de la protection sociale car elle n'a pas d'impact sur les revenus du salarié arrêté.

§. texte conventionnel conclu entre, d'une part, un employeur ou un groupement d'employeurs et, d'autre part, une ou plusieurs organisations syndicales représentatives de salariés, en vue de déterminer l'ensemble des conditions d'emploi, de travail et de formation professionnelle des salariés ainsi que leurs garanties sociales.

1.3 La loi Évin

1.3.1 La loi Évin du 31 décembre 1989

Les faibles indemnités de la sécurité sociale et la courte durée de l'obligation de maintien de salaire par l'employeur rendent les complémentaires indispensables.

Le risque d'arrêt de travail est principalement encadré par la loi Évin du 31 décembre 1989. Elle a été écrite à la suite de la crise de la sidérurgie et de la montée en charge du SIDA afin que les salariés en arrêt de travail continuent à être indemnisés même après la fermeture de leur entreprise.

La loi Évin du 31 décembre 1989 [22] définit la prévoyance comme une opération « ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage ».

Parmi ses objectifs, on retrouve notamment :

- la protection de l'assuré en maintenant les garanties décès aux invalides et aux incapables ;
- l'harmonisation des règles de prévoyance ;
- la fin de la répartition par la constitution de provisions. [14]

Plusieurs obligations sont imposées aux organismes assureurs pour les contrats à adhésion obligatoire :

- L'article 2 de la loi Évin dispose que l'assureur doit prendre en charge les conséquences d'états pathologiques antérieurs. Ainsi, les actifs ont tous le même tarif ;
- Selon l'article 4 de la loi Évin, l'assureur doit continuer à couvrir, sans sélection médicale ni délai de carence, les salariés qui quittent l'entreprise, i.e. les retraités, les invalides, les licenciés, etc. Ainsi, cet article organise une portabilité des droits qui ne nécessite pas de provisionnement car l'assuré continue de payer des cotisations ;
- Au regard de l'article 7 de la Loi Évin, la résiliation est sans effet sur les prestations nées avant la résiliation. Il n'est pas possible d'envisager de faire de la sélection médicale ni d'envisager des prestations différenciées. Les assureurs, quelle que soit leur nature juridique, vont donc constituer des provisions.

1.3.2 L'arrêté du 28 mars 1996

L'arrêté du 28 mars 1996 [23] vient compléter la loi Évin en fixant les règles de provisionnement des prestations d'invalidité et d'incapacité. Lors de l'évaluation des provisions mathématiques d'inventaire, quatre obligations doivent être tenues par les organismes assureurs :

- Le provisionnement de l'incapacité en cours et de l'invalidité en attente, en cas d'incapacité ;
- Le provisionnement de l'invalidité en cours, en cas d'invalidité ;
- L'utilisation des tables de maintien et de passage réglementaires ou d'une table d'expérience certifiée par un actuaire indépendant de l'organisme ;
- L'utilisation d'un taux technique inférieur à 75% du taux moyen d'emprunt d'État (TME). [2, 14]

1.4 Salaire de remplacement en incapacité ¶

1.4.1 En cas d'arrêt non lié à l'activité professionnelle, le salaire de remplacement peut fortement diminuer au cours du temps

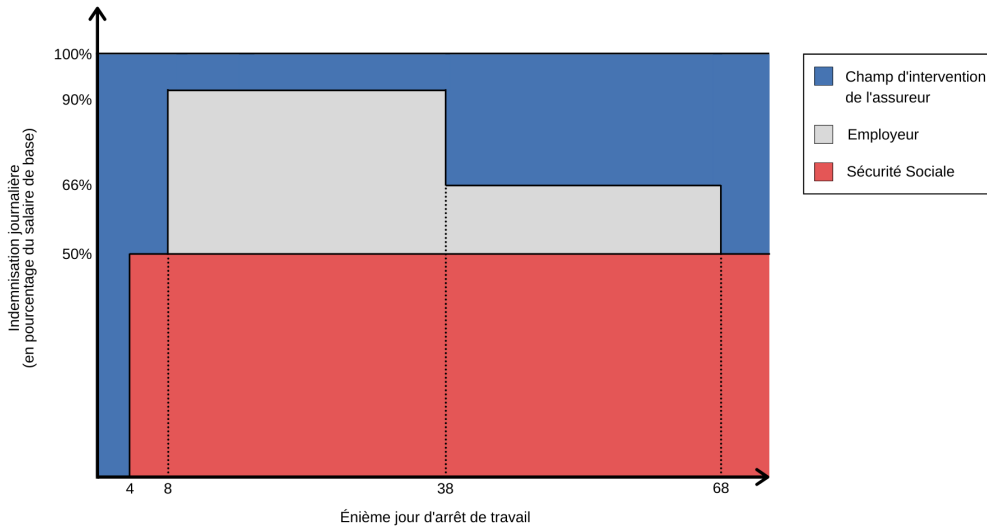


FIGURE 1.1 – Le champ d'intervention de l'assureur est important pour un arrêt non lié à l'activité professionnelle

Par exemple, pour un salarié ayant entre un et cinq ans d'ancienneté, moins de trois enfants à charge et sans convention collective, en cas d'arrêt non lié à l'activité professionnelle, la sécurité sociale indemnise le salarié à partir du 4^e jour d'arrêt à hauteur de 50% du salaire de référence (**partie rouge**). L'employeur doit maintenir le salaire de la personne arrêtée à partir du 8^e jour et à hauteur de 90% pendant un mois puis 66% pendant le mois suivant (**partie grise**). Le champ d'intervention de l'assureur correspond alors à la perte de revenu résiduel de l'assuré (**partie bleue**).

¶. pour un salarié ayant entre 1 et 5 ans d'ancienneté, moins de trois enfants à charge et sans convention collective

1.4.2 En cas d'arrêt lié à l'activité professionnelle, la sécurité sociale indemnise l'assuré de façon plus conséquente

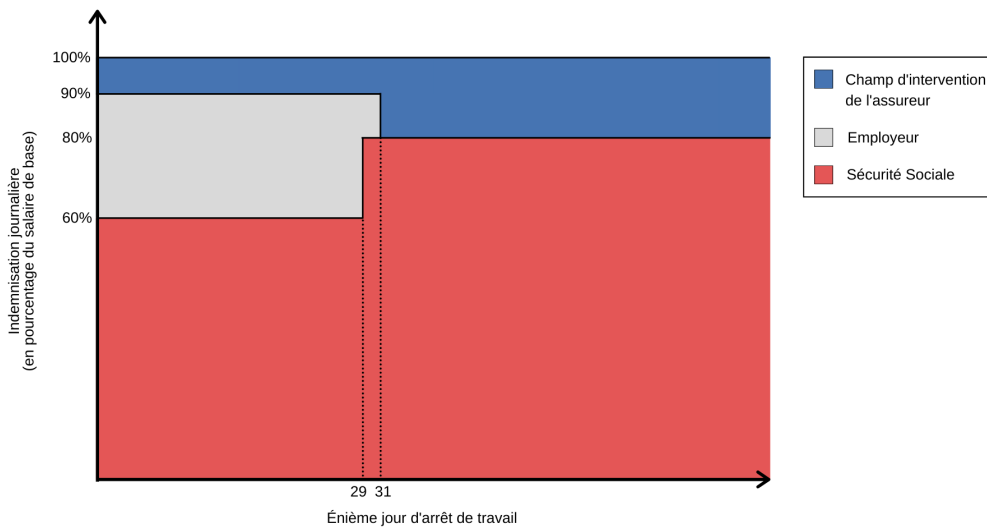


FIGURE 1.2 – Pour un arrêt lié à l'activité professionnelle, le champ d'intervention de l'assureur est inférieur à 20% du salaire

Par exemple, pour un salarié ayant entre un et cinq ans d'ancienneté, moins de trois enfants à charge et sans convention collective, en cas d'arrêt lié à l'activité professionnelle, la sécurité sociale indemnise le salarié dès le 1^{er} jour à hauteur de 60% du salaire de référence pendant 29 jours, puis à hauteur de 80% (**partie rouge**). L'employeur doit maintenir le salaire de la personne arrêtée dès le 1^{er} jour et à hauteur de 90% pendant un mois (**partie grise**). Le champ d'application de l'assureur correspond alors à la perte de revenu résiduel de l'assuré (**partie bleue**).

1.5 Les tables réglementaires de provisionnement

En plus de l'inversion du cycle de production caractérisant le secteur de l'assurance, lors de la survenance d'un arrêt de travail, il est difficile d'évaluer le coût de ce dernier, principalement du fait des risques de maintien en arrêt de travail, de rechute et de passage en invalidité.

Afin de toujours être en mesure de garantir le règlement de ses engagements, l'assureur doit constituer une réserve, appelée provision mathématique et définie par l'article R.343-3 du code des assurances [6] comme la « différence entre les valeurs actuelles des engagements respectivement pris par l'assureur et par les assurés ». Dans notre cas, elle doit « inclure une estimation des frais futurs de gestion qui seront supportés par l'assureur pendant la période de couverture au-delà de la durée de paiement des primes ».

D'après l'article 143-12 du règlement n° 2015-11 du 26 novembre 2015 [4] de l'Autorité des Normes Comptables (ANC) :

« Le calcul des provisions techniques de prestations d'incapacité de travail et d'invalidité est effectué à partir [des] lois de maintien en incapacité de travail et en invalidité indiquées aux articles 600-2, 600-3 et 600-4 [de ce] règlement. Toutefois, il est possible pour une entreprise d'assurances d'utiliser une loi de maintien établie par ses soins et certifiée par un actuaire indépendant de cette entreprise ».

Ces lois correspondent à des tables construites par le Bureau Commun d'Assurances des Collectives (BCAC) que l'on retrouve dans ce règlement :

- L'article 600-2 présente la table de maintien en invalidité en fonction de l'âge de l'assuré à l'entrée en invalidité et de son ancienneté en invalidité en années ;
- L'article 600-3 présente la table de maintien en incapacité temporaire en fonction de l'âge de l'assuré au début de l'arrêt de travail et de son ancienneté en incapacité en mois ;
- L'article 600-4 présente le nombre de passages de l'incapacité temporaire à l'invalidité ^{||} pour 10 000 assurés initialement en incapacité en fonction de l'âge de l'assuré au début de l'arrêt de travail et de son ancienneté en incapacité en mois à la sortie de l'incapacité (et donc à l'entrée de l'invalidité).

^{||}. Il est à noter que le Bureau Commun d'Assurances des Collectives (BCAC) ne fait pas de distinction entre les catégories d'invalidité, mais cela importe pour les compagnies d'assurance car les garanties diffèrent en fonction de la catégorie.

2.1 Table des franchises

Dans le cadre de cette étude, on distingue trois causes d'incapacité de travail (l'accident du travail, l'hospitalisation et la maladie) et trois catégories socioprofessionnelles (cadre, non-cadre ou ensemble du personnel).

Pour chaque contrat sont définies trois franchises applicables en fonction de la cause de l'arrêt de travail (accident du travail, hospitalisation ou maladie). Pour une même entreprise, les franchises diffèrent souvent en fonction de la catégorie socioprofessionnelle et font donc souvent l'objet de deux contrats différents : un contrat pour les cadres et un pour les non-cadres. Dans le cas où il n'y a pas de distinction, on a simplement un contrat relatif à l'ensemble du personnel.

La table des franchises fournit la durée des franchises des contrats de prévoyance collective par contrat, par type d'arrêt et par catégorie socioprofessionnelle. Elle se compose des variables suivantes :

- Nom de l'entreprise ;
- Identifiant du contrat \uparrow ;
- Nom du contrat relatif aux cadres ;
- Nom du contrat relatif aux non-cadres ;
- Convention ;
- Catégorie(s) socioprofessionnelle(s) concernée(s) par le contrat ;
- Durée de la franchise pour les cadres en cas d'un arrêt dû à un accident du travail ;
- Durée de la franchise pour les cadres en cas d'un arrêt dû à une hospitalisation ;
- Durée de la franchise pour les cadres en cas d'un arrêt dû à une maladie ;
- Durée de la franchise pour les non-cadres en cas d'un arrêt dû à un accident du travail ;
- Durée de la franchise pour les non-cadres en cas d'un arrêt dû à une hospitalisation ;
- Durée de la franchise pour les non-cadres en cas d'un arrêt dû à une maladie.

Cette table a été synthétisée en ne gardant que les franchises liées à la population à laquelle le contrat s'applique. Lorsque le contrat s'appliquait à l'ensemble du personnel, la ligne a été dupliquée afin de distinguer les deux catégories socioprofessionnelles. Ainsi, les variables informant sur la durée des franchises deviennent :

- Durée de la franchise en cas d'un arrêt dû à un accident du travail pour la catégorie socioprofessionnelle considérée (cadre ou non-cadre) ;
- Durée de la franchise en cas d'un arrêt dû à une hospitalisation pour la catégorie socioprofessionnelle considérée (cadre ou non-cadre) ;
- Durée de la franchise en cas d'un arrêt dû à une maladie pour la catégorie socioprofessionnelle considérée (cadre ou non-cadre).

Les franchises de moins de trois jours ont été regroupées. Ainsi, par la suite, lorsqu'on parlera de franchise de trois jours, il s'agira en fait des franchises de moins de trois jours.

Par ailleurs, dans notre portefeuille se trouvent des contrats liés à des centres de lutte contre le cancer (CLCC) pour lesquels, à l'inverse du reste du portefeuille, les franchises de 90 jours appliquées sont discontinues.

- La franchise continue consiste en un nombre de jours fixes s'appliquant à chaque arrêt de travail ;
- La franchise discontinue consiste en un nombre de jours cumulés qui s'applique sur l'ensemble des arrêts de travail de l'année civile ou glissante sur une année. Elle « est appréciée au premier jour d'arrêt de travail en décomptant tous les arrêts intervenus au cours des 12 derniers mois. Toutefois, dès lors que la franchise est atteinte, tout nouvel arrêt de travail entrant dans le cadre de l'appréciation du crédit d'indemnisation défini [...] fera l'objet d'une indemnisation dès le premier jour d'arrêt. » ([accord du 25 février 2013 relatif à la prévoyance](#)).

Ainsi, il est important de distinguer les franchises discontinues des autres. Pour ce faire, dans la table, l'observation '90' a été remplacée par '90D' pour la franchise de ces derniers.

2.2 Table des effectifs

La table des effectifs contient les informations disponibles sur les salariés des entreprises clientes en prévoyance collective par année. Elle contient notamment les variables suivantes :

- Année de compétence * ;
- Identifiant de l'entreprise † ;
- Identifiant du contrat † ;
- Numéro de sécurité sociale du salarié † ;
- Nom et prénom du salarié ;
- Catégorie socioprofessionnelle du salarié ;
- Date de naissance du salarié ;
- Sexe du salarié.

Parfois plusieurs numéros de sécurité sociale faisaient référence au même salarié. Dans un premier temps, ces anomalies ont été corrigées en se référant aux divers renseignements disponibles sur les salariés concernés et à la construction du numéro de sécurité sociale.

Par la suite, le sexe et l'âge des salariés ont été corrigés lorsqu'ils ne correspondaient pas à celui du numéro de sécurité sociale. Enfin, les lignes correspondant à des âges aberrants ont été supprimées, i.e. correspondant à des individus de moins de 16 ans ou de plus de 75 ans.

Par convention, l'âge de l'assuré correspond à la différence entre l'année de compétence considérée et l'année de naissance de l'assuré.

Nous avons pu constater que le nombre de clients n'est pas représentatif de la population assurée. En effet, 47% des entreprises clientes ont moins de 50 salariés et, de ce fait, ces 47% ne représentent que 4% des 460 000 assurés entre 2010 et 2018 tandis que seulement 6% des entreprises assurées ont plus de 2000 salariés, mais ces clients sont les plus importants à étudier car ils représentent 68% de la population assurée.

Enfin, nous avons créé une table synthétisant le nombre de salariés par contrat, année, catégorie socioprofessionnelle, sexe et âge.

*. année durant laquelle le contrat est valable

2.3 Table des masses salariales

Dans la table des masses salariales, on retrouve la masse salariale[†] de chaque entreprise cliente en fonction des paramètres suivants :

- le mois et l'année de compétence;
- l'identifiant de l'entreprise †;
- l'identifiant du contrat †;
- la catégorie socioprofessionnelle des salariés.

De même que pour la table des effectifs, nous pouvons synthétiser la masse salariale totale par contrat, année et catégorie socioprofessionnelle. Toutefois, n'ayant pas le salaire de chaque salarié à disposition, le sexe et l'âge des salariés ne peuvent pas être pris en considération.

Cette table a également été jointe à la table des franchises afin de récupérer l'information de la convention et des franchises associées à chaque contrat.

2.4 Table des cotisations appelées

Dans la table des cotisations appelées[‡], nous avons le montant des cotisations appelées de 2010 à 2018 par année, contrat et catégorie socioprofessionnelle.

Cette table est jointe à la table des arrêts.

2.5 Table des clients

La table des clients fournit un ensemble d'informations sur les entreprises clientes, dont notamment :

- l'identifiant de l'entreprise †;
- le département du siège;
- la région du siège;
- la catégorie de l'activité de l'entreprise;
- la sous-catégorie de l'entreprise;
- la direction client de l'activité de l'entreprise;
- le statut juridique de l'entreprise (privé ou non, à but lucratif ou non).

Dans notre étude, la région d'une entreprise est celle dans laquelle se situe son siège. Par ailleurs, en termes de nombre d'entreprises, 58% du portefeuille de Relyens en prévoyance collective se concentre sur quatre régions qui sont : l'Île-de-France (IDF), la Provence-Alpes-Côte d'Azur (PACA), l'Auvergne-Rhône-Alpes (AURA) et l'Occitanie.

Du fait de l'inégale répartition géographique des entreprises clientes, des zones géographiques plus homogènes en termes de nombre de salariés ont été créées :

†. cumul des rémunérations brutes des salariés de l'établissement (hors cotisations patronales) [11]

‡. cotisations payées par les clients ou dont ils doivent s'acquitter

- Île-de-France (IDF) \approx 24%;
- Provence-Alpes-Côte d'Azur (PACA) \approx 23%;
- Grand-Ouest \approx 15% :
 - Pays de la Loire;
 - Bretagne;
 - Normandie;
 - Nouvelle-Aquitaine;
 - Centre-Val de Loire;
- Auvergne-Rhône-Alpes (AURA) \approx 13%;
- Nord-Est \approx 13% :
 - Grand Est;
 - Hauts-de-France;
 - Bourgogne-Franche-Comté;
- Occitanie \approx 12%.

Concernant la catégorie de l'activité des entreprises clientes, on constate que la moitié du portefeuille correspond à des hébergements pour personnes âgées. Toutefois, ce chiffre reste à relativiser du fait que toutes les entreprises n'ont pas la même taille. Aussi, du point de vue du nombre de salariés assurés, on constate que derrière ces 50% se cache en fait 10% des salariés assurés tandis que les Établissements de Santé Privés et Professionnels de la Santé (ESPPS), qui ne représentent que 7% des entreprises, correspondent en fait à 72% des salariés assurés.

La convention semble être un paramètre moins intéressant à étudier car la convention de la Fédération des Établissements Hospitaliers et d'Aide à la Personne (FEHAP) concentre 94% de la population assurée et les 6% restants sont répartis équitablement dans deux autres conventions. On peut toutefois faire une étude ciblée sur les CLCC ou les FHP par exemple.

Malgré le fait que le marché ESPPS ne représente que 22% des clients, il représente 86% de la population assurée tandis que, a contrario, le marché Établissements Sociaux et Médico-Sociaux (ESMS) qui représente 77% des clients ne représente finalement que 14% de la population assurée.

98% de la population assurée travaille dans un organisme privé à but non lucratif et 2% dans un organisme privé à but lucratif.

Cette table est jointe aux autres tables afin d'ajouter ces variables - a priori explicatives - à nos études.

2.6 Tables des arrêts en incapacité et en invalidité

Nous disposons de deux tables détaillant l'ensemble des arrêts de travail du portefeuille en incapacité pour l'une et en invalidité pour l'autre entre 2010 et 2018. Elles se composent des variables suivantes :

- Année de survenance de l'arrêt;
- Identifiant du contrat †;
- Numéro de sécurité sociale du salarié arrêté †;
- Catégorie socioprofessionnelle du salarié arrêté;
- Date de naissance du salarié arrêté;

- Sexe du salarié arrêté;
- Type d'arrêt (accident du travail, hospitalisation ou maladie);
- Date de survenance de l'arrêt;
- Date d'entrée (premier jour d'arrêt indemnisé);
- Date de sortie (dernier jour d'arrêt indemnisé).

Ainsi, les variables renseignant sur le nombre de jours connus de chaque arrêt et le nombre de jours indemnisés de chaque arrêt ont pu être créées.

Enfin, pour les arrêts en incapacité, une variable segmentant les durées des arrêts a été créée selon les tranches suivantes :

- de 0 à 1 an;
- de 1 à 2 ans;
- de 2 à 3 ans.

Ces tranches nous ont notamment permis de redresser le nombre de jours d'arrêts de travail indemnisés et connus des entreprises pour les années récentes (2016 à 2018) dans le cadre de l'étude du taux d'absentéisme (cf. section 3.2).

Parfois, nous avons constaté que les catégories socioprofessionnelles de ces deux tables étaient manquantes ou ne correspondaient pas à celles de la table des franchises. Dans ces cas, nous avons décidé de forcer l'attribution de la catégorie socioprofessionnelle de la table des franchises aux tables des arrêts en incapacité et en invalidité.

Il convenait par ailleurs de supprimer les arrêts de la table des arrêts en incapacité dont la durée dépassait trois ans et qui ne rentraient donc pas dans le cadre de l'incapacité.

La jointure de la table des arrêts en incapacité et de la table des franchises nous a permis de comptabiliser le nombre total de jours d'arrêts indemnisés et connus par année de survenance en fonction du contrat concerné, du type d'arrêt, la durée des arrêts (par tranches), la catégorie socioprofessionnelle des salariés, leur sexe et leur âge.

Hétérogénéité du taux de passage en invalidité et du taux d'absentéisme

3

3.1 Le taux de passage en invalidité

Lors de la survenance d'un arrêt en incapacité, l'assureur est directement engagé pour le risque de passage en invalidité ; il doit d'ailleurs dès lors, effectuer des provisions afin d'être en mesure de respecter cet engagement. Lorsqu'une invalidité fait suite à une incapacité, on lui rattache alors l'année de début de l'arrêt en incapacité.

Le taux de passage en invalidité correspond dans cette partie au rapport entre le nombre de passages en invalidité et l'effectif total d'assurés pour une année (ou une période) et une catégorie d'individus données. Par exemple, un taux de passage en invalidité de 0,2% signifie qu'en moyenne, chaque année, sur 1000 assurés, on observe 2 passages en invalidité. Dès lors, pour une période donnée, on définit le taux de passage en invalidité par :

$$\frac{\text{Nombre de passages en invalidité suivant un arrêt ayant débuté au cours de la période}}{\text{Somme des effectifs moyens annuels} \times \text{Nombre d'années de la période}}$$

3.1.1 Périmètre d'observation

Afin d'étudier le taux d'invalidité, on ne considère que les invalidités suivant un arrêt incapacité ayant débuté il y a plus de 3 ans ; ce dernier est donc totalement observé. Ainsi, on se limite à la période de 2010 à 2017.

3.1.2 Critère de significativité

Plus un risque est fréquent, plus on en observe l'occurrence et plus vite on a une stabilité entre la modélisation et la réalisation. Si on calcule le taux de passage en invalidité sur un échantillon de 10 personnes, on peut passer d'un taux de passage en invalidité de 0% (excellent) à 10% (médiocre) avec un seul passage en invalidité. La taille de l'échantillon considéré, très faible, donne alors trop de poids au comportement d'un seul individu et rend donc le taux d'invalidité réel très volatile, i.e. il peut beaucoup varier au cours d'une année et donc beaucoup s'éloigner de la valeur calculée à partir de la photographie du portefeuille au moment de l'étude. La valeur calculée est alors inutile et on ne peut pas en tirer de conclusion pertinente.

En revanche, si notre échantillon comporte 1 000 000 d'individus, le poids d'un seul individu sur les variations du taux de passage en invalidité du groupe est négligeable. Ainsi, il faudrait beaucoup de passages en invalidité pour avoir une grosse variation du taux de passages en invalidité au cours d'une année ; ce qui est peu probable.

L'objectif de cette section est de déterminer la taille à partir de laquelle il est peu probable qu'une valeur réelle relative à l'invalidité s'éloigne fortement d'une valeur calculée ; et donc à partir de laquelle on peut considérer la valeur calculée comme significative.

Pour ce faire, nous allons calculer le nombre d'individus nécessaire pour que le risque que le nombre moyen de passages en invalidité réel s'éloigne trop (de plus de 10%) du nombre moyen calculé soit suffisamment faible (probabilité inférieure à 5%).

Le nombre d'observations du risque - égalant le produit de la fréquence de survenance et de l'effectif observé - détermine la connaissance que l'on a de ce dernier. Dans le cas du risque d'invalidité, la fréquence de survenance du risque étant d'environ $p \approx 0,4\%$, un effectif élevé s'avère nécessaire pour qu'une valeur obtenue relative à ce risque soit considérée comme significative.

Afin d'évaluer le nombre de salariés nécessaire, on considère la variable aléatoire $N \sim \mathcal{B}(n, p)$ qui représente le nombre de passages en invalidité d'un groupe d'assurés sur la période d'observation, avec n le nombre de salariés dans le groupe et p la fréquence de survenance d'un passage en invalidité du portefeuille.

On considère que le taux de passage en invalidité calculé de ce groupe est significatif si la probabilité que son nombre de passages en invalidité s'éloigne de la moyenne de plus de 10% est inférieure à 5%. Ayant un nombre d'observations suffisant, on applique le théorème central limite afin de faire l'approximation suivante :

$$\frac{N - \mathbb{E}(N)}{\sqrt{\mathbb{V}(N)}} \approx \mathcal{N}(0, 1)$$

Ainsi, on a que

$$\mathbb{P} \left(\frac{|N - \mathbb{E}(N)|}{\mathbb{E}(N)} \geq 0,1 \right) \leq 0,05 \Leftrightarrow n \geq \frac{z_{0,025}^2}{0,1^2} \left(\frac{1-p}{p} \right)$$

où z_ϵ représente le quantile à $1 - \epsilon$ de la loi normale centrée réduite.

Cela nous conduit à fixer un seuil de significativité correspondant à la quasi-totalité des assurés observés.

En conséquence, nous avons déterminé le seuil de significativité correspondant à un écart à la moyenne supérieur à 20% avec une probabilité inférieure à 15%. Une incertitude relativement forte subsiste donc sur les valeurs de taux d'invalidité obtenues.

$$n \geq \frac{z_{0,075}^2}{0,2^2} \left(\frac{1-p}{p} \right) \approx 15\,000 \quad (3.1)$$

Ainsi, l'application de la formule (3.1) conduit à considérer que la significativité est atteinte à partir de 15 000 assurés observés.

3.2 Le taux d'absentéisme

Le taux d'absentéisme, quant à lui, met en relation le nombre de jours d'arrêts de travail connus ou indemnisés et l'effectif assuré pour une année (ou une période) et une catégorie d'individus données. Pour une période donnée, on définit le taux d'absentéisme par :

$$\frac{\text{Nombre de jours d'arrêts connus ou indemnisés rattachés à la période}}{\text{Nombre de jours de la période} \times \text{Effectif assuré}}$$

3.2.1 Périmètre d'observation

Le taux d'absentéisme est étudié entre 2010 et 2018. La connaissance des sinistres en incapacité diffèrent selon la franchise appliquée qui varie entre 0 et 180 jours. Étant donné que plus de 95% des arrêts connus par l'organisme assureur correspondent à des franchises 0 ou 3 jours, on se limite ici à l'étude des taux de passage en invalidité des contrats à franchises courtes.

