

Mémoire de master :

Modélisation de l'absentéisme sur un portefeuille d'assurance collective

Khalil DATSI



Résumé

L'absentéisme est en constante hausse ces dernières années dans le monde de l'entreprise. Cette notion complexe est très importante pour l'assureur du fait qu'il intervient dans la couverture des arrêts de travail.

Pour pouvoir cerner ce phénomène, il faudra d'abord rappeler le cadre de la protection sociale en France, expliciter les différentes notions du risque arrêt de travail et caractériser l'absentéisme en entreprise via les méthodes de mesure, son coût sur les entreprises et les assureurs.

Nous tacherons ensuite d'étudier l'absentéisme sur notre portefeuille collectif, celui de Generali Vie composé de plus de 600 000 assurés, portefeuille qui évolue année après année. Pour ce faire, nous effectuerons dans un premier temps une analyse descriptive du portefeuille prévoyance en étudiant l'évolution de l'absentéisme sur les dernières années, les différentes caractéristiques des arrêts de travail (saisonnalité, type d'arrêt, rechute, ...) et le calcul de taux d'absentéisme sur le portefeuille.

Ensuite, nous modéliserons l'absentéisme via des modèles linéaires généralisés, pour cela il faudra d'abord rappeler le principe théorique de ces modèles, puis construire une base de données appropriée pour notre modélisation, ensuite modéliser l'absentéisme sur le portefeuille et vérifier la cohérence du modèle.

Remerciement

Je tiens tout d'abord à remercier particulièrement Bertrand FONTAINE mon manager pour m'avoir accueilli au sein de l'équipe grands comptes de Generali Vie, pour sa pédagogie, sa disponibilité et pour toutes les connaissances qu'il m'a transmises.

Je tiens ensuite à remercier chaleureusement mon tuteur Selim BELHOUSSE pour ses précieux conseils pour la réalisation de ce mémoire, son suivi et son aide tout au long de l'année sur l'ensemble des missions de souscription qui m'ont été confiés.

Je remercie également mon tuteur pédagogique Denys POMMERET pour ses conseils avisés, sa disponibilité et son suivi.

J'adresse aussi mes remerciements à l'ensemble de mon équipe pour leur bienveillance, leurs conseils et leur bonne humeur au quotidien.

Enfin je tiens à remercier ceux qui sont pour moi ma motivation, mes parents, pour leur amour, leur soutien indéfectible et leurs encouragements depuis toujours.

Table des matières

Résumé	2
Remerciement	3
Introduction	6
Le risque arrêt de travail dans le contexte de l'assurance collective	7
Présentation de la protection sociale en France : régime de base et régime complémentaire.	7
Histoire de la protection sociale.....	7
Organisation fonctionnelle	8
Détails de l'offre proposée par le régime de base.....	8
Panorama de l'offre de couverture complémentaire.....	12
Qu'est-ce que la prévoyance collective ?	14
Le risque arrêt de travail en France	15
Contexte général	15
Environnement juridique.....	16
L'absentéisme en entreprise en France	18
Qu'est-ce que l'absentéisme ?	18
Quelles sont les causes de l'absentéisme ?	18
Comment mesurer l'absentéisme en entreprise ?	20
Quel est son coût réel sur les entreprises ?	22
Etude de la base de données :	23
Présentation de la base données	23
Extraction de la base de données.....	23
Quelle est la structure de la base de données d'étude ?	23
Traitement de la donnée	25
Ciblage sur le risque arrêt de travail	25
Quels choix a-t-on effectué pour le nettoyage de la base de données ?	25
Variables d'étude à calculer	26
Durée de l'arrêt de travail.....	26
Taux d'absentéisme	26
Analyse descriptive	27
Evolution du risque arrêt de travail sur les dernières années	27
Evolution globale du portefeuille arrêt de travail	27
Saisonnalité des arrêts de travail et mise en valeur d'un effet COVID sur 2020.....	28
Effet COVID sur l'année 2020	30
Evolution des types d'arrêts par année de survenance.....	30
Aspect théorique des modèles linéaires généralisés.....	32

Intérêt des modèles linéaires généralisés.....	32
Principe des GLM.....	33
Composant d'un GLM :	33
Modèle gaussien :	34
Modèle de Poisson :	35
Modèle Tweedie :	36
Validation d'un modèle GLM :	37
Validation croisée (k-fold)	37
Validation de la significativité globale du modèle	38
Validation de la significativité individuelle des coefficients de la régression.....	39
Etude des résidus	40
Conclusion – Pistes pour la suite	41

Introduction

Entre 2014 et 2018 l'absentéisme en entreprise a progressé de 39% selon Rehalto, cette problématique touche fortement les assureurs qui interviennent dans la couverture des arrêts de travail. Il est donc important pour les assureurs de comprendre ce risque, de le mesurer et de le prévenir.

Après avoir rappelé le cadre de la protection sociale en France, du risque arrêt de travail et l'absentéisme en entreprise et nous tacherons d'étudier l'absentéisme sur le portefeuille collectif de Generali composé de plus de 600 000 assurés, en effectuant dans un premier temps une analyse descriptive du portefeuille prévoyance puis en modélisant l'absentéisme via des modèles linéaires généralisés.

L'enjeu de ce mémoire est de réussir à modéliser l'absentéisme avec les contraintes inhérentes à l'assurance collective. En effet, l'objectif principal est d'introduire un maximum d'outils et de notions pouvant servir à l'avenir à l'établissement d'un baromètre de l'absentéisme sur notre portefeuille. Cette modélisation est d'autant plus importante que le risque arrêt de travail prend de plus en plus de place dans notre société parmi les risques à considérer par l'assureur.

Le risque arrêt de travail dans le contexte de l'assurance collective

Dans le cadre de l'étude, nous avons pour objectif d'étudier et de modéliser l'absentéisme sur le portefeuille collectif de Generali Vie. Pour introduire ce vaste sujet, il est naturel de tout d'abord présenter le système de protection sociale français, présentation qui nous amènera à aborder les spécificités de la prévoyance collective, ensuite à l'aide d'une définition détaillée du concept d'absentéisme en entreprise et des notions actuarielles autour de l'arrêt de travail, nous expliciterons les enjeux et l'importance de l'établissement d'un baromètre de l'absentéisme sur un portefeuille d'assurance collective.

Présentation de la protection sociale en France : régime de base et régime complémentaire.

La protection sociale désigne l'ensemble des mécanismes qui permettent aux individus de faire face financièrement à certains risques sociaux, à des situations pouvant provoquer une baisse des ressources ou une hausse des dépenses. Ces risques sont les suivants :

- Santé : actes courants, hospitalisation, optique et dentaire : remboursement des frais médicaux
- Arrêt de Travail : incapacité, invalidité, maternité : remboursement d'indemnités journalières pour l'incapacité et la maternité et d'une rente pour l'invalidité
- Décès : versement d'un capital ou d'une rente aux ayants droits
- Retraite : constitution d'un capital ou d'une rente
- Dépendance : versement d'un capital ou d'une rente
- Chômage : versement d'indemnités de chômage

Histoire de la protection sociale

La Sécurité Sociale a été créée en 1945 via les ordonnances du 4 et 19 octobre en fusionnant toutes les anciennes assurances et avec comme triple objectif :

- Unité de la Sécurité Sociale
- Généralisation quant aux personnes
- Extension des risques couverts

L'ordonnance du 4 octobre 1945 prévoit un système coordonné de caisses se substituant à de multiples organismes existants. Les salariés des régimes spéciaux comme les fonctionnaires, les marins, les cheminots, etc vont refuser de s'intégrer dans le nouveau régime général et conserver dans un cadre « transitoire » qui dure encore, leurs régimes spécifiques.

Régime : Ensemble de droits et obligations réciproques des Employés (et leurs « ayant-droit »), des Patrons, et d'une Caisse de Sécurité Sociale.

Les trois grands régimes introduits par les ordonnances des 4 et 19 octobre 1945 sont les suivants :

- Régime général : salariés et travailleurs assimilés à des salariés soit environ 80 % de la population.

- Régime des travailleurs non-salariés non agricoles ou Régime social des indépendants ou RSI : artisans, commerçants et professions libérales.
- Régime agricole : exploitants et salariés agricoles, ainsi que certains secteurs rattachés à l'agriculture (ex l'industrie agro alimentaire). Il est géré par la caisse centrale de la Mutualité Sociale Agricole, MSA.

La protection sociale était à l'origine financée exclusivement par des cotisations. Le financement s'est diversifié dans les années 1990 avec la CSG et d'autres impôts.

Organisation fonctionnelle

D'un point de vue fonctionnel, l'organisation actuelle de la Sécurité Sociale résulte de l'ordonnance de 1967 qui instaure la séparation de la Sécurité Sociale en 4 branches autonomes. Chaque branche est alors responsable de ses ressources et de ses dépenses.

- Branche Maladie : maladie, maternité, paternité, invalidité, décès
- Branche Accidents du travail et Maladies professionnelles
- Branche Vieillesse et veuvage
- Branche Famille : handicap, logement, RSA ...

Détails de l'offre proposée par le régime de base

L'assurance Maladie :

Définition de l'assurance maladie :

L'assurance maladie est définie dans le cadre de l'article L321-1 du CSS :

« L'assurance maladie comporte (...) l'octroi d'indemnités journalières à l'assuré qui se trouve dans l'incapacité physique constatée par le médecin traitant, (...) de continuer ou de reprendre le travail ; l'incapacité peut être également constatée, dans les mêmes conditions, par la sage-femme dans la limite de sa compétence professionnelle et pour une durée fixée par décret (...) ».

Conditions d'ouverture des droits de l'assurance maladie :

Des conditions d'ouverture des droits s'appliquent en fonction de la durée de l'arrêt de travail et de la situation de l'assuré :

- Si l'arrêt de travail est inférieur à 6 mois :
 - Avoir travaillé au moins 150 heures au cours des trois mois précédant l'arrêt de travail
 - Ou avoir cotisé sur un salaire au moins égal à 1 015 fois le montant du SMIC horaire au cours des six mois précédant l'arrêt
- Si l'arrêt de travail est supérieur à 6 mois :
 - Justifier à la date de l'arrêt, de douze mois d'immatriculation en tant qu'assuré social auprès de l'Assurance Maladie
 - Avoir travaillé au moins 600 heures au cours des douze derniers mois
 - Ou avoir cotisé sur un salaire au moins égal à 2 030 fois le montant du SMIC horaire au cours des douze mois précédant l'arrêt de travail, dont au moins 1 015 fois le montant du SMIC horaire au cours des six premiers mois

Les indemnités journalières :

- Durée de service des indemnités journalières :

L'indemnité journalière est accordée à partir d'un délai de carence de 3 jours (1 jour pour les fonctionnaires), ce délai de carence ne s'applique pas dans les cas suivants :

- En cas de reprise d'activité entre deux prescriptions d'arrêt de travail qui ne dépasse pas 48 heures
- L'assuré est en affection de longue durée et les arrêts sont en rapport à cette maladie alors le délai de carence ne s'applique que pour le premier arrêt.

L'indemnité journalière est due au titre de chaque jour, ouvrable ou non.

Les causes de sorties de l'état d'incapacité sont : le rétablissement, le passage en invalidité et le décès.