3.2.2 Redressement des données

Du fait des retards de déclaration, il convient de redresser les données pour les arrêts les plus récents. Pour ce faire, nous avons tronqué les arrêts ayant débuté en 2016 en ne gardant que les arrêts d'une durée inférieure à deux ans, et les arrêts de 2017 et 2018 en ne gardant que les arrêts d'une durée inférieure à un an. Nous avons ensuite prolongé le nombre d'arrêts pour les arrêts d'une durée inférieure à un an (et inférieure à deux ans) à un nombre d'arrêts évalué par Chain Ladder correspondant aux arrêts d'une durée inférieure à trois ans.

Pour toute franchise de i jours ($i \in \{3, 15, 30, 60, 90, 90D, 180\}$), on définit les coefficients de majorations suivants :

$$\begin{aligned} - \text{CM}_{1,2}^i &= \frac{1}{4} \sum_{j=2013}^{2016} \frac{\text{TA}_{i,j}(2)}{\text{TA}_{i,j}(1)}; \\ - \text{CM}_{2,3}^i &= \frac{1}{4} \sum_{j=2012}^{2015} \frac{\text{TA}_{i,j}(3)}{\text{TA}_{i,j}(2)}; \\ - \text{CM}_{1,3}^i &= \text{CM}_{1,2}^i \times \text{CM}_{2,3}^i. \end{aligned}$$

où $\text{TA}_{i,j}(k)$ correspond au taux d'absentéisme de l'année j calculé en ne comptabilisant que les arrêts d'une durée inférieure à k année(s) et avec une franchise appliquée de i jour(s).

Remarque : Les premières années ont été exclues de la détermination des coefficients de majoration car elles ne contiennent pas suffisamment de données et sont considérées comme atypiques.

Enfin, la durée de chaque arrêt a été multipliée par un coefficient de majoration fonction de l'année du début de l'arrêt, de sa durée et de la franchise qui lui a été appliquée :

Année de compétence \ Durée de l'arrêt	Durée de l'arrêt		
	Moins d'un an	De 1 à 2 ans	De 2 à 3 ans
2010 - 2015	1		
2016	$\text{CM}_{2,3}^i$		0
2017 - 2018	$\text{CM}_{1,3}^i$	0	

TABLE 3.1 – Coefficients de majoration appliqués

La méthode de Chain Ladder suppose que les cadences de survenance sont constantes. Pour vérifier empiriquement cette hypothèse, nous avons construit, pour chaque franchise, le triangle des facteurs de développement individuels $f_{i,j} = \frac{\text{TA}_{i,j+1}}{\text{TA}_{i,j}}$ et obtenu, pour chaque colonne une variance inférieure à 5% pour les franchises courtes et 20% pour les autres franchises.

Remarque : Ce redressement a été effectué de manière similaire sur les durées des arrêts indemnisés d'une part et les durées connues d'autre part.

3.2.3 Critère de significativité

De façon analogue à la section 3.1.2, l'objectif de cette partie est de déterminer la taille à partir de laquelle les valeurs relatives au risque d'incapacité que l'on calcule à partir d'un échantillon peuvent être considérées comme significatives. Pour ce faire, nous allons calculer le nombre d'individus nécessaire pour que le risque que le nombre d'arrêts en incapacité réel s'éloigne trop (de plus de 15%) du nombre calculé (l'espérance mathématique) soit suffisamment faible (probabilité inférieure à 5%).

La fréquence d'observation d'un arrêt en incapacité est de $p \approx 40\%$. La fréquence étant élevée, un effectif relativement faible s'avère suffisant afin de considérer une valeur liée au risque d'incapacité significative.

Afin d'évaluer le nombre de salariés nécessaire, on considère la variable aléatoire S représentant le nombre de jours indemnisés d'arrêt total en incapacité d'un groupe d'assurés. On suppose que

$$S = \sum_{i=1}^N X_i$$

avec $N \sim \mathcal{P}(\lambda)$ le nombre d'arrêts en incapacité du groupe d'assurés et (X_i) les durées des arrêts indemnisés (en jour) du groupe. On suppose que ces durées sont deux à deux indépendantes, identiquement distribuées, de moyenne $\mu (\approx 46 \text{ jours})$ et de variance $\sigma^2 (\approx 10\,040)$ et indépendantes de N .

On considère que le taux d'absentéisme indemnisé calculé de ce groupe est significatif si la probabilité que le nombre total de jours d'arrêts indemnisés s'éloigne de la moyenne de plus de 15% est inférieure à 5%. Ayant un nombre d'observations suffisant, on applique le théorème central limite afin de faire l'approximation suivante :

$$\frac{S - \mathbb{E}(S)}{\sqrt{V(S)}} \approx \mathcal{N}(0, 1)$$

Ainsi, on a que

$$\mathbb{P} \left(\frac{|S - \mathbb{E}(S)|}{\mathbb{E}(S)} \geq 0,15 \right) \leq 0,05 \Leftrightarrow \lambda \geq \frac{z_{0,025}^2}{0,1^2} \left(1 + \left(\frac{\sigma}{\mu} \right)^2 \right) \approx 930 \quad (3.2)$$

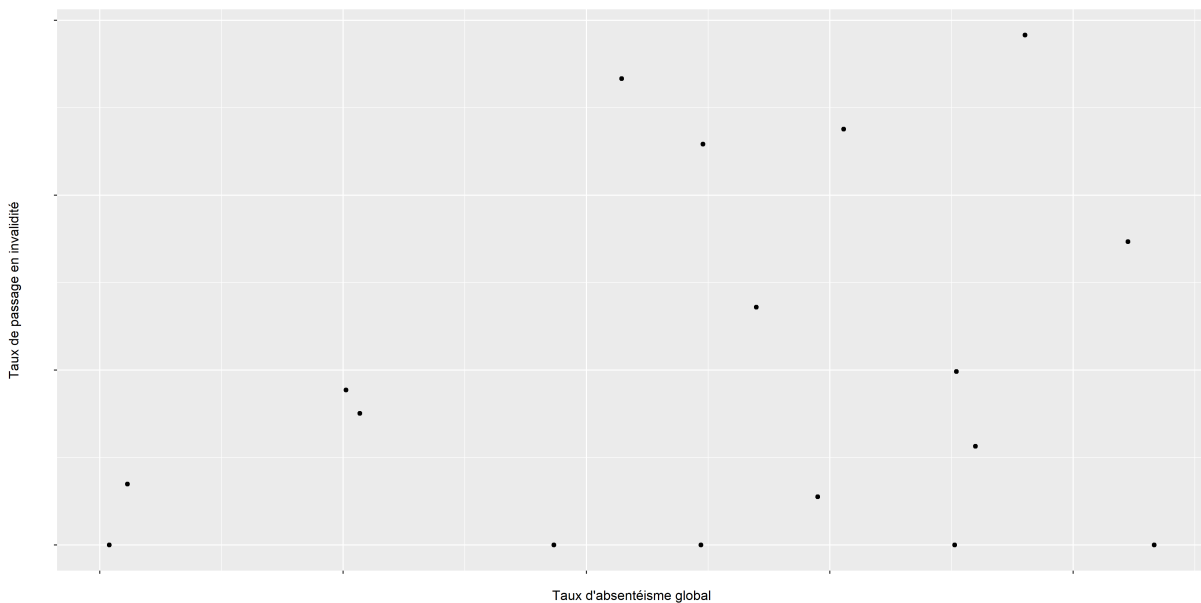
où z_ϵ représente le quantile à $1 - \epsilon$ de la loi normale centrée réduite.

En conséquence, l'application de la formule (3.1) conduit à considérer que la significativité est atteinte à partir de $\frac{\lambda}{p} \approx 2500$ assurés.

3.3 Corrélation entre le taux de passage en invalidité et le taux d'absentéisme

Nous étudions et comparons ici le taux d'absentéisme indemnisé (pour les franchises 0 et 3 jours) et le taux de passage en invalidité globaux avec une population d'environ 200 000 assurés (somme des effectifs au 31/12 de chaque année). Pour cette période, le Taux d'Absentéisme Global (TAG) indemnisé est de 4% et le Taux de passage en Invalidité Global (TIG) de 0,35%.

Pour ce faire, nous commençons par représenter le taux de passage en invalidité en fonction du taux d'absentéisme des entreprises de plus de $\frac{2500}{n}$ salariés observés par an en moyenne, pour une période d'observation de n années (36 entreprises) :



L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.1 – Le taux de passage en invalidité et le taux d'absentéisme sont corrélés positivement

Il semble il y avoir une corrélation linéaire positive entre les deux taux, notamment pour les points avec un taux d'absentéisme non nul.

Le coefficient de corrélation linéaire de Pearson, défini par :

$$\forall n \in \mathbb{N}, \forall X, Y \in \mathbb{R}^n, r(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}}$$

associé aux taux étudiés est 0.599327, ce qui traduit une corrélation linéaire modérée.

Nous effectuons donc une régression linéaire afin de confirmer et quantifier cette relation, i.e. nous cherchons à déterminer β_0, β_1 tels que pour toute entreprise i de plus 2500 salariés :

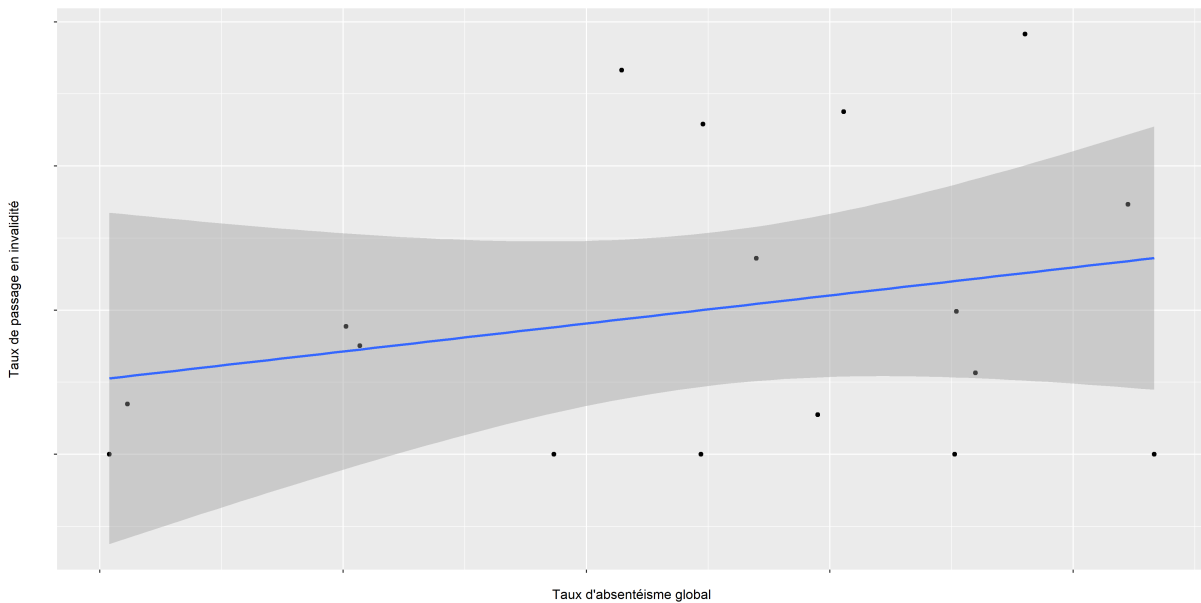
$$\text{TIG}(i) = \beta_0 + \beta_1 \times \text{TAG}(i) + \varepsilon(i), \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

avec

- β_0 (l'ordonnée à l'origine) et β_1 (la pente), les paramètres du modèle;
- ε , l'erreur ou bruit du modèle, est une variable aléatoire suivant une loi normale centrée de variance σ_ε constante. Elle correspond à l'écart entre le taux de passage en invalidité estimé par le modèle et la valeur réelle.

Pour ce faire, nous utilisons le critère des moindres carrés qui consiste à déterminer la droite qui minimise la somme des carrés de la distance entre chaque point et cette droite e_i , i.e. la somme des carrés des erreurs.

Nous obtenons le modèle suivant :



L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.2 – La régression du taux de passage en invalidité en fonction du taux d'absentéisme montre une corrélation linéaire légèrement positive

La droite de régression représentée ci-dessus a pour équation :

$$\text{TIG}(i) = -0.0012735 + 0.0931528 \times \text{TAG}(i) + \varepsilon(i), \quad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

La somme des carrés des erreurs est un bon indicateur de σ_ε qui représente la précision du modèle. Toutefois, elle est difficile à interpréter. On préférera donc utiliser l'erreur quadratique moyenne définie par :

$$\sqrt{\frac{\sum_{i=1}^n e_i^2}{n-2}}$$

où n représente le nombre d'observations.

Dans notre modèle, l'erreur quadratique moyenne est de 0.002011. Compte tenu de l'ordre de grandeur du taux de passage en invalidité global du portefeuille, elle s'avère trop élevée pour considérer que le modèle est acceptable; il n'y a pas de lien linéaire très probant entre ces deux seules valeurs. Il y a tout de même une corrélation positive entre le taux d'absentéisme - mesure pouvant être suivie facilement au cours de la vie d'un contrat - et le risque d'invalidité d'une entreprise ou

d'une catégorie de salariés donnée. Mais, les variables expliquant le taux d'absentéisme et le taux d'invalidité sont-elles les mêmes? Ont-elles la même influence sur ces deux taux?

3.3.1 L'influence de la taille des entreprises (en termes d'effectif)

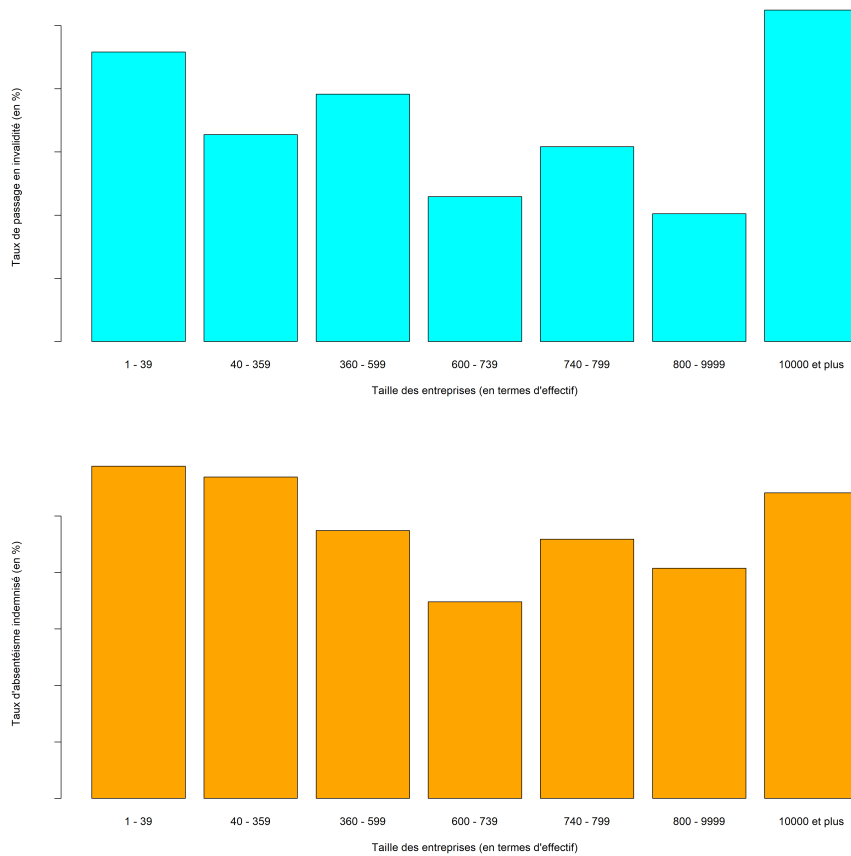
Afin d'étudier l'effet de la taille d'une entreprise (en termes d'effectif) sur le taux d'absentéisme et le taux de passage en invalidité, nous avons créé des tranches de taille d'entreprise. Pour ce faire, nous avons tout d'abord regroupé les entreprises par tranches a priori afin qu'il y ait suffisamment d'effectifs par groupe. Puis, nous avons comparé les taux d'absentéisme moyens des groupes les plus proches en termes de taille d'entreprise avec le test de Mann-Whitney. Ce test non paramétrique consiste à considérer deux échantillons indépendants E_1 et E_2 de moyennes respectives μ_1 et μ_2 et de tailles respectives η_1 et η_2 et à tester l'hypothèse (\mathcal{H}_0) : $\mu_1 = \mu_2$. Pour ce faire, pour chaque valeur de chaque échantillon, on compte le nombre de valeurs de l'autre échantillon qui sont supérieures à cette valeur. Puis, pour chaque échantillon, on somme les valeurs ainsi obtenues et on conserve la somme minimale que l'on note u et que l'on associe à une variable aléatoire U . Si η_1 et η_2 sont suffisamment grands, on suppose que sous (\mathcal{H}_0),

$$U \sim \mathcal{N}(\mu, \sigma^2) \text{ avec}$$

$$\mu = \frac{\eta_1 \eta_2}{2} \text{ et } \sigma = \sqrt{\frac{\eta_1 \eta_2 (\eta_1 + \eta_2 + 1)}{12}}$$

On rejette (\mathcal{H}_0) si le risque d'erreur est supérieur au quantile à 0,85 de la loi normale ($\approx 1,036$) (cf. Annexe 1).[28]

On obtient les tranches suivantes :



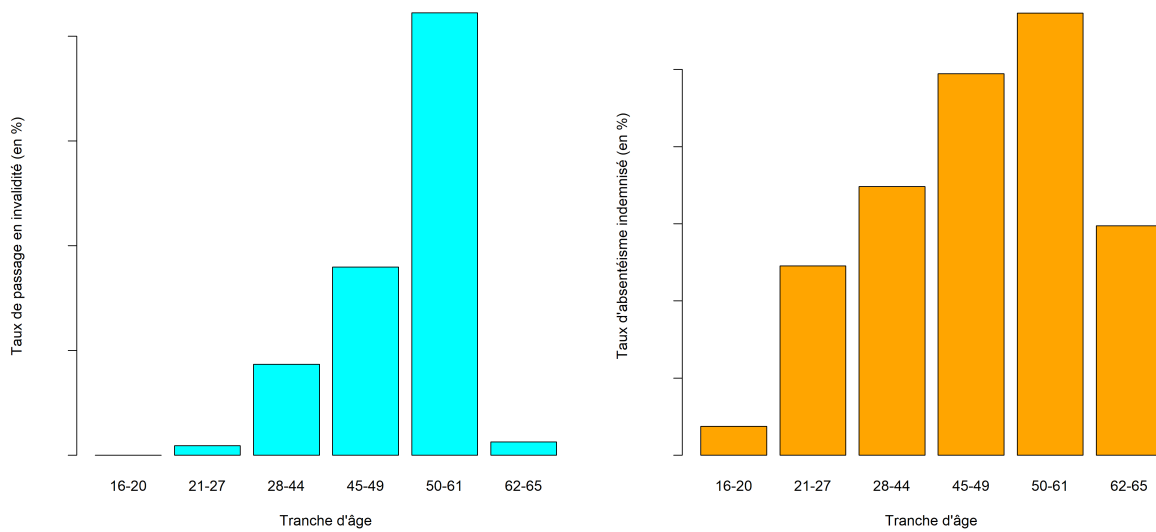
L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.3 – Le taux de passage en invalidité et le taux d'absentéisme évoluent de façon analogue en fonction de la taille des entreprises (en termes d'effectif)

Les deux taux évoluent de façon similaire avec des taux élevés pour les petites et grandes entreprises et des taux plus faibles pour les moyennes entreprises.

3.3.2 L'influence de l'âge des assurés

Après avoir regroupé les âges voisins ayant un taux d'absentéisme global moyen différant de moins de 10%, nous créons des tranches d'âge à partir du test de Mann-Whitney (cf. Annexe 2), en étudiant le nombre de jours d'arrêts de chaque salarié par an. Il est à noter que pour cette variable, quelques assurés ont été exclus de l'analyse car leur date de naissance était non renseignée ou erronée. Nous obtenons les tranches suivantes :

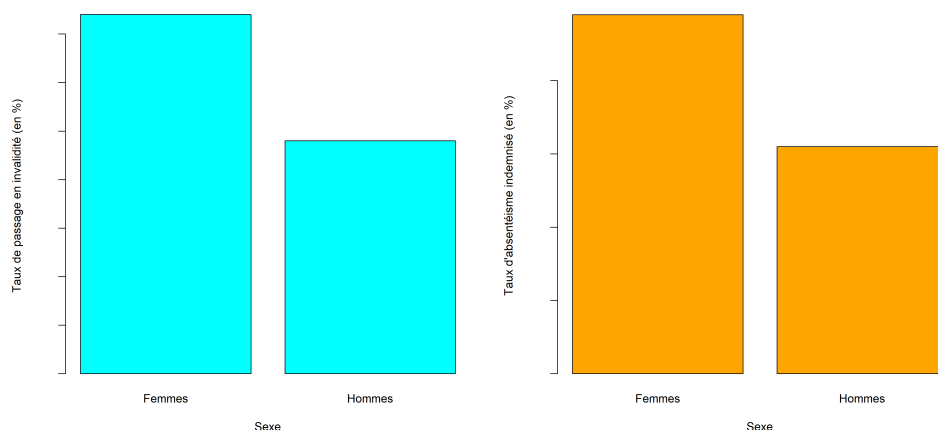


L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.4 – Le taux de passage en invalidité et le taux d'absentéisme augmentent de façon significative avec l'âge des assurés

Entre 28 et 61 ans, l'effectif dans les tranches d'âge est sensiblement constant et suffisant pour considérer le taux d'invalidité significatif. On constate que le taux d'absentéisme, comme le taux de passage en invalidité, augmente avec l'âge mais l'effet de l'âge est beaucoup plus significatif pour le taux de passage en invalidité. En effet, ce dernier est multiplié par 45 lorsque l'on passe de la deuxième à la cinquième tranche. Cette variable, classiquement très significative, pourra être étudiée afin de définir l'âge actuariel moyen des salariés d'une entreprise ou d'un groupe comme variable de segmentation tarifaire.

3.3.3 L'influence du sexe des assurés

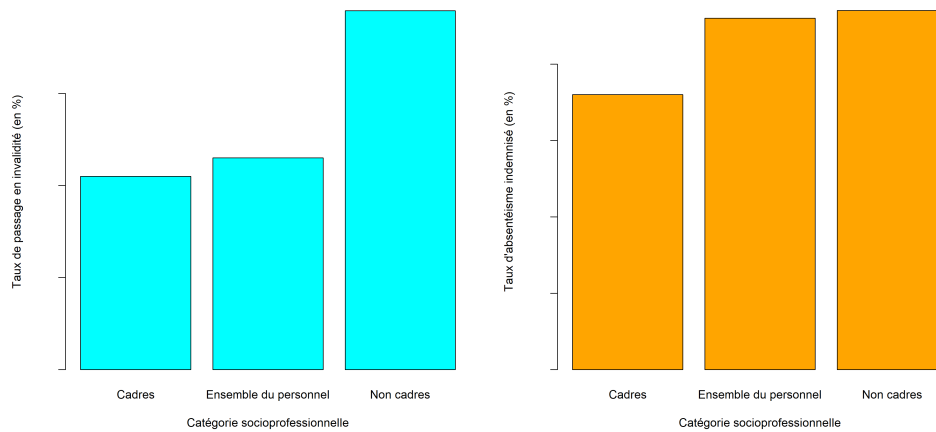


L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.5 – Le taux de passage en invalidité et le taux d'absentéisme doublent pour les femmes

L'effet du sexe sur le taux d'absentéisme et le taux d'invalidité paraît identique. On constate par ailleurs un taux d'absentéisme global et de passage en invalidité presque deux fois plus élevés chez les femmes.

3.3.4 L'influence de la catégorie socioprofessionnelle des assurés

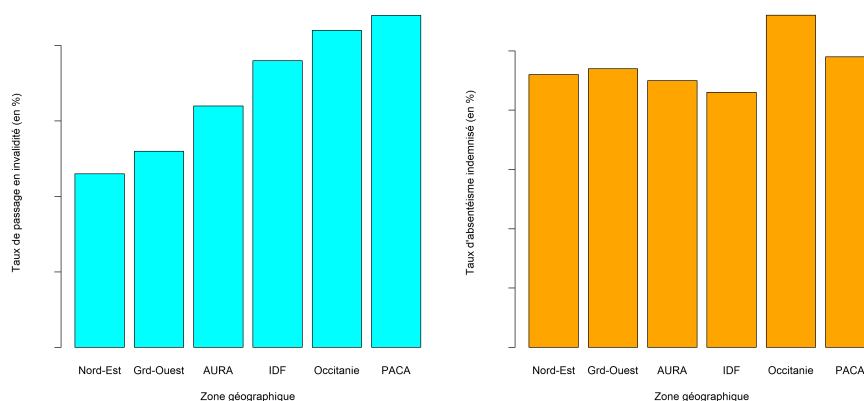


L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.6 – Le taux de passage en invalidité et le taux d'absentéisme sont plus élevés pour les non-cadres

L'influence de la catégorie socioprofessionnelle va dans le même sens pour le taux d'absentéisme et le taux d'invalidité mais semble être plus importante pour le taux d'invalidité qui est deux fois plus élevé pour les non-cadres que pour les cadres.

3.3.5 L'influence de la zone géographique



L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.7 – Le taux de passage en invalidité et le taux d'absentéisme par rapport à la zone géographique ne sont pas corrélés

On peut voir qu'il n'y a pas de corrélation entre les deux taux par rapport à la région. Le taux d'absentéisme tourne autour de la moyenne pour toutes les régions tandis que le taux de passage

en invalidité varie du simple au double. La région semble davantage jouer sur ce dernier. On peut tout de même voir que l'Occitanie et la région PACA présentent un risque particulièrement élevé.

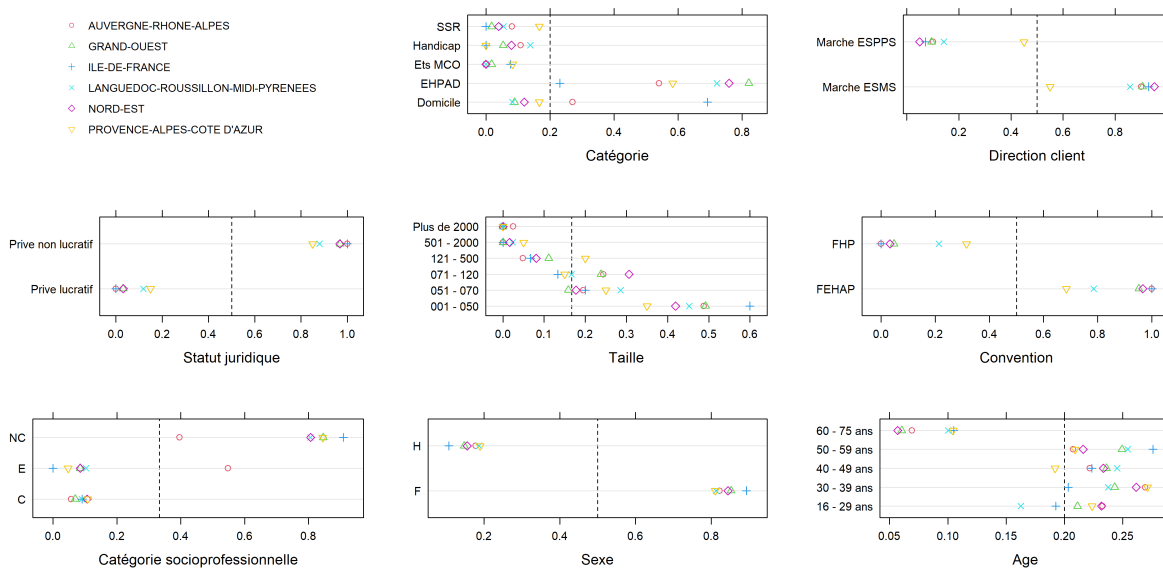


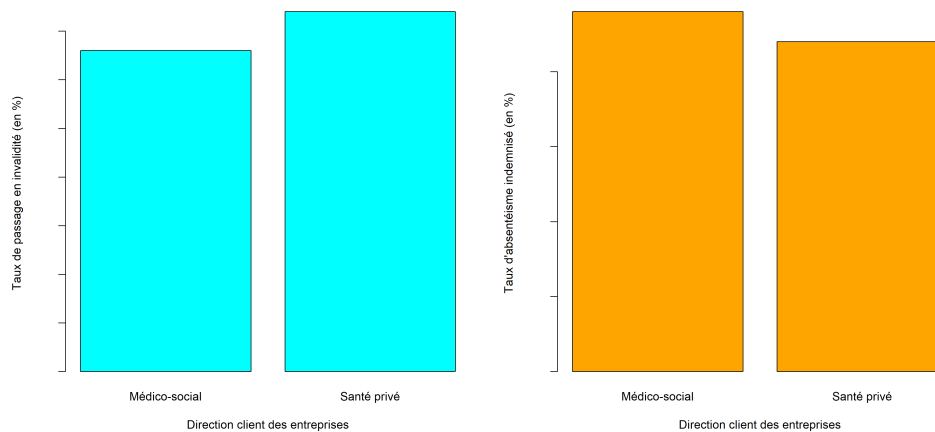
FIGURE 3.8 – Les proportions des modalités de chaque variable sont similaires d’une région à une autre

La répartition des modalités de chaque variable au sein de chaque région est représentée. Ainsi, sur chaque graphique, la somme des marqueurs représentant une même région est égale à 1. De cette façon, on peut visualiser la répartition géographique de l’ensemble des modalités significatives de l’ensemble des variables. Graphiquement, on constate une répartition géographique inégale des modalités au sein d’une région mais une répartition analogue des modalités d’une région à l’autre, notamment pour la direction client (sauf pour la région PACA), le statut juridique, la taille, la convention (sauf pour la région PACA et l’Occitanie), la catégorie socioprofessionnelle (sauf pour l’AURA), le sexe et l’âge.

3.3.6 L'influence de la catégorie de l'activité des entreprises

Nous avons une bonne vision du risque des hébergements pour personnes âgées - avec un taux de passage en invalidité dans la moyenne basse - car l'effectif est suffisant et du fait de la diversité des entreprises clientes. Pour les catégories d'Hospitalisation à Domicile (HAD), handicap, domicile et Soins de Suite et de Réadaptation (SSR), l'effectif est insuffisant pour considérer le taux de passage en invalidité significatif. En ce qui concerne les catégories d'activité Médecine, Chirurgie, Obstétrique (MCO) et Établissements de Santé Privés et Professionnels de la Santé (ESPPS), bien que l'effectif soit suffisant, elles représentent quasi exclusivement le comportement d'un unique client et donc les informations tirées de leur observation ne sont pas généralisables au secteur d'activité concerné.

3.3.7 L'influence de la direction client des entreprises

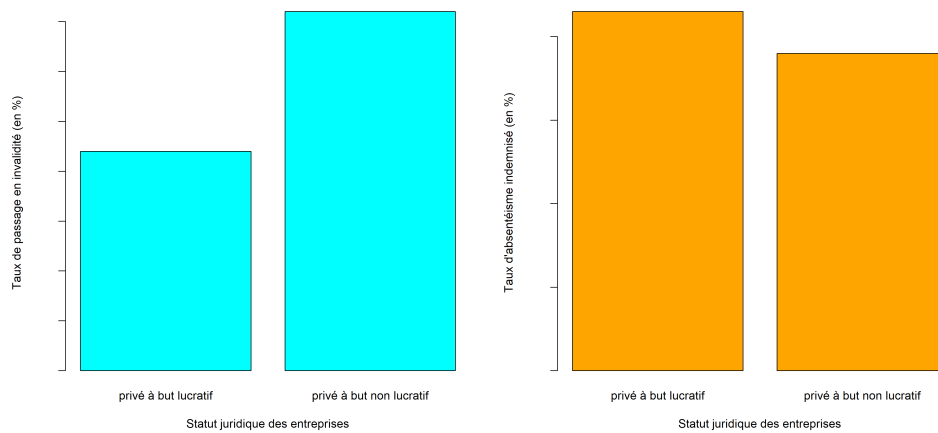


L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.9 – Le taux de passage en invalidité et le taux d'absentéisme des deux directions client sont similaires

Le taux d'absentéisme est plus important pour le marché du médico-social – qui inclut les Établissements d'Hébergement pour Personnes Âgées Dépendantes (EHPAD) – tandis que le taux de passage en invalidité est légèrement plus faible pour cette même catégorie.

3.3.8 L'influence du statut juridique des entreprises

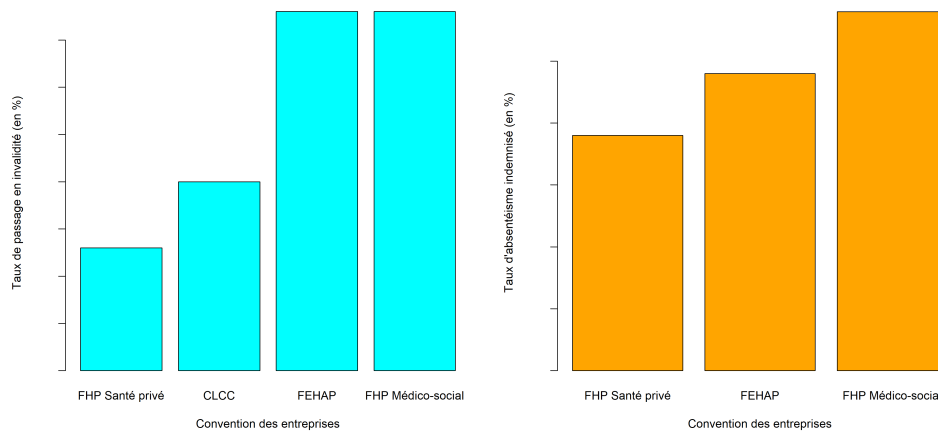


L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.10 – Le taux de passage en invalidité et le taux d'absentéisme par rapport au statut juridique sont antagoniques

Les entreprises privées à but lucratif ont un taux d'absentéisme et un taux de passage en invalidité plus faible mais il est à noter que les entreprises à but non lucratif représentent environ 90% de la population salariale étudiée.

3.3.9 L'influence de la convention



L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 3.11 – Le taux de passage en invalidité et le taux d'absentéisme par rapport à la convention sont analogues

Les CLCC ne sont pas représentés pour le taux d'absentéisme car ont très peu d'effectifs bénéficiant d'un contrat pour les franchises 3 jours. L'effet de la convention sur les deux taux est analogue.

**ÉTUDE DES LOIS DE MAINTIEN EN INCAPACITÉ
ET D'INVALIDITÉ EN ATTENTE**

Remarque : Cette partie a été effectuée avec le logiciel R, grâce aux bibliothèques DescTools, dplyr, gdata, reshape2, survival et survminer.

Il pourrait s'avérer intéressant de comparer les tables de maintien en incapacité et de passage en invalidité du BCAC avec le comportement des assurés du portefeuille. Pour ce faire, il convient de construire des tables similaires basées sur notre portefeuille.

4.1 Méthodologie appliquée

Les tables de maintien en incapacité et de passage en invalidité étant mensuelles, afin de faciliter la comparaison - et par la suite d'avoir des résultats moins volatiles, nous travaillons de même avec des durées mensuelles. Pour ce faire, un mois étant égal à $\frac{365}{12}$ jours, nous avons multiplié nos durées en jours par $\frac{12}{365}$ puis ajouté une unité à la valeur obtenue avant de la tronquer à l'unité près. Ainsi, nous obtenons la valeur du i ème mois d'arrêt au cours duquel l'incapacité se termine.

Du fait des différentes franchises appliquées, certains arrêts en incapacité ne sont connus qu'après plusieurs jours suivant le début de l'arrêt de travail. Nous appelons ici l'ancienneté d'entrée, la durée entre le début de l'arrêt de travail et la prise en charge de l'arrêt par l'assureur.

En vue de traiter les arrêts de travail dont la sortie en incapacité est totalement observée, on ne prend en compte que les arrêts ayant débuté avant le 30 avril 2017 et donc il y a plus de trois ans. Ainsi, selon la définition de la sécurité sociale de l'incapacité, elle est nécessairement terminée.

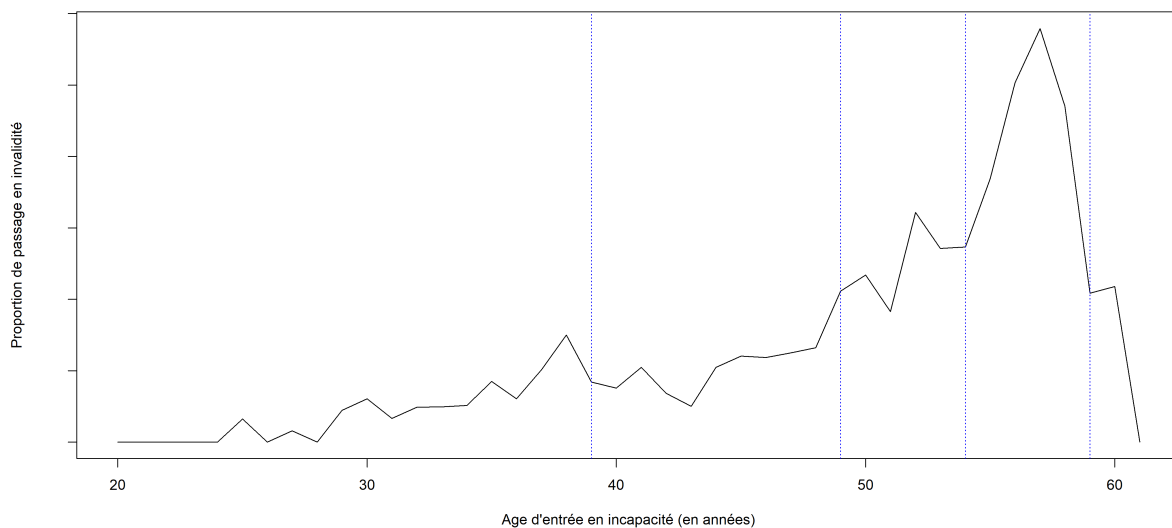
Afin d'étudier le passage en invalidité, pour chaque assuré, on joint à la ligne correspondant à l'arrêt de travail le plus récent en incapacité la date de début de l'arrêt de travail le plus ancien en invalidité, s'il existe. Dans ce cas, on fixe la valeur de l'indicateur de passage en invalidité égale à 1; sinon, il vaudra 0.

En cas de passage en invalidité, l'ancienneté de sortie en mois correspond à la différence entre la date d'entrée en invalidité et la survenance de l'arrêt en incapacité. Autrement, elle correspond à la différence entre la date de sortie et la survenance de l'arrêt en incapacité. Dans le cas où l'ancienneté de sortie en incapacité se retrouve alors supérieure à 36 mois, on impose sa valeur à 36 mois, de telle sorte à ce que l'arrêt en incapacité dure au plus 36 mois. Dans le cas où l'ancienneté à l'entrée serait inférieure à la franchise contractuelle, on assigne à l'ancienneté d'entrée en incapacité la valeur de la franchise contractuelle.

Enfin, on arrondit la valeur de l'ancienneté d'entrée et de l'ancienneté de sortie de l'incapacité à l'unité près afin qu'elles correspondent à une valeur entière comme dans les tables du BCAC. Ainsi, la durée indemnisée de l'arrêt en incapacité correspond à l'ancienneté de sortie en incapacité (en mois) à laquelle on soustrait l'ancienneté d'entrée en incapacité (en mois); et la durée totale de l'arrêt en incapacité correspond simplement à l'ancienneté de sortie en incapacité.

Par ailleurs, les âges à l'entrée du risque incapacité allant de 20 ans à 61 ans dans les tables du BCAC, nous ne considérons que les assurés ayant un âge d'entrée dans le risque incapacité compris entre 20 ans et 61 ans. Enfin, nous décidons de créer des tranches d'âge afin d'obtenir une variable de type

facteur représentant l'âge de l'assuré. Pour ce faire, nous calculons le nombre moyen de passages de l'incapacité à l'invalidité par âge que nous représentons graphiquement. Nous constatons qu'entre 20 et 39 ans, la proportion de passage en invalidité est très faible et croît très lentement avec l'âge, puis entre 40 et 49 ans, la vitesse de croissance augmente et atteint son paroxysme vers 50 ans avec une croissance quasi linéaire. La proportion des patients en invalidité atteint son pic à 57 ans puis diminue brutalement par la suite.



L'échelle est masquée pour des raisons de confidentialité des données

FIGURE 4.1 – Proportion de passages de l'incapacité à l'invalidité par âge

Après avoir tracé la proportion de passage en invalidité par âge d'entrée, nous avons décidé de remanier les tranches d'âge pour mieux répartir le nombre d'arrêts par tranche : en regroupant notamment les plus jeunes qui ont un taux plus faible d'arrêts et en divisant en deux la tranche des quinquagénaires qui présentent un risque plus élevé avec un pic à 57 ans.

On observe une croissance exponentielle de la proportion de passage en invalidité en fonction de l'âge, avec notamment :

- Un pic à 30 ans correspondant à l'âge de la maternité ;
- Une croissance particulièrement forte à partir de 45 ans ;
- Puis une chute du taux à partir de 57 ans pouvant s'expliquer par des départs à retraite anticipés.

Cette analyse graphique nous a conduits à choisir les tranches d'âge suivantes :

- Entre 20 et 39 ans ;
- Entre 40 et 49 ans ;
- Entre 50 et 54 ans ;
- Entre 55 et 59 ans ;
- Plus de 60 ans.

Afin de comparer nos données aux tables du BCAC, nous construisons des tables avec un format similaire. Pour ce faire, nous calculons le nombre de sorties en incapacité par ancienneté dans ce risque, le nombre de passages en invalidité en fonction du nombre de mois passés en incapacité

ainsi que le nombre d'entrées en incapacité par mois d'ancienneté de l'arrêt de travail en incapacité - car, du fait des franchises, l'entrée dans le risque pour les assureurs en incapacité ne correspond pas toujours au début de l'arrêt de travail.

De là, nous pouvons calculer le nombre d'individus maintenus en incapacité en fonction de l'ancienneté de leur arrêt de travail en mois r_k ; ce qui correspond, pour un mois donné, au nombre de personnes maintenues en incapacité au début du mois précédent r_{k-1} auquel on soustrait le nombre de sorties de l'incapacité s_{k-1} au cours du mois précédent et on ajoute le nombre de personnes entrées dans le risque incapacité pour l'assureur au cours du mois e_k .

$$r_k := \begin{cases} 0 & \text{si } k = 0 \\ r_{k-1} - s_{k-1} + e_k & \text{si } k \in \llbracket 1, 36 \rrbracket \end{cases}$$

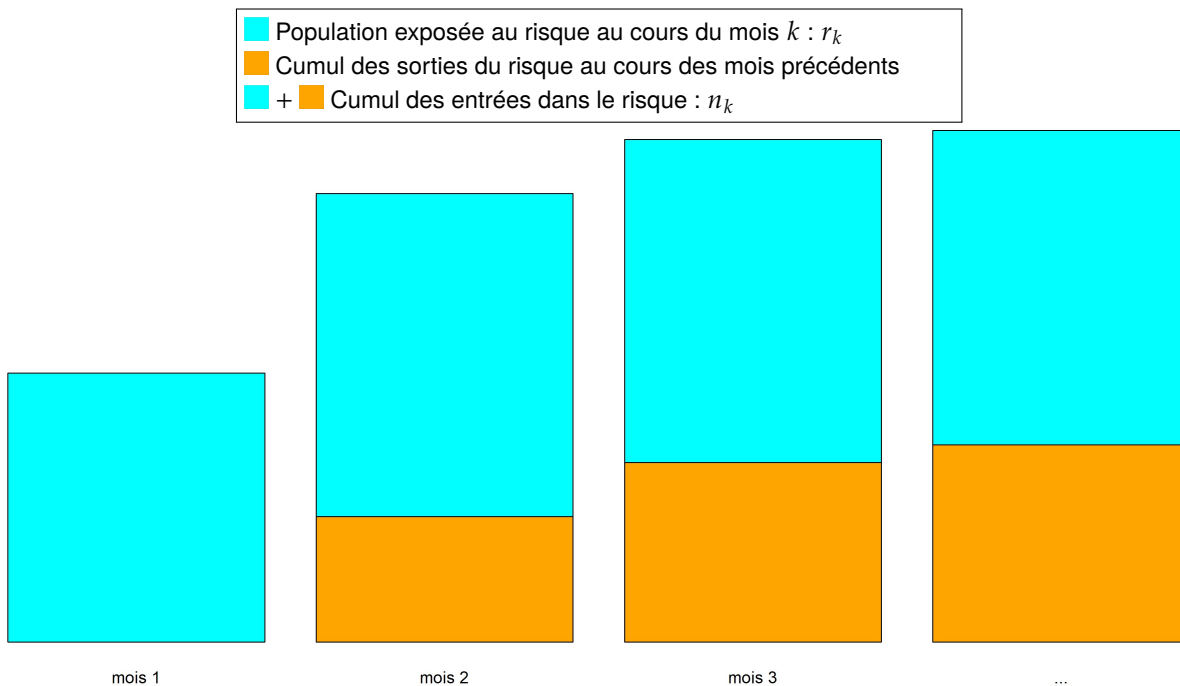


FIGURE 4.2 – Schéma explicatif du calcul de l'exposition

Chaque ligne des tables du BCAC représente un âge d'entrée dans le risque incapacité, tandis que chaque colonne représente une ancienneté dans le risque. Afin de supprimer le biais de l'âge dans la comparaison des tables du BCAC et de notre portefeuille, nous décidons de pondérer les tables du BCAC par la répartition des âges de notre portefeuille. En effet, les tables du BCAC considèrent un effectif de 10 000 individus à l'entrée du risque incapacité pour chaque âge - homogénéité s'éloignant fortement de la répartition des âges dans notre portefeuille. Pour ce faire, nous calculons la proportion de chaque âge à l'entrée dans le risque incapacité que nous divisons par la proportion obtenue pour l'âge le moins représenté dans notre portefeuille. Ainsi, nous obtenons un facteur multiplicateur supérieur ou égal à 1 pour chaque âge. Enfin, nous multiplions chaque ligne au coefficient associé à l'âge qu'elle représente.

Nous sommes donc à présent à même d'étudier la survie en incapacité et le passage en invalidité de notre portefeuille et du BCAC.

4.2 Variables étudiées

Pour chaque variable, les modalités non suffisamment représentées, i.e. les modalités pour lesquelles on observe moins de 10 passages en invalidité, sont ignorées. Ainsi subsistent les modalités suivantes :

- Catégorie de l'activité de l'entreprise : Domicile, Établissement Médecine, Chirurgie, Obstétrique (MCO), Handicap, Établissements d'Hébergement pour Personnes Âgées Dépendantes (EHPAD), Soins de Suite et de Réadaptation (SSR);
- Direction client de l'entreprise : Marché des Établissements Sociaux et Médico-Sociaux (ESMS), Marché des Établissements de Santé Privés et Professionnels de la Santé (ESPPS);
- Statut juridique de l'entreprise : Privé à but lucratif, Privé à but non lucratif (y compris associatif);
- Convention de l'entreprise : Fédération des Établissements Hospitaliers et d'Aide à la Personne (FEHAP), Fédération Hospitalière Privée (FHP);
- Catégorie socioprofessionnelle : Cadre (C), Non-cadre (NC);
- Sexe : Femme (F), Homme (H);
- Tranche d'âge : 20-39 ans, 40-49 ans, 50-54 ans, 55-59 ans;
- Zone géographique : Auvergne-Rhône-Alpes (AURA), Grand-Ouest, Île-de-France (IDF), Nord-Est, Occitanie, Provence-Alpes-Côte d'Azur (PACA);
- Franchise : 0 jour, 3 jours et 90 jours;
- Tranche de taille d'entreprise : 40-359 salariés, 360-599 salariés, 600-739 salariés, 800-9 999 salariés, 10 000 salariés et plus.

Toutes ces variables ont été étudiées, toutefois, seuls les résultats pertinents seront présentés dans ce mémoire.

4.3 Notations

Soit $T^{\text{inc}} \in \llbracket 0, 36 \rrbracket$ la variable aléatoire représentant le mois de fin d'un arrêt de travail en incapacité d'un individu tronquée au mois près.

Par exemple, si l'arrêt se termine au cours du 5^e mois d'arrêt, la valeur de T^{inc} sera 5.

Enfin, dans notre modèle, $T^{\text{inc}} := 0$ correspond à un arrêt de 0 jour.

T^{inc} est une variable aléatoire discrète car son support $\llbracket 1, 36 \rrbracket$ est fini, de plus, il est **bien ordonné**, i.e. toute partie non vide de son support possède un plus petit élément.

Pour tout $k \in \mathbb{N}$, on note :

- $p_k := \mathbb{P}(T^{\text{inc}} = k)$: la probabilité que cet arrêt dure k mois;
- $F_{\text{inc}}(k) := \mathbb{P}(T^{\text{inc}} \leq k) = \sum_{m=0}^k p_m$: la fonction de répartition de T^{inc} en k , i.e. la probabilité que l'arrêt dure au plus k mois;

Remarque : T étant discrète et de support bien ordonné, on peut définir sa fonction de répartition sur \mathbb{R} par :

$$\forall k \in \llbracket 0, 35 \rrbracket, \forall t \in]k, k + 1], F_{\text{inc}}(t) = F_{\text{inc}}(k) \text{ et } \forall t > 36, F_{\text{inc}}(t) = 1$$

— $S_{\text{inc}}(k) := \mathbb{P}(T^{\text{inc}} > k) = 1 - F_{\text{inc}}(k) = \sum_{m \geq k+1} p_m$: la fonction de survie de T^{inc} en k , i.e. la probabilité que l'arrêt dure au moins k mois.

Pour estimer les $(p_k)_{k \in \llbracket 0, 36 \rrbracket}$ à partir de nos données, il suffit, pour chaque mois, de diviser le nombre de sorties en incapacité s_k par le nombre total d'arrêts de travail ayant débuté avec une ancienneté d'au plus k :

$$n_k := \begin{cases} 0 & \text{si } k = 0 \\ n_{k-1} + e_k & \text{si } k \in \llbracket 1, 36 \rrbracket \end{cases}$$

Ainsi, on estime les $(p_k)_{k \in \llbracket 0, 36 \rrbracket}$ par

$$\begin{cases} p_0 = 0 \\ \forall k \in \llbracket 1, 36 \rrbracket, \hat{p}_k := \frac{s_k}{n_k} \end{cases}$$

4.3.1 Application : Fonction de survie en incapacité

Afin de comparer l'historique de Relyens aux tables réglementaires pour le risque incapacité, nous traçons la probabilité de survie en incapacité :

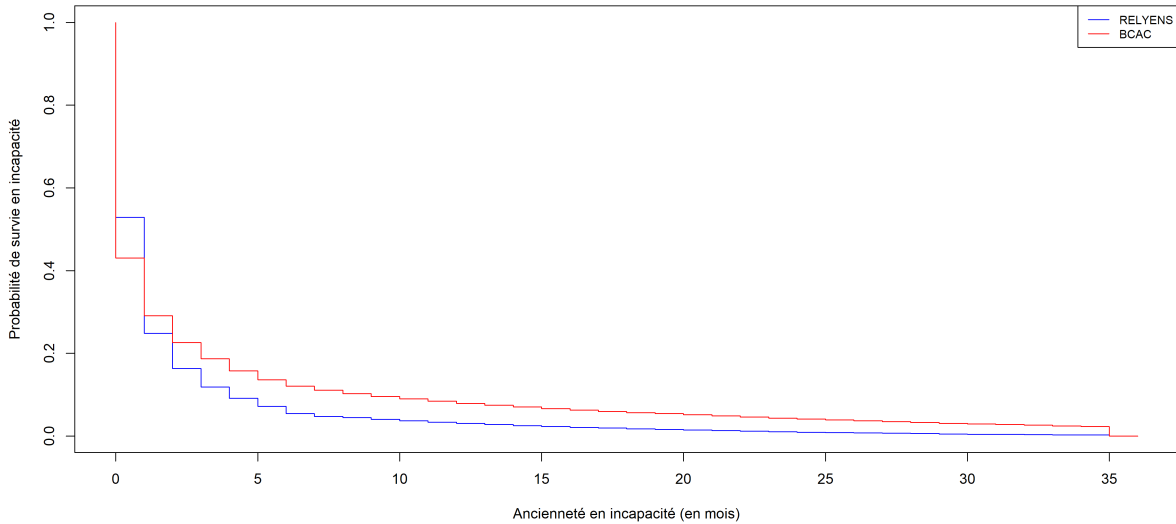


FIGURE 4.3 – Comparaison de la fonction de survie en incapacité du portefeuille de l'entreprise et du BCAC

On observe que la probabilité qu'un arrêt en incapacité dure plus d'un mois est plus grande pour le portefeuille de l'entreprise. Toutefois, la probabilité de survie en incapacité du portefeuille du BCAC est supérieure à celle de l'entreprise à partir du deuxième mois d'arrêt.

Pour mieux visualiser l'écart entre les deux courbes précédentes, nous traçons le rapport entre les deux fonctions :

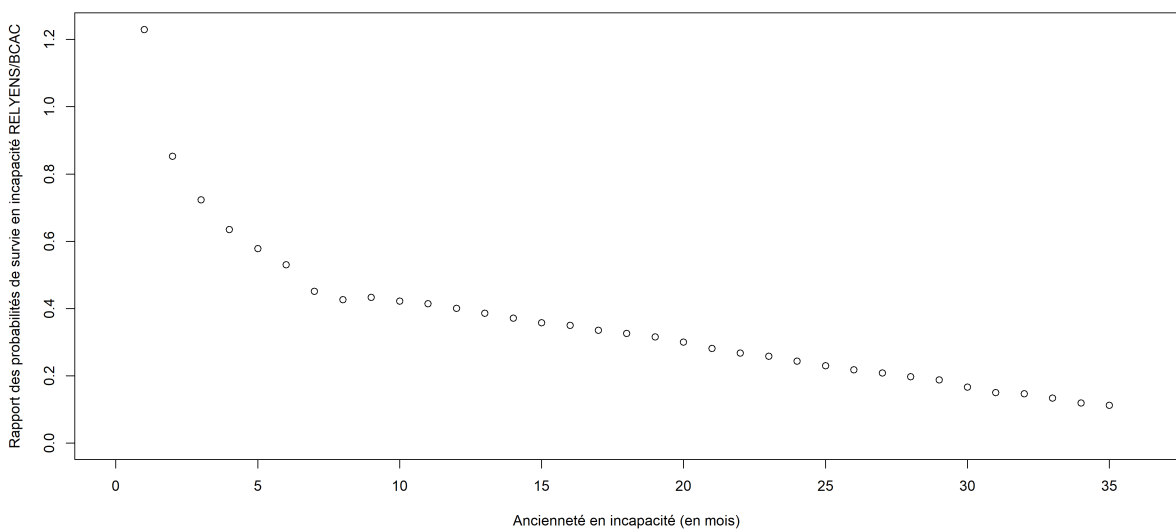


FIGURE 4.4 – Ratio entre la fonction de survie en incapacité du portefeuille de l'entreprise et celle du BCAC

L'écart entre les deux courbes décroît au cours du temps mais le rapport entre les deux fonctions de survie s'éloigne de plus en plus de 1 avec une décroissance exponentielle au cours des 6 premiers mois, puis linéaire à partir du septième mois.