- Montant de l'indemnité journalière :

Le montant de l'indemnité journalière versée par la Sécurité Sociale est calculé via le salaire journalier de base dont le calcul est détaillé ci-dessous et est égale à 50% du salaire journalier de base.

Salaire journalier de base : ce salaire est la moyenne des salaires bruts des 3 derniers mois travaillés précédant l'arrêt, plafonné à 1.8 SMIC (3 022,11 € au 1^{er} août 2022), c'est-à-dire :

$$\text{Salaire journalier de base} = \min \left(\frac{\text{Somme des salaires bruts des 3 derniers mois}}{91,25}; 1,8 * \text{SMIC} \right) \quad (\text{Décret n° 2010-1305 -29/10/10})$$

L'indemnité peut notamment être revalorisée en cas d'augmentation générale des salaires de l'entreprise durant l'arrêt.

L'assurance Maternité et congé Paternité

Conditions d'ouverture des droits de l'assurance maternité/paternité :

Les conditions d'ouverture des droits à l'assurance Maternité et le congé Paternité sont les suivantes :

- Pour l'assurance maternité :
 - Être immatriculée en tant qu'assurée sociale depuis au moins 10 mois, à la date prévue de l'accouchement
 - Avoir cotisé sur un salaire au moins égal à 1 015 fois la valeur du SMIC horaire au cours des six mois civils précédant la date du début de la grossesse ou du début
 - Ou avoir effectué au moins 150 heures de travail au cours des trois mois précédant l'arrêt de travail, à la date du début de la grossesse ou du début du congé prénatal
- Pour le congé paternité :
 - Justifier d'au moins 10 mois d'immatriculation en tant qu'assuré social à la date de début du congé
 - Avoir travaillé au moins 150 heures au cours des trois mois précédant la date de début du congé ou avoir cotisé sur un salaire au moins égal à 1 015 fois le montant du SMIC horaire au cours des six mois précédant le début du congé

Les indemnités journalières :

- Durée de service des indemnités journalières :

Le niveau de l'indemnité journalière due au titre du congé maternité/paternité est d'abord limité par la durée légale du congé :

- Congé Maternité : 16 semaines (peut être plus longue selon le nombre d'enfants de l'assuré et le nombre d'enfants attendus)

- Congé Paternité : 11 jours consécutifs maximum, 18 en cas de naissance multiple, ce congé est à prendre par le père dans les 4 mois suivants la naissance

Il n'y a aucun délai de carence.

- Montant de l'indemnité journalière :

- L'indemnité journalière Maternité : moyenne des salaires nets des prélèvements légaux et conventionnels et de la CSG (déduction forfaitaire 21%), limités à la Tranche A (= 1PMSS : 3 428 € en 2022) des trois mois précédant le congé prénatal.
- L'indemnité journalière Paternité : moyenne des salaires nets des prélèvements légaux et conventionnels et de la CSG, limités à la Tranche A des trois mois précédant le congé prénatal.

L'assurance Invalidité

Définition l'assurance invalidité :

L'assurance invalidité est définie dans le cadre des articles suivants comme étant le droit d'avoir accès à une pension lorsque l'assuré présente une invalidité réduisant sa capacité de travail ou de gains dans des proportions déterminées :

Article L341-1 du CSS : « L'assuré a droit à une pension d'invalidité lorsqu'il présente une invalidité réduisant dans des proportions déterminées sa capacité de travail ou de gain, c'est-à-dire le mettant hors d'état de se procurer un salaire supérieur à une fraction de la rémunération soumise à cotisations et contributions sociales qu'il percevait dans la profession qu'il exerçait avant la date de l'interruption de travail suivie d'invalidité ou la date de la constatation médicale de l'invalidité. »

Article R341-2 du CSS : « Pour l'application des dispositions de l'article L. 341-1 :

- 1) L'invalidité que présente l'assuré doit réduire au moins des 2/3 sa capacité de travail ou de gain
- 2) Le salaire de référence ne doit pas être supérieur au tiers de la rémunération normale mentionnée audit article. »

Article L341-4 du CSS : « 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. »

Conditions d'ouverture des droits de l'assurance Invalidité :

Les conditions d'ouverture des droits à l'assurance Invalidité sont les suivantes :

- Être immatriculé depuis au moins 12 mois au moment de l'arrêt de travail suite à l'invalidité ou au moment de la constatation de l'invalidité par le médecin conseil de la caisse d'Assurance Maladie ;
- Justifier, au cours des 12 mois qui précèdent l'arrêt de travail pour invalidité ou constatation médicale de l'invalidité,
 - Soit avoir effectué au moins 600 heures de travail salarié ;
 - Soit avoir cotisé sur un salaire au moins égal à 2 030 fois le SMIC horaire.

La pension d'invalidité :

- Durée de service des indemnités journalières :

La pension invalidité est interrompue en totalité ou en partie en cas de reprise du travail, sinon elle prend fin à l'âge de la retraite prévu par la loi, elle est alors remplacée par la pension de vieillesse allouée en cas d'inaptitude au travail.

- Montant de la pension invalidité :

Le montant de la pension d'invalidité versée par la Sécurité Sociale est calculé via le salaire annuel moyen égal à la moyenne des salaires annuels bruts plafonnés à la tranche A des dix meilleures années d'activité. Le montant de la pension d'invalidité est le suivant :

- 1ère catégorie = 30 % du salaire annuel moyen
- 2ème catégorie = 50 % du salaire annuel moyen
- 3ème catégorie = 50 % du salaire annuel moyen majoré de l'allocation tierce personne

L'assurance Décès

Définition et condition d'ouverture des droits de l'assurance Décès :

D'après l'article L361-1 du CSS, l'assurance décès garantit aux ayants droit de l'assuré le paiement d'un capital égal à un multiple du gain journalier de base lorsque l'assuré exerçait une activité salariée moins de trois mois avant le décès, s'il était titulaire d'une pension d'invalidité ou d'une rente allouée en vertu de la législation sur les accidents du travail et maladies professionnelles ou lorsqu'il bénéficiait, au moment du décès, du maintien de ses droits à l'assurance décès.

Montant du capital décès :

Depuis la loi de financement de la sécurité social de 2015, il est égal à un montant forfaitaire fixé par décret et revalorisé chaque année (3 539€ en 2022).

L'assurance Accident du travail et Maladie Professionnelle

Définitions :

D'après l'article L411-1 du CSS, est considéré comme accident du travail un accident survenu par le fait ou à l'occasion du travail à toute personne salariée ou travaillant, à quelque titre ou en quelque lieu que ce soit, pour un ou plusieurs employeurs ou chefs d'entreprise. De plus, est aussi considéré comme accident du travail l'accident qui est survenu à un travailleur, pendant le trajet d'aller et de retour, entre :

- Tout lieu où le travailleur se rend de façon habituelle (résidence principale, résidence secondaire, ...) et le lieu de travail.
- Le lieu où le travailleur prend habituellement ses repas dans la mesure où le parcours n'a pas été interrompu ou détourné pour un motif dicté par l'intérêt personnel et étranger aux nécessités essentielles de la vie courante ou indépendant de l'emploi et le lieu de travail.

Pour ce qui est des maladies professionnelles, d'après l'article L411-1, est présumée d'origine professionnelle toute maladie désignée dans un tableau de maladies professionnelles et contractée dans les conditions mentionnées à ce tableau.

Aucune condition n'est formulée pour l'ouverture des droits.

Les indemnités journalières :

- Délai de carence :

Aucun délai de carence n'est applicable : l'indemnité est due à partir du 1er jour suivant l'arrêt du travail, pendant toute la période d'incapacité de travail qui précède soit la guérison complète, soit la consolidation de l'état ou le décès.

- **Montant de l'indemnisation :**

Ce montant est calculé via le salaire journalier de base dont le calcul a été détaillé précédemment, le montant de l'indemnité est le suivant :

- 60 % du salaire journalier de base jusqu'au 28ème jour
- 80 % du salaire journalier de base à partir du 29ème jour d'arrêt

De plus, une des spécificités de cette assurance est le taux d'incapacité permanente, taux qui est proposé par un médecin conseil lors d'un contrôle médical organisé par la caisse d'Assurance Maladie de l'assuré.

Le montant de la prestation est alors dicté par ce taux :

- Si taux < 10% : Un capital est versé en fonction du taux d'incapacité de la victime sur la base d'un barème forfaitaire.
- Si taux ≥ 10% : Versement d'une rente viagère égale au salaire annuel multiplié par le taux d'incapacité qui peut être réduit ou augmenté en fonction de la gravité de celle-ci

Panorama de l'offre de couverture complémentaire

Organisation du marché de l'assurance complémentaire

Le marché de l'assurance complémentaire est composé de quatre acteurs principaux :

- **Les organismes assureur :** mutuelle, institut de prévoyance, société d'assurance. Ces organismes sont régis par des réglementations différentes et ont un mode de fonctionnement qui leur est propre et chaque organisme à son domaine d'intervention privilégié.
- **Les courtiers :** c'est un travailleur indépendant qui n'est pas lié à une compagnie d'assurance, il est un intermédiaire au contrat.
- **Les gestionnaires :** soit rattachés de près ou de loin à un cabinet de courtage, soit indépendants, ces entreprises sont dédiées à la gestion administrative des contrats de Prévoyance, Santé, contrats Emprunteurs, ...
- **Les entreprises de conseil :** elles portent assistance à l'entreprise et possède des domaines de compétences et d'interventions souvent variés (Retraite, Santé, Prévoyance, ...)

Les garanties de prévoyance complémentaire

Dans un premier temps, il faut distinguer les deux secteurs principaux du périmètre de la protection sociale complémentaire en France, ces deux secteurs sont : L'épargne retraite et la Prévoyance.

Les garanties collectives d'un tel contrat ont objectif de prévoir pour les salariés, les anciens salariés et leurs ayants droit, la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité, des risques d'incapacité de travail ou d'invalidité, des risques d'inaptitude et du risque chômage, ainsi que la constitution d'avantages sous forme de pensions de retraite, d'indemnités ou de primes de départ en retraite ou de fin de carrière.

- **Les Frais de Santé**

La garantie complémentaire frais de santé intervient en complément de la Sécurité Sociale, voire même dans certains cas lorsque la Sécurité Sociale n'intervient pas.