4.4 Fonction de hasard

Nous souhaitons maintenant étudier la variable T^{inv} représentant, pour un individu donné en arrêt de travail en incapacité, la durée écoulée entre le début de son arrêt de travail en incapacité et son passage en invalidité, ou encore son ancienneté (en mois) en incapacité au moment de son passage en invalidité. Ainsi,

$$T^{\text{inv}} \begin{cases} \in \mathbb{N} & \text{en cas de passage en invalidité;} \\ = +\infty & \text{si l'individu ne passe jamais en invalidité.} \end{cases}$$

Dans ce cas, la fonction de répartition représente la probabilité de passer en invalidité au bout d'une ancienneté en incapacité d'au plus k mois et la fonction de survie représente la probabilité de non-passage en invalidité avant k mois. Ces deux quantités sont plus difficilement interprétables lors de l'évaluation d'un risque d'occurrence d'un évènement. On leur préférera donc la fonction de hasard (ou le taux de risque instantané) définie par :

$$\forall k \in \mathbb{N}, h_{\text{inv}}(k) := \mathbb{P}(T^{\text{inv}} = k | T^{\text{inv}} \geq k)$$

La fonction de hasard représente la probabilité d'apparition d'un évènement dans un intervalle de temps $[t, t + \Delta t]$ sachant que l'évènement ne s'est pas encore réalisé.[8] Elle peut donc permettre d'évaluer le risque de passage en invalidité d'un individu avec une ancienneté en incapacité donnée au cours du mois entamé. Ainsi, pour une ancienneté k donnée, plus $h_{\text{inv}}(k)$ est grand, plus la probabilité de passage en invalidité dans le mois est élevée.

On établit un lien entre la fonction de hasard et la fonction de survie par la formule suivante :

$$\forall k \in \mathbb{N}^*, h_{\text{inv}}(k) = 1 - \frac{S_{\text{inv}}(k)}{S_{\text{inv}}(k-1)} \quad (4.1)$$

et

$$h_{\text{inv}}(0) = 0.$$

ou encore :

$$S_{\text{inv}}(k) = \prod_{m=0}^k (1 - h_{\text{inv}}(m)) \quad (4.2)$$

La fonction de hasard définit donc totalement la loi de passage en invalidité.

Enfin, on note par ailleurs que

$$\forall k \in \mathbb{N}^*, h_{\text{inv}}(k) = \frac{\Delta F_{\text{inv}}(k-1)}{1 - F_{\text{inv}}(k-1)} \quad (4.3)$$

avec pour toute suite $(u_k)_{k \in \mathbb{N}}$ et pour tout $k \in \mathbb{N}$, $\Delta u_k := u_{k+1} - u_k$.

4.5 Censures et troncatures de l'information

Il arrive qu'un arrêt de travail ne soit plus observable avant un potentiel passage en invalidité, i.e. $T^{\text{inc}} < T^{\text{inv}}$.

On parle de censure à droite lorsque, pour un arrêt de travail i donné, au lieu d'observer directement la durée entre son passage en incapacité et son passage en invalidité, T_i^{inv} , on observe le couple de variables aléatoires (X_i, D_i) :

$$\begin{cases} X_i := T_i^{\text{inv}} \wedge C_i \\ D_i := \mathbb{1}_{\{T_i^{\text{inv}} \leq C_i\}} = \mathbb{1}_{\{X_i = T_i^{\text{inv}}\}} \end{cases}$$

où C_i est la durée après laquelle il nous est impossible d'observer l'occurrence du risque étudié pour l'individu i , et représente donc la censure à droite.

- X_i représente alors la durée d'observation du risque de passage en invalidité;
- D_i , quant à elle, nous indique si le passage en invalidité a été observé (dans ce cas, $D_i = 1$) ou non (dans ce cas, $D_i = 0$), et donc, si l'information est complète ou non.

Dans notre étude, cette censure a deux sources principales.

4.5.1 Censure à droite de type I

Premièrement, nous avons une censure législative : un arrêt de travail en incapacité durant au plus 3 ans, soit 36 mois, nous ne pouvons pas observer le passage en invalidité si $T^{\text{inv}} > 36$. Cette censure, $C^1 := 36$, étant fixe et déterministe, on parle de censure à droite de type I.

4.5.2 Censure à droite de type III

Souvent, l'arrêt de travail en incapacité s'interrompt pour une autre cause que le passage en invalidité : le rétablissement (et donc la reprise du travail) ou encore le décès. Ces causes d'interruption de l'observation du risque de passage en invalidité constituent une censure à droite stochastique que l'on représente par une variable aléatoire C_i^3 pour un arrêt de travail i donné. On parle ici de censure à droite de type III.

En tout, pour tout arrêt i , on a en fait une censure $C_i = C^1 \wedge C_i^3$.

4.5.3 Troncature à gauche

Par ailleurs, tous les arrêts de travail ne sont pas observables à partir de la même date, du fait des différentes franchises appliquées qui dépendent du contrat et de la cause de l'arrêt de travail (accident du travail, hospitalisation ou maladie). En effet, les arrêts de travail ne sont déclarés par les entreprises - et donc observable par l'assureur - que s'ils sont éligibles à une indemnisation par l'assureur et donc si leur durée excède celle de la franchise.

Il faut donc attendre que la durée d'un arrêt excède la franchise pour commencer à l'observer et, dans le cas où sa durée est inférieure à la franchise, il n'est pas du tout observable. Pour un arrêt de travail i donné, soumis à une franchise τ_i telle que $T_i \geq \tau_i$, au lieu d'observer X_i , on observe en fait

la durée résiduelle de l'arrêt de travail conditionnellement au fait que sa durée soit supérieure à la franchise écoulée :

$$Y_i := X_i - \tau_i | X_i \geq \tau_i$$

τ_i étant déterminée par le contrat (et donc déterministe).

Estimation des fonctions de hasard

5.1 L'estimateur de Kaplan-Meier

Il est usuel d'estimer la fonction de répartition par la fonction de répartition empirique définie par :

$$\forall t \in \mathbb{R}, F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{T_i^{\text{inv}} \leq t\}}$$

avec n le nombre total d'observations.

Toutefois, en présence de censure, la fonction de répartition empirique est sous-estimée. Afin de prendre en compte la censure, on utilise usuellement l'estimateur de Kaplan-Meier (KM).

5.1.1 Construction de l'estimateur de Kaplan-Meier

On suppose que :

- les variables (T_i^{inv}) sont deux à deux indépendantes et identiquement distribuées de même loi que la variable T^{inv} ;
- les variables (C_i) sont deux à deux indépendantes et identiquement distribuées de même loi que la variable C ;
- les variables (T_i^{inv}) sont indépendantes des variables de censure (C_i) .

Il est à noter que rien ne nous permet de vérifier l'indépendance entre la variable d'intérêt et la variable de censure. Il faudrait donc se demander si cette hypothèse est raisonnable.

Afin d'estimer empiriquement la fonction de répartition de T^{inv} , nous exprimons F_{inv} en fonction de quantités facilement estimables empiriquement, i.e. dépendant de la variable observée X , à savoir, pour tout $k \in \llbracket 0, 36 \rrbracket$:

- $G(k) := \mathbb{P}(X \leq k)$ estimée par $\widehat{G}(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq k\}} = 1 - \frac{r_k}{n_k}$
- $G_1(k) := \mathbb{P}(\{X \leq k\} \cap \{\delta = 1\})$ estimée par $\widehat{G}_1(k) = \frac{1}{n} \sum_{i=1}^n \delta_i \mathbb{1}_{\{Y_i \leq k\}} = \sum_{j=1}^k \frac{d_j}{n_j}$
- $G_0(k) := \mathbb{P}(\{X \leq k\} \cap \{\delta = 0\}) = G(k) - G_1(k)$

On peut estimer la fonction de répartition de T^{inv} par l'estimateur de Kaplan-Meier :

$$\forall k \in \mathbb{N}, \widehat{F}^{\text{KM}}(k) = 1 - \prod_{m=1}^k \left(1 - \frac{d_m}{r_m}\right) \quad (5.1)$$

ou de manière équivalente :

$$\forall k \in \mathbb{N}^*, \widehat{F}^{\text{KM}}(k) = \sum_{m=1}^k \frac{d_m}{r_m} \widehat{S}^{\text{KM}}(m-1) \quad (5.2)$$

5.1.2 Application : Estimation de la fonction de hasard du passage en invalidité par Kaplan-Meier

À partir de l'expression de la fonction de hasard relativement à la fonction de répartition (4.3), la fonction de hasard du passage en invalidité est estimée par Kaplan-Meier (5.1).

Ainsi, nous pouvons comparer l'évolution du risque instantané de passage en invalidité du BCAC et de Relyens en fonction de l'ancienneté en incapacité :

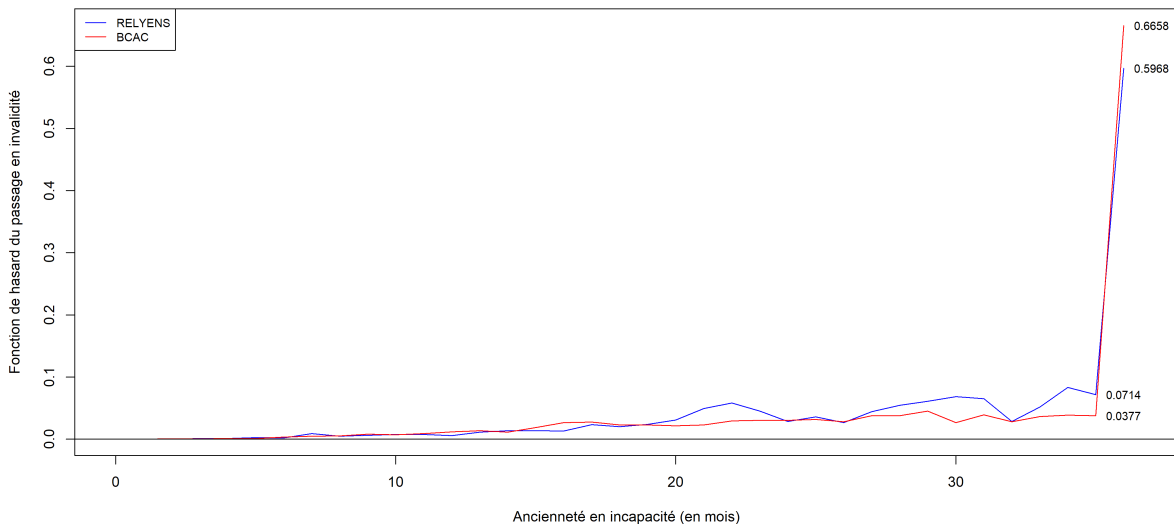


FIGURE 5.1 – Les fonctions de hasard de passage en invalidité du portefeuille de l'entreprise et du BCAC estimées par Kaplan-Meier passent de moins de 10% à environ 60% au cours du dernier mois

La méthode de Kaplan-Meier estime la probabilité de passage en invalidité au 35^e mois à moins de 4%, puis au 36^e mois à 60%. On constate également des sauts de plus importants à partir d'un an d'observation. On peut supposer une surestimation de la probabilité de passage en invalidité à la fin de la période d'observation. Cette surestimation peut avoir deux causes possibles :

Le taux de censure trop important à partir d'une certaine ancienneté

Pour une ancienneté $k \in \llbracket 1, 36 \rrbracket$, si le taux de censure, $\frac{\sum_{i=1}^k s_i - d_i}{n_k}$, est trop proche de 1, l'estimateur évalue les observations censurées comme si elles se comportaient exactement comme les observations complètes, pour lesquelles la date de passage en invalidité est inférieure à 36 mois.

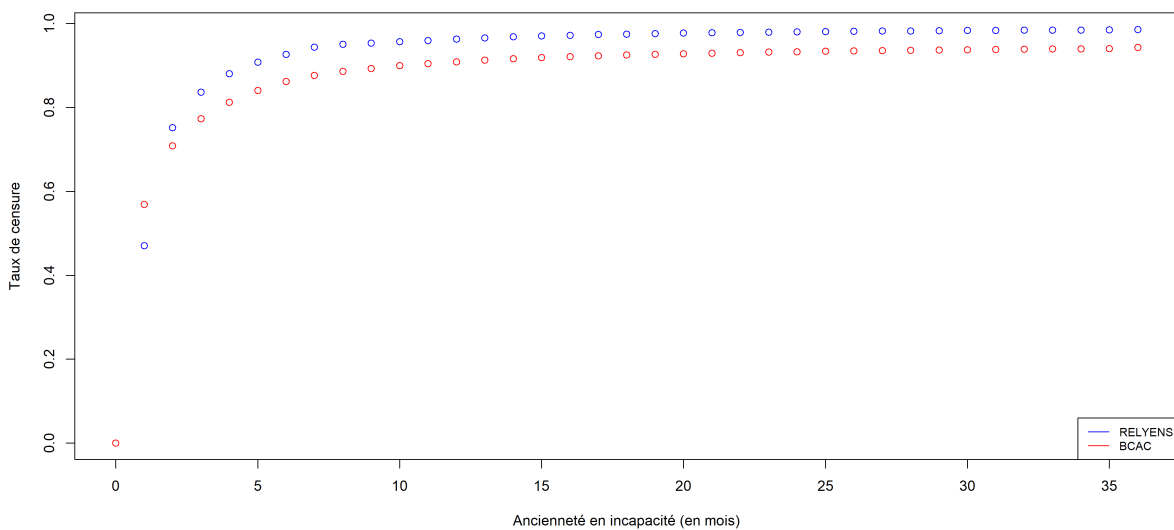


FIGURE 5.2 – Les taux de censure des portefeuilles de l’entreprise et du BCAC sont très élevés dès le 8^e mois

C’est le cas dès le huitième mois d’ancienneté en incapacité. À partir de cet instant, on se rapproche donc plus de la probabilité de passage en invalidité conditionnelle au maintien en incapacité au bout de $k \in \llbracket 1, 36 \rrbracket$ mois.

La dépendance entre la variable d’intérêt et la variable de censure

Il convient de discuter du lien entre la variable d’intérêt (durée entre le début de l’arrêt en incapacité et l’instant de passage en invalidité) et la variable de censure composée de trois sources de censure (le rétablissement, le décès, la durée maximale légale de l’incapacité).

Comme le passage en invalidité, le rétablissement (et donc la reprise du travail) et le décès (hors décès accidentel et homicide) sont liés à l’intégrité physique de l’individu. En effet,

- si l’état de santé d’un assuré s’améliore, sa probabilité de passer en invalidité diminue et sa probabilité de se rétablir (et de reprendre le travail) augmente ;
- si l’état de santé d’un assuré se dégrade, sa probabilité de passer en invalidité augmente, son espérance de vie résiduelle peut diminuer et sa probabilité de reprendre le travail diminue ;
- si un assuré décède, sa probabilité de passer en invalidité s’annule.

Il y a donc une dépendance assez forte entre le risque de passage en invalidité et les sources aléatoires de censure. On peut donc rejeter l’hypothèse d’indépendance sous-jacente à la méthode de Kaplan-Meier.

5.2 L'estimateur de Kalbfleisch et Prentice

5.2.1 Les risques compétitifs

Les évènements concurrentiels correspondent aux évènements qui, lors de leur survenance, empêchent ou modifient la probabilité d'observer l'évènement étudié; en leur présence, les résultats sont alors biaisés, voire non interprétables.

Dans le modèle de Kaplan-Meier, les évènements concurrentiels ne sont pas pris en compte : ils sont considérés comme des censures. Dans le cas où les risques compétitifs et le risque étudié sont indépendants, cela ne pose pas de problème. Toutefois, dans le cas contraire, la fonction de répartition de la variable d'intérêt est surestimée (cf. *Gouley et al., 1999*) étant donné qu'il considère que les individus pour lesquels un évènement concurrentiel est survenu en amont de l'évènement étudié sont soumis au même risque que les individus encore observables. Il convient d'utiliser une méthode prenant en compte l'information apportée par la survenance d'évènements concurrentiels avant l'observation de l'évènement d'intérêt, telle que celle de Kalbfleisch et Prentice (KP). [12, 13]

5.2.2 Construction de l'estimateur de Kalbfleisch et Prentice

L'estimateur de Kalbfleisch et Prentice se base sur l'écriture sous forme de somme de l'estimateur de Kaplan-Meier (5.2), appliquée à la variable aléatoire correspondant à la durée d'observation d'un individu; soit, ici la sortie en incapacité que l'on note T^{inc} :

$$\forall k \in \mathbb{N}^*, \widehat{F}_{\text{inc}}^{\text{KM}}(k) = \sum_{m=1}^k \frac{s_m}{r_m} \widehat{S}_{\text{inc}}^{\text{KM}}(m-1)$$

Puis, il décompose le nombre de sorties (en incapacité) s_m en la somme des sorties dues à la survenance de l'évènement d'intérêt (l'entrée en invalidité) d_m et des sorties dues à la survenance d'un évènement concurrentiel c_m :

$$\begin{aligned} \forall k \in \mathbb{N}^*, \widehat{F}_{\text{inc}}^{\text{KM}}(k) &= \sum_{m=1}^k \frac{d_m + c_m}{r_m} \widehat{S}_{\text{inc}}^{\text{KM}}(m-1) \\ &= \underbrace{\sum_{m=1}^k \frac{d_m}{r_m} \widehat{S}_{\text{inc}}^{\text{KM}}(m-1)}_{\widehat{F}_{\text{inv}}^{\text{KP}}(k)} + \underbrace{\sum_{m=1}^k \frac{c_m}{r_m} \widehat{S}_{\text{inc}}^{\text{KM}}(m-1)}_{\widehat{F}_{\text{c}}^{\text{KP}}(k)} \end{aligned}$$

Ainsi, on obtient l'estimateur de Kalbfleisch et Prentice, aussi appelé Fonction d'incidence cumulée (CIF) - de l'anglais *cumulative incidence function* :

$$\boxed{\forall k \in \mathbb{N}^*, \widehat{F}_{\text{inv}}^{\text{KP}}(k) = \sum_{m=1}^k \frac{d_m}{r_m} \widehat{S}_{\text{inc}}^{\text{KM}}(m-1)} \quad (5.3)$$

5.2.3 Application : Estimation de la fonction de hasard du passage en invalidité par Kalbfleisch et Prentice

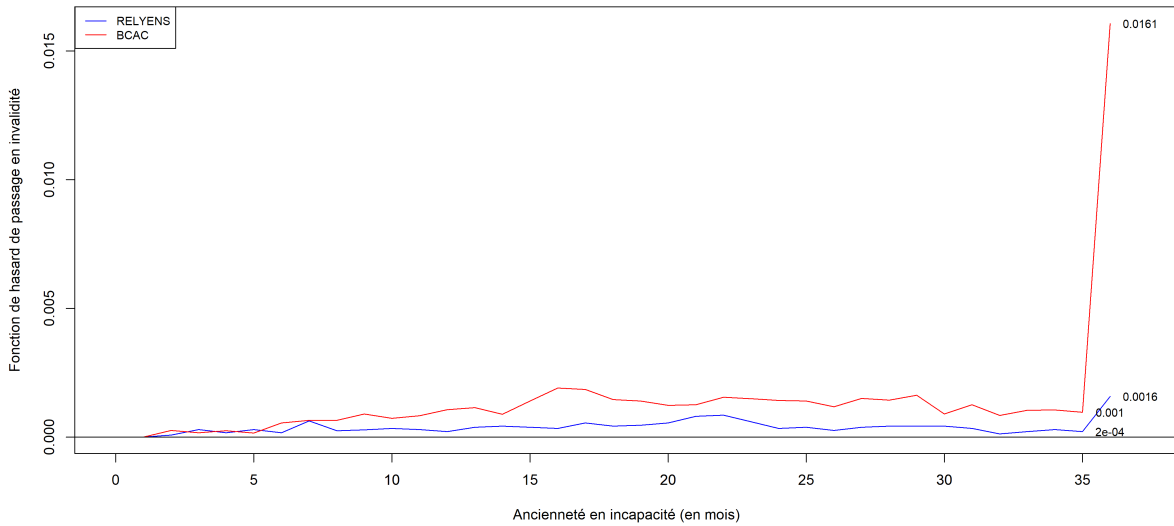


FIGURE 5.3 – La fonction de hasard du passage de l'incapacité à l'invalidité du BCAC estimées par Kalbfleisch et Prentice s'éloigne de celle du portefeuille de l'entreprise dès le 8^e mois

On observe que les risques instantanés de passage en invalidité avant le septième mois de la table du BCAC et de la table de Relyens sont très proches.

À partir du septième mois, le risque instantané de passage en invalidité du BCAC commence à augmenter tandis que celui de la mutuelle est quasi constant. Par ailleurs, on constate que, dans les deux cas, le risque instantané est multiplié par 10 entre le 35^e mois et le 36^e mois.

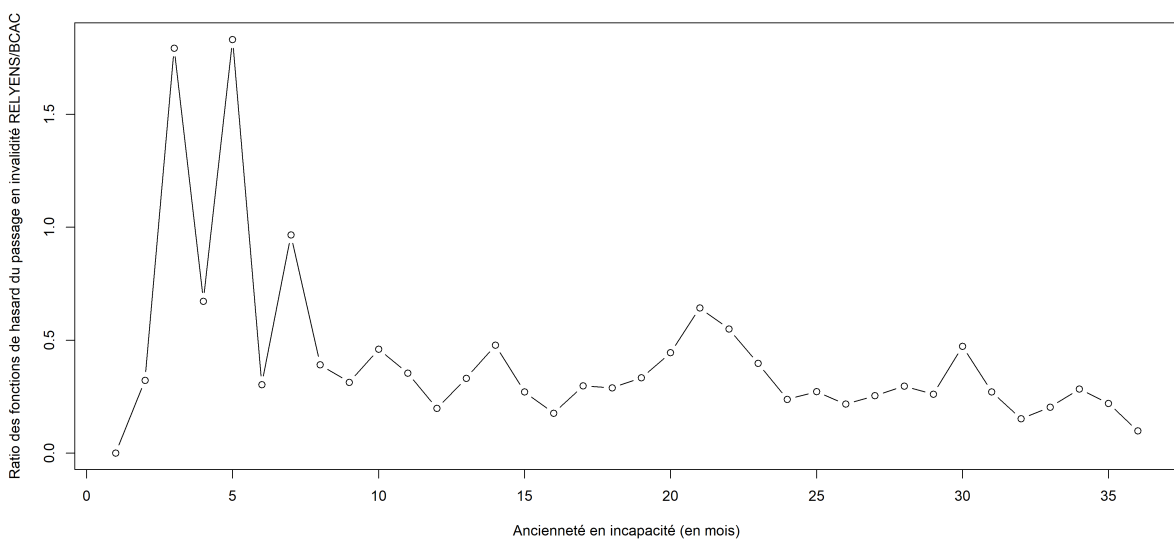


FIGURE 5.4 – Le rapport entre la fonction de hasard par Kalbfleisch et Prentice du passage en invalidité du portefeuille de l'entreprise et celle du BCAC converge vers 0 à partir du 8^e mois

En faisant le rapport du risque instantané de passage en invalidité issu de la table du BCAC par celui du portefeuille de Relyens, on constate une forte volatilité de ce ratio avant le 8^e mois. Le ratio se stabilise par la suite et semble osciller autour de 30% mais converge lentement vers 0.

5.3 Hétérogénéité de la fonction de hasard du risque d'invalidité en attente

Afin d'avoir une première idée de l'influence des modalités sur la probabilité de passage en invalidité au cours d'un arrêt de travail, on peut tracer les fonctions de hasard issues des courbes d'incidence cumulée segmentées par modalité pour chaque variable.

5.3.1 Statut juridique

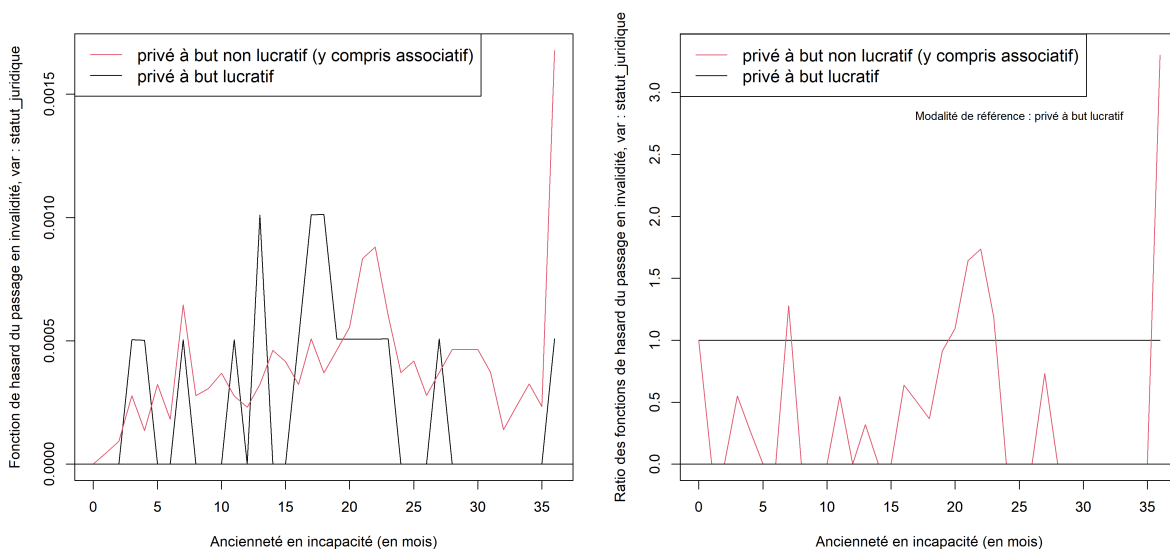


FIGURE 5.5 – Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction du statut juridique

Le risque instantané de passage en invalidité au cours d'un arrêt en incapacité estimé est plus volatil pour les entreprises à but lucratif du fait de leur effectif cumulé plus faible. On constate tout de même un risque globalement plus faible pour les entreprises et associations à but non lucratif, sauf pour le dernier mois.

5.3.2 Convention

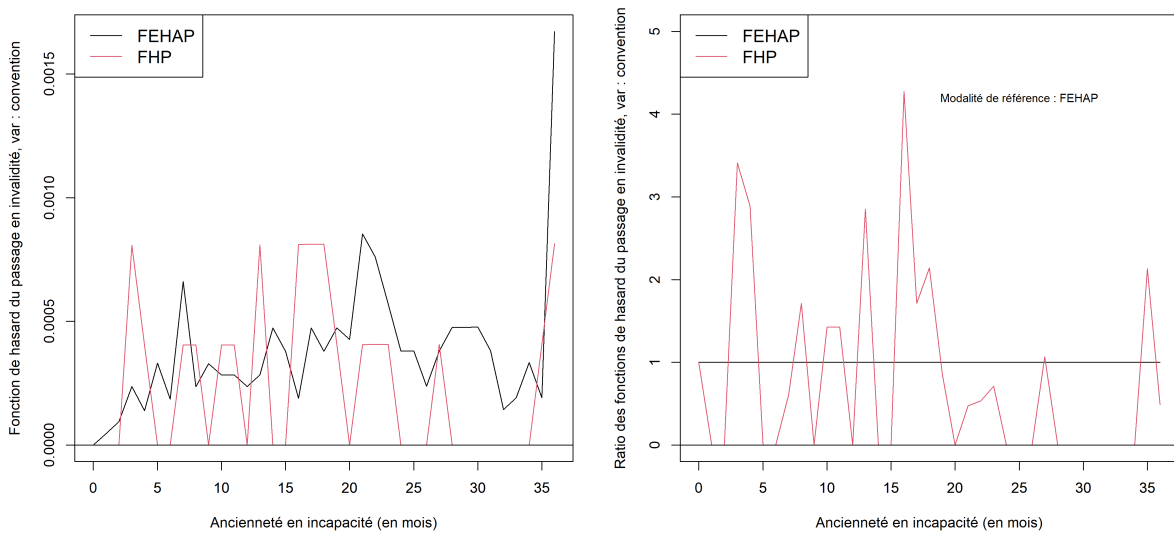


FIGURE 5.6 – Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction de la convention

L'évolution du risque instantané des entreprises de la convention FEHAP est très proche de celle des entreprises à but non lucratif car la majorité du portefeuille dépend de cette convention. À partir du 19^e mois, le risque instantané de passage en invalidité de la convention FHP est plus faible.

5.3.3 Catégorie socioprofessionnelle

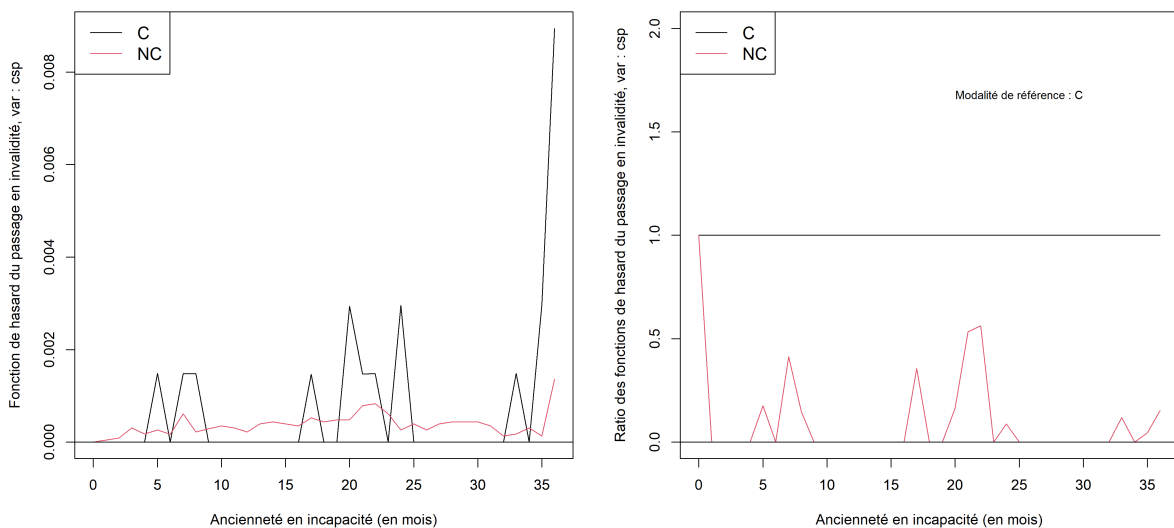


FIGURE 5.7 – Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction de la catégorie socioprofessionnelle

Après une analyse du taux d'absentéisme, nous avons conclu que le risque invalidité est plus faible pour les cadres, ainsi que leur risque d'arrêt en incapacité ; avec un taux d'invalidité deux fois plus faible que celui des non-cadres sur l'ensemble des assurés. Cependant, les cadres arrêtés ont un risque instantané de passage en invalidité plus élevé que les non-cadres arrêtés.

5.3.4 Âge

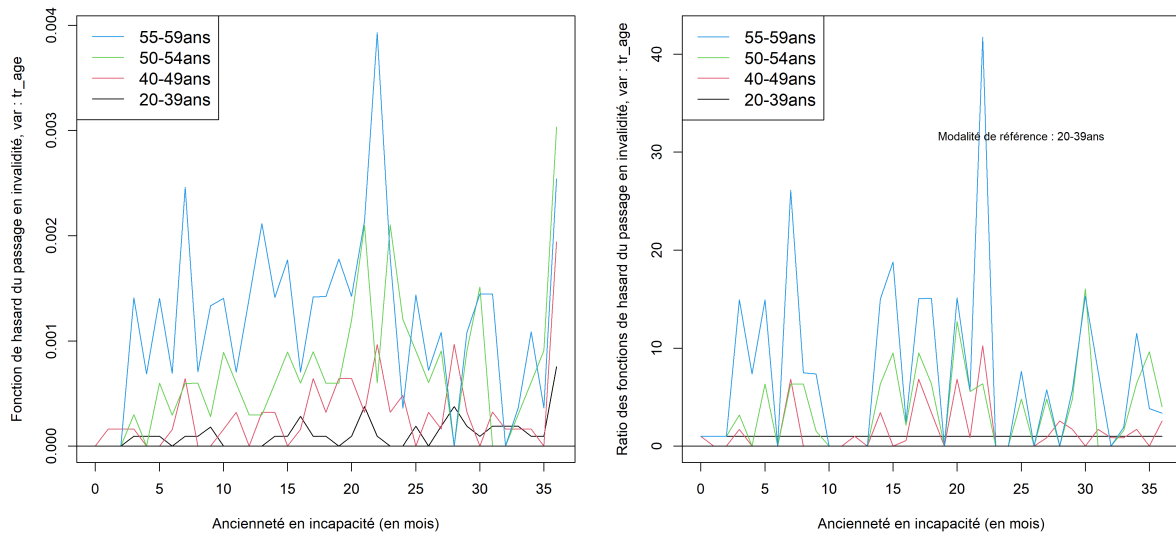


FIGURE 5.8 – Comparaison des fonctions de hasard du passage en invalidité du portefeuille de l'entreprise en fonction de l'âge

On observe clairement une corrélation positive entre le risque instantané de passage en invalidité et l'âge.

Par ailleurs, nous pouvons également comparer les fonctions de hasard du portefeuille et du BCAC par tranche d'âge :

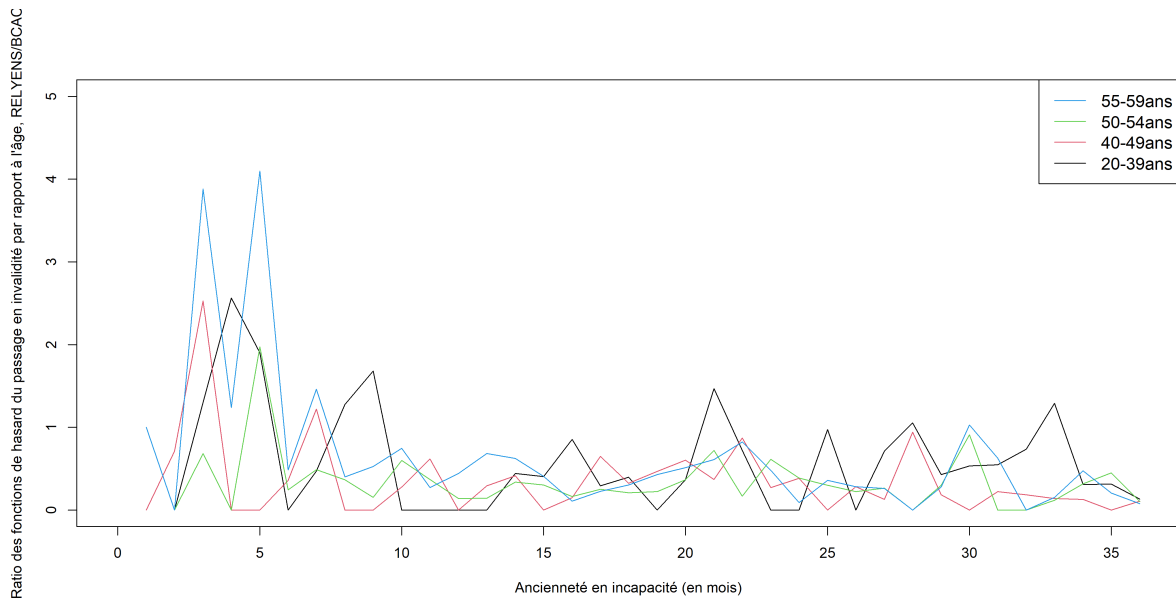


FIGURE 5.9 – Rapport entre la fonction de hasard par Kalbfleisch et Prentice du passage en invalidité du portefeuille de l'entreprise et celle du BCAC par âge

On ne dénote pas de comportement différencié fonction de l'âge entre le portefeuille du BCAC et de Relyens.

5.4 Comparaison des courbes de survie

5.4.1 Le test du log-rank

Le test du log-rank est un test non paramétrique qui consiste à comparer plusieurs courbes de survie estimées en contrôlant l'hypothèse nulle (\mathcal{H}_0) selon laquelle les distributions de survie sont statistiquement identiques, i.e. les proportions de passage en invalidité des groupes sont égales au cours du temps. Tout comme l'estimation de Kaplan-Meier, le test du log-rank n'est valide que sous l'hypothèse de censure non informative et en l'absence d'effet de période. Il compare le nombre d'évènements observés au nombre d'évènements attendus sous l'hypothèse nulle.[25]

On considère $k \in \mathbb{N}$ échantillons E_1, E_2, \dots, E_k . Sous (\mathcal{H}_0), pour tout échantillon E_j et tout instant t_1, \dots, t_N , d_{ij} suit la loi hypergéométrique $\mathcal{H}\left(r, d_i, \frac{r_{ij}}{r_i}\right)$. Pour chaque échantillon E_j , on considère la statistique ϕ_j , fonctions des $\delta_{ij} := d_{ij} - \mathbb{E}(d_{ij})$ [14] :

$$\phi_j = \frac{\left(\sum_{i=1}^N \delta_{ij}\right)^2}{\sum_{i=1}^N \mathbb{V}(\delta_{ij})} \sim \chi_{k-1}^2$$

5.4.2 Application du test du log-rank

En effectuant le test du log-rank sur chaque variable étudiée, nous obtenons les valeurs p suivantes :

Variable	Valeur p
Tranche d'âge	$1,16 \times 10^{-10}$
Zone géographique	0,0016
Taux d'absentéisme global	0,0024
Franchise appliquée	0,0225
Taille de l'entreprise	0,0307
Convention	0,0366
Direction client	0,3497
Catégorie de l'activité	0,4028
Statut juridique	0,4063
Sexe	0,6976
Catégorie socioprofessionnelle	0,8641

Ainsi, les courbes de survie diffèrent significativement en fonction de la tranche d'âge, la zone géographique du siège de l'entreprise, le taux d'absentéisme global, la franchise appliquée, la taille de l'entreprise ou sa convention.

6.1 Présentation du modèle de Cox

Le modèle de Cox est l'un des modèles les plus utilisés en analyse de survie. Il s'agit d'un modèle semi-paramétrique qui permet de calculer l'effet cumulé de covariables sur une durée de survie. Le modèle de Cox est un modèle à risques proportionnels, i.e. il requiert l'hypothèse de proportionnalité de la fréquence instantanée d'occurrence de l'évènement. Autrement dit, graphiquement, les courbes de loi de survie découlant du modèle de Cox ne se croisent pas.[16]

Dans le modèle de Cox, la fonction de hasard - ou risque instantané - s'écrit :

$$h(t|X) = h_0(t) \exp(\beta'X)$$

où, en posant $p \in \mathbb{N}$ le nombre de covariables du modèle :

- $X = (X_1, \dots, X_p)'$ est le vecteur des covariables du modèle;
- h_0 est la fonction de hasard de base qui ne dépend que du temps (elle ne dépend pas des modalités des covariables) qui est inconnue;
- $\beta' = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$ est le vecteur des coefficients associés aux covariables du modèle, qui est également inconnu.[27]

Dans ce modèle, la relation entre la fonction de hasard et les covariables est log-linéaire.[26]

L'objectif du modèle de Cox est d'estimer - par maximum de vraisemblance - le vecteur β qui est indépendant du temps et définit totalement l'effet des covariables sur le risque instantané. [8]

Lorsque le coefficient associé à une covariable est négatif, cela signifie que cette covariable réduit la probabilité de survenance du changement d'état et donc qu'elle a un effet positif sur la durée de survie. Inversement, s'il est positif, cela signifie que cette covariable augmente la probabilité de survenance du changement d'état et donc qu'elle a un effet négatif sur la durée de survie. Enfin, s'il est nul, cela signifie que cette covariable n'a pas d'influence sur l'évènement considéré.[16]

Par ailleurs, le rapport des risques instantanés de deux individus, i.e. le risque relatif de survenance de l'évènement étudié d'un individu par rapport à un autre, est indépendant du temps :

$$\frac{h(t|X)}{h(t|Y)} = \frac{\exp(\beta'X)}{\exp(\beta'Y)} = \exp(\beta'(X - Y))$$

6.2 Tests de validation du modèle

6.2.1 Tests de significativité globale des coefficients de régression

Les tests classiquement utilisés afin de tester l'hypothèse de nullité globale des coefficients - à savoir le **test de Wald**, le **test du rapport des maxima de vraisemblance** ou encore le **test de score** - sont des tests asymptotiques équivalents qui se fondent sur le maximum de vraisemblance, qui, pour le modèle de Cox, se définit par :

$$\mathcal{L}(\beta) = \prod_{i=1}^p \frac{\exp(-\beta' X_i)}{\sum_{j=1}^p \exp(-\beta' X_j) \mathbb{1}_{\{T_i \leq T_j\}}}$$

Ils permettent de tester l'hypothèse nulle : $(\mathcal{H}_0) : \beta = 0$.

Dans le test du rapport des maxima de vraisemblance, on note $\hat{\beta}$, l'estimateur du maximum de vraisemblance du modèle estimé et $\hat{\beta}_0$, l'estimateur du maximum de vraisemblance sous (\mathcal{H}_0) . Il s'agit de rejeter l'hypothèse (\mathcal{H}_0) si $\hat{\beta}$ et $\hat{\beta}_0$ sont proches et de la rejeter sinon. Pour ce faire, on considère la statistique :

$$\xi = 2 \left(\ln \mathcal{L}(\hat{\beta}) - \mathcal{L}(\hat{\beta}_0) \right)$$

Sous (\mathcal{H}_0) , ξ converge vers une loi χ_p^2 .

6.2.2 Test de l'hypothèse de hasard proportionnel

Ce test se base sur les **résidus de Schoenfeld** définis pour chaque individu i par :

$$r_i = d_i \times \left(X_i - \frac{\sum_{j \in R_i} \exp(-\beta' X_j) X_j}{\sum_{j \in R_i} \exp(-\beta' X_j)} \right)$$

avec $R_i = \sum_{j=1}^n \mathbb{1}_{\{T_i \leq T_j\}}$ [14]

Il consiste à tester l'hypothèse nulle de constance des coefficients de régression au cours du temps : $\mathcal{H}_0 : \forall t > 0, \beta(t) = \beta(t = 0) = \beta$. Le modèle est validé si la valeur p est supérieure au seuil (0,05).

6.3 Application du modèle de Cox

Les variables explicatives intégrées dans ce modèle correspondent à l'ensemble des variables significatives selon le test du log-rank effectué dans la partie précédente, à savoir : la tranche d'âge, la zone géographique, le taux d'absentéisme global, la franchise appliquée, la taille de l'entreprise et la convention.

6.3.1 Coefficients de régression

On obtient les coefficients de régression suivants :

Covariable	β
Tranche d'âge = 40-49 ans	0,1816
Tranche d'âge = 50-54 ans	0,5769
Tranche d'âge = 55-59 ans	0,8747
Tranche d'âge = Plus de 60 ans	0,0575
Zone géographique = Grand-Ouest	-0,3990
Zone géographique = Nord-Est	-0,1883
Zone géographique = Occitanie	0,1921
Zone géographique = PACA	0,2480
Zone géographique = AURA	0,6151
Taux d'Absentéisme Global (TAG)	0,0996
Franchise appliquée	-0,0017
Tranche de taille de l'entreprise = 1-39 salariés	1,6403
Tranche de taille de l'entreprise = 40-359 salariés	-0,2671
Tranche de taille de l'entreprise = 360-599 salariés	0,0299
Tranche de taille de l'entreprise = 600-739 salariés	0,0138
Tranche de taille de l'entreprise = 740-799 salariés	-0,1535
Tranche de taille de l'entreprise = 800-9999 salariés	-0,3540
Convention = FHP	-0,4628

Les modalités de référence sont la tranche d'âge des 20-39 ans, la zone géographique de l'Île-de-France (IDF), de la convention FEHAP et la tranche de taille d'entreprise des plus de 10000 salariés.

Selon ce modèle, toute chose égale par ailleurs, plus l'âge est élevé plus le risque d'invalidité en attente est élevé ; sauf pour les plus de 60 ans qui dérogent à cette règle du fait de la possibilité de partir à la retraite plutôt que de passer en invalidité. Le risque d'invalidité en attente est plus faible pour les entreprises du Grand-Ouest et du Nord-Est. Lorsque la durée de franchise appliquée est plus élevée, le risque semble légèrement plus faible. La taille de l'entreprise n'a pas un effet monotone sur le risque. Le risque des entreprises d'effectif intermédiaires est plus faible que celui des grandes et petites entreprises. Enfin, le risque est plus élevé pour les entreprises de la convention FEHAP.

6.3.2 Validation du modèle

Pour les tests de significativité globale des coefficients de régression, nous obtenons les valeurs p suivantes :

Test	Valeur p
Rapport des maxima de vraisemblance	9×10^{-9}
Wald	5×10^{-9}
Score	8×10^{-10}

D'après ces tests, le modèle est donc globalement significatif.

Pour le test de l'hypothèse de hasard proportionnel, nous obtenons les valeurs p suivantes :

Variable	Valeur p
Tranche d'âge	$1,1 \times 10^{-8}$
Zone géographique	0,28
Taux d'Absentéisme Global (TAG)	0,38
Franchise appliquée	0,30
Taille de l'entreprise	0,11
Convention	0,22

La valeur p de l'âge est inférieure à 0,05 ; ce qui signifie que l'hypothèse de constance des coefficients de régression des modalités de la variable âge au cours du temps n'est pas vérifiée.

On peut également étudier ce test de manière graphique :

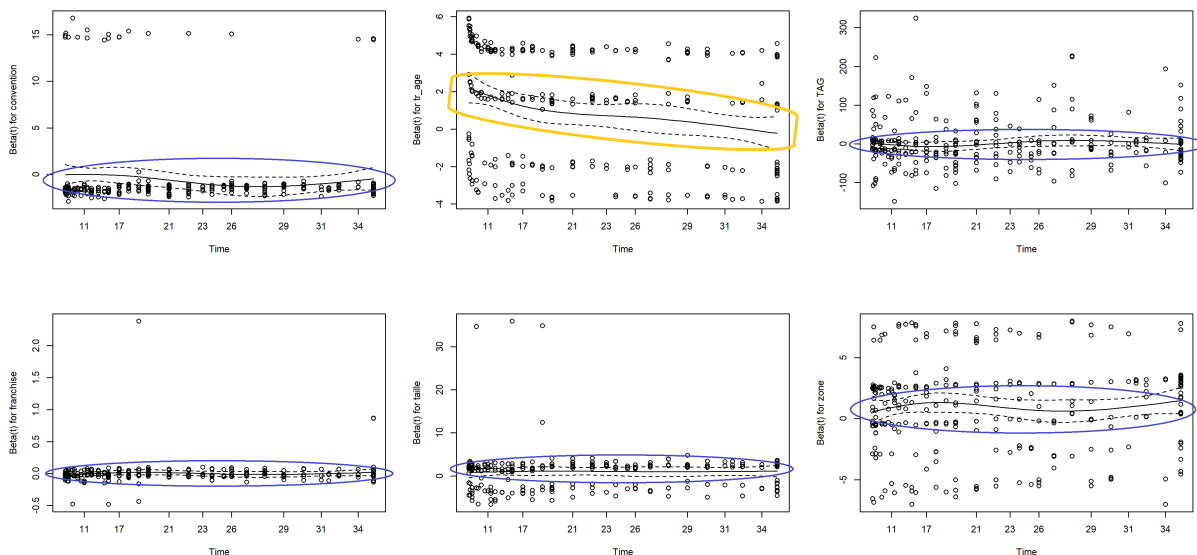


FIGURE 6.1 – Test de l'hypothèse de hasard proportionnel du modèle de Cox

On observe également graphiquement que la droite de régression des coefficients β de l'âge n'est pas horizontale ; ce qui signifie qu'ils dépendent du temps.

Ainsi, d'après ce test, l'hypothèse de hasard proportionnel est vérifiée, sauf pour la variable tranche d'âge.

6.4 Modèle de Cox stratifié

6.5 Présentation du modèle

Le non-respect de l'hypothèse de proportionnalité nous conduit à stratifier le modèle selon la variable en cause. Ainsi, la variable tranche d'âge ne sera plus une variable explicative du modèle mais la variable selon laquelle sont définies des sous-populations à partir desquelles sont estimés les coefficients de régression, qui sont les mêmes pour tous les sous-groupes. La variable tranche d'âge étant très discriminante, son effet sur le risque, déjà connu, est pris en considération à part entière.

6.5.1 Application du modèle

On obtient les coefficients de régression suivants :

Covariable	β
Zone géographique = Grand-Ouest	-0.4228
Zone géographique = Nord-Est	-0.2233
Zone géographique = Occitanie	0.2111
Zone géographique = PACA	0.2191
Zone géographique = AURA	0.6460
Taux d'Absentéisme Global (TAG)	0.6119
Franchise appliquée	-0.0032
Tranche de taille de l'entreprise = 1-39 salariés	1.6991
Tranche de taille de l'entreprise = 40-359 salariés	-0.2824
Tranche de taille de l'entreprise = 360-599 salariés	0.0778
Tranche de taille de l'entreprise = 600-739 salariés	0.2654
Tranche de taille de l'entreprise = 740-799 salariés	-0.1110
Tranche de taille de l'entreprise = 800-9999 salariés	-0.3851
Convention = FHP	-0.4410

Les conclusions du modèle stratifiées sont très proches de celles du modèle précédent. Selon ce modèle, toute chose égale par ailleurs, le risque d'invalidité en attente est plus faible pour les entreprises du Grand-Ouest et du Nord-Est. Lorsque la durée de franchise appliquée est plus élevée, le risque semble légèrement plus faible. La taille de l'entreprise n'a pas un effet monotone sur le risque. Enfin, le risque est plus élevé pour les entreprises de la convention FEHAP.

6.5.2 Validation du modèle

Pour les tests de significativité globale des coefficients de régression, nous obtenons les valeurs p suivantes :

Test	Valeur p
Rapport des maxima de vraisemblance	3×10^{-4}
Wald	2×10^{-4}
Score	7×10^{-5}

D'après ces tests, le modèle est donc globalement significatif.

Pour le test de l'hypothèse de hasard proportionnel, nous obtenons les valeurs p suivantes :

Variable	Valeur p
Zone géographique	0,132
Taux d'Absentéisme Global (TAG)	0,700
Franchise appliquée	0,335
Taille de l'entreprise	0,095
Convention	0,434

Graphiquement, on a :

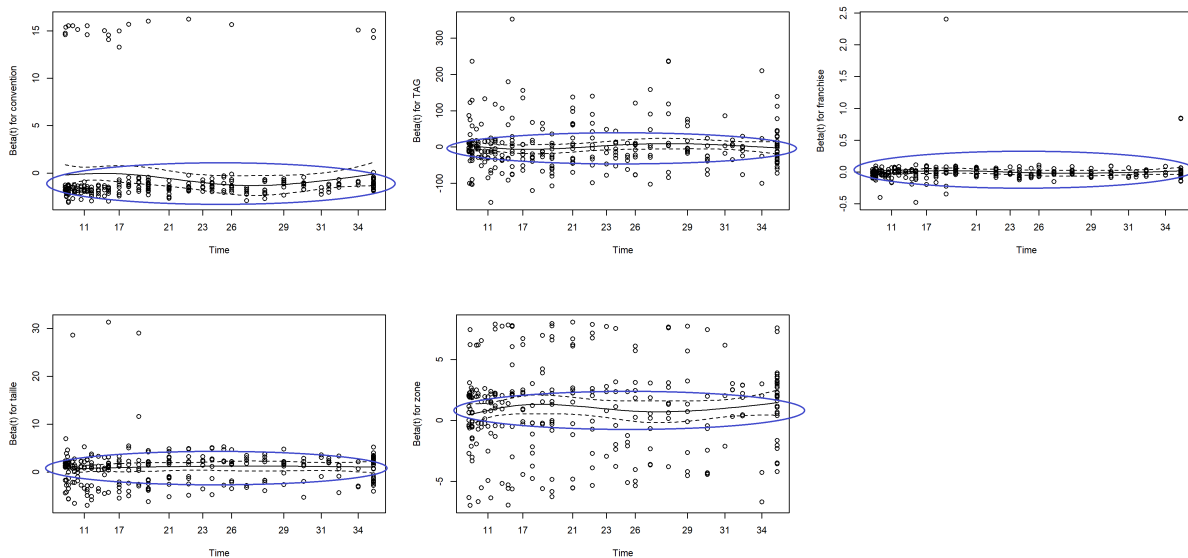


FIGURE 6.2 – Test de l'hypothèse de hasard proportionnel du modèle de Cox stratifié

D'après ce test, l'hypothèse de hasard proportionnel est vérifiée.

**CLASSIFICATION DU RISQUE D'ENTRÉE EN
INVALIDITÉ**

Remarque : Cette partie a été effectuée avec le logiciel R, grâce aux bibliothèques `dplyr`, `gmodels`, `pROC`, `rpart` et `rpart.plot`.

7.1 Contexte

L'étude des lois de maintien en incapacité et de passage en invalidité nous a permis d'avoir une vision temporelle des risques étudiés et d'identifier les facteurs principaux influant sur le tarif en incapacité et en invalidité. Ces derniers sont liés à la population assurée et à leur environnement de travail :

- La tranche d'âge ;
- La zone géographique ;
- Le taux d'absentéisme global ;
- La franchise appliquée ;
- La taille de l'entreprise ;
- La convention.

7.2 Variables étudiées

Il paraît judicieux de construire un **arbre de décision** résumant simplement l'information en regroupant les arrêts en fonction des facteurs influents et par niveau de risque de passage en invalidité.

Pour ce faire, pour chaque arrêt en incapacité terminé, nous étudions la variable indiquant le passage en invalidité ou non. Ainsi, nous pouvons estimer l'espérance de passage en invalidité (i.e. la moyenne des probabilités de passer en invalidité) de sous-groupes d'employés en incapacité en fonction de l'ensemble des variables disponibles :

- La zone géographique de l'entreprise ;
- Le Taux d'Absentéisme Global (TAG) de l'entreprise ;
- La catégorie de l'activité de l'entreprise ;
- La direction client de l'entreprise ;
- Le statut juridique de l'entreprise ;
- La tranche d'âge du salarié étudié ;
- La taille de l'entreprise ;
- La convention de l'entreprise ;
- La catégorie socioprofessionnelle du salarié étudié ;
- Le sexe du salarié étudié ;
- La franchise appliquée à l'arrêt étudié.

Afin de vérifier la pertinence de notre arbre final et notamment éviter le surapprentissage, nous divisons notre jeu de données en un jeu d'apprentissage - qui nous permettra de créer le modèle - et un jeu de validation représentant $\frac{1}{3}$ des observations disponibles.

7.3 Structure d'un arbre de décision

La construction d'un arbre de décision est une méthode d'apprentissage supervisée. En effet, elle consiste en l'analyse d'un jeu de données afin de produire un modèle pouvant par la suite être utilisé sur d'autres jeux de même structure afin de prédire une ou plusieurs valeurs.[15]

Facile à interpréter, l'arbre de décision pourra être utilisé par la Direction Technique comme outil de communication puissant auprès des services avec lesquels elle collabore afin de leur permettre de mieux visualiser le risque de passage en invalidité.

Un arbre de décision se compose :

- Premièrement, d'une racine, au sommet, comportant l'ensemble de la population à segmenter, à savoir les arrêts de travail en incapacité ayant débuté avant le 30 avril 2017 ;
- De là se succèdent les branches - ou ramifications - définissant les règles binaires de segmentation disjointes des observations, fonctions des modalités d'une unique variable ;
- D'où naissent enfin les feuilles contenant les sous-populations construites appartenant au même sous-groupe de modalités pour les variables explicatives et dont les valeurs prédites de la variable d'intérêt sont proches.