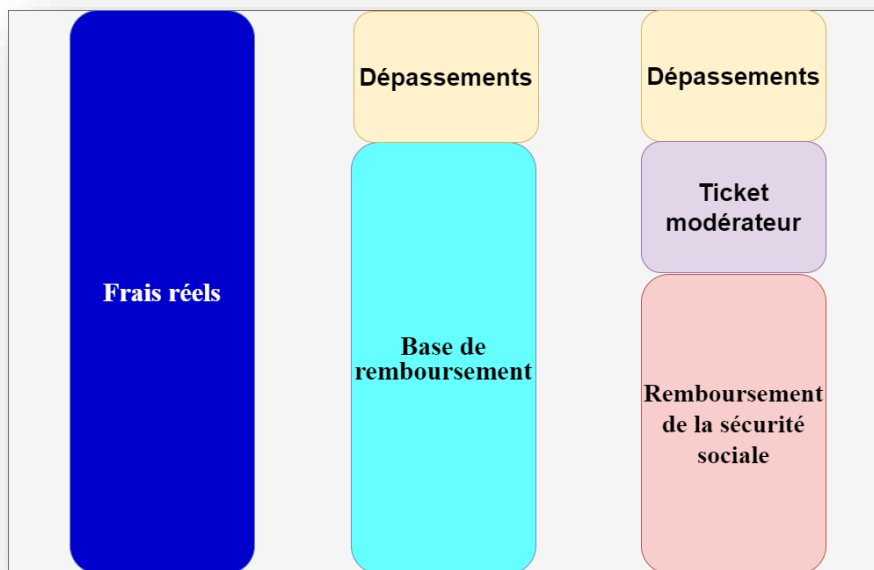


Figure 1 : Intervention de la Sécurité Sociale sur les frais de santé

- La Prévoyance Lourde

- Capitaux Décès :

Le capital décès est versé par la complémentaire en cas de décès de l'assuré mais peut aussi être versée par anticipation en cas d'IAD (Invalidité Absolue et Définitive). Son montant dépend de la rémunération annuelle brute de l'assuré.

Ce capital est parfois accompagné de garanties annexes comme une garantie frais d'obsèques, majoration en cas décès accidentel, garantie « double effet » (en cas de décès du conjoint survivant, versement d'un capital aux enfants à charge), capital « pré-décès », ...

- Rente de Conjoint :

La rente de Conjoint est destinée à compenser la disparition d'un revenu immédiat et/ou différé au sein de la famille, elle peut être de nature viagère ou temporaire dans l'attente de la pension de réversion des régimes de retraite obligatoire.

Elle est généralement calculée soit en fonction des droits de réversion acquis auprès des régimes de retraite complémentaire AGIRC-ARRCO, soit en fonction du dernier salaire de l'assuré décédé.

- Rente éducation :

La rente d'éducation est destinée à assurer le versement d'une rente aux enfants à charge de l'assuré décédé, au-delà d'un certain âge, elle peut être conditionnée à la poursuite des études par le bénéficiaire.

Le montant de la rente est généralement calculé au prorata du dernier salaire de l'assuré et est généralement exprimée par palier d'âge croissant du bénéficiaire.

- Incapacité de travail :

La garantie incapacité de travail permet au salarié en arrêt de travail de percevoir des indemnités journalières afin de compenser sa perte de salaire, ces indemnités viennent en complément de la Sécurité Sociale ainsi que du complément de revenu versé par l'employeur. Les principaux paramètres définissant une garantie incapacité sont :

- La franchise : type et durée (ferme ou continue, rétroactive, discontinue)
- Le montant : en pourcentage du salaire de l'assuré
- L'assiette de calcul

L'employeur est tenu de verser un complément d'indemnisation appelé « Maintien de salaire » sous certaines conditions à un salarié en arrêt de travail. Cette obligation a été mise en place par la loi de mensualisation de 1978 puis améliorée par l'ANI du 11 janvier 2008. Les conditions pour bénéficier du maintien de salaire pour un assuré sont :

- Condition d'ancienneté : le salarié doit avoir au moins 1 an d'ancienneté
- Franchise : 7 jours sauf en cas d'arrêt consécutif à un arrêt de travail ou une maladie professionnelle
- L'Employeur doit compléter les indemnités versées par la Sécurité Sociale à hauteur de 90 %, puis de 66,66 % de la rémunération brute que le salarié aurait gagnée s'il avait continué à travailler, et ce pour les durées définies ci-dessous :

Ancienneté	90% du salaire brut	66,6% du salaire brut
1 – 6 années	30 jours	30 jours
6 – 11 années	40 jours	40 jours
11 – 16 années	50 jours	50 jours
16 – 21 années	60 jours	60 jours
21 – 26 années	70 jours	70 jours
26 – 31 années	80 jours	80 jours
Plus de 31 années	90 jours	90 jours

➤ Invalidité :

La rente d'invalidité compense, en totalité ou en partie, la perte de revenu d'un assuré déclaré invalide. Elle vient compléter la pension d'invalidité versée par la Sécurité sociale. Le montant de cette rente peut-être fixe ou bien correspondre à un pourcentage de son salaire, elle dépend souvent aussi de la catégorie d'invalidité définie par la Sécurité sociale.

Qu'est-ce que la prévoyance collective ?

Dans le cadre de cette étude, nous avons la possibilité d'étudier les caractéristiques en matière d'absentéisme d'un portefeuille de prévoyance collective, celui de Generali Vie. La « prévoyance collective » est la couverture d'un groupe de personnes qui ont un lien objectif entre elles, ce groupe étant représenté par une personne qui va signer ce contrat, généralement l'entreprise dans laquelle travail un assuré.

Les différents intervenants dans un contrat de prévoyance collective sont le souscripteur, les affiliés, les assurés et les bénéficiaires.

- **Le souscripteur** : la personne morale qui signe le contrat et paie les cotisations (entreprise, association, établissement de crédit)
- **Les affiliés** : l'ensemble des personnes appartenant au groupe assurable
- **Les assurés** : l'ensemble des personnes soumises au risque
- **Les bénéficiaires** : l'ensemble des personnes susceptibles de recevoir des prestations (exemple : les enfants dans le cadre d'une rente éducation)

Le risque arrêt de travail en France

Contexte général

Pour l'assureur, la définition des prestations et de son intervention dans la prise en charge, est alignée sur la reconnaissance par la Sécurité Sociale d'un arrêt de travail vue précédemment. Les règles de base qui vont donc conduire à définir ce qu'est un arrêt de travail, avec la distinction entre l'incapacité et l'invalidité, sont donc celles de la Sécurité Sociale. On désigne en pratique par « incapacité » l'incapacité temporaire de travail et par « invalidité » l'incapacité permanente de travail.

✂ *Que désigne-t-on par la notion de risque arrêt de travail ?*

Ce que nous désignerons dans cette étude par la notion de « risque arrêt de travail » constitue en pratique pour l'assureur l'association du risque incapacité et du risque invalidité. Cette association est nécessaire car l'incapacité temporaire de travail peut devenir permanente, soit par aggravement du facteur incapacitant, soit après la durée maximum légale de l'incapacité (36 mois), d'ailleurs cela constitue un réel enjeu pour l'actuaire qui doit prendre en compte la probabilité de passage en invalidité pour le calcul des provisions ITIP via les tables de maintien en incapacité.

Malgré qu'en pratique il aurait été plus simple de ne considérer qu'un seul type d'arrêt, ces deux risques sont très différents notamment via leur statut clairement et leur durée définis par la loi.

Parmi les différents types d'arrêts, on distingue les maladies ordinaires, que l'on classera dans la partie dédiée à l'absentéisme parmi les arrêts difficilement évitable, et le reste des types d'arrêts que nous trouverons dans le graphique ci-dessous issu du baromètre sur les arrêts de travail établie par REHALTO en 2019 et intitulé « Comprendre pour agir » :

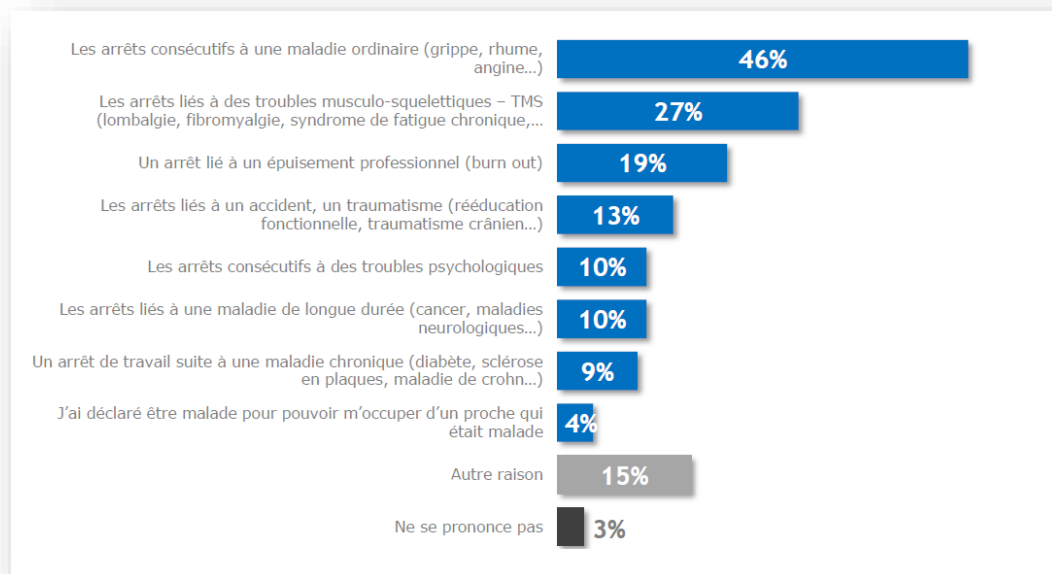


Figure 2 : Proportion des types d'arrêts connu par des salariés en entreprise

Environnement juridique

L'environnement juridique autour de la notion du risque arrêt de travail est défini par les textes juridiques suivants :

- **Loi Evin (31 décembre 1989) :**

Son titre exact est : « Loi 89-1009 du 31 décembre 1989 renforçant les garanties aux personnes assurées contre certains risques »

C'est la première loi spécifique de Prévoyance, elle crée le premier ensemble de règles applicables à toutes les familles d'assureurs (les institutions de prévoyances, les mutuelles, les assureurs). Elle impose les trois points suivants :

- Évaluation des provisions techniques au niveau atteint.
- Reprise des encours (et évaluation des provisions pour revalorisation).
- Maintien des garanties décès aux invalides et aux incapables (décret du 17 juillet 2001).

Cette loi oblige l'ensemble des familles d'assureurs à provisionner les sinistres en cours pour les couvertures du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité, des risques d'incapacité de travail, d'invalidité et du risque chômage.

- **Loi du 8 août 1994 :**

Cette loi arriva en complément de la loi Evin, dans le but de transposer dans le droit français une directive européenne en matière d'assurance, cette loi crée le premier corps de règles relatives à la protection sociale des salariés.

Cette loi pose les fondements juridiques de toute couverture collective, elle impose notamment une période de réexamen d'au plus 5 années pour les conventions collectives de branche ou

d'entreprise qui désigne un organisme assureur et, en complément à la loi Evin, elle ajoute l'obligation d'organiser la poursuite de la revalorisation des prestations en cours de service en cas de résiliation du contrat ainsi que le maintien de la garantie décès.

▪ **Arrêté du 28 mars 1996 :**

Cet arrêté a pour objectif la définition de règles sur l'évaluation des provisions mathématiques d'inventaire sur les risques incapacité et invalidité :

- Concernant le provisionnement incapacité : une provision de l'incapacité en cours doit être complétée par une provision de passage en invalidité ou encore appelée provision d'invalidité en attente.
- Pour l'invalidité, une provision mathématique d'invalidité en cours doit être constituée
- Les provisions mathématiques doivent être calculées en utilisant des tables réglementaires de maintien du BCAC, ou d'une table certifiée et en utilisant un taux actuariel inférieur à 75% du TME, majoré par 4,5%

L'absentéisme en entreprise en France

Qu'est-ce que l'absentéisme ?