Construire un arbre de décision afin de modéliser le risque de passage en invalidité peut s'avérer assez périlleux compte tenu de sa très faible probabilité de survenance ($\approx 1,4\%$). En effet, la méthode classique d'élaboration d'un arbre de décision conduit à un arbre composé d'une seule branche - et donc d'une unique feuille - avec une précision de 98% ne prévoyant aucun passage en invalidité. L'évaluation de ce risque nécessite donc des méthodes de classification adaptées à la prédiction de la survenance d'un évènement rare.

8.1 Méthode 1 : Sous-échantillonnage

Afin d'augmenter artificiellement la fréquence de survenance de l'évènement rare (i.e. la fréquence de passage en invalidité) du jeu d'entraînement, nous supprimons aléatoirement des observations pour lesquelles on n'observe pas de passage en invalidité de manière à ce que le taux de passage en invalidité soit égal à $x\%$. [16]

Pour déterminer x , pour toute graine aléatoire $i \in \llbracket 1, N \rrbracket$ donnée, on détermine $x_i \in \llbracket 1, 50 \rrbracket$ telle qu'une fréquence de $x_i\%$ maximise l'AUC* (avec $N \in \mathbb{N}$ très grand).

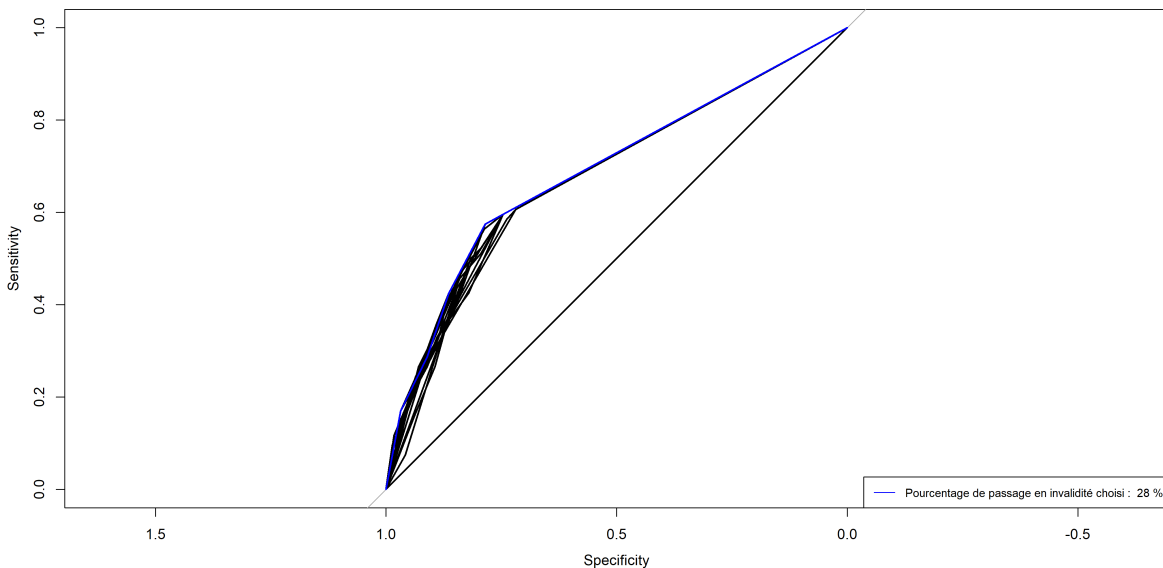


FIGURE 8.1 – Exemple : Détermination du taux de passage en invalidité x_i pour la graine aléatoire $i = 150$ pour le sous-échantillonnage

Par exemple, pour la graine aléatoire $i = 150$, l'AUC est maximisée pour un taux de passage en invalidité fixé à 28%.

On obtient $\frac{1}{N} \sum_{i=1}^N x_i = 31,87$. On choisit donc $x = 32$.

*. Mesure couramment utilisée pour valider un modèle de classification, présentée en section 9.3.3

On obtient l'arbre suivant :

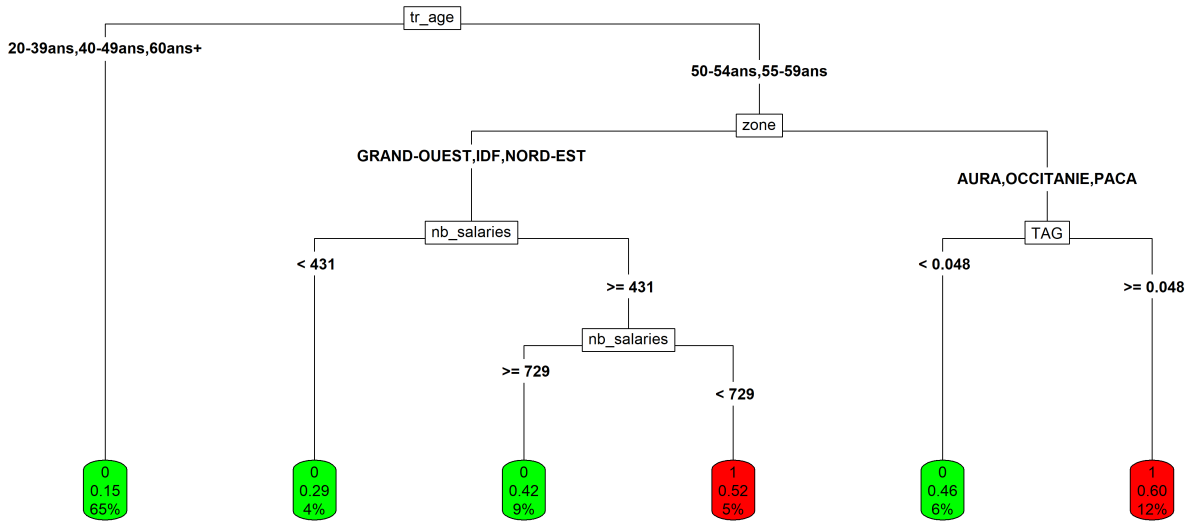


FIGURE 8.2 – Méthode 1 : Arbre de décision par sous-échantillonnage

D'après ce modèle, le risque de passage en invalidité concerne principalement les salariés entre 50 et 60 ans. Les régions Auvergne-Rhône-Alpes (AURA), Occitanie et Provence-Alpes-Côte d'Azur (PACA) se démarquent par un risque particulièrement important lorsque le taux d'absentéisme est supérieur à 4,8%. Pour le Grand-Ouest, l'Île-de-France (IDF) et le Nord-Est, c'est la taille de l'entreprise qui s'avère discriminante avec un risque plus élevé pour les entreprises ayant entre 430 et 730 salariés.

8.2 Méthode 2 : Modification de la probabilité a priori

Afin de classer les observations, pour chaque modalité d'une variable, on calcule une probabilité de passage en invalidité conditionnelle : $\mathbb{P}(X_n^{\text{inv}} = 1 | Y_n^i = y_{i,j})$, valeur augmentant avec la probabilité a priori de passage en invalidité : $\mathbb{P}(X_n^{\text{inv}} = 1)$,

où :

- n représente la quantième observation ;
- X^{inv} est une variable aléatoire valant 1 en cas de passage en invalidité et 0 sinon ;
- Y^i représente la i^{e} variable explicative et $y_{i,j}$ sa j^{e} modalité.

Cette méthode consiste à modifier arbitrairement la probabilité a priori de passage en invalidité afin d'augmenter de nouveau la fréquence de passage en invalidité dans le modèle.

Pour choisir cette valeur, nous avons :

- testé toutes les valeurs entre 0 et 0,5 avec un pas de 0,05 ;
- éliminé les valeurs conduisant à un arbre avec une seule feuille ou à des modèles en doublons ;
- gardé le modèle maximisant l'AUC (cf. section 9.3.3).

Nous obtenons une probabilité arbitraire a priori de 0,5.

On obtient l'arbre suivant :

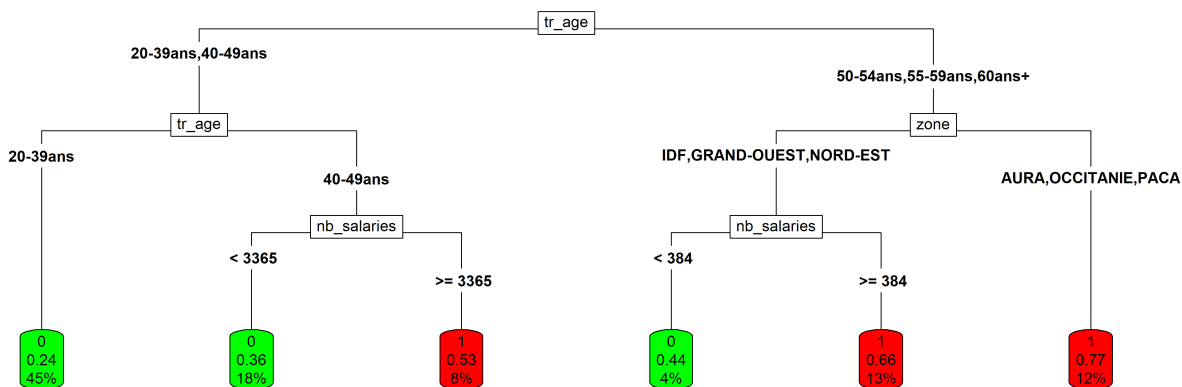


FIGURE 8.3 – Méthode 2 : Arbre de décision par modification de la probabilité a priori

D'après ce modèle, le risque de passage en invalidité concerne principalement les salariés de plus de 50 ans, et notamment ceux dont l'entreprise se situe en Auvergne-Rhône-Alpes (AURA), en Occitanie ou en Provence-Alpes-Côte d'Azur (PACA). Pour les autres régions, les entreprises de plus de 385 salariés se démarquent par un risque particulièrement important. Pour les salariés entre 40 et 49 ans, c'est encore la taille de l'entreprise qui s'avère particulièrement déterminante avec un risque moins bon pour les entreprises de plus de 3365 salariés.[16]

8.3 Autre méthode : Spécification d'une matrice de perte

Une troisième méthode consiste à définir une matrice de perte (ou *loss matrix* en anglais) qui associe des coûts à une mauvaise classification d'un passage (ou non passage) en invalidité, modifiant ainsi la gravité d'un classement d'un non-passage en invalidité en passage en invalidité en comparaison à celle d'un classement d'un passage en invalidité en non-passage en invalidité, en pénalisant

plus fortement un des deux types d'erreurs de classification. Toutefois, cette méthode, n'ayant pas conduit à des résultats structurants, ne sera pas présentée en détail.[16]

Après avoir construit les arbres de décision, il convient de les appliquer au jeu de données d'apprentissage afin de vérifier leur pertinence, i.e. si les prédictions faites sur le jeu d'apprentissage se rapprochent suffisamment de l'expérience, i.e. des passages en invalidité ou non-passages en invalidité observés. Une prédiction de passage en invalidité correspond en général (et ici) à une probabilité estimée supérieure à 0,5.

9.1 Matrice de confusion

Pour ce faire, une première approche consiste à construire une matrice de confusion, i.e. un tableau de contingence des bonnes et mauvaises prédictions :

Observation \ Prédiction	0	1
	0	Vrai négatif (VN)
1	Faux négatif (FN)	Vrai positif (VP)

TABLE 9.1 – Structure d'une matrice de confusion

Où 1 représente le passage en invalidité prédit ou observé et 0 le non-passage en invalidité prédit ou observé.

Sur la diagonale, on retrouve le nombre d'observations classées correctement tandis que sur l'antidiagonale, se trouvent les observations mal classées : un faux négatif correspondant à un passage en invalidité classé en non-passage en invalidité et un faux positif à un non-passage en invalidité classé en passage en invalidité.

On obtient les matrices de confusion suivantes :

Observation \ Prédiction	0	1
	0	7072
1	59	35

TABLE 9.2 – Matrice de confusion du modèle 1 (par sous-échantillonnage)

Avec le 1^{er} modèle, on obtient 89,4% de vrais négatifs et 9,4% de faux positifs.

Observation \ Prédiction	0	1
	0	5283
1	29	65

TABLE 9.3 – Matrice de confusion du modèle 2 (par modification de la probabilité a priori)

Avec le 2^e modèle, on obtient 66,7% de vrais négatifs et 32% de faux positifs.

De plus, le taux de faux négatifs est deux fois plus élevé dans le modèle 1, tandis que le taux de vrais positifs est près de deux fois plus élevé dans le modèle 2.

On observe que le nombre de faux positifs est particulièrement élevé pour le deuxième modèle : près d'un tiers des arrêts n'aboutissant pas à un passage en invalidité ont été mal classés. Toutefois, cela ne pose pas trop problème car cela montre que le premier modèle est plus prudent et attribue une probabilité de passage en invalidité supérieure à la réalité.

9.2 Précision du modèle

9.2.1 Expression

La précision d'un modèle correspond au taux de classifications correctes :

$$\frac{VN + VP}{VN + VP + FN + FP}$$

On obtient les précisions suivantes :

	Précision
Modèle 1	0,8982558
Modèle 2	0,6759353

90% des prédictions du 1^{er} modèle sont correctes, tandis que seulement 68% de celles du 2^e modèle sont correctes.

Ainsi, le premier modèle est meilleur selon le critère de la précision.

9.2.2 Limites

Cette mesure présente trois limites majeures :

- Premièrement, elle dépend d'un seuil d'affectation permettant de classer les prévisions : en général, on prédit un positif lorsque la probabilité de survenance de l'évènement étudié est supérieure à 0,5 - seuil qui n'est pas toujours optimal, notamment dans notre cas où la survenance de l'évènement étudié est rare ;
- En deuxième lieu, dans la formule de la précision, le coût d'un faux positif est le même que celui d'un faux négatif. Ainsi, dans notre cas, cela dégrade fortement la précision du deuxième modèle, ce qui pourrait conduire à l'abandonner, alors que le premier modèle est bien moins prudent et semble se rapprocher du modèle qui ne prédirait aucun passage en invalidité ;
- Enfin, le modèle qui ne prédirait aucun passage en invalidité aurait quant à lui une précision très élevée (98%) alors que l'erreur de prévision est très grande car il ne permet pas d'estimer le risque de survenance de l'évènement rare. En effet, le taux de faux négatifs - que l'on souhaite minimiser - du 1^{er} modèle est plus élevé, tandis que son taux de vrais positifs - que l'on souhaite maximiser - est plus faible.

9.3 Sensibilité et spécificité du modèle

9.3.1 Expression

La sensibilité d'un modèle correspond au taux de vrais négatifs, i.e. de non-passages en invalidité bien prédits :

$$\frac{VN}{VN + FN}$$

La spécificité d'un modèle correspond au taux de vrais positifs, i.e. de passages en invalidité bien prédits :

$$\frac{VP}{VP + FP}$$

On obtient les indices suivants :

	Sensibilité	Spécificité
Modèle 1	0,3723404	0,9045792
Modèle 2	0,6914894	0,6757483

Le 1^{er} modèle conduit à un taux de vrais négatifs plus élevé tandis que le 2^e modèle a un meilleur taux de vrais positifs.

9.3.2 Taux de faux négatifs en fonction du seuil d'affectation

L'idée est de calculer les x^e centiles $c(x)$, $x \in \llbracket 0, 100 \rrbracket$, des probabilités de passage en invalidité estimées (p_i) par chaque modèle. Puis, pour chaque quantile, on calcule le taux de faux négatifs que l'on aurait si l'on avait fixé le seuil d'affectation à $c(x)$, autrement dit, le taux de passage en invalidité observé pour les arrêts auxquels une probabilité inférieure au quantile à été affectée :

$$y = \frac{\sum_{i=1}^N \mathbb{1}_{\{p_i \leq c(x)\}} \times \mathbb{1}_{\text{invalidité}(i)}}{\sum_{i=1}^N \mathbb{1}_{\{p_i \leq c(x)\}}} = \frac{FN}{VN + FN} = 1 - \text{sensibilité}(c(x))$$

où N représente le nombre d'arrêts de travail observés.

Enfin, nous traçons le taux de faux négatifs associé à $c(x)$ en fonction de x :

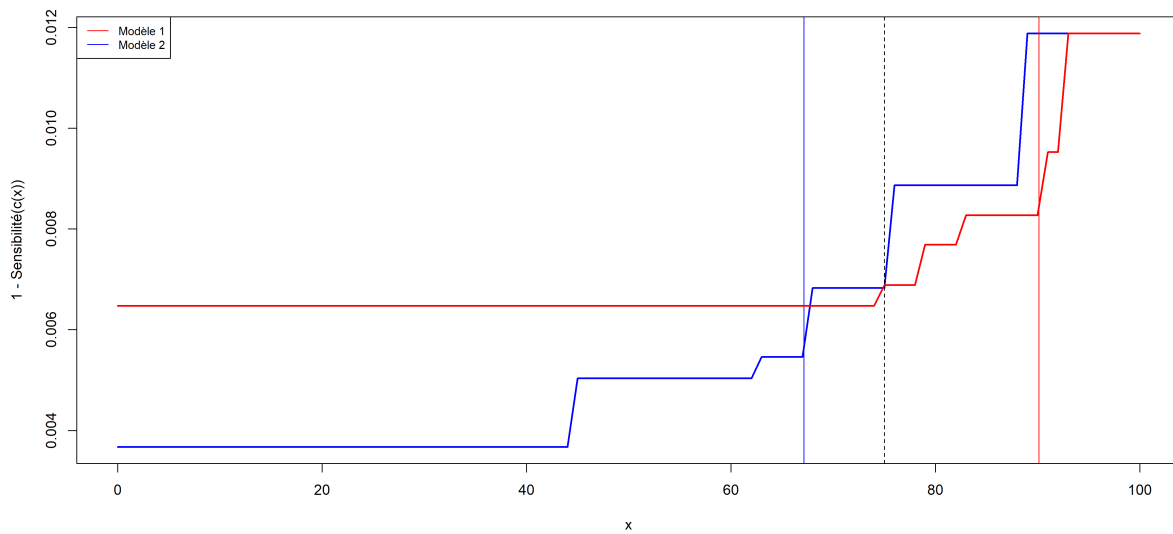


FIGURE 9.1 – Taux de faux négatifs en fonction du taux d'affectation négative

La droite verticale rouge (respectivement bleue) correspond à la valeur $x \in \llbracket 0, 100 \rrbracket$ telle que $c(x) = 0,5$ pour le premier modèle (respectivement deuxième modèle). On constate qu'un seuil d'affectation à 0,5 n'est pas optimal, notamment pour le premier modèle qui attribue une probabilité de passage en invalidité minimale à 75% des arrêts du jeu de validation et pour lequel un tel seuil correspond à un taux de faux négatif proche du maximum.

On constate que pour les deux modèles, le taux de faux négatifs commence à fortement augmenter à partir du 75^e centile (droite verticale noire en pointillés).

En traçant le taux de faux négatifs associé à $c(x)$ en fonction de $c(x)$, nous obtenons :

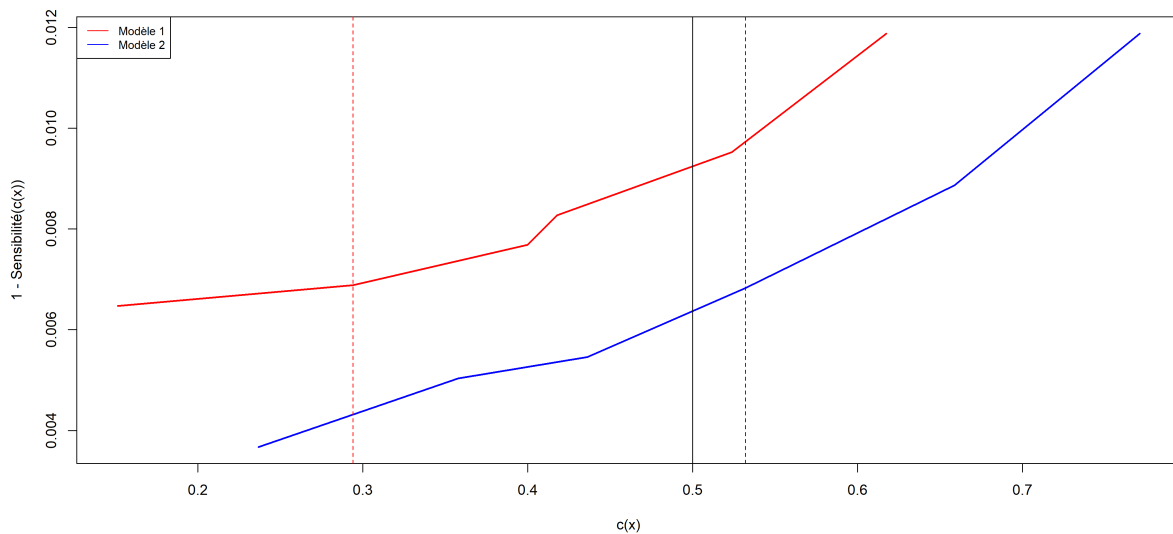


FIGURE 9.2 – Taux de faux négatifs en fonction du seuil d'affectation

La droite verticale noire correspond au seuil d'affectation généralement utilisé : $c(x) = 0,5$; tandis

que la droite verticale rouge (respectivement bleue) en pointillés correspond à la valeur $c(75)$ associée à la distribution des probabilités prédites par le premier modèle (respectivement deuxième modèle).

Nous constatons que les distributions des modèles sont décalées et ne sont pas réparties de la même façon. Si l'on fixait un taux d'affectation à 75% plutôt que de fixer un seuil d'affectation à 0,5, on obtiendrait les indices suivants :

	Précision	Sensibilité	Spécificité
Modèle 1	0,7903185	0,5425532	0,7932975
Modèle 2	0,7606168	0,5638298	0,7629829

De la sorte, les indices de comparaison et validation des deux modèles sont extrêmement proches, ce qui rend difficile un choix et met d'autant plus en évidence la limite de ces trois indices du fait du choix arbitraire d'un seuil d'affectation qu'elles requièrent.

9.3.3 Courbe ROC et AUC

La fonction d'efficacité du récepteur représente, pour tout seuil d'affectation donné, la sensibilité (i.e. le taux de vrais négatifs) en fonction du taux de faux positifs (i.e. $1 - \text{spécificité}$).

La représentation graphique de la fonction d'efficacité du récepteur, couramment appelée courbe ROC (de l'anglais *Receiver Operating Characteristic*), permet de visualiser pour tous les seuils d'affectation si le taux de faux positifs est suffisamment inférieur au taux de vrais négatifs pour considérer que le modèle est pertinent. Plus le nombre de faux positifs est faible (et donc plus la courbe est vers la gauche) et plus le nombre de vrais négatifs est élevé (et donc plus la courbe est vers le haut) et plus le modèle considéré est fiable. Ainsi, plus la courbe s'éloigne de la première diagonale (qui représente un modèle affectant les observations par pile ou face) et se rapproche du coin supérieur gauche, plus le modèle est performant.

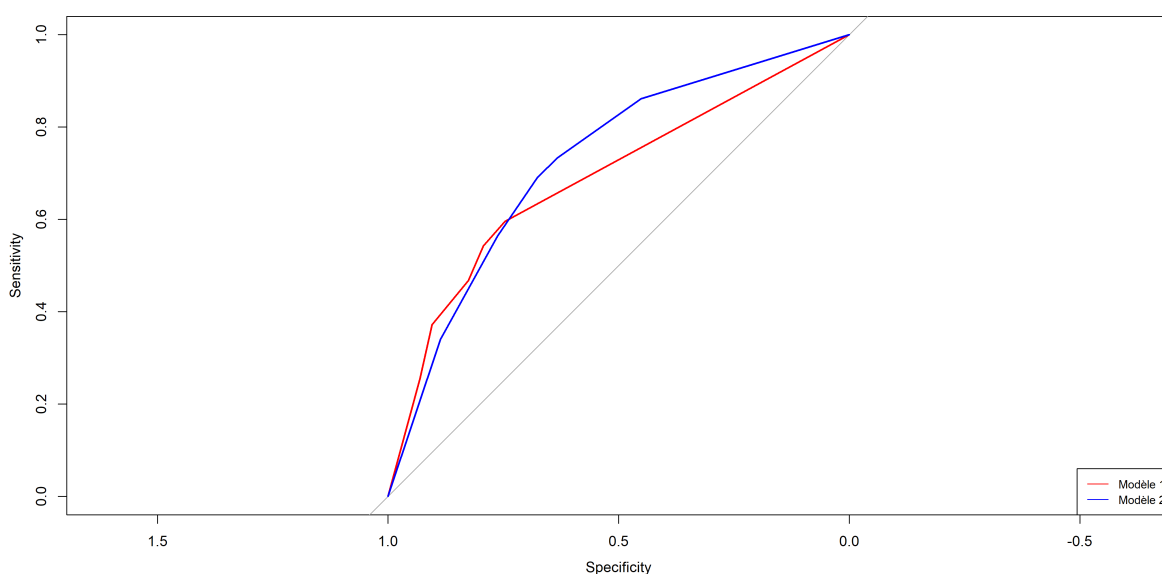


FIGURE 9.3 – Courbes ROC

Lorsque les courbes se croisent, il est difficile de déterminer à l'œil laquelle est la plus proche du coin supérieur gauche. Pour la déterminer, nous calculons l'aire sous chaque courbe (ou AUC, de l'anglais *Area Under the Curve*) par intégration - le modèle parfait ayant une AUC égale à 1.[16]

L'AUC est un indice de performance ayant deux avantages majeurs :

- Premièrement, il ne dépend pas d'un seuil d'affection fixé ;
- De plus, elle permet de comparer les indices de sensibilité et de spécificité au lieu d'exploiter directement ces valeurs pour évaluer la qualité du modèle.[10]

Nous obtenons les valeurs suivantes :

	AUC
Modèle 1	0,6888
Modèle 2	0,7254

Ainsi, le modèle par modification de la probabilité a priori est meilleur selon le critère de l'AUC.

Ce mémoire porte sur l'étude des arrêts de travail et de la typologie des profils clients en prévoyance collective. Il vise d'une part à affiner le processus de tarification, et d'autre part, à améliorer la connaissance de Sham de ses clients.

Son objectif est de répondre à la problématique suivante : Comment mettre en exergue l'hétérogénéité du portefeuille de Sham vis-à-vis des risques de maintien en incapacité et d'invalidité en attente ? Ainsi, il est question d'être à même de mieux interpréter les informations disponibles lors de la tarification ou du renouvellement et d'identifier les facteurs aggravants.

Partie 1 : Contexte de l'étude, données et analyse descriptive

Afin d'étudier le risque incapacité, nous avons utilisé la table répertoriant les arrêts survenus entre 2010 et 2018. Sur cette période, les contrats avec une franchise de 0 ou 3 jours concernent environ 98 500 salariés et cumulent 36 600 arrêts, ce qui représente environ 1,7 millions de jours indemnisés. Ainsi, les franchises courtes représentent 97% des arrêts. On a considéré qu'une valeur relative à l'incapacité est significative si elle concerne un échantillon de plus de 2 500 assurés.

Pour le risque invalidité, nous avons considéré les invalidités faisant suite à un arrêt en incapacité ayant débuté il y a plus de trois ans, et donc de survenance entre 2010 et 2017. Pendant cette période, un cumul d'environ 99 500 salariés et 400 invalidités ont été observés. On a considéré qu'une valeur relative à l'invalidité est significative si elle concerne un échantillon de plus de 15 000 assurés.

Partie 2 : Étude des lois de maintien en incapacité et d'invalidité en attente

Premièrement, nous avons étudié la corrélation entre le taux de passage en invalidité et le taux d'absentéisme. La régression linéaire du taux de passage en invalidité par rapport au taux d'absentéisme nous a permis de mettre en évidence l'existence d'une corrélation positive entre les deux taux. Il s'en est suivi la comparaison de l'influence de diverses variables sur ces deux taux. Nous avons pu observer des comportements similaires pour la plupart des variables avec :

- Des taux plus élevés pour les petites et grandes entreprises (en termes d'effectif) et des taux plus faibles pour les moyennes entreprises ;
- Des taux croissants en fonction de l'âge ;
- Des taux plus élevés pour les femmes et les non-cadres ;
- Des taux plus élevés pour les régions Occitanie et Provence-Alpes-Côte d'Azur (PACA).
- Le classement des conventions, avec par ordre croissant des taux : FHP Santé privé, FEHAP et FHP Médico-social.