L'absentéisme est une notion dont la définition n'est pas unanime selon les sources et diffère en fonction de la nature de l'absence, selon l'encyclopédie Larousse, l'absentéisme est le fait d'être absent du lieu de travail ou de tout lieu où pour des raisons de travail, de participation à une action ou autre, la présence est obligatoire, en entreprise on distingue les absences non évitables comme par exemple les congés maternité, pour événements familiaux, etc, et les arrêts évitables qui dépendrait autant de comportements personnels que de facteurs sociaux.

Selon le rapport de l'ANACT, l'absentéisme en entreprise désigne toute absence qui aurait pu être évitée par une prévention suffisante, faite en amont sur les conditions de travail au sens large, c'est-à-dire les conditions qui concernent les conditions physiques de travail (position de travail, matériel adapté à disposition du salarié, ...), l'organisation du travail, la conciliation des temps professionnel et privé, et tout autre facteur pouvant améliorer la qualité de vie au travail. Selon ce rapport, les absences comme les congés de formation ou maternité font parti des absences qui ne sont pas de l'absentéisme. En bref, le choix de ce que l'on désigne ou pas comme de l'absentéisme appartient en pratique au responsable qui choisit ce que l'on considère ou non dans le calcul du taux d'absentéisme, qui est un indicateur que l'on détaillera dans la suite.

Quelles sont les causes de l'absentéisme ?

Les causes de l'absentéisme sont nombreuses et de nature très variées, en passant par des causes fréquentes et évidentes comme des conditions de travail pénibles ou inconfortable (ex : port de charge lourdes, mauvais matériel,...) et des causes un peu plus cachés comme l'ambiance au travail ou les exigences du travail qui peuvent avoir un impact psychologique important sur les salariés et ainsi être la source de troubles psychosociaux, nous allons à présent tenter de présenter certaines causes importante d'absentéisme en entreprise :

- **Les conditions matérielles de travail :** il faut dans un premier temps considérer les conditions matérielles de réalisation du travail comme étant une cause très importante de l'absentéisme en entreprise, cela passe par la configuration et l'état des locaux de travail, les horaires de travail, ces différents éléments augmentent la charge physique des salariés et sont souvent une des causes d'apparition de troubles comme les troubles musculo squelettiques ou psychosociaux. Hormis les maladies ordinaires, les troubles musculo squelettiques (TMS) et les troubles psychosociaux sont les causes principales d'arrêt de travail en France, selon Rehalto, respectivement 27% et 29%. Les TMS sont à l'origine de gênes et de douleurs dans la vie quotidienne et dans le travail, ils sont qualifiés par l'OMS comme des « maladies liées au travail » (« work-related diseases ») car la nature, le milieu et les conditions de travail ont une part importante dans l'étiologie des maladies multifactorielles. En effet, on identifie dans la littérature de nombreux facteurs d'origine professionnelle et personnelle, dont des facteurs non modifiables comme l'âge, le sexe ou la taille, et d'autres facteurs que l'on considérera évitable c'est-à-dire que l'on pourrait atténuer par de la prévention comme la consommation de tabac, la sédentarité ou la posture adoptée au travail.

L'entreprise a une responsabilité quant à la prévention des facteurs de risques de TMS de nature professionnelle ce sont en général des facteurs de risques biomécanique comme des mouvements trop répétitifs, des vibrations ou encore des séquelles dues au port de charges lourdes et des facteurs de risques psychosociaux.

Les risques psychosociaux sont réputés pour favoriser certaines pathologies de mal-être au travail comme par exemple le burn out qui est synonyme d'épuisement lié à une charge de travail et d'une pression trop intense. La notion de risques psychosociaux est vaste et concerne l'ensemble des éléments inhérent impactant la charge mentale de l'employé comme les situations de stress, les violences internes ou externes à l'entreprise qui peuvent générer un mal-être de l'employé au travail puis dérivé sur des pathologies menant à de l'absentéisme.

- **L'engagement et le désengagement au travail :** Les causes de l'absentéisme ne sont pas uniquement dû à la santé physique ou psychologique des salariés, l'engagement au travail définit par LAWLER E.E. et HALL D.T. dans leur ouvrage *Journal of Applied Psychology* comme étant : « le degré selon lequel une personne perçoit son travail comme étant une partie importante de sa vie et de son identité, grâce aux opportunités qu'il offre de satisfaire des besoins importants » joue un rôle très important dans la vie d'une entreprise car il traduit la santé de la relation entre une personne et son employeur, particulièrement son degré d'implication dans son travail.

Comment mesurer l'absentéisme en entreprise ?

L'absentéisme comporte de multiples aspects qui s'entrecroisent : administratifs, économiques, sociaux et sanitaires. Il n'y a pas de cause unique à l'absentéisme des salariés. L'absentéisme est un révélateur du fonctionnement de l'organisation, de l'attrait que celle-ci recèle pour les salariés (en favorisant l'engagement au travail), mais aussi de l'état de santé global d'une population et de ses caractéristiques (âge et genre, etc.). Dans tous les cas, il est nécessaire, avant d'agir, de procéder à un bon diagnostic de la situation. Ensuite, un débat nourri entre les directions, l'encadrement, le personnel et ses représentants sera indispensable pour mettre en place un programme d'action efficace.

Pour pouvoir effectuer des actions de prévention sur l'absentéisme en entreprise, il faut d'abord être en mesure de le quantifier et de le caractériser.

Pour caractériser de l'absentéisme, il faut être en mesure d'identifier la nature des absences, la durée, la fréquence, potentiellement la rechute, etc.

Nous définissons ensuite les indicateurs que nous pouvons utiliser en pratique pour quantifier l'absentéisme en entreprise.

Le taux d'absentéisme est un indicateur fréquemment utilisé pour quantifier l'absentéisme, c'est le rapport entre nombre de jours arrêté sur une année et le nombre de jours de travail total :

$$\text{Taux d'absentéisme} = \frac{\text{Nombre de jours d'absence total sur la période d'étude}}{\text{Nombre de jours sur la période d'étude}}$$

	Définition	Rôles
Nombre de jours d'arrêts	Volume total des jours d'arrêts sur l'année	<ul style="list-style-type: none"> ▪ Prise de conscience des volumes ▪ Identification des « masses » → focaliser l'action
Taux d'absentéisme	Part des jours de l'année passés en arrêt de travail sur toute la population	<ul style="list-style-type: none"> ▪ Pilotage global ▪ Comparaison à la concurrence
Prévalence	Part des salariés ayant eu au moins un arrêt dans l'année	<ul style="list-style-type: none"> ▪ Indicateur du contexte socio-économique d'une entreprise et des risques psychosociaux
Fréquence	Nombre moyen d'arrêts par salarié dans l'année	<ul style="list-style-type: none"> ▪ Analyse et comparaison des populations entre elles ▪ Analyse et comparaison des arrêts par durée
Durée moyenne annuelle	Nombre moyen de jours d'arrêt dans l'année pour un salarié arrêté au moins une fois	<ul style="list-style-type: none"> ▪ Indicateur notamment de la santé des salariés et du rôle de l'âge

Quel est son coût réel sur les entreprises ?

L'absentéisme engendre des coûts complexes en pratique pour l'entreprise. Il existe deux types de coûts selon qu'ils soient directs ou indirects :

- **Les coûts directs :**

Ce sont les plus simples à évaluer, puisqu'ils sont constitués principalement des montants de salaires et des cotisations sociales associés à la période d'absence d'un salarié. Ainsi, il s'agit de considérer le complément patronal durant l'absence du salarié par rapport aux indemnités journalières de la Sécurité sociale, et le maintien du salaire pendant le délai de carence. Ce coût dépend essentiellement de la politique RH de l'entreprise et de la convention collective : en effet, une entreprise peut décider si oui ou non elle souhaite maintenir le salaire d'un employé pendant le délai de carence en cas de maladie lorsque son ancienneté est inférieure à un an.

- **Les coûts indirects**

Il existe certains coûts, dits *indirects*, reliés principalement à des risques sociaux engendrés par l'absence d'un salarié susceptibles de faire augmenter significativement le coût de l'absentéisme. Parmi ceux-ci peuvent être nommés :

- ✎ **Les coûts de remplacement** : un salarié absent peut être remplacé par un autre salarié (CDD, intérim,...) ce qui engendrera un coût supplémentaire à l'entreprise.
- ✎ **Les coûts de gestion** : la gestion des absences reste un processus long et coûteux pour les ressources humaines.
- ✎ **Les coûts liés aux dysfonctionnements organisationnels** : une absence provoque effectivement un dysfonctionnement dans son équipe puisque les tâches qui lui reviendraient en temps normal doivent être redistribuées à d'autres salariés. A ce titre, elles peuvent être redistribuées à des collègues, ce qui pourra engendrer des heures supplémentaires, ou elles peuvent être relayées à un remplaçant qu'il faudra alors former.
- ✎ **Les coûts d'improductivité** : Une personne remplaçant un salarié absent aura besoin de temps pour se former aux différentes tâches qui lui seront confiées, et seront moins efficaces que le salarié remplacé.
- ✎ **Les coûts sociaux** : les dépenses et la réorganisation du travail engendrées par une absence peuvent venir dégrader le climat social puisque les salariés présents peuvent voir leur charge de travail augmenter significativement.
- ✎ **Les coûts d'image** : le risque de perdre en productivité étant très prononcé.

Etude de la base de données :

Présentation de la base données

Extraction de la base de données

Les données que nous utilisons pour cette étude sont des données de prestation en prévoyance collective du portefeuille de Generali VIE fournies par la direction de la Techniques Assurance. Cette base de données est alimentée par les équipes de gestion, actualisée et mise à jour chaque mois puis importé dans le logiciel de gestion AVT, utilisé par l'ensemble des équipes de souscription et surveillance de portefeuille pour leurs opérations au quotidien. Cette base de données nous donne accès à l'ensemble des sinistres ITIP survenu sur le périmètre collectif depuis 2012. Nous allons à présent nous focaliser dans le détail de la structure de cette base de données et des informations qu'elle pourra nous apporter dans notre démarche.

Quelle est la structure de la base de données d'étude ?