On a également observé des comportements antagoniques pour le classement des autres régions, de la direction client et du statut juridique, avec notamment un taux d'absentéisme plus faible pour les entreprises privées de la santé et celles à but non lucratif.

Afin d'étudier l'évolution du risque de passage en invalidité au cours d'un arrêt en incapacité, nous avons tout d'abord comparé les lois de maintien en incapacité et d'invalidité en attente que Sham utilise pour la tarification et le provisionnement de ces risques, au comportement réel de son portefeuille. Nous avons ainsi remarqué que l'évolution du risque de passage en invalidité du portefeuille est très proche de celle des tables réglementaires pendant les huit premiers mois d'arrêt en incapacité puis, on a par ailleurs observé que le rapport du risque de l'entreprise par rapport à celui du BCAC oscille autour de 30% avant de décroître progressivement.

Il s'en est suivi une analyse descriptive permettant, la comparaison, pour chaque variable, des fonctions de hasard du passage en invalidité du portefeuille, estimées par Kalbfleisch et Prentice, selon les différents paramètres. La comparaison était souvent difficile du fait de la forte volatilité du risque instantané pour certaines modalités, comme les cadres et les entreprises à but lucratif, du fait de leur effectif plus faible dans le portefeuille. On retient tout de même que l'âge est le critère le plus discriminant pour ce risque et que ce dernier est plus important pour les cadres.

Enfin, nous avons réalisé un test du log-rank qui nous a permis de déterminer les variables pour lesquelles les courbes de survie diffèrent significativement : la tranche d'âge, la zone géographique du siège de l'entreprise, le taux d'absentéisme global, la franchise appliquée, la taille de l'entreprise et sa convention.

Nous avons poursuivi notre étude en construisant un modèle de Cox afin d'exprimer le risque instantané de passage en invalidité au cours d'un arrêt en fonction du temps et des variables significatives identifiées par le test du log-rank.

L'hypothèse forte de hasard proportionnel du modèle de Cox ayant été rejetée pour l'âge, nous avons stratifié notre modèle selon cette variable; c'est-à-dire que nous avons inclus l'effet de l'âge dans la fonction de hasard de base.

Selon ce modèle, toute chose égale par ailleurs, le risque de l'Auvergne-Rhône-Alpes (AURA) est deux fois plus élevé que celui de l'Île-de-France (IDF). La Provence-Alpes-Côte d'Azur (PACA) et l'Occitanie ont un risque supérieur d'environ 25%. Tandis que les régions du Grand-Ouest et du Nord-Est ont un risque légèrement plus faible. De plus, toute chose égale par ailleurs, le risque observé augmente de 10% lorsque le taux d'absentéisme augmente d'un point, et il décroît légèrement avec la franchise appliquée. On constate également que le risque des petites entreprises est particulièrement élevé. Pour les autres tranches de tailles, il est difficile de conclure car l'exponentiel des coefficients est assez proche de 1 et le risque en fonction de la taille de l'entreprise n'est pas monotone. Enfin, la convention Fédération Hospitalière Privée (FHP) a un risque inférieur de 35% à celui de convention Fédération des Établissements Hospitaliers et d'Aide à la Personne (FEHAP).

Partie 3 : Classification du risque d'entrée en invalidité

Dans la dernière partie de cette étude, nous souhaitions avoir une vision du risque invalidité pur, c'est-à-dire sans prendre en compte le temps passé en incapacité. Il nous paraissait intéressant de dresser un arbre de classification car cela permettrait de résumer l'information de façon facilement interprétable.

La première méthode est celle du sous-échantillonnage qui consistait à supprimer aléatoirement de la base d'apprentissage, un certain nombre d'observations pour lesquelles il n'y a pas eu de passage en invalidité de façon à augmenter artificiellement le taux de passage en invalidité. Cette méthode a l'avantage d'être très simple à mettre en œuvre, mais elle implique une perte d'information et risque de supprimer du jeu de données d'apprentissage des individus importants pour le modèle.

La deuxième méthode consistait à augmenter arbitrairement la probabilité a priori de passage en invalidité. Cette méthode, également simple à mettre en place, a l'avantage d'utiliser toute l'information disponible, cependant, elle se base tout de même sur un choix arbitraire.

Afin de déterminer le modèle le plus cohérent avec nos données, nous avons comparé les résultats des deux méthodes en calculant premièrement leur précision, leur sensibilité et leur spécificité. Il était difficile de conclure car ce n'était pas toujours le même modèle qui ressortait et ces mesures ont l'inconvénient de se baser sur un seuil d'affectation arbitraire.

Nous avons donc privilégié le critère de l'AUC qui a l'avantage de ne pas dépendre d'un seuil d'affectation arbitraire. Et qui consiste à comparer le taux de faux positifs de la base de validation au taux de vrais négatifs pour tous les seuils d'affectation. Ainsi, cela nous a amenés à retenir le modèle par modification de la probabilité a priori.

D'après ce modèle, le risque de passage en invalidité concerne principalement les salariés de plus de 50 ans, et notamment ceux dont l'entreprise se situe en Auvergne-Rhône-Alpes (AURA), en Occitanie ou en Provence-Alpes-Côte d'Azur (PACA). Pour les autres régions, les entreprises de plus de 384 salariés se démarquent par un risque particulièrement important. Pour les salariés entre 40 et 49 ans, c'est encore la taille de l'entreprise qui s'avère particulièrement déterminante avec un risque moins bon pour les entreprises de plus de 3365 salariés.

Conclusion

Le risque d'arrêt de travail est en partie couvert par les indemnités de la sécurité sociale (régime obligatoire) qui s'avèrent cependant insuffisantes pour maintenir un niveau de vie équivalent à celui d'une personne active. Une couverture complémentaire est généralement souscrite au travers d'un contrat de prévoyance, cofinancé par l'employeur et le salarié.

Les tables réglementaires du Bureau Commun d'Assurances des Collectives (BCAC) fournissent des lois de maintien en incapacité et de passage de l'incapacité à l'invalidité de référence, notamment pour le provisionnement de ces risques par les assureurs.

Avec un démarrage en 2010, l'activité de prévoyance collective est encore récente chez Sham. Ce mémoire avait pour objectif de mettre en évidence les variables pouvant expliquer l'hétérogénéité des risques de maintien en incapacité et de passage en invalidité dans le portefeuille de Relyens.

Afin de répondre à cette problématique, les variables suivantes ont été étudiées : la catégorie de l'activité des entreprises clientes, leur direction client, leur statut juridique, leur taille (en termes d'effectif), la convention à laquelle elles sont soumises, la catégorie socioprofessionnelle des salariés, leur sexe, leur âge, la zone géographique du siège des entreprises et la franchise appliquée aux arrêts observés.

Premièrement, la régression du taux de passage en invalidité en fonction du taux d'absentéisme - indicateur permettant de suivre l'évolution du risque incapacité - des entreprises clientes permet de mettre en évidence la corrélation positive entre ces deux taux. La comparaison des taux de passage en invalidité et d'absentéisme en fonction de différentes variables conduit aux conclusions suivantes : ces taux évoluent de façon similaire en fonction de la taille des entreprises, de l'âge, du sexe, de la catégorie socioprofessionnelle et de la convention. Les taux se comportent de façon antagoniste en fonction de la direction client et du statut juridique. Enfin, aucune corrélation ne semble se dessiner entre les taux en fonction de la zone géographique.

Il s'ensuit l'expression de la fonction de hasard du passage en invalidité par l'estimateur de Kalbfleisch et Prentice se basant sur l'estimateur de Kaplan-Meier, qui dans un premier temps, permet de mettre en évidence un risque instantané plus faible au cours d'un arrêt en incapacité de passer en invalidité pour le portefeuille étudié que celui présenté par les tables réglementaires. La segmentation de la fonction de hasard selon les différentes variables exploitées permet en particulier de dénoter que le risque instantané de passage en invalidité au cours d'un arrêt en incapacité est plus faible pour les non-cadres que pour les cadres ; de ce résultat étonnant a été tirée la conclusion que bien que les cadres ont une probabilité plus faible d'être arrêtés, leurs arrêts ont une probabilité plus élevée d'aboutir à une invalidité et nécessitent donc une attention particulière lorsqu'ils surviennent.

Le test du log-rank effectué dégage les variables par lesquelles la segmentation des courbes de survie en incapacité présente un intérêt : l'âge, la zone géographique, le taux d'absentéisme, la franchise, la taille et la convention.

L'effet de l'interférence de ces variables sur les lois de maintien en incapacité et de passage en invalidité est résumé grâce au modèle de Cox : les régions sont ainsi classées avec, par ordre croissant

de risque, le Grand-Ouest, puis le Nord-Est, ensuite l'Île-de-France (IDF), suivie de l'Occitanie et de la région Provence-Alpes-Côte d'Azur (PACA) et enfin, l'Auvergne-Rhône-Alpes (AURA). Le taux d'absentéisme s'avère ici avoir une corrélation positive très forte avec le risque instantané de passage en invalidité; contrairement à la franchise contractuelle dont l'effet est faiblement négatif. Enfin, la convention de la Fédération Hospitalière Privée (FHP) présente un risque plus faible selon le modèle de Cox.

Pour finir, deux arbres de décision, construits selon deux méthodes différentes - à savoir premièrement le sous-échantillonnage des observations pour lesquelles l'évènement étudié ne survient pas puis la modification de la probabilité a priori de survenance de l'évènement étudié - permettent d'avoir une vision globale et facilement interprétable du risque de passage en invalidité. Ces modèles confirment la corrélation positive du risque de passage en invalidité avec l'âge et le taux d'absentéisme. Toutefois, le lien entre la taille de l'entreprise et ce risque reste difficile à déceler.

Il convient de garder à l'esprit les limites de travaux effectués et de la méthodologie utilisée. Premièrement, une incertitude non négligeable subsiste sur les données relatives à l'invalidité du fait du manque d'observations. Par ailleurs, peu d'entreprises disposent d'un effectif suffisant ou ont été observées suffisamment longtemps pour considérer leur taux de passage en invalidité comme étant significatif et donc leur taux d'absentéisme comme un indicateur statistiquement fiable pour prédire le risque de passage en invalidité. Par la suite, l'indépendance niabile entre la variable représentant la durée de survie en incapacité précédant le passage en invalidité et la variable de censure ainsi que le taux de censure très élevé à partir du huitième mois d'observation des arrêts de travail en incapacité nous a amenés à remettre en question la fiabilité de l'estimateur de Kaplan-Meier, auquel nous avons préféré l'estimateur de Kalbfleisch et Prentice prenant davantage en considération l'influence des risques compétitifs. Le test du log-rank, quant à lui, se base sur des hypothèses similaires à celles de l'estimateur de Kaplan-Meier et s'avère donc discutable.

Le modèle de Cox permet d'étudier l'influence croisée des variables explicatives sur les risques de maintien en incapacité et de changement d'état. Cependant, l'hypothèse de hasard proportionnel sur laquelle il se base n'est pas vérifiée pour la variable la plus discriminante, à savoir l'âge. De ce fait, il ne peut expliquer à lui seul le comportement d'un groupe d'assuré face aux risques étudiés.

Enfin, les méthodes de construction des arbres de décision nécessitent d'effectuer des choix empiriques pouvant influencer sur le résultat obtenu, et donc le biaiser.

Pour conclure, cette étude permet de s'aiguiller sur les différences de provisionnement qu'engendrait la construction de tables mieux adaptées à la typologie des personnes assurées. Par ailleurs, elle fournit des outils qui permettront de suivre l'évolution du risque d'incapacité et d'invalidité des acteurs de la santé, du social et du médico-social en première ligne face à la crise sanitaire actuelle.

Cette étude évalue les risques liés à l'arrêt de travail uniquement de manière probabiliste. Elle pourrait donc être prolongée en étudiant l'hétérogénéité du coût d'un arrêt de travail en incapacité et en invalidité puis évaluer les provisions nécessaires en se basant sur le portefeuille de Relyens. Dès lors, lorsque les données disponibles seront suffisantes pour construire des tables de provisionnement âge par âge telles que celles du BCAC, le provisionnement avec les tables du BCAC et celles issues de l'expérience de Relyens pourront être simulés et comparés.

En outre, on pourrait également compléter ces travaux par une étude du risque de décès en incapacité et en invalidité, afin d'avoir une vision complète du niveau de provisionnement d'expérience par rapport aux tables du BCAC.

Enfin, nous pouvons d'ores et déjà nous interroger sur la résistance de l'analyse suite aux conséquences de la crise sanitaire ayant particulièrement affecté la survenance du risque incapacité. Plus tard, des questions se poseront sur l'invalidité. En effet, par exemple, les effets du confinement sur la santé psychique ou encore le renoncement aux soins pendant cette période pourrait augmenter la sinistralité en invalidité des prochaines années.

Les outils de suivi de ces risques développés dans le cadre de ce mémoire pourront servir de base à une analyse plus poussée de l'évolution des risques de passage en invalidité et de maintien en invalidité lorsque les données seront suffisantes.

Bibliographie

- [1] AG2R LA MONDIALE (2016). *La garantie arrêt de travail - mensualisation / maintien de salaire*. URL : <https://www.ag2rlamondiale.fr> (cité à la page 6).
- [2] — (2020). *Les obligations de la Loi Évin*. URL : <https://www.ag2rlamondiale.fr/culture-branches/mise-en-place-d-un-accord/les-obligations-de-la-loi-evin> (cité à la page 7).
- [3] APICIL (2019). *PMSS 2020 : le nouveau plafond de la Sécurité Sociale*. URL : <https://pro.apicil.com/actualites/pmss-2020/> (cité à la page 5).
- [4] AUTORITÉ DES NORMES COMPTABLES (2015). *Règlement n° 2015-11 du 26 novembre 2015 relatif aux comptes annuels des entreprises d'assurance* (cité à la page 10).
- [5] CODE DE LA SÉCURITÉ SOCIALE (2015). *Article L321-1*. URL : <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006073189&idArticle=LEGIARTI000006742902&dateTexte=&categorieLien=cid> (cité à la page 4).
- [6] — (2019). *Article R343-3*. URL : <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006073984&idArticle=LEGIARTI0000030576271&dateTexte=&categorieLien=cid> (cité à la page 10).
- [7] — (2020). *Article L341-4*. URL : https://www.legifrance.gouv.fr/affichCode.do?jsessionid=C332FC19DEA0C48FF9994503D0AFBA64.tplgfr38s_1?idSectionTA=LEGISCTA000006172604&cidTexte=LEGITEXT000006073189&dateTexte=20200608 (cité à la page 5).
- [8] DÉPARTEMENT BIOMÉTRIE DE FOVEA (2020). In : *Traitement statistique des études cliniques* (cité aux pages 37, 50).
- [9] DREES (2019). *Infographie : La protection sociale en France et en Europe - Les chiffres clés*. URL : <https://drees.solidarites-sante.gouv.fr> (cité à la page 4).
- [10] GOOGLE DEVELOPERS (2020). *Classification : ROC et AUC*. URL : <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=fr> (cité à la page 68).
- [11] INSEE (2019). *Masse salariale*. URL : <https://www.insee.fr/fr/metadonnees/definition/c1503> (cité à la page 13).
- [12] Sandrine KATSAHIAN (2012). *Incidence cumulée et risques compétitifs : ça veut dire quoi?* URL : <http://horizonshemato.com/wp-content/uploads/2017/10/Horizons-Hemato-Juin-2012-108-109-Katsahian-Varet.pdf> (cité à la page 43).
- [13] Melania PINTILIE (2011). « An Introduction to Competing Risks Analysis ». In : *Revista Española de Cardiología*. DOI : [10.1016/j.rec.2011.03.016](https://doi.org/10.1016/j.rec.2011.03.016) (cité à la page 43).
- [14] PLANCHET, F. et THÉRON, P. (2006). *Modèles de Durée - Applications actuarielles*. Economica (cité aux pages 7, 49, 51).
- [15] Gabriela QUEIROZ (2020). *Tree-Based Models in R*. DataCamp (cité à la page 58).
- [16] Heidi SEIBOLD (2020). *Survival Analysis in R*. DataCamp (cité aux pages 50, 59, 61, 62, 68).
- [17] SERVICE PUBLIC (2019). *Qu'est-ce que l'invalidité au sens de la sécurité sociale?* URL : <https://www.service-public.fr/particuliers/vosdroits/F13093> (cité à la page 5).
- [18] — (2020a). *Arrêt maladie : indemnités journalières versées au salarié*. URL : <https://www.service-public.fr/particuliers/vosdroits/F3053> (cité aux pages 5, 6).
- [19] — (2020b). *Maladie professionnelle : indemnités journalières pendant l'arrêt de travail*. URL : <https://www.service-public.fr/particuliers/vosdroits/F32148> (cité à la page 6).

- [20] SERVICE PUBLIC (2020c). *Pension d'invalidité de la Sécurité sociale*. URL : <https://www.service-public.fr/particuliers/vosdroits/F672> (cité à la page 5).
- [21] TEXTE LÉGISLATIF ET RÉGLEMENTAIRE (1978). *Loi n° 78-49 du 19 janvier 1978 relative à la mensualisation et à la procédure conventionnelle* (cité à la page 6).
- [22] — (1989). *Loi n° 89-1009 du 31 décembre 1989 renforçant les garanties offertes aux personnes assurées contre certains risques*. URL : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000709057> (cité à la page 7).
- [23] — (1996). *Arrêté du 28 mars 1996 fixant les règles de provisionnement des garanties d'invalidité et d'incapacité* (cité à la page 7).
- [24] — (2008). *Loi n° 2008-596 du 25 juin 2008 portant modernisation du marché du travail* (cité à la page 6).
- [25] TIMSIT, J.F., ALBERTI, C. et CHEVRET, S. (2005a). « Analyse de survie : le test de logrank ». In : *Rev Mal Respir* (cité à la page 49).
- [26] — (2005b). « Le modèle de Cox ». In : *Rev Mal Respir* (cité à la page 50).
- [27] Thomas Guillon VERNE (2015). *Construction de tables de mortalité d'expérience sur de petits échantillons pour l'estimation de la sinistralité décès* (cité à la page 50).
- [28] Stéphane VINATIER (2007). *Tests non paramétriques*. URL : https://www.unilim.fr/pages_perso/stephane.vinatier/Biologie/ctd10.pdf (cité à la page 23).

Annexes

1 Création des tranches de taille d'entreprise par Maan-Whitney

Tranches 1		Groupe 17 : 00340 - 00359		Groupe 33 : 00716 - 00716	
Groupe 1 : 00001 - 00039	n1 = 42 n2 = 14 u1.2 = 228 u1.3 = 294 s1.2 = 52.85 e1.2 = -1.23 groupes différents	n17 = 6 n18 = 2 u17.18 = 20 u17.19 = 6 s17.18 = 3.00 e17.18 = 1.33 groupes différents	n33 = 0 n34 = 0 u33.34 = 0 s33.34 = 0 e33.34 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 2 : 00041 - 00059	n3 = 14 n4 = 12 u3.4 = 90 u3.5 = 84 s3.4 = 19.44 e3.4 = 0.31 groupes semblables	n18 = 2 n19 = 5 u18.19 = 7 u19.19 = 4 s18.19 = 2.58 e18.19 = 0.77 groupes semblables	n34 = 0 n35 = 1 u34.35 = 1 s34.35 = 0 e34.35 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 3 : 00060 - 00076	n5 = 12 n6 = 9 u5.6 = 60 u5.7 = 54 s5.6 = 14.07 e5.6 = 0.43 groupes semblables	n19 = 5 n20 = 6 u19.20 = 20 u20.20 = 15 s19.20 = 5.43 e19.20 = 0.81 groupes semblables	n35 = 1 n36 = 0 u35.36 = 0 s35.36 = 0 e35.36 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 4 : 00082 - 00098	n4 = 9 n5 = 9 u4.5 = 47 u4.6 = 42 s4.5 = 11.32 e4.5 = 0.57 groupes semblables	n20 = 6 n21 = 3 u20.21 = 12 u21.21 = 9 s20.21 = 3.87 e20.21 = 0.77 groupes semblables	n36 = 0 n37 = 0 u36.37 = 0 s36.37 = 0 e36.37 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 5 : 00102 - 00117	n5 = 9 n6 = 18 u5.6 = 85 u5.7 = 81 s5.6 = 19.44 e5.6 = 0.21 groupes semblables	n21 = 3 n22 = 5 u21.22 = 11 u22.22 = 7 s21.22 = 3.35 e21.22 = 1.04 groupes différents	n37 = 0 n38 = 0 u37.38 = 0 s37.38 = 0 e37.38 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 6 : 00121 - 00139	n6 = 18 n7 = 6 u6.7 = 59 u6.8 = 54 s6.7 = 15.00 e6.7 = 0.33 groupes semblables	n22 = 5 n23 = 3 u22.23 = 11 u23.23 = 7 s22.23 = 3.35 e22.23 = 1.04 groupes différents	n38 = 0 n39 = 0 u38.39 = 0 s38.39 = 0 e38.39 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 7 : 00142 - 00157	n7 = 6 n8 = 10 u7.8 = 30 u7.9 = 30 s7.8 = 8.22 e7.8 = 0.00 groupes semblables	n23 = 3 n24 = 1 u23.24 = 2 u24.24 = 1 s23.24 = 1.5 e23.24 = 1.12 e24.24 = 0.45 groupes semblables	n39 = 0 n40 = 1 u39.40 = 0 s39.40 = 0 e39.40 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 8 : 00160 - 00177	n8 = 10 n9 = 5 u8.9 = 31 u8.10 = 25 s8.9 = 8.16 e8.9 = 0.73 groupes semblables	n24 = 1 n25 = 2 u24.25 = 1 u25.25 = 1 s24.25 = 0.83 e24.25 = 0.00 groupes semblables	n40 = 1 n41 = 0 u40.41 = 0 s40.41 = 0 e40.41 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 9 : 00180 - 00194	n9 = 5 n10 = 6 u9.10 = 14 u9.10 = 15 s9.10 = 5.48 e9.10 = -0.18 groupes semblables	n25 = 1 n26 = 1 u25.26 = 1 u26.26 = 1 s25.26 = 0.82 e25.26 = 0.00 groupes semblables	n41 = 0 n42 = 0 u41.42 = 0 s41.42 = 0 e41.42 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 10 : 00200 - 00216	n10 = 6 n11 = 10 u10.11 = 37 u10.11 = 30 s10.11 = 9.22 e10.11 = 0.76 groupes semblables	n26 = 1 n27 = 3 u26.27 = 2 u27.27 = 1 s26.27 = 1.12 e26.27 = 0.45 groupes semblables	n42 = 0 n43 = 0 u42.43 = 0 s42.43 = 0 e42.43 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 11 : 00224 - 00239	n11 = 10 n12 = 5 u11.12 = 32 u11.12 = 25 s11.12 = 8.16 e11.12 = 0.86 groupes semblables	n27 = 3 n28 = 0 u27.28 = 0 u27.28 = 0 s27.28 = 0.00 e27.28 = 0.00 une seule entreprise dans l'un des groupes	n43 = 0 n44 = 0 u43.44 = 0 s43.44 = 0 e43.44 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 12 : 00242 - 00255	n12 = 5 n13 = 9 u12.13 = 17 u12.13 = 22.5 s12.13 = 7.50 e12.13 = -0.73 groupes semblables	n28 = 0 n29 = 2 u28.29 = 1 u28.29 = 0 s28.29 = 0.00 e28.29 = 0.00 une seule entreprise dans l'un des groupes	n44 = 0 n45 = 0 u44.45 = 0 s44.45 = 0 e44.45 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 13 : 00260 - 00274	n13 = 9 n14 = 10 u13.14 = 54 u13.14 = 45 s13.14 = 12.29 e13.14 = 0.73 groupes semblables	n29 = 2 n30 = 0 u29.30 = 1 u29.30 = 0 s29.30 = 0.00 e29.30 = 0.00 une seule entreprise dans l'un des groupes	n45 = 0 n46 = 0 u45.46 = 0 s45.46 = 0 e45.46 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 14 : 00280 - 00299	n14 = 10 n15 = 2 u14.15 = 8 u14.15 = 10 s14.15 = 4.85 e14.15 = -0.21 groupes semblables	n30 = 0 n31 = 3 u30.31 = 1 u30.31 = 0 s30.31 = 0.00 e30.31 = 0.00 une seule entreprise dans l'un des groupes	n46 = 0 n47 = 0 u46.47 = 0 s46.47 = 0 e46.47 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 15 : 00300 - 00317	n15 = 2 n16 = 6 u15.16 = 6 u15.16 = 6 s15.16 = 3.00 e15.16 = -0.67 groupes semblables	n31 = 3 n32 = 2 u31.32 = 4 u31.32 = 3 s31.32 = 1.73 e31.32 = 0.58 groupes semblables	n47 = 0 n48 = 0 u47.48 = 0 s47.48 = 0 e47.48 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 16 : 00326 - 00339	n16 = 6 n17 = 6 u16.17 = 22 u16.17 = 18 s16.17 = 6.24 e16.17 = 0.64 groupes semblables	n32 = 2 n33 = 0 u32.33 = 0 u32.33 = 0 s32.33 = 0.00 e32.33 = 0.00 une seule entreprise dans l'un des groupes	n48 = 0 n49 = 0 u48.49 = 0 s48.49 = 0 e48.49 = 0.00 une seule entreprise dans l'un des groupes		
Groupe 17 : 00340 - 00359		n33 = 0 n34 = 2 u33.34 = 0 s33.34 = 0 e33.34 = 0.00 une seule entreprise dans l'un des groupes	n49 = 0 n50 = 0 u49.50 = 0 s49.50 = 0 e49.50 = 0.00 une seule entreprise dans l'un des groupes		