La base de données d'étude est structurée par ligne, chacune d'elle correspond à un sinistre prévoyance, voici une liste non-exhaustive des garanties couvertes :

Nom de la garantie	Fréquence
ITT	183090
ITT_HPRO	144514
INC	21805
CDC_ASS	17817
ITT_PRO	17305
ITTM	14273
ITTM_HPR	11815
ITT_HOSP	6249
ITTA_PRO	6207
IPT_HPRO	5172
ALL_OBS	4653
ITTA	4561
IPT	3650
ITT_ATMC	3303
ITT_MO	3243
ITT_AT	2985
INV	2354
ITM_HPRO	1977
ITP	1874
IPP_HPRO	1598
REDU_DC	1532
ITT_MAT	1041

La base de données initiale est composée de 63 variables pour chacun des sinistres, nous avons sélectionné parmi celles-ci, les variables suivantes :

Nom des variables	Description
NUM_CONTRAT	Numéro de contrat
NOM_ASSURE	Nom de l'assuré
PRENOM_ASSURE	Prénom de l'assuré
SOC_APERITRICE	Nom de la société apéritrice
GAR_ELEMENTAIRE	Nom de la garantie élémentaire
MOTIF_SITUATION	Motif du sinistre
DT_EFFECT_SITUATION_DOS	Date d'effet du sinistre
DAT_NAISSANCE_ASS	Date de naissance de l'assuré
DAT_SURVENANCE_SIN	Date de survenance du sinistre
DEB_COUVERTURE	Date de début de couverture
FIN_COUVERTURE	Date de fin de couverture
DEB_PER_REGLEMENT	Date de début de la période de règlement
FIN_PER_REGLEMENT	Date de fin de la période de règlement
MNT_BASE	Montant de base ou montant de la rente annualisée
MNT_TOT_REGLEMENT	Montant total réglé
MNT_REVALORISE	Montant total revalorisé
NB_J_INDEMNISES	Nombre de jours indemnisés
DT_MISE_INVALIDITE	Date de mise en invalidité dans le cas d'un dossier passé en invalidité
MNT_BASE_IJ	Montant de l'indemnité journalière
MNT_REVALORISATION_IJ	Montant de l'IJ revalorisée
MNT_SS_IJ	Montant l'IJ versé par la Sécurité Sociale
MNT_TIERS_IJ	Montant tiers payant de l'IJ
LIB_ASSIETTE_SALAIRE	Libellé de l'assiette de salaire
VALEUR_SALAIRE	Montant du salaire brut annuel de l'assuré
LIB_CHOIX_DE_PRESTATION	Libellé de la prestation
CD_OPTION	Code option

Après avoir sélectionné les informations intéressantes pour notre étude, un travail de data-cleaning était nécessaire pour différentes raisons, d'abord pour homogénéiser les données car elles ont été saisies par différentes équipes de gestions qui n'ont pas les mêmes process et les mêmes normes depuis 2012 donc il s'agira d'effectuer un travail d'homogénéisation des données, de plus il faut pouvoir être capable de capter les erreurs de saisie et de faire des choix quant au nettoyage ou au remplacement de celles-ci afin d'avoir la base de données la plus « propre » possible.

Traitement de la donnée

Ciblage sur le risque arrêt de travail

Pour étudier le risque arrêt de travail, il s'agit de sélectionner dans la base de données uniquement les sinistres de type ITT : Incapacité temporaire de travail. Pour ce faire, nous avons sélectionné dans la suite uniquement les sinistres dont la garantie élémentaire couverte contient les mentions "ITT" ou "INV" ou tout autres garanties incapacité/invalidité. De plus, parmi les données sélectionnées en incapacité, nous souhaitons ne considérer que les sinistres d'une durée inférieure à 3 ans, durée maximum légale de l'incapacité, c'est pourquoi nous avons calculé des variables à partir des données initiales pour en tirer les informations utiles à notre étude, ces variables vont être présentées et expliquées dans la suite.

Quels choix a-t-on effectué pour le nettoyage de la base de données ?

Dans le processus de nettoyage d'une base de données différents cas de figures peuvent exister lorsque l'on parle d'une donnée que l'on juge impropre, celle-ci peut être erroné, décalée à cause d'une erreur de saisie manuelle par exemple, ou tout simplement manquante. Dans ces différents cas de figure plusieurs façons de résoudre le problème existent, on peut soit essayer de corriger les erreurs de saisie ce qui en pratique peut s'avérer très compliqué lorsque l'on travaille avec une base de données conséquente (notre base de données initiale contient 697000 lignes), soit trouver la source de l'erreur lorsqu'elle est automatique et la réparer, soit simplement supprimer la donnée. Dans le cas de notre étude, l'objectif était d'avoir la base de données la plus précise et donc pour ce faire nous avons fait les choix suivants :

- Supprimer les lignes qui possédaient des informations importantes manquantes (ex : le code option)
- Homogénéiser et simplifier les informations de certaines variables (ex : les garanties élémentaires, les codes options), par exemple en regroupant, dans le cas des garanties élémentaires, l'ensemble des garanties faisant référence à de l'incapacité sous la mention « ITT »
- Corriger lorsque c'était possible les erreurs de saisie : par exemple lors de la saisie du nom et du prénom de l'assuré, dans de nombreux cas le nom et le prénom étaient saisis dans le même champ, celui du nom, ce qui fait qu'un grand nombre de prénoms d'assuré étaient manquants dans la variable prénom car saisi dans le champ nom. Cette anomalie par exemple était corrigeable pour la plupart des dossiers

Variables d'étude à calculer

Durée de l'arrêt de travail

L'une des informations importantes que nous devons capter grâce à cette base de données pour notre étude est la durée de l'arrêt de travail, pour cela nous devons procéder avec les informations que nous possédons et en particulier grâce à la période de couverture, ainsi on calcule la durée d'un sinistre incapacité de la façon suivante :

$$\text{Durée de l'arrêt} = \text{FIN_COUVERTURE} - \text{DEB_COUVERTURE} + 1$$

D'ailleurs nous ne possédons pas l'information sur la durée de la franchise pour chaque arrêt dans nos données donc il faut tenir compte que cette façon de calculer la durée d'un arrêt de travail n'est pas totalement exacte car elle ignore cette période.

Taux d'absentéisme

Cette variable à calculer est très importante et représente un enjeu important de cette étude, notamment pour la modélisation GLM dans la suite, elle est cependant assez complexe à calculer dans le contexte de l'étude avec les données en notre possession. La difficulté de ce calcul provient du fait que nous avons besoin de l'information du nombre de jours arrêté total sur une année pour chaque assuré et que cette base référence uniquement les sinistres un à un, la solution que nous avons choisie pour calculer ce taux dans la suite est la création d'une nouvelle base de données qui cette fois-ci sera une base de données agrégée par assurés et par année de survenance.

Rappelons à présent la formule de calcul de ce taux présenté dans la première partie :

$$\text{Taux d'absentéisme} = \frac{\text{Nombre de jours d'absence total sur la période d'étude}}{\text{Nombre de jours sur la période d'étude}}$$

Analyse descriptive

Dans cette partie nous tacherons d'analyser nos données en nous intéressant d'abord à la taille de notre portefeuille d'étude, puis à l'évolution du nombre d'arrêts de travail et sa saisonnalité.

Evolution du risque arrêt de travail sur les dernières années

Evolution globale du portefeuille arrêt de travail

Nous nous intéresserons dans un premier temps à l'évolution du nombre d'arrêts de travail sans différencier incapacité, invalidité et sans prendre en compte l'évolution du portefeuille collectif. En effet l'évolution du portefeuille par année est une information importante pour que l'on puisse avoir des données comparables entre années.

Année de survenance	Nombre de sinistre arrêt de travail
2013	21611
2014	31537
2015	36926
2016	38107
2017	40444
2018	42368
2019	47319
2020	49680
2021	40877

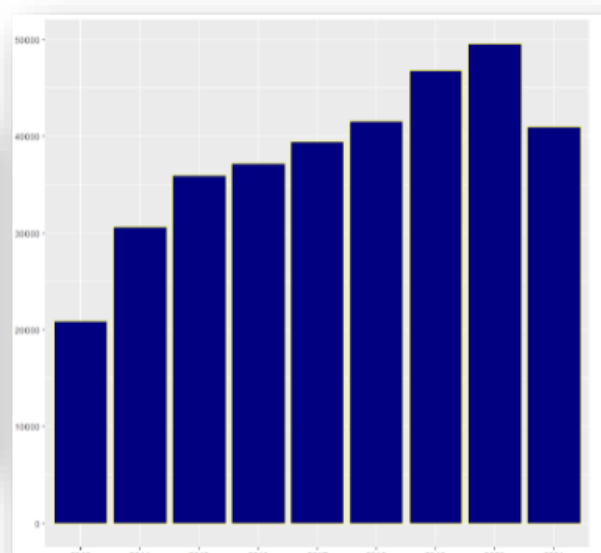


Figure 3 : Evolution du nombre d'arrêts par année de survenance

On observe une croissance relativement constante du nombre d'arrêt de travail entre 2014 et 2018 puis une forte hausse sur 2019/2020 suivi d'un retour à la normal en 2021. Nous verrons dans la suite lorsque nous étudierons la saisonnalité que la forte hausse sur 2020 est très probablement dû à un « effet COVID » car on observe une hausse significative des nombres d'arrêts durant les périodes de vague d'infection. La hausse en 2019 est sûrement dû à un agrandissement du portefeuille collectif.

En divisant le nombre d'arrêts par la taille du portefeuille collectif, nous obtenons alors un taux d'arrêts maladie par salarié, ce qui semble très intéressant pour comparer les différentes années.

Saisonnalité des arrêts de travail et mise en valeur d'un effet COVID sur 2020

Il s'agit ici de s'intéresser à la saisonnalité de la survenance des arrêts de travail sur notre portefeuille afin de constater les tendances sur les dernières années.

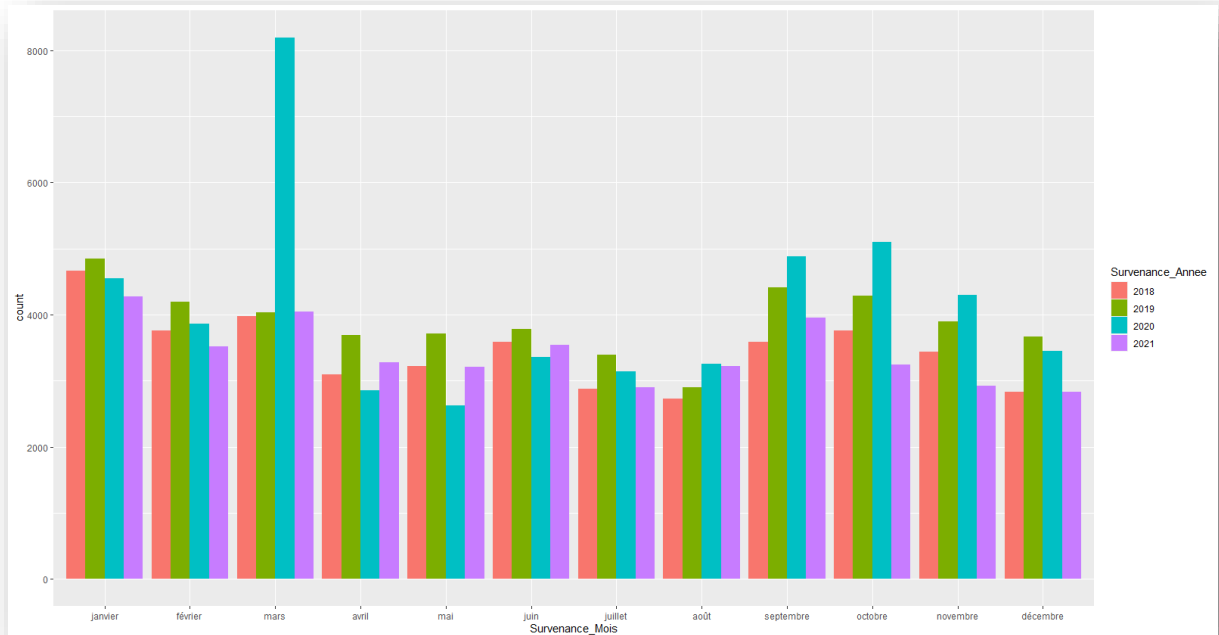


Figure 4 : Saisonnalité des arrêts de travail entre 2018 et 2021

Ce graphique est très intéressant car il nous permet de constater une réelle saisonnalité du nombre d'arrêts par année si l'on omet l'année 2020 et son caractère exceptionnel, nous nous intéresserons à cette année dans la suite mais dans un premier temps, nous allons considérer un nouveau graphique sans cette année :