Tranches 2		Tranches 3		Tranches finales	
Groupe 1 : 00001 - 00039	n1 = 42 n2 = 152 u1,2 = 2490 μ1,2 = 3192 s1,2 = 322,09 e1,2 = -2,18 groupes différents	Groupe 1 : 00001 - 00039	n1 = 42 n2 = 152 n3 = 44 u1,2 = 2490 μ1,2 = 3192 s1,2 = 322,09 e1,2 = -2,18 groupes différents	1 - 39 40 - 399 360 - 599 600 - 739 740 - 799 800 - 999 10000 et plus	
Groupe 2 : 00041 - 00359	n2 = 152 n3 = 19 u2,3 = 1179 μ2,3 = 1444 s2,3 = 203,46 e2,3 = -1,30 groupes différents	Groupe 2 : 00041 - 00359	n2 = 152 n3 = 44 u2,3 = 2629 μ2,3 = 3344 s2,3 = 331,35 e2,3 = -2,16 groupes différents		
Groupe 3 : 00366 - 00438	n3 = 19 n4 = 5 u3,4 = 60 μ3,4 = 47,5 s3,4 = 14,07 e3,4 = 0,89 groupes semblables	Groupe 3 : 00366 - 00592	n3 = 44 n4 = 9 u3,4 = 130 μ3,4 = 198 s3,4 = 42,21 e3,4 = -1,61 groupes différents		
Groupe 4 : 00440 - 00456	n4 = 5 n5 = 14 u4,5 = 38 μ4,5 = 35 s4,5 = 10,80 e4,5 = 0,28 groupes semblables	Groupe 4 : 00612 - 00721	n4 = 9 n5 = 2 u4,5 = 9 μ4,5 = 9 s4,5 = 4,24 e4,5 = 0,00 groupes semblables		
Groupe 5 : 00463 - 00553	n5 = 14 n6 = 3 u5,6 = 21 μ5,6 = 21 s5,6 = 7,94 e5,6 = 0,00 groupes semblables	Groupe 5 : 00740 - 00771	n5 = 2 n6 = 11 u5,6 = 16 μ5,6 = 11 s5,6 = 5,07 e5,6 = 0,59 groupes semblables		
Groupe 6 : 00574 - 00592	n6 = 3 n7 = 7 u6,7 = 5 μ6,7 = 10,5 s6,7 = 4,39 e6,7 = -1,25 groupes différents	Groupe 6 : 00845 - 03309	n6 = 11 n7 = 1 u6,7 = 4 μ6,7 = 5,5 s6,7 = 3,45 e6,7 = -0,43 groupes semblables		
Groupe 7 : 00612 - 00659	n7 = 7 n8 = 1 u7,8 = 2 μ7,8 = 3,5 s7,8 = 2,29 e7,8 = -0,65 groupes semblables	Groupe 7 : 13205 - 15776			
Groupe 8 : 00716 - 00721	n8 = 1 n9 = 2 u8,9 = 2 μ8,9 = 1 s8,9 = 0,82 e8,9 = 1,22 groupes différents				
Groupe 9 : 00740 - 00771	n9 = 2 n10 = 2 u9,10 = 4 μ9,10 = 2 s9,10 = 1,29 e9,10 = 1,55 groupes différents				
Groupe 10 : 00845 - 00885	n10 = 2 n11 = 1 u10,11 = 1 μ10,11 = 1 s10,11 = 0,82 e10,11 = 0,00 groupes semblables				
Groupe 11 : 00936 - 00938	n11 = 1 n12 = 2 u11,12 = 1 μ11,12 = 1 s11,12 = 0,82 e11,12 = 0,00 groupes semblables				
Groupe 12 : 01008 - 01153	n12 = 2 n13 = 3 u12,13 = 3 μ12,13 = 3 s12,13 = 1,79 e12,13 = 0,00 groupes semblables				
Groupe 13 : 01432 - 03309	n13 = 3 n14 = 1 u13,14 = 0 μ13,14 = 1,5 s13,14 = 1,12 e13,14 = -1,34 groupes différents				
Groupe 14 : 13205 - 15776					

Franchises 2																	
Étape 1 : 19-19	<table border="1"> <tr><td>n2 =</td><td>30</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,2 =</td><td>62</td></tr> <tr><td>p2,2 =</td><td>120</td></tr> <tr><td>e2,2 =</td><td>27,34</td></tr> <tr><td>e2,2 =</td><td>-5,97</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	30	n3 =	8	u2,2 =	62	p2,2 =	120	e2,2 =	27,34	e2,2 =	-5,97	groupes différents			
n2 =	30																
n3 =	8																
u2,2 =	62																
p2,2 =	120																
e2,2 =	27,34																
e2,2 =	-5,97																
groupes différents																	
Étape 2 : 20	<table border="1"> <tr><td>n2 =</td><td>3</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,3 =</td><td>19</td></tr> <tr><td>p2,3 =</td><td>904</td></tr> <tr><td>e2,3 =</td><td>24,63</td></tr> <tr><td>e2,3 =</td><td>-2,42</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	3	n3 =	26	u2,3 =	19	p2,3 =	904	e2,3 =	24,63	e2,3 =	-2,42	groupes différents			
n2 =	3																
n3 =	26																
u2,3 =	19																
p2,3 =	904																
e2,3 =	24,63																
e2,3 =	-2,42																
groupes différents																	
Étape 3 : 21-23	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,4 =</td><td>276</td></tr> <tr><td>p2,4 =</td><td>332</td></tr> <tr><td>e2,4 =</td><td>24,64</td></tr> <tr><td>e2,4 =</td><td>-1,12</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,4 =	276	p2,4 =	332	e2,4 =	24,64	e2,4 =	-1,12	groupes différents			
n2 =	26																
n3 =	26																
u2,4 =	276																
p2,4 =	332																
e2,4 =	24,64																
e2,4 =	-1,12																
groupes différents																	
Étape 4 : 24-26	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>89</td></tr> <tr><td>u2,5 =</td><td>89</td></tr> <tr><td>p2,5 =</td><td>904</td></tr> <tr><td>e2,5 =</td><td>24,63</td></tr> <tr><td>e2,5 =</td><td>-0,61</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	26	n3 =	89	u2,5 =	89	p2,5 =	904	e2,5 =	24,63	e2,5 =	-0,61	groupes semblables			
n2 =	26																
n3 =	89																
u2,5 =	89																
p2,5 =	904																
e2,5 =	24,63																
e2,5 =	-0,61																
groupes semblables																	
Étape 5 : 27	<table border="1"> <tr><td>n2 =</td><td>8</td></tr> <tr><td>n3 =</td><td>13</td></tr> <tr><td>u2,6 =</td><td>24</td></tr> <tr><td>p2,6 =</td><td>360</td></tr> <tr><td>e2,6 =</td><td>21,94</td></tr> <tr><td>e2,6 =</td><td>-1,44</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	8	n3 =	13	u2,6 =	24	p2,6 =	360	e2,6 =	21,94	e2,6 =	-1,44	groupes différents			
n2 =	8																
n3 =	13																
u2,6 =	24																
p2,6 =	360																
e2,6 =	21,94																
e2,6 =	-1,44																
groupes différents																	
Étape 6 : 28-31	<table border="1"> <tr><td>n2 =</td><td>33</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,7 =</td><td>476</td></tr> <tr><td>p2,7 =</td><td>455</td></tr> <tr><td>e2,7 =</td><td>0,31</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	33	n3 =	26	u2,7 =	476	p2,7 =	455	e2,7 =	0,31	groupes semblables					
n2 =	33																
n3 =	26																
u2,7 =	476																
p2,7 =	455																
e2,7 =	0,31																
groupes semblables																	
Étape 7 : 32-34	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>17</td></tr> <tr><td>u2,8 =</td><td>309</td></tr> <tr><td>p2,8 =</td><td>211</td></tr> <tr><td>e2,8 =</td><td>40,26</td></tr> <tr><td>e2,8 =</td><td>-0,30</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	26	n3 =	17	u2,8 =	309	p2,8 =	211	e2,8 =	40,26	e2,8 =	-0,30	groupes semblables			
n2 =	26																
n3 =	17																
u2,8 =	309																
p2,8 =	211																
e2,8 =	40,26																
e2,8 =	-0,30																
groupes semblables																	
Étape 8 : 35-36	<table border="1"> <tr><td>n2 =</td><td>17</td></tr> <tr><td>n3 =</td><td>17</td></tr> <tr><td>u2,9 =</td><td>126</td></tr> <tr><td>p2,9 =</td><td>144,3</td></tr> <tr><td>e2,9 =</td><td>29,03</td></tr> <tr><td>e2,9 =</td><td>-0,46</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	17	n3 =	17	u2,9 =	126	p2,9 =	144,3	e2,9 =	29,03	e2,9 =	-0,46	groupes semblables			
n2 =	17																
n3 =	17																
u2,9 =	126																
p2,9 =	144,3																
e2,9 =	29,03																
e2,9 =	-0,46																
groupes semblables																	
Étape 9 : 37-38	<table border="1"> <tr><td>n2 =</td><td>17</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,10 =</td><td>246</td></tr> <tr><td>p2,10 =</td><td>246</td></tr> <tr><td>e2,10 =</td><td>22,1</td></tr> <tr><td>e2,10 =</td><td>40,26</td></tr> <tr><td>e2,10 =</td><td>-0,27</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	17	n3 =	26	u2,10 =	246	p2,10 =	246	e2,10 =	22,1	e2,10 =	40,26	e2,10 =	-0,27	groupes semblables	
n2 =	17																
n3 =	26																
u2,10 =	246																
p2,10 =	246																
e2,10 =	22,1																
e2,10 =	40,26																
e2,10 =	-0,27																
groupes semblables																	
Étape 10 : 39-41	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,11 =</td><td>26</td></tr> <tr><td>p2,11 =</td><td>332</td></tr> <tr><td>e2,11 =</td><td>24,64</td></tr> <tr><td>e2,11 =</td><td>-1,12</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,11 =	26	p2,11 =	332	e2,11 =	24,64	e2,11 =	-1,12	groupes différents			
n2 =	26																
n3 =	26																
u2,11 =	26																
p2,11 =	332																
e2,11 =	24,64																
e2,11 =	-1,12																
groupes différents																	
Étape 11 : 42-44	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>33</td></tr> <tr><td>u2,12 =</td><td>290</td></tr> <tr><td>p2,12 =</td><td>420</td></tr> <tr><td>e2,12 =</td><td>68,57</td></tr> <tr><td>e2,12 =</td><td>-2,43</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	33	u2,12 =	290	p2,12 =	420	e2,12 =	68,57	e2,12 =	-2,43	groupes différents			
n2 =	26																
n3 =	33																
u2,12 =	290																
p2,12 =	420																
e2,12 =	68,57																
e2,12 =	-2,43																
groupes différents																	
Étape 12 : 45-48	<table border="1"> <tr><td>n2 =</td><td>33</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,13 =</td><td>82</td></tr> <tr><td>p2,13 =</td><td>360</td></tr> <tr><td>e2,13 =</td><td>31,04</td></tr> <tr><td>e2,13 =</td><td>-1,91</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	33	n3 =	8	u2,13 =	82	p2,13 =	360	e2,13 =	31,04	e2,13 =	-1,91	groupes différents			
n2 =	33																
n3 =	8																
u2,13 =	82																
p2,13 =	360																
e2,13 =	31,04																
e2,13 =	-1,91																
groupes différents																	
Étape 13 : 49	<table border="1"> <tr><td>n2 =</td><td>8</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,14 =</td><td>43</td></tr> <tr><td>p2,14 =</td><td>904</td></tr> <tr><td>e2,14 =</td><td>24,63</td></tr> <tr><td>e2,14 =</td><td>-2,42</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	8	n3 =	26	u2,14 =	43	p2,14 =	904	e2,14 =	24,63	e2,14 =	-2,42	groupes différents			
n2 =	8																
n3 =	26																
u2,14 =	43																
p2,14 =	904																
e2,14 =	24,63																
e2,14 =	-2,42																
groupes différents																	
Étape 14 : 50-53	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>13</td></tr> <tr><td>u2,15 =</td><td>439</td></tr> <tr><td>p2,15 =</td><td>455</td></tr> <tr><td>e2,15 =</td><td>68,57</td></tr> <tr><td>e2,15 =</td><td>-0,13</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	26	n3 =	13	u2,15 =	439	p2,15 =	455	e2,15 =	68,57	e2,15 =	-0,13	groupes semblables			
n2 =	26																
n3 =	13																
u2,15 =	439																
p2,15 =	455																
e2,15 =	68,57																
e2,15 =	-0,13																
groupes semblables																	
Étape 15 : 54-56	<table border="1"> <tr><td>n2 =</td><td>33</td></tr> <tr><td>n3 =</td><td>120</td></tr> <tr><td>u2,16 =</td><td>360</td></tr> <tr><td>p2,16 =</td><td>904</td></tr> <tr><td>e2,16 =</td><td>24,63</td></tr> <tr><td>e2,16 =</td><td>-0,61</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	33	n3 =	120	u2,16 =	360	p2,16 =	904	e2,16 =	24,63	e2,16 =	-0,61	groupes semblables			
n2 =	33																
n3 =	120																
u2,16 =	360																
p2,16 =	904																
e2,16 =	24,63																
e2,16 =	-0,61																
groupes semblables																	
Étape 16 : 57	<table border="1"> <tr><td>n2 =</td><td>8</td></tr> <tr><td>n3 =</td><td>17</td></tr> <tr><td>u2,17 =</td><td>90</td></tr> <tr><td>p2,17 =</td><td>66</td></tr> <tr><td>e2,17 =</td><td>11,13</td></tr> <tr><td>e2,17 =</td><td>-0,47</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	8	n3 =	17	u2,17 =	90	p2,17 =	66	e2,17 =	11,13	e2,17 =	-0,47	groupes semblables			
n2 =	8																
n3 =	17																
u2,17 =	90																
p2,17 =	66																
e2,17 =	11,13																
e2,17 =	-0,47																
groupes semblables																	
Étape 17 : 58-59	<table border="1"> <tr><td>n2 =</td><td>17</td></tr> <tr><td>n3 =</td><td>17</td></tr> <tr><td>u2,18 =</td><td>360</td></tr> <tr><td>p2,18 =</td><td>144,3</td></tr> <tr><td>e2,18 =</td><td>29,03</td></tr> <tr><td>e2,18 =</td><td>-0,30</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	17	n3 =	17	u2,18 =	360	p2,18 =	144,3	e2,18 =	29,03	e2,18 =	-0,30	groupes semblables			
n2 =	17																
n3 =	17																
u2,18 =	360																
p2,18 =	144,3																
e2,18 =	29,03																
e2,18 =	-0,30																
groupes semblables																	
Étape 18 : 60-61	<table border="1"> <tr><td>n2 =</td><td>17</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,19 =</td><td>126</td></tr> <tr><td>p2,19 =</td><td>246</td></tr> <tr><td>e2,19 =</td><td>22,1</td></tr> <tr><td>e2,19 =</td><td>40,26</td></tr> <tr><td>e2,19 =</td><td>-0,27</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	17	n3 =	26	u2,19 =	126	p2,19 =	246	e2,19 =	22,1	e2,19 =	40,26	e2,19 =	-0,27	groupes différents	
n2 =	17																
n3 =	26																
u2,19 =	126																
p2,19 =	246																
e2,19 =	22,1																
e2,19 =	40,26																
e2,19 =	-0,27																
groupes différents																	
Étape 19 : 62-64	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,20 =</td><td>47</td></tr> <tr><td>p2,20 =</td><td>904</td></tr> <tr><td>e2,20 =</td><td>24,63</td></tr> <tr><td>e2,20 =</td><td>-1,12</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	8	u2,20 =	47	p2,20 =	904	e2,20 =	24,63	e2,20 =	-1,12	groupes différents			
n2 =	26																
n3 =	8																
u2,20 =	47																
p2,20 =	904																
e2,20 =	24,63																
e2,20 =	-1,12																
groupes différents																	
Étape 20 : 65	<table border="1"> <tr><td>n2 =</td><td>8</td></tr> <tr><td>n3 =</td><td>13</td></tr> <tr><td>u2,21 =</td><td>24</td></tr> <tr><td>p2,21 =</td><td>360</td></tr> <tr><td>e2,21 =</td><td>21,94</td></tr> <tr><td>e2,21 =</td><td>-1,44</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	8	n3 =	13	u2,21 =	24	p2,21 =	360	e2,21 =	21,94	e2,21 =	-1,44	groupes différents			
n2 =	8																
n3 =	13																
u2,21 =	24																
p2,21 =	360																
e2,21 =	21,94																
e2,21 =	-1,44																
groupes différents																	

Franchises 3															
Étape 1 : 19-20	<table border="1"> <tr><td>n2 =</td><td>30</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,2 =</td><td>62</td></tr> <tr><td>p2,2 =</td><td>207</td></tr> <tr><td>e2,2 =</td><td>74,68</td></tr> <tr><td>e2,2 =</td><td>-5,90</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	30	n3 =	8	u2,2 =	62	p2,2 =	207	e2,2 =	74,68	e2,2 =	-5,90	groupes différents	
n2 =	30														
n3 =	8														
u2,2 =	62														
p2,2 =	207														
e2,2 =	74,68														
e2,2 =	-5,90														
groupes différents															
Étape 2 : 21-23	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,3 =</td><td>261</td></tr> <tr><td>p2,3 =</td><td>420</td></tr> <tr><td>e2,3 =</td><td>68,57</td></tr> <tr><td>e2,3 =</td><td>-1,06</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,3 =	261	p2,3 =	420	e2,3 =	68,57	e2,3 =	-1,06	groupes différents	
n2 =	26														
n3 =	26														
u2,3 =	261														
p2,3 =	420														
e2,3 =	68,57														
e2,3 =	-1,06														
groupes différents															
Étape 3 : 24-27	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,4 =</td><td>488</td></tr> <tr><td>p2,4 =</td><td>511,3</td></tr> <tr><td>e2,4 =</td><td>85,13</td></tr> <tr><td>e2,4 =</td><td>-1,46</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,4 =	488	p2,4 =	511,3	e2,4 =	85,13	e2,4 =	-1,46	groupes différents	
n2 =	26														
n3 =	26														
u2,4 =	488														
p2,4 =	511,3														
e2,4 =	85,13														
e2,4 =	-1,46														
groupes différents															
Étape 4 : 28-31	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>89</td></tr> <tr><td>u2,5 =</td><td>841,1</td></tr> <tr><td>p2,5 =</td><td>1527,5</td></tr> <tr><td>e2,5 =</td><td>100,11</td></tr> <tr><td>e2,5 =</td><td>-0,70</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	26	n3 =	89	u2,5 =	841,1	p2,5 =	1527,5	e2,5 =	100,11	e2,5 =	-0,70	groupes semblables	
n2 =	26														
n3 =	89														
u2,5 =	841,1														
p2,5 =	1527,5														
e2,5 =	100,11														
e2,5 =	-0,70														
groupes semblables															
Étape 5 : 32-41	<table border="1"> <tr><td>n2 =</td><td>8</td></tr> <tr><td>n3 =</td><td>13</td></tr> <tr><td>u2,6 =</td><td>1074</td></tr> <tr><td>p2,6 =</td><td>1127</td></tr> <tr><td>e2,6 =</td><td>140,24</td></tr> <tr><td>e2,6 =</td><td>-0,55</td></tr> <tr><td colspan="2">groupes semblables</td></tr> </table>	n2 =	8	n3 =	13	u2,6 =	1074	p2,6 =	1127	e2,6 =	140,24	e2,6 =	-0,55	groupes semblables	
n2 =	8														
n3 =	13														
u2,6 =	1074														
p2,6 =	1127														
e2,6 =	140,24														
e2,6 =	-0,55														
groupes semblables															
Étape 6 : 42-44	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,7 =</td><td>230</td></tr> <tr><td>p2,7 =</td><td>455</td></tr> <tr><td>e2,7 =</td><td>68,57</td></tr> <tr><td>e2,7 =</td><td>-2,41</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,7 =	230	p2,7 =	455	e2,7 =	68,57	e2,7 =	-2,41	groupes différents	
n2 =	26														
n3 =	26														
u2,7 =	230														
p2,7 =	455														
e2,7 =	68,57														
e2,7 =	-2,41														
groupes différents															
Étape 7 : 45-48	<table border="1"> <tr><td>n2 =</td><td>33</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,8 =</td><td>8</td></tr> <tr><td>p2,8 =</td><td>360</td></tr> <tr><td>e2,8 =</td><td>31,04</td></tr> <tr><td>e2,8 =</td><td>-1,81</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	33	n3 =	8	u2,8 =	8	p2,8 =	360	e2,8 =	31,04	e2,8 =	-1,81	groupes différents	
n2 =	33														
n3 =	8														
u2,8 =	8														
p2,8 =	360														
e2,8 =	31,04														
e2,8 =	-1,81														
groupes différents															
Étape 8 : 49	<table border="1"> <tr><td>n2 =</td><td>8</td></tr> <tr><td>n3 =</td><td>107</td></tr> <tr><td>u2,9 =</td><td>124</td></tr> <tr><td>p2,9 =</td><td>428</td></tr> <tr><td>e2,9 =</td><td>36,57</td></tr> <tr><td>e2,9 =</td><td>-1,01</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	8	n3 =	107	u2,9 =	124	p2,9 =	428	e2,9 =	36,57	e2,9 =	-1,01	groupes différents	
n2 =	8														
n3 =	107														
u2,9 =	124														
p2,9 =	428														
e2,9 =	36,57														
e2,9 =	-1,01														
groupes différents															
Étape 9 : 50-61	<table border="1"> <tr><td>n2 =</td><td>107</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,10 =</td><td>246</td></tr> <tr><td>p2,10 =</td><td>1471,5</td></tr> <tr><td>e2,10 =</td><td>211,25</td></tr> <tr><td>e2,10 =</td><td>-5,01</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	107	n3 =	26	u2,10 =	246	p2,10 =	1471,5	e2,10 =	211,25	e2,10 =	-5,01	groupes différents	
n2 =	107														
n3 =	26														
u2,10 =	246														
p2,10 =	1471,5														
e2,10 =	211,25														
e2,10 =	-5,01														
groupes différents															
Étape 10 : 62-65	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,11 =</td><td>26</td></tr> <tr><td>p2,11 =</td><td>332</td></tr> <tr><td>e2,11 =</td><td>24,64</td></tr> <tr><td>e2,11 =</td><td>-1,12</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	8	u2,11 =	26	p2,11 =	332	e2,11 =	24,64	e2,11 =	-1,12	groupes différents	
n2 =	26														
n3 =	8														
u2,11 =	26														
p2,11 =	332														
e2,11 =	24,64														
e2,11 =	-1,12														
groupes différents															

Franchises 4															
Étape 1 : 19-20	<table border="1"> <tr><td>n2 =</td><td>30</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,2 =</td><td>62</td></tr> <tr><td>p2,2 =</td><td>207</td></tr> <tr><td>e2,2 =</td><td>74,68</td></tr> <tr><td>e2,2 =</td><td>-5,90</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	30	n3 =	8	u2,2 =	62	p2,2 =	207	e2,2 =	74,68	e2,2 =	-5,90	groupes différents	
n2 =	30														
n3 =	8														
u2,2 =	62														
p2,2 =	207														
e2,2 =	74,68														
e2,2 =	-5,90														
groupes différents															
Étape 2 : 21-23	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,3 =</td><td>261</td></tr> <tr><td>p2,3 =</td><td>420</td></tr> <tr><td>e2,3 =</td><td>68,57</td></tr> <tr><td>e2,3 =</td><td>-1,06</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,3 =	261	p2,3 =	420	e2,3 =	68,57	e2,3 =	-1,06	groupes différents	
n2 =	26														
n3 =	26														
u2,3 =	261														
p2,3 =	420														
e2,3 =	68,57														
e2,3 =	-1,06														
groupes différents															
Étape 3 : 24-27	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,4 =</td><td>488</td></tr> <tr><td>p2,4 =</td><td>511,3</td></tr> <tr><td>e2,4 =</td><td>85,13</td></tr> <tr><td>e2,4 =</td><td>-1,46</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	26	u2,4 =	488	p2,4 =	511,3	e2,4 =	85,13	e2,4 =	-1,46	groupes différents	
n2 =	26														
n3 =	26														
u2,4 =	488														
p2,4 =	511,3														
e2,4 =	85,13														
e2,4 =	-1,46														
groupes différents															
Étape 4 : 28-44	<table border="1"> <tr><td>n2 =</td><td>152</td></tr> <tr><td>n3 =</td><td>44</td></tr> <tr><td>u2,5 =</td><td>1522</td></tr> <tr><td>p2,5 =</td><td>1644</td></tr> <tr><td>e2,5 =</td><td>211,13</td></tr> <tr><td>e2,5 =</td><td>-2,47</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	152	n3 =	44	u2,5 =	1522	p2,5 =	1644	e2,5 =	211,13	e2,5 =	-2,47	groupes différents	
n2 =	152														
n3 =	44														
u2,5 =	1522														
p2,5 =	1644														
e2,5 =	211,13														
e2,5 =	-2,47														
groupes différents															
Étape 5 : 45-49	<table border="1"> <tr><td>n2 =</td><td>44</td></tr> <tr><td>n3 =</td><td>107</td></tr> <tr><td>u2,6 =</td><td>1050</td></tr> <tr><td>p2,6 =</td><td>1254</td></tr> <tr><td>e2,6 =</td><td>246,26</td></tr> <tr><td>e2,6 =</td><td>-1,62</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	44	n3 =	107	u2,6 =	1050	p2,6 =	1254	e2,6 =	246,26	e2,6 =	-1,62	groupes différents	
n2 =	44														
n3 =	107														
u2,6 =	1050														
p2,6 =	1254														
e2,6 =	246,26														
e2,6 =	-1,62														
groupes différents															
Étape 6 : 50-61	<table border="1"> <tr><td>n2 =</td><td>107</td></tr> <tr><td>n3 =</td><td>26</td></tr> <tr><td>u2,7 =</td><td>800</td></tr> <tr><td>p2,7 =</td><td>1471,5</td></tr> <tr><td>e2,7 =</td><td>211,25</td></tr> <tr><td>e2,7 =</td><td>-5,01</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	107	n3 =	26	u2,7 =	800	p2,7 =	1471,5	e2,7 =	211,25	e2,7 =	-5,01	groupes différents	
n2 =	107														
n3 =	26														
u2,7 =	800														
p2,7 =	1471,5														
e2,7 =	211,25														
e2,7 =	-5,01														
groupes différents															
Étape 7 : 62-65	<table border="1"> <tr><td>n2 =</td><td>26</td></tr> <tr><td>n3 =</td><td>8</td></tr> <tr><td>u2,8 =</td><td>26</td></tr> <tr><td>p2,8 =</td><td>332</td></tr> <tr><td>e2,8 =</td><td>24,64</td></tr> <tr><td>e2,8 =</td><td>-1,12</td></tr> <tr><td colspan="2">groupes différents</td></tr> </table>	n2 =	26	n3 =	8	u2,8 =	26	p2,8 =	332	e2,8 =	24,64	e2,8 =	-1,12	groupes différents	
n2 =	26														
n3 =	8														
u2,8 =	26														
p2,8 =	332														
e2,8 =	24,64														
e2,8 =	-1,12														
groupes différents															

Franchises finales	
n2 =	30-30
n3 =	21-27
n2 =	26-44
n3 =	15-49
n2 =	50-61
n3 =	62-65

Lexique

A

AUC correspond à l'aire sous chaque courbe, de l'anglais *Area Under the Curve*. 59, 61, 68, 71

AURA Auvergne-Rhône-Alpes. 13, 14, 27, 34, 52, 54, 60, 61, 70, 71, 73

B

BCAC Bureau Commun d'Assurances des Collectives. 10, 31–33, 36, 44, 45, 72, 73

C

CIF Fonction d'incidence cumulée. 43

CLCC Centre de Lutte Contre le Cancer. 12, 14, 29

E

EHPAD Établissements d'Hébergement pour Personnes Âgées Dépendantes. 28, 34

ESMS Établissements Sociaux et Médico-Sociaux. 14, 34

ESPPS Établissements de Santé Privés et Professionnels de la Santé. 14, 28, 34

F

FEHAP Fédération des Établissements Hospitaliers et d'Aide à la Personne. 14, 34, 46, 52, 54, 69, 70

FHP Fédération Hospitalière Privée. 34, 46, 52, 54, 69, 70, 73

FN Faux négatif. 63

FP Faux positif. 63

G

Grand-Ouest correspond au regroupement des régions Pays de la Loire, Bretagne, Normandie, Nouvelle-Aquitaine et Centre-Val de Loire. 14, 34, 52, 54, 60, 70, 73

H

HAD Hospitalisation à Domicile. 28

I

IDF Île-de-France. 13, 14, 34, 52, 60, 70, 73

K

KM Kaplan-Meier. 40

KP Kalbfleisch et Prentice. 43

M

MCO Médecine, Chirurgie, Obstétrique. 28, 34

N

Nord-Est correspond au regroupement des régions Grand Est, Hauts-de-France et Bourgogne-Franche-Comté. 14, 34, 52, 54, 60, 70, 73

P

PACA Provence-Alpes-Côte d'Azur. 13, 14, 27, 34, 52, 54, 60, 61, 69–71, 73

PMSS Plafond Mensuel de la Sécurité Sociale. 5

R

ROC correspond à la fonction d'efficacité du récepteur, de l'anglais *Receiver Operating Characteristic*.
67

S

SSR Soins de Suite et de Réadaptation. 28, 34

T

TAG Taux d'Absentéisme Global. 20, 52–55, 57

TIG Taux de passage en Invalidité Global. 20

V

VN Vrai négatif. 63

VP Vrai positif. 63