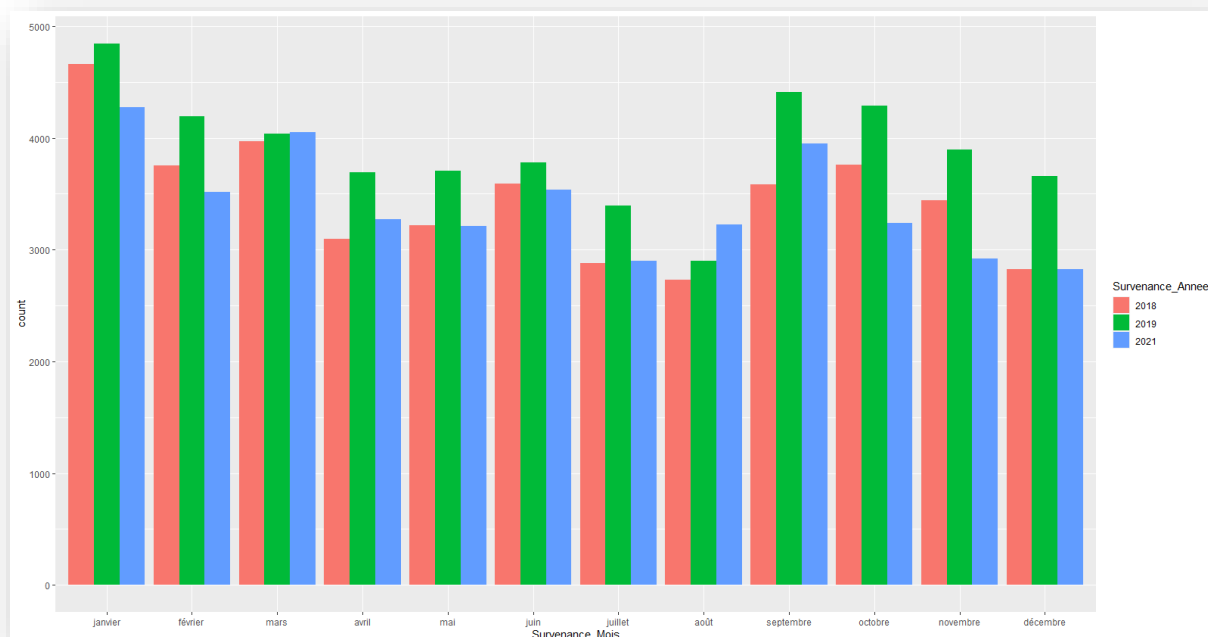


Figure 5 : Saisonnalité des arrêts de travail sur 2018, 2019 et 2021

On remarque une baisse significative du nombre d'arrêts entre la période automne-hiver et la période printemps-été, cette baisse est bien évidemment dû en partie à la présence accrue des maladies ordinaires comme par exemple le rhûme ou la grippe en hiver en sachant notamment que les maladies ordinaires représentent une partie importante du nombre d'arrêts maladie déclarés chaque année, 46% des arrêts maladie selon l'étude effectuée par REHALTO. Pour bien se rendre compte de cette différence, considérons le graphique suivant qui présente le nombre d'arrêts de travail par saison :

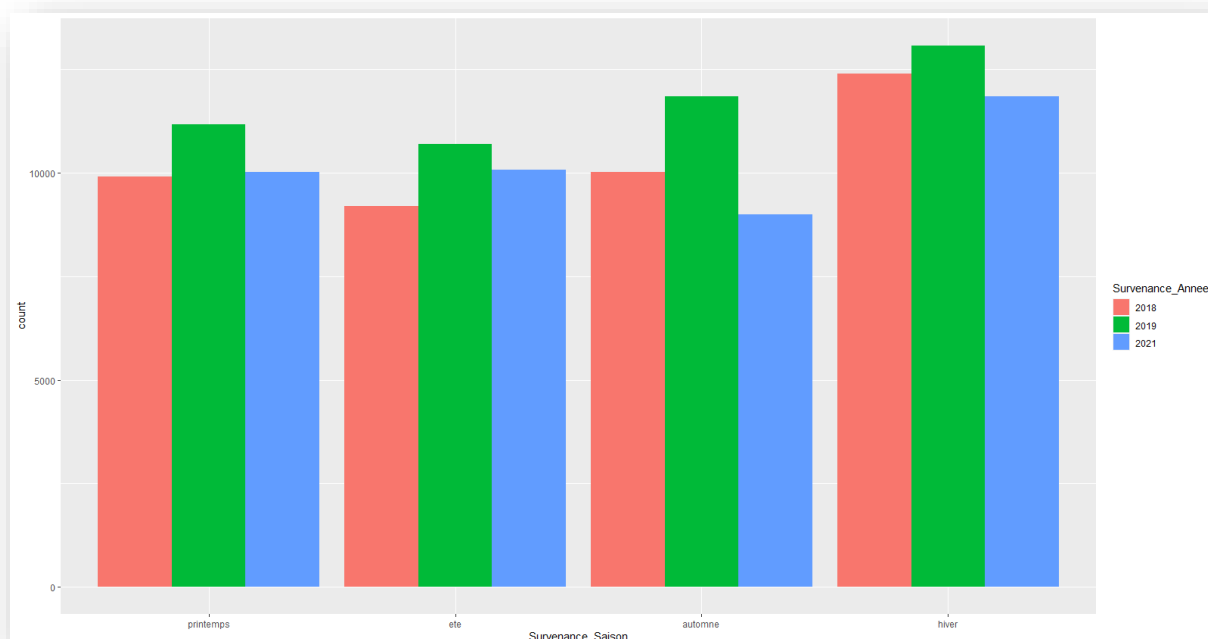


Figure 6 : Nombre d'arrêts de travail par saison

Sur 2018, nous observons une hausse de 34,7% du nombre d'arrêts entre l'été (=9196) et l'hiver (=12387), une légère baisse de 7,2% entre printemps et été et une hausse 9% entre été et l'automne.

Intéressons-nous à présent à l'année 2020 et à l'effet COVID que l'on observe bien sur nos données.

Effet COVID sur l'année 2020

Nous reconsidérons à présent le graphique d'évolution du nombre d'arrêts de travail par année de survenance depuis 2018 et sans omettre d'année :

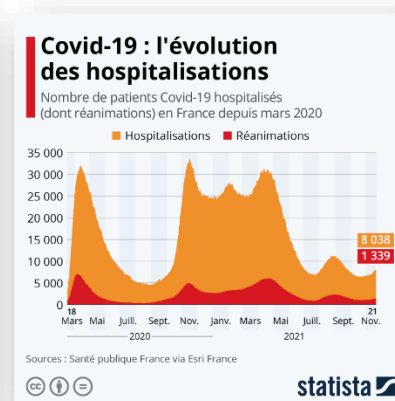
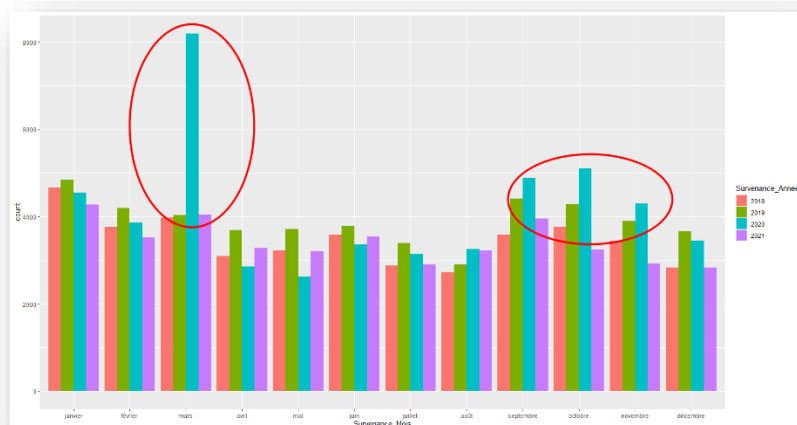


Figure 7 : Mise en valeur d'un « effet COVID » en comparaison avec la courbe d'évolution des hospitalisations sur 2020

Nous observons un énorme bond du nombre d'arrêts intervenus en mars 2020, date de la première grosse vague d'infection de COVID-19 en France et du premier confinement. Puis avec le confinement et les mesures gouvernementales, nous observons un nombre d'arrêts inférieur à la moyenne sur la période avril-juin, ce qui correspond à la période du premier confinement. Ensuite, nous observons une hausse significative du nombre d'arrêts sur la période septembre-novembre, ce qui semble correspondre avec la seconde grosse vague d'infection intervenue en 2020 et qui s'explique aussi par les mesures prises par le gouvernement pour simplifier la mise en arrêt de travail à la suite d'une infection au COVID.

Evolution des types d'arrêts par année de survenance

Nous allons à présent distinguer les différents types d'arrêts :

- Arrêt court : arrêt inférieur à 10 jours
- Arrêt moyen : arrêt durant entre 10 et 90 jours
- Arrêt long : arrêts d'une durée supérieure à 90 jours

Il semble très intéressant d'étudier l'évolution du type d'arrêt des salariés notamment pendant l'année 2020, mais aussi pour vérifier certaines tendances que nous pouvons lire dans la littérature comme par exemple le fait que les jeunes ont de plus en plus tendance à faire des arrêts court.

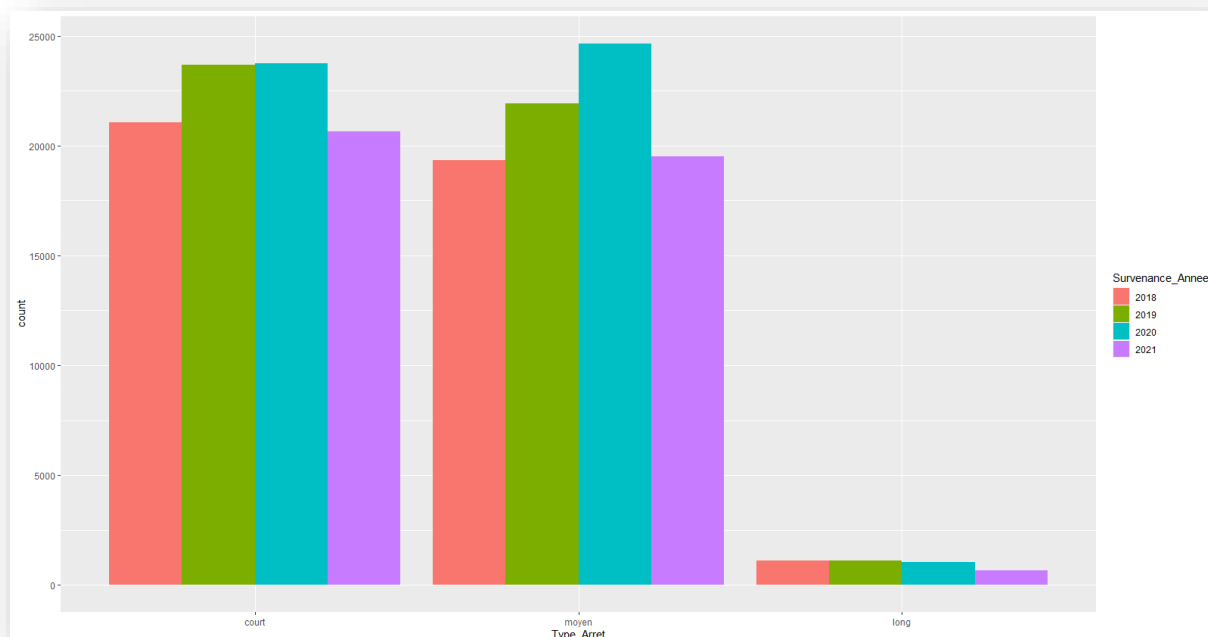


Figure 8 : Evolution du nombre d'arrêt par type

Nous observons ici une hausse significative du nombre d'arrêts moyen sur 2020, cette hausse est encore une fois sûrement dû aux dispositifs d'indemnisations des salariés lors des interruptions de travaux à la suite d'un test positif au COVID-19, dans ces cas il était très fréquent que la période d'arrêt dépassait 10 jours qui correspondait au temps moyen d'atténuation des symptômes sauf pour les cas plus graves de COVID comme les COVID long.

Aspect théorique des modèles linéaires généralisés

Les modèles linéaires généralisés ont été introduits par Nelder et Wedderburn en 1972, ils constituent la base de référence pour modéliser l'effet des variables de segmentation sur un tarif en assurance. Ces modèles sont adaptés à de nombreuses problématiques courante dans le domaine de la statistique et de l'actuariat.

Intérêt des modèles linéaires généralisés

Les modèles linéaires généralisés sont des modèles régulièrement utilisés en assurance que ce soit en assurance santé (remboursements soins, frais d'hospitalisation), en assurance auto (dommage matériel, vol, ...), en assurance MRH (incendie, vol, dégâts des eaux, ...).

Les GLM permettent de :

- Modéliser des réponses diverses
- Intégrer toute type d'information exogène susceptible d'influer sur la variable dépendante (réponse Y)
- Quantifier l'impact des facteurs de risque X (sens/intensité)
- Résidus hétéroscédastiques (la loi varie par profil)

Cependant, leur mise en place nécessite d'introduire deux hypothèses fondamentales :

- Les données que l'on cherche à expliquer sont indépendants entre elles
- Les variables explicatives X sont indépendantes deux à deux

Ces hypothèses sont très importantes notamment car l'un des intérêts des modèles linéaires généralisés en assurance est la tarification et donc le calcul d'une prime pure, calculée généralement de la façon suivante avec une approche fréquence-sinistre : $\mathbb{E}[S|X] = \mathbb{E}[N|X] * \mathbb{E}[Y_i|X]$

Principe des GLM

Composant d'un GLM :

Les différents composants d'un modèle linéaire généralisé sont :

- **La loi de la réponse aléatoire Y_i** : par hypothèse la distribution de cette loi appartient à la famille exponentielle

Rappel : La densité de probabilité d'une loi appartenant à la famille exponentielle s'écrit de la façon suivante :

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right)$$

Avec :

- ❖ Avec θ le paramètre naturel de la dispersion,
- ❖ ϕ le paramètre de dispersion, b une fonction définie sur \mathbb{R} deux fois dérivable et de dérivée première injective,
- ❖ c une fonction définie sur \mathbb{R}^2

De plus on a en particulier avec les lois de cette famille :

$$\mathbb{E}[Y] = b'(\theta) \text{ ainsi que } \mathbb{V}[Y] = b''(\theta)\phi$$

- **Le prédicteur** noté η_i et défini tel que $\eta_i = \sum_{j=1}^J \beta_j X_{ij}$ est linéaire et déterministe : il est calculé via les facteurs de risques explicatifs
- **La fonction de lien g** : elle est par définition monotone, dérivable et inversible. Elle est définie et utilisée en pratique de la façon suivante :

$$g(\mathbb{E}[Y_i]) = \eta_i$$

En pratique, il faut adapter la fonction de lien au domaine de définition de Y , voici un tableau qui présente les différentes fonction lien utilisées en pratique :

Loi	Lien naturel	Moyenne	Utilisation
$\mathcal{N}(\mu, \sigma^2)$	Id : $\eta = \mu$	$\mu = X\beta$	Rég. lin.
$\mathcal{B}(\mu)$	logit : $\eta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(X\beta)}{1+\exp(X\beta)}$	Taux
$\mathcal{P}(\mu)$	log : $\eta = \ln(\mu)$	$\mu = \exp(X\beta)$	Fréquence
$\mathcal{G}(\alpha, \beta)$	inverse : $\eta = \frac{1}{\mu}$	$\mu = (X\beta)^{-1}$	Sévérité
$\mathcal{IN}(\mu, \lambda)$	inverse ² : $\eta = -\frac{1}{\mu^2}$	$\mu = (X\beta)^{-2}$	Sévérité

La fonction de lien canonique est la fonction lien qui associe la moyenne μ au paramètre canonique θ . Elle est telle que :

$$g(\mu_i) = \theta_i$$

On trouvera ci-dessous la fonction de lien canonique associée à certaines lois usuelles :

Loi de probabilité	Fonction de lien canonique
Normale	$\eta = \mu$ (identité)
Poisson	$\eta = \ln(\mu)$ (log)
Gamma	$\eta = \frac{1}{\mu}$ (inverse)
Gaussienne inverse	$\eta = \frac{1}{\mu^2}$
Binomiale	$\eta = \ln\left(\frac{\mu}{1-\mu}\right)$ (logistique)

A présent, présentons quelques modèles fréquemment utilisés dans le contexte assurantiel :

Modèle gaussien :

Dans le cas d'un échantillon gaussien, les densités d'une famille de lois $N(\mu_i, \sigma^2)$ s'écrivent :

$$f(y_i, \mu_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right]$$

Et appartiennent à la famille exponentielle en posant :

$$\begin{aligned}\theta_i &= E(Y_i) = \mu_i \\ a(\theta_i) &= \exp\left[-\frac{\mu_i^2}{2\sigma^2}\right] \\ b(\theta_i) &= \frac{\theta_i}{\phi} = \frac{\mu_i}{\sigma^2} \\ c(y_i) &= \exp\left[-\frac{y_i^2}{2\sigma^2} - \frac{\ln(2\pi\sigma^2)}{2}\right]\end{aligned}$$

On remarque avec la première égalité que la famille gaussienne se met sous forme canonique

Modèle de Poisson :

Le modèle de Poisson est un modèle de comptage pour le définir, on considère n variables indépendantes Y_i de loi de Poisson de paramètre $\mu_i = E(Y_i)$. Les Y_i sont par exemple les effectifs d'une table de contingence. Ces variables admettent pour densités :

$$f(y_i, \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} = \exp(-\mu_i) * \frac{1}{y_i!} \exp(y_i \ln(\mu_i!))$$

Avec : $\theta_i = \ln(\mu_i)$

Et donc la fonction de lien canonique de ce modèle est la fonction logarithme népérien.

Modèle Tweedie :

Les variables de réponses Y_i admettent dans ce modèle la densité suivante :

$$f(y, \mu, \phi) = a(y, \phi) * \exp\left(\frac{1}{\phi} [y * \theta(\mu) - \kappa(\theta(\mu))]\right)$$

$$\text{Avec : } \theta(\mu) = \begin{cases} \frac{\mu^{1-p}}{1-p} & \text{si } p \neq 1 \\ \ln(\mu) & \text{si } p = 1 \end{cases} \quad \kappa(\theta(\mu)) = \begin{cases} \frac{\mu^{2-p}}{2-p} & \text{si } p \neq 2 \\ \ln(\mu) & \text{si } p = 2 \end{cases}$$

Avec cette formalisation, $\mathbb{E}[Y_i] = \mu_i$ et $\text{Var}[Y_i] = \psi \mu_i^p$ avec ψ un facteur de dispersion positif.

Dans le cadre de ce modèle, on appelle p le paramètre d'indice, il est choisi en fonction de l'application avant d'estimer μ et ϕ et définit le type de distribution :

- ↻ $p < 0$: réalisations dans \mathbb{R} ;
- ↻ $p = 0$: loi gaussienne
- ↻ $0 < p < 1$: pas de distribution
- ↻ $p = 1$ avec $\psi = 1$: loi de Poisson
- ↻ $1 < p < 2$: loi composée Poisson-Gamma
- ↻ $2 < p < 3$ ou $p > 3$: positive stable distributions
- ↻ $p = 2$: loi Gamma
- ↻ $p = 3$: loi inverse gaussienne.

Validation d'un modèle GLM :

Pour valider le modèle linéaire généralisé étudié, il existe les différentes méthodes suivantes :

Validation croisée (k-fold)

Le découpage de l'échantillon en apprentissage et en validation induit une perte d'un certain volume de données pour calculer les estimateurs mais permet de calculer les erreurs moyenne de l'estimateur, cette segmentation de l'échantillon se fait via les étapes suivantes :

- On coupe aléatoirement l'échantillon de données en k sous-échantillons de taille égale
- On sélectionne un échantillon de validation et k-1 échantillon d'apprentissage
- On fait tourner k fois les étapes précédentes pour que chaque sous-échantillon ait été utilisé une fois en validation
- On calcule la moyenne sur les k estimation pour obtenir l'estimateur final
- On calcule l'erreur moyenne sur les échantillons de validation

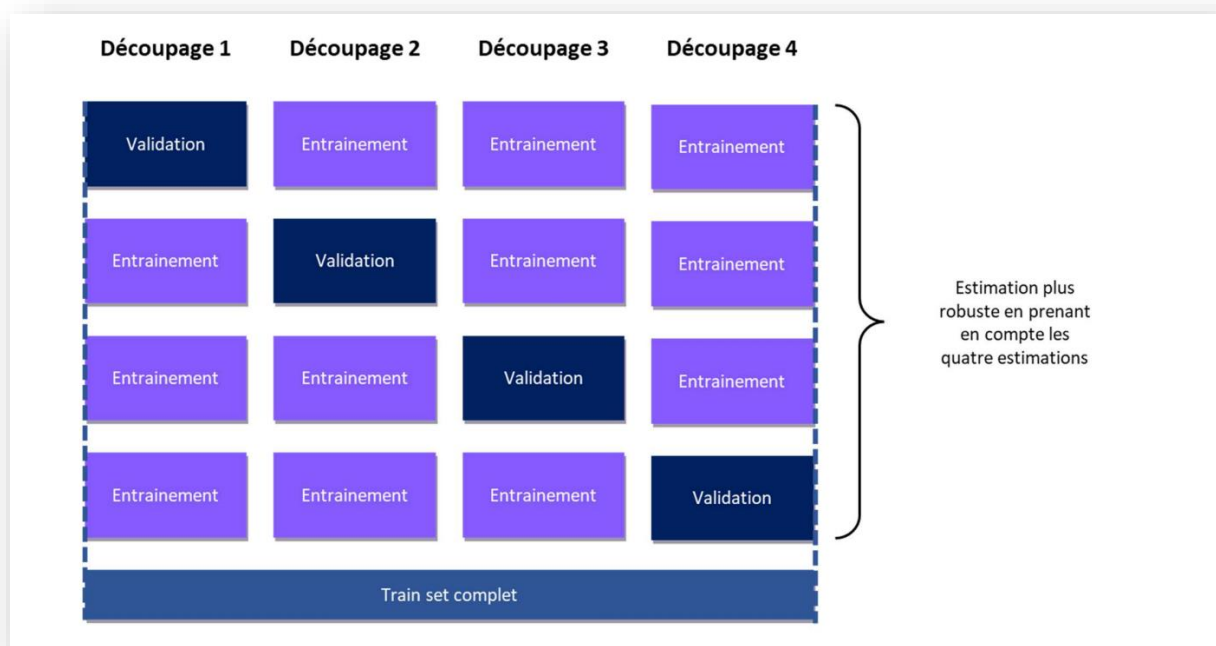


Figure 9 : Exemple illustré d'une validation croisée (4-fold)

Validation de la significativité globale du modèle

Pour valider la significativité globale du modèle, il est judicieux de travailler avec des critères de qualité d'ajustement du modèle comme la déviance ou la statistique de Pearson que nous définissons ci-dessous :

☞ La déviance :

Pour définir la déviance du modèle, nous introduisons la notion de modèle saturé qui est le modèle qui possède autant de paramètres que d'observations. Il est caractérisé par l'égalité entre les facteurs estimés et les observations, c'est-à-dire : $\hat{\mu}_i = y_i$. Puis nous comparerons ce modèle avec le modèle d'étude pour calculer la déviance.

Il faut aussi introduire la log-vraisemblance pour calculer la déviance :

Soit $Y = (Y_1, \dots, Y_n)$ le vecteur à expliquer (appartenant à la famille exponentielle) dont la densité s'écrit sous la forme :

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right)$$

La log-vraisemblance de ce modèle s'écrit :

$$\begin{aligned} l(y, \theta, \phi) &= \sum_{i=1}^n \ln(f(y_i|\theta_i, \phi)) \\ &= \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi} + \sum_{i=1}^n c_i(y_i, \phi) \end{aligned}$$

Nous utiliserons comme mesure de la qualité d'ajustement du modèle la statistique du rapport de vraisemblance suivante :

$$\lambda = \frac{L_{SAT}}{L} \quad \text{ou} \quad \ln(\lambda) = \ln(L_{SAT}) - \ln(L)$$

Car le modèle décrit bien les données lorsque $L \simeq L_{SAT}$

Enfin, on définit la statistique de déviance réduite ou normalisée de la façon suivante :

$$D = 2 \ln(\lambda) = 2(\ln(L_{SAT}) - \ln(L))$$

La déviance non réduite est :

$$D^* = \phi D$$

De plus, $D \sim \chi_{n-p-1}^2$, c'est-à-dire que le modèle est de mauvaise qualité si $D_{obs} > \chi_{n-p-1; 1-\alpha}^2$

☞ La statistique de Pearson :

La statistique du test de significativité globale de Pearson est définie de la façon suivante :

$$X^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{Var(y_i)} = \frac{1}{\phi} \sum_{i=1}^n \omega_i \frac{(y_i - \mu_i)^2}{V(\mu_i)}$$

Comme pour la Déviance, nous avons $X^2 > \chi_{n-p-1; 1-\alpha}^2$

Validation de la significativité individuelle des coefficients de la régression

Pour valider ou non le modèle en étudiant la significativité individuelle des coefficients de la régressions, nous pouvons utiliser des tests d'hypothèses sur les paramètres, pour cela nous introduirons l'hypothèse H_0 de la façon suivante :

$$H_0: C\beta = r$$

Avec : C une matrice connue possédant q lignes et r un ensemble de valeurs testées.

Ensuite nous pourrions considérer trois approches pour valider ou non cette hypothèse :

☞ Test du rapport de vraisemblance :

Nous avons défini le rapport de vraisemblance précédemment : λ mais dans ce cas nous n'utiliserons pas un modèle saturé mais un modèle sans contrainte de vraisemblance \hat{L} et un modèle sous contrainte de vraisemblance \tilde{L} .

Ainsi nous le rapport de vraisemblance s'exprimera : $\lambda = \frac{\hat{L}}{\tilde{L}}$

La statistique de ce test est :

$$D = 2 \ln(\lambda) \sim \chi_q^2$$

☞ Test de Wald :

Pour définir la statistique du test de Wald, il faut dans un premier temps définir l'estimateur du maximum de vraisemblance de $\beta, \hat{\beta}$.

$$\hat{\beta} \sim \mathcal{N}(\beta, \phi(X'WX)^{-1})$$

Avec W la matrice diagonale de pondération définie telle que :

$$[W]_{i,i} = \frac{1}{Var(Y_i)} \left(\frac{\partial y_i}{\partial \mu_i} \right)^2$$

Et donc avec les notations précédentes, sous H_0 , $C\hat{\beta} - r \sim \mathcal{N}(0, \phi C(X'WX)^{-1}C')$

Enfin, la statistique du test de Wald est définie comme :

$$(C\hat{\beta} - r)' [\phi C(X'WX)^{-1}C']^{-1} (C\hat{\beta} - r) \sim \chi_q^2$$

En pratique, nous pouvons tester un seul paramètre :

$$H_0: \beta_j = r_j$$

On a alors, $C_i = 0$, pour $i \neq j$ et $C_j = 1$ et $Var(\hat{\beta}_j) = \phi\psi_j$.

La statistique du test de Wald est dans ce cas égale à : $\frac{(\hat{\beta}_j - r_j)^2}{Var(Y_i)} \sim \chi_1^2$

☞ Test du Score :

Ce test est basé sur la dérivée de la fonction log-vraisemblance définie précédemment, cette dérivée est appelée le Score, on la définit en pratique dans ce cas par :

$$l'(\beta) = \phi^{-1} X' W G (y - \mu)$$

Avec G une matrice diagonale composée des éléments $g'(\mu_i)$ et W une matrice d'éléments $\left[(g'(\mu_i))^2 V(\mu_i) \right]^{-1}$. De plus, on peut montrer que $\mathbb{E}[l'(\beta)] = 0$ et $\text{Var}(l'(\beta)) = \phi^{-1} X' W X$

Et enfin la statistique du test du score est donnée par :

$$(l'(\beta))^t [\text{Var}(l'(\beta))]^{-1} l'(\beta) \sim \chi_q^2$$

Etude des résidus

Pour évaluer l'erreur du modèle, nous pouvons calculer les résidus du modèle de différentes manières, les méthodes que l'on utilisera dans cette étude sont les principales utilisées pour calculer des résidus, les résidus de Pearson et les résidus de déviations définis tels que :

- Les résidus de Pearson :

Le résidu de Pearson est défini comme :

$$r_i^p = \frac{\sqrt{\omega_i}(y_i - \mu_i)}{\sqrt{V(\mu_i)}}$$

- Les résidus de Déviance :

On peut considérer que chaque observation y_i contribue à une quantité d_i à la déviance ($D = \sum_{i=1}^n d_i$), le résidu de déviance est défini comme :

$$r_i^D = \sqrt{d_i} \times \text{signe}(y_i - \mu_i)$$

La somme des carrés des résidus dans les deux cas est asymptotiquement une statistique du Khi-2 à n-p-1 degrés de liberté.

Conclusion – Pistes pour la suite

L'étude du risque arrêt de travail est de plus en plus importante de nos jours, le phénomène très coûteux d'absentéisme doit être appréhender, évaluer et gérer par les assureurs mais aussi par les entreprises en général.

Dans le cadre de ce mémoire, nous avons pu étudier l'absentéisme dans la globalité sur le portefeuille collectif de Generali Vie, d'abord en expliquant les notions clés à comprendre pour appréhender le phénomène de l'absentéisme en entreprise, puis en analysant le portefeuille, les tendances sur les arrêts de travail et l'impact du COVID-19 sur l'absentéisme.

La phase de préparation des données a été primordiale dans cette étude, car l'objectif principal de ce mémoire était de fournir les premiers outils pour analyser l'absentéisme de notre portefeuille collectif. Pour la poursuite de ce mémoire, l'objectif est d'aller plus loin en étudiant l'absentéisme individuel des salariés, d'abord en construisant une base de données agrégée par année pour chaque salarié, puis d'analyser ce phénomène, de le comprendre et d'étudier les méthodes pour gérer ce risque en pratique.

Grâce à ce mémoire, nous avons pu constater les limites que le manque d'informations sur les assurés du périmètre collectif crée pour la modélisation de l'absentéisme avec des modèles linéaires généralisés. Ainsi nous avons pu fournir une piste d'amélioration pour l'étude de l'absentéisme sur notre portefeuille qui est la création d'une base de données sur les bénéficiaires des contrats en y indiquant par exemple leur situation professionnelle, leur secteur d'activité, leur position géographique.

Bibliographie

- AXA. (2019). "Quels sont les risques de demain?". Récupéré sur <https://www.axa.com/fr/magazine/rapport-risques-futurs-2019#:~:text=Selon%20le%20rapport%2C%20les%20principaux,et%20%C3%A0%20l'instabilit%C3%A9%20g%C3%A9opolitique>.
- Brami, L., Damart, S., & Kletz, F. (2013). SANTÉ AU TRAVAIL ET TRAVAIL EN SANTÉ. LA PERFORMANCE DES ÉTABLISSEMENTS DE SANTÉ FACE À L'ABSENTÉISME ET AU BIEN-ÊTRE DES PERSONNELS SOIGNANTS.
- Chaupain-Guillot, S., & Guillot, O. (2007). Les absences au travail : une analyse à partir des données françaises du Panel européen des ménages. *INSEE*.
- Fouquet, N., Chérié-Challine, L., Rubion, É., Descatha, A., & Roquelaure, Y. (2020). TROUBLES MUSCULO-SQUELETTIQUES LIÉS AU TRAVAIL : NOMBRE DE CAS ÉVITABLES.
- Humanis, M. (2020). Rapport annuel sur l'absentéisme.
- Milhaud, X. (2022). Compléments de tarification en assurance: problématiques opérationnelles.
- Planchet, F. (2020). *Ressources actuarielles*. Récupéré sur Ressources actuarielles: [http://www.ressources-actuarielles.net/EXT/ISFA/fp-isfa.nsf/0/1430AD6748CE3AFFC1256F130067B88E/\\$FILE/Seance7.pdf?OpenElement](http://www.ressources-actuarielles.net/EXT/ISFA/fp-isfa.nsf/0/1430AD6748CE3AFFC1256F130067B88E/$FILE/Seance7.pdf?OpenElement)
- Rehalto. (2019). *Baromètre sur les arrêts de travail « Comprendre pour agir » 5ème édition*. Récupéré sur <https://entreprise-rh.com/wp-content/uploads/2019/10/R%C3%A9halto-BVA-Barom%C3%A8tre-arr%C3%Aats-de-travail-2019-4.pdf>